Expert Systems With Applications

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

--Manuscript Draft--

Manuscript Number:	ESWA-D-25-28448R1
Article Type:	Full length article
Section/Category:	1.1 AI / ML model and method
Keywords:	Conflict risk modeling; Manifold similarity; mixed traffic flow; Spatiotemporal characteristics; Adaptive graph networks
Corresponding Author:	Zongshi Liu Tongji University CHINA
First Author:	Zongshi Liu
Order of Authors:	Zongshi Liu
	Guojian Zou
	Ting Wang
	Meiting Tu
	Hongwei Wang
	Ye Li
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.

November 18, 2025

Dear Prof. Peter Werner Eklund,

Please find attached our revised version of the manuscript titled " Learning and Predicting Traffic Conflicts in Mixed Traffic: A Spatiotemporal Graph Neural Network with Manifold Similarity Learning" that we would like you to consider for publication in Expert Systems With Applications. This paper, for the first time, proposes manifold similarity spatiotemporal graph network (MS-STGNet) to predict conflict potential for mixed CAV-HDV traffic on freeways. The contributions of this paper are outlined as follows:

- 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.
- 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.
- 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.
- 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.

Based on all comments and suggestions, the authors conducted revisions to the paper, which are summarized as follows:

- Clarified the methodological novelty and positioning of MS-STGNet relative to existing manifold-based traffic models and STGAT-type adaptive graph networks, including expanded discussions in Sections 1, 2.2, 2.3, 5.3.1, and 6.5.
- Strengthened the justification of our simulation-based framework and data choices, explaining the limitations of current mixed CAV–HDV datasets and detailing our “real-trajectory calibration + large-scale simulation” strategy, together with an explicit statement of validation limitations and future real-world testing plans in the Conclusion.
- Enhanced the analysis of model performance, stability, and sensitivity by

reporting mean ± standard deviation over five independent runs (with statistical significance at the 5% level), clarifying the posterior-probability analyses (Section 6.8), and adding new sensitivity-style results on traffic volume, CAV penetration, and speed separation in Section 6.9 and Appendix C.

- Added a computational-cost analysis (Section 6.6) comparing GPU memory usage and parameter counts across all deep learning baselines and MS-STGNet, thereby demonstrating that our model achieves superior performance with a computational burden comparable to or lower than STGCN and STGAT.
- Provided a clearer discussion of influential traffic features for conflict prediction (e.g., CAV penetration, traffic volume, speed dispersion patterns, and vehicle composition), and how MS-STGNet aligns its predicted risk with these structures.
- Streamlined and updated the literature review, reorganized some technical details into appendices, refined mathematical notation, and polished the language throughout to improve readability and consistency.
- Expanded the discussion of limitations and future research directions, especially regarding scenario generalizability, online manifold updating, graded conflict severity modelling, and large-scale real-world deployment.

We hope these changes will strengthen our manuscript. Thank you for your consideration.

With my kindest regards,

Corresponding Author: Ye Li
Affiliation: College of Transportation Engineering, Tongji University, Shanghai 201804, PR China
E-mail: JamesLI@tongji.edu.cn

Modification and Responses to the Reviewers' Comments and Suggestions
Manuscript ESWA-D-25-28448

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

Zongshi Liu, Guojian Zou, Ting Wang, Meiting Tu, Hongwei Wang, Ye Li

AE: There are conflicting review reports, even the recommendation to reject the paper. Against this backdrop it is crucial to carefully address all the comments, especially the critical ones.

Dear Professor Eklund,
We are very grateful to the Editor and the Reviewers for spending precious time reviewing our work and providing useful comments. We have carefully considered and responded to each of the comments from the reviewers. Below we briefly summarize the major revisions that have been made to improve the quality and clarity of the manuscript:

- Clarified the methodological novelty and positioning of MS-STGNet relative to existing manifold-based traffic models and STGAT-type adaptive graph networks, including expanded discussions in Sections 1, 2.2, 2.3, 5.3.1, and 6.5.
- Strengthened the justification of our simulation-based framework and data choices, explaining the limitations of current mixed CAV–HDV datasets and detailing our "real-trajectory calibration + large-scale simulation" strategy, together with an explicit statement of validation limitations and future real-world testing plans in the Conclusion.
- Enhanced the analysis of model performance, stability, and sensitivity by reporting mean ± standard deviation over five independent runs (with statistical significance at the 5% level), clarifying the posterior-probability analyses (Section 6.8), and adding new sensitivity-style results on traffic volume, CAV penetration, and speed separation in Section 6.9 and Appendix C.
- Added a computational-cost analysis (Section 6.6) comparing GPU memory usage and parameter counts across all deep learning baselines and MS-STGNet, thereby demonstrating that our model achieves superior performance with a computational burden comparable to or lower than STGCN and STGAT.
- Provided a clearer discussion of influential traffic features for conflict prediction (e.g., CAV penetration, traffic volume, speed dispersion patterns, and vehicle composition), and how MS-STGNet aligns its predicted risk with these structures.
- Streamlined and updated the literature review, reorganized some technical details into appendices, refined mathematical notation, and polished the language throughout to improve readability and consistency.
- Expanded the discussion of limitations and future research directions, especially regarding scenario generalizability, online manifold updating, graded conflict severity modelling, and large-scale real-world deployment.

Explanations of what we have changed in response to the reviewers' concerns are given point by point in the following pages. The changes in the revised manuscript have highlighted in blue. We hope these changes will strengthen our manuscript.

Comments from Reviewer #1:

Comment 1:

Firstly, the novelty of the proposed framework is somewhat limited. Although the integration of manifold similarity into a spatiotemporal graph neural network is an interesting idea, the concept of using manifold learning for traffic state modeling is not entirely new and has been explored in previous studies. The manuscript needs to provide a more detailed comparison with existing methods to clearly highlight the unique contributions of the proposed MS-STGNet framework.

Response to Comment 1:

We sincerely thank the reviewer for these thoughtful comments on the novelty of MS-STGNet and its relation to existing manifold-based traffic models and STGAT-type adaptive graph networks.

In the revised manuscript, we have clarified our contributions at both the problem and method levels. From the problem perspective, we emphasize that our primary goal is real-time conflict prediction in mixed CAV–HDV freeway traffic, a setting where existing work is still limited. To ensure robustness in this new safety-critical application, we deliberately build on mature components (residual CNN, TCN, spatiotemporal GNNs) while introducing a manifold-similarity graph as a physically meaningful prior rather than proposing an entirely new architecture for its own sake.

From the methodological perspective, we now explicitly distinguish our approach from prior manifold-learning studies and STGAT-type adaptive adjacency mechanisms. Section 2.3 has been expanded to include recent manifold-based traffic-flow and safety studies and to clarify that these works mainly use manifold embeddings for clustering, visualization, or as features in conventional models, without embedding manifold-based traffic-state similarity into a spatiotemporal GNN for online conflict prediction. In contrast, MS-STGNet integrates a pre-computed manifold similarity matrix, derived from historical traffic states, as an interpretable prior that constrains adaptive adjacency learning for mixed CAV–HDV conflicts.

We also revise Section 2.2 and Section 5.3.1 to clearly contrast our manifold-similarity graph with standard STGAT mechanisms: instead of learning adjacency solely from instantaneous node features at each time step, MS-STGNet initializes the graph from manifold distances computed offline and then performs lightweight adaptive refinement. This design ties the learned graph to physically meaningful traffic-state geometry while keeping the per-iteration computational cost comparable to standard STGNNs. Finally, in Section 6.5 we explicitly highlight that the manifold-similarity prior contributes to the reduction of false alarm rates and improved robustness compared with STGCN and STGAT, especially at medium-to-high CAV penetration rates.

We hope these revisions and clarifications make the unique contributions and methodological positioning of MS-STGNet more evident. Below is the revised version:

Revised:

In Section 1 (Page 2 Line 45—49, Page 3 Line 10—14):

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.

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.

In Section 2.2 (Page 4 Line 41—44):

In addition, existing spatiotemporal graph-based safety models typically define spatial dependencies through fixed adjacency matrices or adaptive attention mechanisms in the original feature space, and rarely exploit manifold-based traffic-state similarity as an explicit prior, particularly in mixed CAV–HDV traffic environments.

In Section 2.3 (Page 5 Line 26—29):

Additionally, few studies have attempted to integrate the concept of state transitions in manifold learning into deep learning frameworks, and, to the best of our knowledge, none has embedded manifold-based traffic-state similarity into a spatiotemporal graph neural network for real-time conflict prediction in mixed CAV–HDV traffic.

In Section 5.3.1 (Page 12 Line 1—10):

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.

In Section 6.5 (Page 18 Line 31—43):

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.

Comment 2:

Secondly, the experimental setup and validation process lack sufficient rigor. The simulation environment, while realistic, is based on predefined parameters and assumptions that may not fully capture the complexities of real-world mixed traffic conditions. The manuscript should include more comprehensive validation using real-world traffic data to demonstrate the practical applicability and robustness of the proposed model.

Response to Comment 2:

We sincerely appreciate this important comment. At the early stage of this study, our initial intention was indeed to develop and validate MS-STGNet directly on real-world mixed CAV–HDV data. However, after a thorough review of existing datasets, we found that currently available data sources cannot simultaneously meet the two core requirements of our problem: 1) truly mixed CAV–HDV traffic, and 2) long, spatially continuous freeway segments with macroscopic measurements (flow, speed, occupancy) suitable for segment-level conflict prediction over continuous time series.

On the one hand, classical trajectory datasets such as NGSIM and highD contain only human-driven vehicles and therefore do not match our target mixed-traffic scenario. Nevertheless, to ensure that our simulation is not based on purely theoretical assumptions, we calibrated the human-driven car-following model directly on highD freeway trajectories, so that HDV behaviour in the simulation reflects empirically observed acceleration, deceleration, and headway patterns rather than arbitrary parameter choices.

On the other hand, recent autonomous-vehicle datasets such as the Lyft Level 5 AV Dataset (Houston et al., 2021), nuScenes (Caesar et al., 2020), and the Waymo Open Dataset (Sun et al., 2020) do provide mixed traffic with AVs/CAVs, but their structure is not well suited to our macroscopic conflict-prediction task. As summarized in Table 1 of Hu et al. (2022), these AV datasets are organized into short trajectory segments: Waymo comprises 1,000 segments with a temporal resolution of 0.1 s and a typical segment length of 20 s; Lyft Level 5 contains 366 segments at 0.2 s resolution and 25–45 s duration; and nuScenes includes 1,000 segments with 0.5 s resolution and 20 s duration. These segments are collected from the viewpoint of individual AVs and are neither spatially contiguous along a single freeway facility nor

temporally continuous over long periods. Hu et al. (2022) explicitly note that substantial preprocessing and reconstruction are required even to obtain usable car-following trajectories from these segment-based recordings, and that the resulting data remain fragmented in space and time for macroscopic analyses.

Table 1. Overview of three AV trajectory dataset.

Dataset	Number of segments	Resolution (s)	Length of each segment (s)
Waymo	1000	0.1	20
Lyft	366	0.2	25–45
nuScenes	1000	0.5	20

[Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... & Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11621-11631).

Houston, J., Zuidhof, G., Bergamini, L., Ye, Y., Chen, L., Jain, A., ... & Ondruska, P. (2021, October). One thousand and one hours: Self-driving motion prediction dataset. In Conference on Robot Learning (pp. 409-418). PMLR.

Hu, X., Zheng, Z., Chen, D., Zhang, X., & Sun, J. (2022). Processing, assessing, and enhancing the Waymo autonomous vehicle open dataset for driving behavior research. *Transportation Research Part C: Emerging Technologies*, *134*, 103490.

Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., ... & Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2446-2454).]

Similarly, Zhang et al. (2025) show that such AV trajectory datasets are particularly suitable for microscopic behaviour analysis and data-driven stochastic fundamental-diagram modelling, but they are inherently sparse in space and time and therefore not directly aligned with macroscopic, segment-level modelling of traffic states over extended freeway sections. In our study, however, the prediction target is whether a given freeway segment will experience a conflict within a continuous time series, based on macroscopic indicators (flow, speed, occupancy) monitored along a 14 km stretch. This requires long, contiguous observations along one facility, which current AV datasets do not provide.

[Zhang, X., Yang, K., Sun, J., & Sun, J. (2025). Stochastic fundamental diagram modeling of mixed traffic flow: A data-driven approach. *Transportation Research Part C: Emerging Technologies*, *179*, 105279.]

These limitations are consistent with the broader challenges identified in recent reviews of machine-learning-based crash prediction. For example, Ali et al. (2024) point out that empirical safety studies for mixed CAV–HDV environments are still scarce, that most real-time crash and conflict prediction models are developed for conventional freeway or urban networks without CAVs, and that many studies necessarily rely on simulation or indirectly inferred data due to the lack of suitable field datasets. Against this background, we adopted a hybrid strategy that combines empirical calibration with large-scale simulation.

[Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis & Prevention*, *194*, 107378.]

Given these data limitations, we adopted a hybrid “real-trajectory calibration + simulation-

based testing" strategy. At the microscopic level, we construct mixed CAV–HDV traffic with multiple penetration rates (10%, 30%, 50%, 70%, and 90%) and generate trajectories at 0.2s resolution. Using these trajectories, we compute widely used surrogate safety measures TTC, DRAC, and DDR—to identify both longitudinal and lateral conflicts. At the macroscopic level, we aggregate the simulated data along a 14 km four-lane freeway with ramps and extract continuous time series of flow, speed, and occupancy for each segment, paired with the conflict/non-conflict labels derived from TTC/DRAC/DDR. This design allows us to study conflict prediction as a segment-level time-series classification problem under a wide range of demand levels and CAV penetration rates, while keeping driver behavior anchored in real freeway observations and defining conflicts via physically interpretable criteria.

We fully acknowledge that this hybrid validation cannot replace large-scale field testing on real mixed CAV–HDV networks. In the revised Conclusion, we now explicitly state that 1) the current evaluation is conducted in a calibrated freeway simulation, 2) the direct transferability to urban or suburban networks is therefore limited, and 3) the lack of suitable real-world mixed-traffic datasets is a key limitation of the present work. We also clarify that, once continuous macroscopic observations of mixed CAV–HDV traffic over long freeway segments become available, we plan to retrain and evaluate MS-STGNet on those data and to systematically compare its performance with other state-of-the-art safety models in real-world environments. Below is the revised version:

Revised (Page 25 Line 25, Page 26 Line 1—21):

The proposed framework has several practical implications. It can be embedded as a safety prediction component in CAV cloud management systems for freeway corridors and urban expressways, integrated into freeway traffic management centers and ramp control or variable speed limit systems to support mixed CAV–HDV operations, and used within regional expressway operation platforms to provide real-time conflict or crash risk warnings at bottlenecks and merging/diverging areas, thereby enhancing the safety management and visualization of freeway networks. The limitations of this study are summarized as follows: 1) The model is calibrated and evaluated in a microscopic simulation of a four-lane freeway segment with motorized traffic only. Although the simulation is grounded in highD trajectory data, we do not yet validate MS-STGNet on large-scale field observations of mixed CAV–HDV traffic, and the direct transferability of the results to urban or suburban road networks with signalized intersections, pedestrians, and non-motorized vehicles is therefore limited. 2) The current experiments focus on a single 14 km corridor with specific demand patterns; additional facilities and more diverse demand scenarios would further test the generalizability of the framework. 3) The predefined manifold similarity matrix remains static over time, preventing the model from capturing previously unseen traffic state transitions unless it is retrained. 4) The proposed framework currently focuses on binary conflict/non-conflict prediction. Although the sigmoid activation in the output layer produces continuous risk 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.

Comment 3:

Additionally, the performance metrics used for evaluation are standard, but the manuscript does not provide a thorough analysis of the model's performance under different traffic conditions and scenarios. A more detailed sensitivity analysis and comparison with state-of-the-art models in various settings would strengthen the manuscript.

Response to Comment 3:

We appreciate this constructive suggestion. In the revised manuscript, we have clarified and strengthened the analysis of MS-STGNet under different traffic conditions and scenarios, and its comparison with state-of-the-art baselines. First, Section 6.1 (Data preparation) clearly explains that the simulation covers 500 hours of mixed CAV–HDV traffic with three representative demand levels (low, medium, high traffic volume) and five CAV penetration rates (10%, 30%, 50%, 70%, 90%), which define the set of operating scenarios used throughout the experiments (See Page 14 Line 10—16). Section 6.5 has been updated to report all metrics in Table 4 as mean ± standard deviation over five independent runs with different random seeds, and to note that the improvements of MS-STGNet over the baseline models are statistically significant at the 5% level ($p < 0.05$), thereby quantifying cross-run stability in a scenario-wise comparison.

Second, to provide a more detailed sensitivity and scenario analysis, we have explicitly highlighted several complementary results: 1) Section 6.8 (Posterior probability analyses) examines how the distributions of predicted conflict probabilities evolve with penetration rate for MS-STGNet versus STGCN and STGAT, showing that MS-STGNet maintains better separation between conflict and non-conflict classes as class imbalance increases. 2) Section 6.10, together with the new Appendix C, investigates the impact of traffic volume and speed dispersion on conflict risk and model behavior, using trajectory plots at low/medium/high volumes to illustrate how higher demand intensifies speed oscillations and conflicts, and how MS-STGNet aligns its risk predictions with these patterns more effectively than the baselines. 3) Section 6.6 now includes a computational cost analysis (GPU memory usage and parameter counts) for all deep learning baselines and MS-STGNet, complementing the accuracy-based comparison with an efficiency perspective. Taken together, these additions provide a more thorough sensitivity analysis across penetration rates, demand levels, disturbance patterns, and computational cost in line with the reviewer's recommendation. Below is the revised version:

Revised:

In Section 6.2 (Page 14 Line 34—35):

To reduce the impact of randomness and evaluate the stability of each method, all models

are trained and evaluated five times with different random seeds orders.

In Section 6.5 (Page 16 Line 38—43, Page 17, Page 18 Line 1—43):

To further assess cross-run stability, each entry in Table 4 is reported as the mean ± 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 ($p < 0.05$).

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 23% and 17.8% lower, respectively, than those of the proposed MS-STGNet. Additionally, their false alarm rates increased by 27.9% and 20.0%, AUC values decreased by 24.3% and 18.9%, and accuracy was reduced by 26.4% and 20.5%. These results emphasize the importance of extracting nonlinear correlations for traffic conflict prediction.

The introduction of deep learning methods significantly improved model performance. CNN and LSTM-CNN outperformed SVM and XGBoost across all metrics, demonstrating the importance of capturing spatial dependencies and temporal correlations in conflict prediction. However, deep learning methods relying on CNNs to capture spatial dependencies face a notable limitation: they cannot model spatial similarities in unconnected grid fields. This highlights the advantage of leveraging graph neural networks (GNNs), such as STGCN and STGAT, to model semantic spatial dependencies, further enhancing performance. For instance, under a 30% penetration rate, STGAT and STGCN 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% and 1.5%, and improved accuracy by 4.4% and 3.2%, respectively, compared to LSTM-CNN. These results underscore the advanced capability of utilizing the inherent graph structure of road networks to extract spatial dependencies related to conflict risks. GNNs are particularly well-suited for capturing complex relationships between road segments, 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 with spatial proximity and contextual similarity, better representing spatial dependencies.

Building on prior advancements in graph-based models, the proposed MS-STGNet model demonstrated robust performance across all penetration rate scenarios. For instance, under a 50% penetration rate, MS-STGNet outperformed the next-best models by 4.9% in recall, reduced false alarm rates by 3.3%, improved AUC by 5.3%, and increased accuracy by 3.3%. Notably, as shown in Table 4, MS-STGNet achieved a significant reduction in false alarm rates, with improvements of 23.9%, 24.0%, and 23.8% under 50%, 70%, and 90% penetration rates, respectively. This improvement can be attributed to the manifold similarity module, which reduces misjudgments in conflict-prone areas of traffic flow—a point further analyzed in subsequent sections.

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 ± 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.

Table 4
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**↑1.92% ± 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**↑9.09% ± 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**↑2.11% ± 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**↑3.01% ± 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**↑2.12% ± 0.017
30%	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**↑3.86% ± 0.010
	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**↑3.50% ± 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**↑3.58% ± 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**↑2.76% ± 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**↑1.84% ± 0.012
50%	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**↑5.92% ± 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**↑23.91% ± 0.017
	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**↑6.36% ± 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**↑3.85% ± 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**↑5.22% ± 0.011
70%	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**↑3.42% ± 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**↑24.00% ± 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**↑4.62% ± 0.015
	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**↑7.42% ± 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**↑3.86% ± 0.014
90%	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**↑1.24% ± 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**↑23.77% ± 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**↑3.37% ± 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**↑2.63% ± 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**↑2.86% ± 0.013

In Section 6.10 (Page 23 Line 28—29, Page 24 Line 1—17):

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

In Appendix C (Page 28 Line 7—13, Page 29):

Appendix C. Supplementary vehicle position–speed trajectories

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

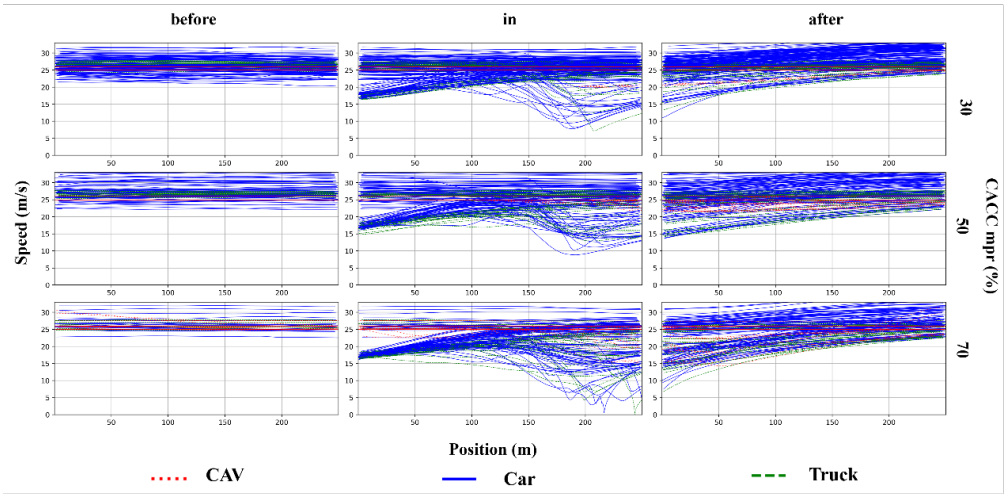

Fig. C1. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 3000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

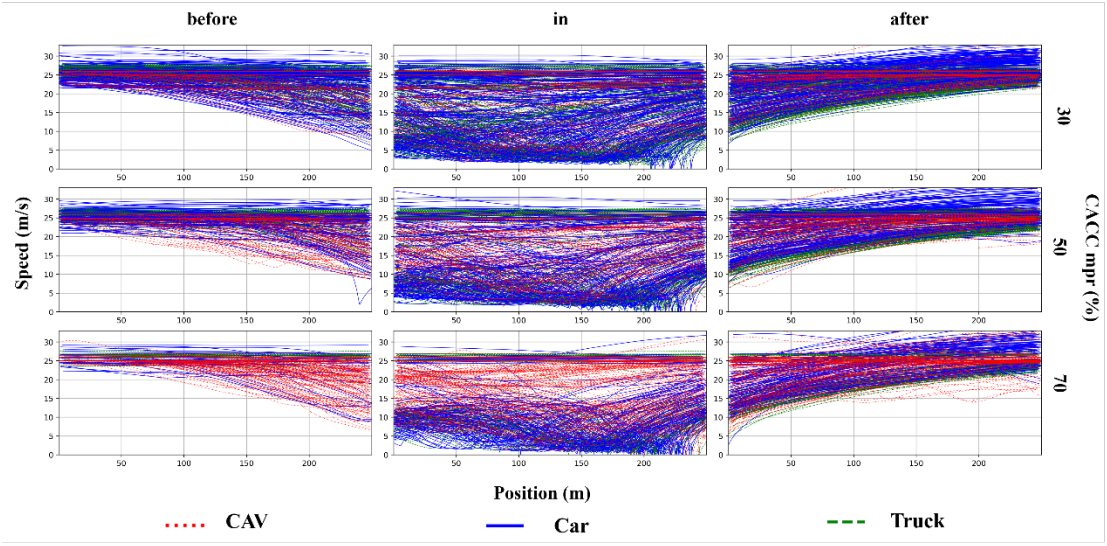

Fig. C2. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 6000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

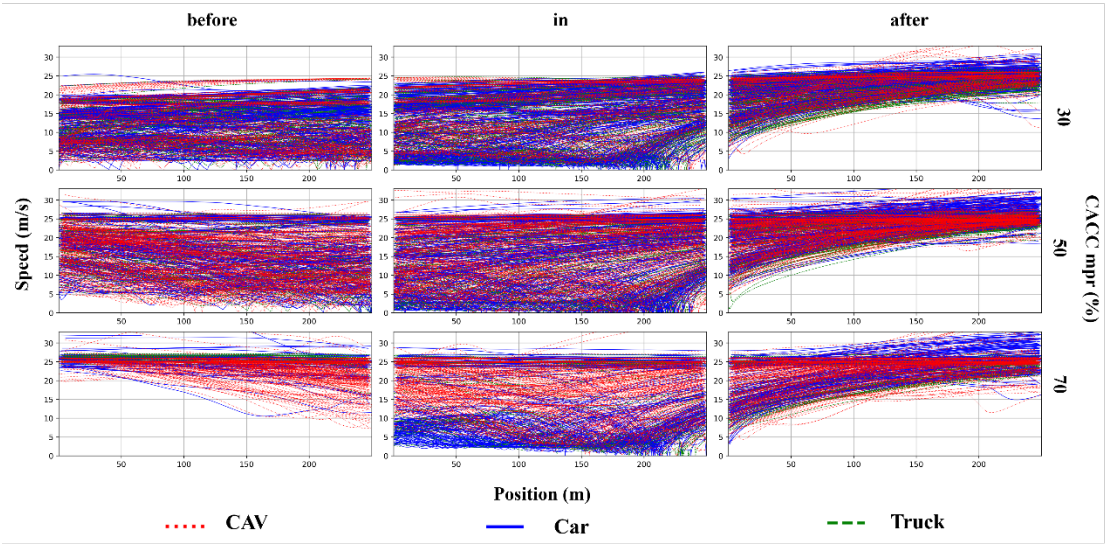

Fig. C3. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 9000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

In Section 7 (Page 25 Line 25, Page 26 Line 1—21):

The proposed framework has several practical implications. It can be embedded as a safety prediction component in CAV cloud management systems for freeway corridors and urban expressways, integrated into freeway traffic management centers and ramp control or variable speed limit systems to support mixed CAV–HDV operations, and used within regional expressway operation platforms to provide real-time conflict or crash risk warnings at bottlenecks and merging/diverging areas, thereby enhancing the safety management and visualization of freeway networks. The limitations of this study are summarized as follows: 1) The model is calibrated and evaluated in a microscopic simulation of a four-lane freeway

segment with motorized traffic only. Although the simulation is grounded in highD trajectory data, we do not yet validate MS-STGNet on large-scale field observations of mixed CAV–HDV traffic, and the direct transferability of the results to urban or suburban road networks with signalised intersections, pedestrians, and non-motorised vehicles is therefore limited. 2) The current experiments focus on a single 14 km corridor with specific demand patterns; additional facilities and more diverse demand scenarios would further test the generalizability of the framework. 3) The predefined manifold similarity matrix remains static over time, preventing the model from capturing previously unseen traffic state transitions unless it is retrained. 4) The proposed framework currently focuses on binary conflict/non-conflict prediction. Although the sigmoid activation in the output layer produces continuous risk 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.

In Section 6.6 (Page 18 Line 44—52, Page 19 Line 1—16):

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 5 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 5. 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.

Table 5
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**

Overall, these results indicate that MS-STGNet achieves superior predictive performance (as shown in Table 4) 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.

Comment 4:

Lastly, the manuscript could benefit from a more in-depth discussion of the potential limitations and future work. The proposed framework, while showing promising results, may face challenges in real-time implementation and scalability, which should be addressed in the manuscript.

Response to Comment 4:

We thank the reviewer for this helpful suggestion. In the revised manuscript, we have expanded the Conclusion section (Section 7) to more explicitly discuss the limitations and future work, including issues related to real-time implementation and scalability. Specifically, we now 1) clarify the scope of our current evaluation (a simulated 14 km multilane freeway calibrated on highD data, without urban/suburban networks or multimodal road users), 2) acknowledge methodological constraints such as the use of a static manifold-similarity matrix and a binary conflict/non-conflict output, and 3) outline concrete future directions, including validation on emerging mixed CAV–HDV field datasets, extension to more diverse networks

and contextual variables, development of online/adaptive manifold learning, and graded conflict-severity modelling. These additions better delimit the applicability of the present framework and directly address the reviewer's concerns about its potential real-time deployment and scalability. Below is the revised version:

Revised (Page 25 Line 25, Page 26 Line 1—21):

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

Comments from Reviewer #2:

Comment 1:

Please provide specific recent (<5-year) prior works supporting Section 2.3.

Response to Comment 1:

Thank you for this helpful suggestion.In the revised manuscript, we have added several recent references that explicitly exploit manifold-based representations for traffic state modelling and traffic safety analysis. Specifically, in Section 2.3, we now discuss:

- Su et al. (2020), who use a convolutional variational auto-encoder to extract low-dimensional manifold representations of daily urban traffic flow for clustering;
- Seo (2023), who applies Uniform Manifold Approximation and Projection (UMAP), a manifold-learning-based dimension reduction method, to visualize large-scale network traffic states and identify distinct congestion regimes;
- Liu et al. (2022), who incorporate manifold characteristics of traffic flow into a transfer-learning-based highway crash risk evaluation model and show that manifold features improve the discrimination between high- and low-risk traffic states.

[Liu, Q., Li, C., Jiang, H., Nie, S., & Chen, L. (2022). Transfer learning-based highway crash risk evaluation considering manifold characteristics of traffic flow. *Accident Analysis & Prevention*, *168*, 106598.

Seoa, T. (2023). Understanding large-scale traffic flow using model-based and data-driven dimension reduction: with COVID-19 and Olympic-Paralympic case study. *EU Science Hub*, 124.

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*(2).]

These additions strengthen the motivation of Section 2.3 and provide up-to-date support for the use of manifold similarity in our proposed MS-STGNet framework. Below is the revised version:

Revised (Page 5 Line 13—23):

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

Comment 2:

Does Fig. 8 present only a subset of the complete vehicle position-speed trajectories corresponding to the silhouette in Fig. 1? Please specify how lane position is defined and measured (e.g., lane index vs. lateral offset in meters)?

Response to Comment 2:

Thank you very much for this helpful comment. In the revised version, we have clarified in Section 6.9 that Fig. 8 shows the vehicle position–speed trajectories for a 250 m subsegment of the on-ramp merging area depicted in Fig. 1, selected because this location exhibits the most pronounced speed oscillations under high CAV penetration and thus better illustrates the speed separation between CAVs and HDVs. We also now explicitly state that the lateral axis in Fig. 8 represents the lateral offset in meters (rather than a discrete lane index), measured across the cross-section of the entire roadway. Since our analysis focuses on speed disturbances within the CAV and HDV systems rather than lane-by-lane differences, we deliberately avoid lane coloring and instead emphasize the contrast in speed fluctuation patterns between the two vehicle groups. Below is the revised version:

Revised (Page 23 Line 5—8):

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.

Comment 3:

The formula (15) seems wrong. The ReLU takes a single tensor as input, the comma notation is non-standard and ambiguous.

Response to Comment 3:

We thank the reviewer for pointing out the issue in Eq. (15). In the original manuscript, ReLU was mistakenly written as $\mathrm{ReLU}(\mathbf{M}_{\mathrm{lt}}, \mathbf{M}_{\mathrm{rt}})$ which is non-standard and indeed ambiguous, since ReLU should take a single tensor as input. We have revised the notation in the manuscript accordingly to avoid ambiguity, and in response to comments from other reviewers, this formula has been moved to Appendix B. Below is the revised version:

Revised (Page 28 Appendix B, B.5):

$$\widetilde{\mathbf{A}^{\star}} = \mathbf{I_N} + \mathrm{softmax}\big(\mathrm{ReLU}(\mathbf{M}_{\mathrm{lt}}\mathbf{M}_{\mathrm{rt}})\big)$$

Comment 4:

In formula (16), the placement of b_k outside the summation is confusing.

Response to Comment 4:

We appreciate the reviewer's insightful comment regarding the placement of b_k in Eq. (16). In the original manuscript, the bias term b_k was written outside the summation while still indexed by k, which is indeed confusing and mathematically ambiguous. As correctly pointed out by the reviewer, if the bias depends on k, it should appear inside the summation. In our implementation, each order has its own learnable bias associated with the corresponding weights. We have revised the notation in the manuscript accordingly to avoid ambiguity. After the article structure was readjusted, it is now Eq. (11). Below is the revised version:

Revised (Page 12 Eq.11):

$$\mathbf{Z}_{\mathbf{t}}^{*} = \sum_{k=1}^{K} \left((\mathbf{P}_{\mathbf{f}}^{*})^{k} \mathbf{X}^{\mathbf{t}} \mathbf{W}_{\mathbf{k},1} + (\mathbf{P}_{\mathbf{b}}^{*})^{k} \mathbf{X}^{\mathbf{t}} \mathbf{W}_{\mathbf{k},2} + \widetilde{\mathbf{A}_{\star}} \mathbf{X}^{\mathbf{t}} \mathbf{W}_{\mathbf{k},3} + \mathbf{b}_{\mathbf{k}} \right)$$

Comment 5:

There are several types of f in Eq. (18) and Eq. (19). What's the difference?

Response to Comment 5:

We thank the reviewer for pointing out the ambiguity regarding the different notations of f in Eqs. (18) and (19). In our model, all these symbols refer to learnable 1D convolution kernels, but they play slightly different roles:

- In Eq. (18), $\mathbf{f}^{l,k} \in \mathbb{R}^{C}$ denotes the generic 1D convolution kernel of the l-th TCN layer and the k-th output channel, and $\mathbf{f}^{l,k}(m)$ is its m-th coefficient. This equation defines the general form of a dilated causal convolution.
- In Eq. (19), $\mathbf{f}_k^{(0)}$ and $\mathbf{f}_k^{(1)}$ denote the convolution kernels used in the first and second dilated convolution within the k-th residual TCN block, respectively. They are specific instances of the generic kernel used in Eq. (18), and we use the superscripts (0) and (1) to distinguish the two convolutional layers inside a block.

To avoid confusion, we have revised the manuscript to explicitly clarify these roles. In particular, we now 1) add an explicit description of $\mathbf{f}^{l,k}$ below Eq. (18), and 2) clarify below Eq. (19) that $\mathbf{f}_k^{(0)}$ and $\mathbf{f}_k^{(1)}$ are the kernels of the two dilated convolutions in the k-th residual TCN block. We believe this resolves the ambiguity about the different types of f in these equations. Below is the revised version:

Revised (Page 12 Line 38—39, Page 13 Line 2—4):

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 the 1D convolution kernel of the l-th TCN layer and the k-th output channel.

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 k-th residual TCN block, respectively. They are specific instances of the generic kernel $\mathbf{f}^{l,k}$ defined in Eq. (18), but we use superscripts (0) and (1) to

distinguish the two convolutional layers within each block;

Comment 6:

In formula (24), the definitions of L_{LDAM} *and* Δ_y *are set consecutively without a separator, which can be misread as a single expression.*

Response to Comment 6:

We appreciate the reviewer's comment regarding the readability of Eq. (24). In the original manuscript, the definitions of $\mathcal{L}_{\mathrm{LDAM}}$ and Δ_y were typeset consecutively in the same display without any separator, which could indeed be misread as a single expression. To avoid this ambiguity, we have revised Eq. (24) to clearly separate the definitions. After the article structure was readjusted, it is now Eq. (19). Below is the revised version:

Revised (Page 13 Eq.19):

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

Comments from Reviewer #3:

Comment 1:

The idea of using a manifold-based similarity to construct adaptive graphs is well motivated and implemented cleanly. However, the formulation is conceptually close to existing adaptive adjacency or attention mechanisms in STGAT-type models. A clearer discussion of how the manifold metric differs in principle or in computational benefit would strengthen the methodological message. I think, this is the weakest point, also being positioned as a selling point.

Response to Comment 1:

We sincerely thank the reviewer for these thoughtful comments on the novelty of MS-STGNet and its relation to existing manifold-based traffic models and STGAT-type adaptive graph networks.

In the revised manuscript, we have clarified our contributions at both the problem and method levels. From the problem perspective, we emphasize that our primary goal is real-time conflict prediction in mixed CAV–HDV freeway traffic, a setting where existing work is still limited. To ensure robustness in this new safety-critical application, we deliberately build on mature components (residual CNN, TCN, spatiotemporal GNNs) while introducing a manifold-similarity graph as a physically meaningful prior rather than proposing an entirely new architecture for its own sake.

From the methodological perspective, we now explicitly distinguish our approach from prior manifold-learning studies and STGAT-type adaptive adjacency mechanisms. Section 2.3 has been expanded to include recent manifold-based traffic-flow and safety studies and to clarify that these works mainly use manifold embeddings for clustering, visualization, or as features in conventional models, without embedding manifold-based traffic-state similarity into a spatiotemporal GNN for online conflict prediction. In contrast, MS-STGNet integrates a pre-computed manifold similarity matrix, derived from historical traffic states, as an interpretable prior that constrains adaptive adjacency learning for mixed CAV–HDV conflicts.

We also revise Section 2.2 and Section 5.3.1 to clearly contrast our manifold-similarity graph with standard STGAT mechanisms: instead of learning adjacency solely from instantaneous node features at each time step, MS-STGNet initializes the graph from manifold distances computed offline and then performs lightweight adaptive refinement. This design ties the learned graph to physically meaningful traffic-state geometry while keeping the per-iteration computational cost comparable to standard STGNNs. Finally, in Section 6.5 we explicitly highlight that the manifold-similarity prior contributes to the reduction of false alarm rates and improved robustness compared with STGCN and STGAT, especially at medium-to-high CAV penetration rates.

We hope these revisions and clarifications make the unique contributions and methodological positioning of MS-STGNet more evident. Below is the revised version:

Revised:

In Section 1 (Page 2 Line 45—49, Page 3 Line 10—14):

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.

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.

In Section 2.2 (Page 4 Line 41—44):

In addition, existing spatiotemporal graph-based safety models typically define spatial dependencies through fixed adjacency matrices or adaptive attention mechanisms in the original feature space, and rarely exploit manifold-based traffic-state similarity as an explicit prior, particularly in mixed CAV–HDV traffic environments.

In Section 2.3 (Page 5 Line 26—29):

Additionally, few studies have attempted to integrate the concept of state transitions in manifold learning into deep learning frameworks, and, to the best of our knowledge, none has embedded manifold-based traffic-state similarity into a spatiotemporal graph neural network for real-time conflict prediction in mixed CAV–HDV traffic.

In Section 5.3.1 (Page 12 Line 1—10):

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.

In Section 6.5 (Page 18 Line 31—43):

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.

Comment 2:

The simulation framework and evaluation are comprehensive, very solid! The ablation study (Section 6.6) is exemplary and demonstrates the incremental gain from each module. Nevertheless, the validation remains limited to simulated data; any test on partially real or hybrid datasets (like MITRA, HDSim, etc.) would raise the practical impact substantially.

Response to Comment 2:

We sincerely appreciate this important comment. At the early stage of this study, our initial intention was indeed to develop and validate MS-STGNet directly on real-world mixed CAV–HDV data. However, after a thorough review of existing datasets, we found that currently available data sources cannot simultaneously meet the two core requirements of our problem: 1) truly mixed CAV–HDV traffic, and 2) long, spatially continuous freeway segments with macroscopic measurements (flow, speed, occupancy) suitable for segment-level conflict prediction over continuous time series.

On the one hand, classical trajectory datasets such as NGSIM and highD contain only human-driven vehicles and therefore do not match our target mixed-traffic scenario. Nevertheless, to ensure that our simulation is not based on purely theoretical assumptions, we calibrated the human-driven car-following model directly on highD freeway trajectories, so that HDV behaviour in the simulation reflects empirically observed acceleration, deceleration, and headway patterns rather than arbitrary parameter choices.

With respect to the specific datasets mentioned by the reviewer, we have carefully examined their suitability for our study:

MITRA is a high-resolution drone-based trajectory dataset on a 900m urban freeway segment in Milan, covering all traffic states with ramps and lane changes. It is highly valuable for microscopic analysis of human-driven behavior. However, the current release contains only HDVs (no explicit CAV logic). Thus, MiTra is well suited for refining microscopic models but does not directly provide the mixed CAV–HDV, segment-level macroscopic data needed for our conflict prediction framework.

[Chaudhari, A. A., Treiber, M., & Okhrin, O. (2025). Mitra: A drone-based trajectory data for an all-traffic-state inclusive freeway with ramps. Scientific Data, 12(1), 1174.]

HDSim is a cognitively inspired human-like traffic simulation framework that generates realistic microscopic scenario for testing autonomous driving systems. Conceptually, it is closer to our own SUMO/Plexe-based simulator than to a ready-made macroscopic dataset: it focuses on scene-level microscopic interaction rather than delivering continuous freeway-level traffic-state time series with conflict labels. Using HDSim would therefore require porting our entire controller, detection, surrogate safety (TTC/DRAC/DDR), and aggregation pipeline to another engine, which we regard as valuable for future cross-simulator robustness studies but beyond the scope of the present work.

[Li, W., Wu, H., Gao, H., Mao, B., Xu, F., & Zhong, S. (2025). LLM-based Human-like Traffic Simulation for Self-driving Tests. arXiv preprint arXiv:2508.16962.]

On the other hand, recent autonomous-vehicle datasets such as the Lyft Level 5 AV Dataset (Houston et al., 2021), nuScenes (Caesar et al., 2020), and the Waymo Open Dataset (Sun et al., 2020) do provide mixed traffic with AVs/CAVs, but their structure is not well suited to our macroscopic conflict-prediction task. As summarized in Table 1 of Hu et al. (2022), these AV datasets are organized into short trajectory segments: Waymo comprises 1,000 segments with a temporal resolution of 0.1 s and a typical segment length of 20 s; Lyft Level 5 contains 366 segments at 0.2 s resolution and 25–45 s duration; and nuScenes includes 1,000 segments with 0.5 s resolution and 20 s duration. These segments are collected from the viewpoint of individual AVs and are neither spatially contiguous along a single freeway facility nor temporally continuous over long periods. Hu et al. (2022) explicitly note that substantial preprocessing and reconstruction are required even to obtain usable car-following trajectories from these segment-based recordings, and that the resulting data remain fragmented in space and time for macroscopic analyses.

Table 1. Overview of three AV trajectory dataset.

Dataset	Number of segments	Resolution (s)	Length of each segment (s)
Waymo	1000	0.1	20
Lyft	366	0.2	25–45
nuScenes	1000	0.5	20

[Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... & Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11621-11631).

Houston, J., Zuidhof, G., Bergamini, L., Ye, Y., Chen, L., Jain, A., ... & Ondruska, P. (2021, October). One thousand and one hours: Self-driving motion prediction dataset. In Conference on Robot Learning (pp. 409-418). PMLR.

Hu, X., Zheng, Z., Chen, D., Zhang, X., & Sun, J. (2022). Processing, assessing, and enhancing the Waymo autonomous vehicle open dataset for driving behavior research. *Transportation Research Part C: Emerging Technologies*, *134*, 103490.

Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., ... & Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2446-2454).]

Similarly, Zhang et al. (2025) show that such AV trajectory datasets are particularly suitable for microscopic behaviour analysis and data-driven stochastic fundamental-diagram modelling, but they are inherently sparse in space and time and therefore not directly aligned

with macroscopic, segment-level modelling of traffic states over extended freeway sections. In our study, however, the prediction target is whether a given freeway segment will experience a conflict within a continuous time series, based on macroscopic indicators (flow, speed, occupancy) monitored along a 14 km stretch. This requires long, contiguous observations along one facility, which current AV datasets do not provide.

[Zhang, X., Yang, K., Sun, J., & Sun, J. (2025). Stochastic fundamental diagram modeling of mixed traffic flow: A data-driven approach. *Transportation Research Part C: Emerging Technologies*, *179*, 105279.]

These limitations are consistent with the broader challenges identified in recent reviews of machine-learning-based crash prediction. For example, Ali et al. (2024) point out that empirical safety studies for mixed CAV–HDV environments are still scarce, that most real-time crash and conflict prediction models are developed for conventional freeway or urban networks without CAVs, and that many studies necessarily rely on simulation or indirectly inferred data due to the lack of suitable field datasets. Against this background, we adopted a hybrid strategy that combines empirical calibration with large-scale simulation.

[Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis & Prevention*, *194*, 107378.]

Given these data limitations, we adopted a hybrid "real-trajectory calibration + simulation-based testing" strategy. At the microscopic level, we construct mixed CAV–HDV traffic with multiple penetration rates (10%, 30%, 50%, 70%, and 90%) and generate trajectories at 0.2s resolution. Using these trajectories, we compute widely used surrogate safety measures TTC, DRAC, and DDR—to identify both longitudinal and lateral conflicts. At the macroscopic level, we aggregate the simulated data along a 14 km four-lane freeway with ramps and extract continuous time series of flow, speed, and occupancy for each segment, paired with the conflict/non-conflict labels derived from TTC/DRAC/DDR. This design allows us to study conflict prediction as a segment-level time-series classification problem under a wide range of demand levels and CAV penetration rates, while keeping driver behavior anchored in real freeway observations and defining conflicts via physically interpretable criteria.

We fully acknowledge that this hybrid validation cannot replace large-scale field testing on real mixed CAV–HDV networks. In the revised Conclusion, we now explicitly state that 1) the current evaluation is conducted in a calibrated freeway simulation, 2) the direct transferability to urban or suburban networks is therefore limited, and 3) the lack of suitable real-world mixed-traffic datasets is a key limitation of the present work. We also clarify that, once continuous macroscopic observations of mixed CAV–HDV traffic over long freeway segments become available, we plan to retrain and evaluate MS-STGNet on those data and to systematically compare its performance with other state-of-the-art safety models in real-world environments. Below is the revised version:

Revised (Page 25 Line 25, Page 26 Line 1—21):

The proposed framework has several practical implications. It can be embedded as a safety prediction component in CAV cloud management systems for freeway corridors and urban expressways, integrated into freeway traffic management centers and ramp control or variable speed limit systems to support mixed CAV–HDV operations, and used within regional

expressway operation platforms to provide real-time conflict or crash risk warnings at bottlenecks and merging/diverging areas, thereby enhancing the safety management and visualization of freeway networks. The limitations of this study are summarized as follows: 1) The model is calibrated and evaluated in a microscopic simulation of a four-lane freeway segment with motorized traffic only. Although the simulation is grounded in highD trajectory data, we do not yet validate MS-STGNet on large-scale field observations of mixed CAV–HDV traffic, and the direct transferability of the results to urban or suburban road networks with signalized intersections, pedestrians, and non-motorized vehicles is therefore limited. 2) The current experiments focus on a single 14 km corridor with specific demand patterns; additional facilities and more diverse demand scenarios would further test the generalizability of the framework. 3) The predefined manifold similarity matrix remains static over time, preventing the model from capturing previously unseen traffic state transitions unless it is retrained. 4) The proposed framework currently focuses on binary conflict/non-conflict prediction. Although the sigmoid activation in the output layer produces continuous risk 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.

Comment 3:

Improvements over STGCN and STGAT are numerically evident but modest. Reporting standard deviations or statistical significance across runs would help to substantiate the claimed stability.

Response to Comment 3:

We thank the reviewer for this helpful comment. In the original manuscript, our notion of "stability" mainly referred to performance consistency across different CAV penetration rates, as highlighted by the persistent reduction in false alarm rates compared with STGCN and STGAT (See Page 18 Line 21—24). We agree that cross-run variability and statistical significance should also be evaluated.

In the revised version, we retrain all models five times with different random seeds. Section 6.2 has been updated to describe this protocol, and Table 4 now reports each metric as mean ± standard deviation over these five runs. Because the task is a binary conflict/non-conflict prediction on a large dataset, the standard deviations are generally small across all models. Nonetheless, MS-STGNet consistently achieves higher recall and AUC and lower false alarm rates than STGCN and STGAT under all penetration scenarios, with improvements clearly exceeding the corresponding standard deviations in most cases.

We further conducted paired t-tests across the five runs and found that, for all penetration rates and all reported metrics, the improvements of MS-STGNet over STGCN and STGAT are statistically significant at the 5% level ($p < 0.05$). Section 6.5 has been revised accordingly to highlight both the cross-condition stability and the cross-run robustness of MS-STGNet. Below is the revised version:

Revised:

In Section 6.2 (Page 14 Line 34—35):

To reduce the impact of randomness and evaluate the stability of each method, all models are trained and evaluated five times with different random seeds orders.

In Section 6.5 (Page 16 Line 38—43, Page 17, Page 18 Line 1—43):

To further assess cross-run stability, each entry in Table 4 is reported as the mean ± 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 ($p < 0.05$).

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 23% and 17.8% lower, respectively, than those of the proposed MS-STGNet. Additionally, their false alarm rates increased by 27.9% and 20.0%, AUC values decreased by 24.3% and 18.9%, and accuracy was reduced by 26.4% and 20.5%. These results emphasize the importance of extracting nonlinear correlations for traffic conflict prediction.

The introduction of deep learning methods significantly improved model performance. CNN and LSTM-CNN outperformed SVM and XGBoost across all metrics, demonstrating the importance of capturing spatial dependencies and temporal correlations in conflict prediction. However, deep learning methods relying on CNNs to capture spatial dependencies face a notable limitation: they cannot model spatial similarities in unconnected grid fields. This highlights the advantage of leveraging graph neural networks (GNNs), such as STGCN and STGAT, to model semantic spatial dependencies, further enhancing performance. For instance, under a 30% penetration rate, STGAT and STGCN 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% and 1.5%, and improved accuracy by 4.4% and 3.2%, respectively, compared to LSTM-CNN. These results underscore the advanced capability of utilizing the inherent graph structure of road networks to extract spatial dependencies related to conflict risks. GNNs are particularly well-suited for capturing complex relationships between road segments, 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 with spatial proximity and contextual similarity, better representing spatial dependencies.

Building on prior advancements in graph-based models, the proposed MS-STGNet model demonstrated robust performance across all penetration rate scenarios. For instance, under a 50% penetration rate, MS-STGNet outperformed the next-best models by 4.9% in recall,

reduced false alarm rates by 3.3%, improved AUC by 5.3%, and increased accuracy by 3.3%. Notably, as shown in Table 4, MS-STGNet achieved a significant reduction in false alarm rates, with improvements of 23.9%, 24.0%, and 23.8% under 50%, 70%, and 90% penetration rates, respectively. This improvement can be attributed to the manifold similarity module, which reduces misjudgments in conflict-prone areas of traffic flow—a point further analyzed in subsequent sections.

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 ± 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.

Table 4
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**↑1.92% ±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**↑9.09% ±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**↑2.11% ±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**↑3.01% ±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**↑2.12% ±0.017
30%	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**↑3.86% ±0.010
	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**↑3.50% ±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**↑3.58% ±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**↑2.76% ±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**↑1.84% ±0.012
50%	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**↑5.92% ±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**↑23.91% ±0.017
	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**↑6.36% ±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**↑3.85% ±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**↑5.22% ±0.011
70%	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**↑3.42% ±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**↑24.00% ±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**↑4.62% ±0.015
	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**↑7.42% ±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**↑3.86% ±0.014
90%	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**↑1.24% ±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**↑23.77% ±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**↑3.37% ±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**↑2.63% ±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**↑2.86% ±0.013

Comment 4:

As the model will be of interest to traffic-safety practitioners, visualization of the learned manifold-similarity matrices or attention weights could make the results more transparent and explain which spatial-temporal interactions dominate conflict risk.

Response to Comment 4:

We appreciate this constructive suggestion. We agree that visualizing the learned spatial relationships can help practitioners better understand which segment pairs play a dominant role in conflict risk. In response, we have added Appendix A, where we present the learned manifold-similarity matrices for flow, occupancy, and speed. Each matrix is large, so we show the top-left 5×5 block together with the last row and last column, using ellipses to indicate continuation. This tabular layout makes it easier to identify which mainline and ramp segments exhibit strong manifold-based similarity and thus exert greater influence on the learned spatial interactions in MS-STGNet. Below is the revised version:

Revised (Page 26 Line 27—32, Page 27 Line 1—3):

Appendix A. Visualization of the learned manifold-similarity matrices

Appendix A presents the learned manifold-similarity matrices for flow, occupancy, and speed, denoted by $\boldsymbol{Matrices}^{(flow)}$, $\boldsymbol{Matrices}^{(occupancy)}$, *and* $\boldsymbol{Matrices}^{(speed)}$, *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.*

$$\mathbf{Matrices}^{(\mathrm{flow})} = \begin{bmatrix} 1.000 & 0.277 & 0.268 & 0.274 & 0.745 & \cdots & 0.686 \\ 0.277 & 1.000 & 0.701 & 0.285 & 0.689 & \cdots & 0.279 \\ 0.268 & 0.701 & 1.000 & 0.707 & 0.693 & \cdots & 0.699 \\ 0.274 & 0.285 & 0.707 & 1.000 & 0.688 & \cdots & 0.759 \\ 0.745 & 0.689 & 0.693 & 0.688 & 1.000 & \cdots & 0.696 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.686 & 0.279 & 0.699 & 0.759 & 0.696 & \cdots & 1.000 \end{bmatrix}, \quad \mathbf{Matrices}^{(\mathrm{flow})} \in \mathbb{R}^{108\times108}$$

$$\mathbf{Matrices}^{(\mathrm{occupancy})} = \begin{bmatrix} 1.000 & 0.365 & 0.316 & 0.276 & 0.353 & \cdots & 0.250 \\ 0.365 & 1.000 & 0.367 & 0.327 & 0.302 & \cdots & 0.314 \\ 0.316 & 0.367 & 1.000 & 0.390 & 0.283 & \cdots & 0.358 \\ 0.276 & 0.327 & 0.390 & 1.000 & 0.237 & \cdots & 0.499 \\ 0.353 & 0.302 & 0.283 & 0.237 & 1.000 & \cdots & 0.016 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.250 & 0.314 & 0.358 & 0.499 & 0.016 & \cdots & 1.000 \end{bmatrix}, \quad \mathbf{Matrices}^{(\mathrm{occupancy})} \in \mathbb{R}^{108\times108}$$

$$\mathbf{Matrices}^{(\text{speed})} = \begin{bmatrix} 1.000 & 0.602 & 0.491 & 0.523 & 0.600 & \cdots & 0.396 \\ 0.602 & 1.000 & 0.537 & 0.566 & 0.561 & \cdots & 0.441 \\ 0.491 & 0.537 & 1.000 & 0.503 & 0.450 & \cdots & 0.402 \\ 0.523 & 0.566 & 0.503 & 1.000 & 0.474 & \cdots & 0.512 \\ 0.600 & 0.561 & 0.450 & 0.474 & 1.000 & \cdots & 0.344 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.396 & 0.441 & 0.402 & 0.512 & 0.344 & \cdots & 1.000 \end{bmatrix}, \quad \mathbf{Matrices}^{(\text{speed})} \in \mathbb{R}^{108\times108}$$

Comment 5:

The paper is well written but somewhat lengthy in the literature review. Sections 2.1-2.3 could be tightened without loss of content.

Response to Comment 5:

We thank the reviewer for pointing out that the literature review was somewhat lengthy. We agree that it can be streamlined without losing essential context. Accordingly, we have compressed Sections 2.1–2.3 by 1) merging overlapping descriptions of mixed-traffic simulation and CAV/HDV modelling, 2) summarizing traditional statistical and deep-learning-based safety models more concisely, and 3) grouping early manifold-learning applications into a shorter, synthesized paragraph while retaining detailed discussion of the most relevant recent work. These changes reduce redundancy and improve readability, while still providing sufficient background to motivate MS-STGNet and the proposed manifold-similarity module. Below is the revised version:

Revised:

In Section 2.1 (Page 3 Line 23—43):

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

A notable gap in these studies is the insufficient distinction between CAVs and HDVs,

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

In Section 2.2 (Page 3 Line 45—47, Page 4 Line 1—34):

Predicting traffic accidents has long been a critical topic in mobility management research. Early studies predominantly employed traditional statistical methods such as regression models (Caliendo et al., 2007; Bergel-Hayat et al., 2013), Bayesian networks (Martin et al., 2009; Hossain and Muromachi, 2012), and tree-based algorithms (Wang et al., 2010; Lin et al., 2015). These approaches provided initial insights into accident patterns, particularly in small geographical areas, but their ability to capture nonlinear relationships and dynamic dependencies between road segments was limited (Zhang et al., 2014a). Moreover, they often analyzed accident data in isolation, neglecting critical interdependencies between locations, which restricted their applicability to citywide analyses with large datasets (Wang et al., 2021).

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

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

zero inflation by under sampling to balance risky and non-risky segments.

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

In Section 2.3 (Page 5 Line 4—12):

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

Comment 6:

Would it be possible to compare with some (Safe-)RL model?

Response to Comment 6:

We appreciate the reviewer's suggestion to relate our work to (Safe-)RL approaches. We agree that Safe-RL is highly relevant for designing CAV control policies that explicitly account for safety.

In our setting, the task is to estimate, given observed mixed CAV–HDV traffic states, the probability of near-future traffic conflicts at each segment and time step. This is a supervised spatiotemporal prediction problem, for which it is natural to compare against other prediction-based baselines such as STGCN and STGAT. In contrast, (Safe-)RL methods typically learn a control policy (e.g., longitudinal or lane-change decisions, ramp metering, or signal control) by interacting with a simulation environment and optimizing a long-term reward that may include safety-related terms. Their outputs are control actions rather than explicit conflict probabilities, and their performance is evaluated in terms of overall system-level outcomes under the learned

policy.

Because of this fundamental difference in problem formulation and evaluation, a direct, quantitative baseline comparison between MS-STGNet and a Safe-RL controller would not be straightforward or necessarily meaningful: it would require designing a complete CAV control framework, specifying actions and reward functions, and coupling it with a microscopic simulation environment, which goes beyond the scope of the present study. We appreciate the reviewer's suggestion and will consider integrating MS-STGNet into a Safe-RL-based control framework as an important direction for future research.

Comment 7:

Some equations (11-15) may move to an appendix.

Response to Comment 7:

We thank the reviewer for this helpful suggestion. We agree that the detailed mathematical expressions originally given in Eqs. (11)–(15) can be moved to an appendix to improve the readability of the main text. In the revised manuscript, the Gaussian-kernel similarity and AICc-based bandwidth selection from Section 5.3.1 have been relocated to Appendix B.1 (Eqs. (B.1)–(B.2)), and the SVD-based initialization and adaptive adjacency formulation from Section 5.3.2 have been moved to Appendix B.2 (Eqs. (B.3)–(B.5)). The main text now contains concise, high-level descriptions of these steps with explicit references to Appendix B for readers interested in the full derivations. Below is the revised version:

Revised:

In Section 5.3.1 (Page 11 Line 25—30):

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)).

In Section 5.3.2 (Page 12 Line 12—19):

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, $\boldsymbol{M}_{lt}$ and $\boldsymbol{M}_{rt}$, 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 $\widetilde{\boldsymbol{A}^{\star}}$ that balances flexibility and interpretability. The complete mathematical formulation of this SVD-based initialization and adaptive update, including the definitions of $\boldsymbol{M}_{lt}$, $\boldsymbol{M}_{rt}$, and $\widetilde{\boldsymbol{A}^{\star}}$, is given in Appendix B (Eqs. (B.3)–(B.5)).

In Appendix B (Page 27 Line 4—29, Page 28 Line 1—4):

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

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:*

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

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

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

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

B.2. SVD-based initialization and adaptive adjacency

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 (Guo et al., 2015; Zou et al., 2024), and $\mathbf{A}^*$ *can be expressed as the product of three distinct matrices, as follows:*

$$\mathbf{A}^* = \mathbf{U}^*\mathbf{\Sigma}^*\mathbf{V}^{*\top} \tag{B.3}$$

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

$$\mathbf{A}^\star = ReLU(\mathbf{M}_{lt}\mathbf{M}_{rt}) \tag{B.4}$$

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

$$\tilde{\mathbf{A}}^\star = \mathbf{I}_N + \mathrm{softmax}\left(\mathrm{ReLU}(\mathbf{M}_{lt}\mathbf{M}_{rt})\right) \tag{B.5}$$

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

Comment 8:

English is fluent; minor stylistic tightening would suffice

Response to Comment 8:

We sincerely appreciate the reviewer's positive assessment of the overall English quality. Following your suggestion, we have carefully re-read the entire manuscript and carried out minor stylistic refinements throughout, including polishing sentence structures, improving wording for clarity and conciseness, and harmonizing terminology and notation. We hope that the revised version reads more smoothly and meets the journal's language standards.

Comments from Reviewer #4:

Comment 1:

In Section 4.1, please explain the reasons for selecting a 14-kilometer four-lane highway for simulation.

Response to Comment 1:

We thank the reviewer for this helpful comment. Our simulation setup is closely linked to the calibration of the Enhanced Intelligent Driver Model (EIDM), whose parameters are estimated from the highD dataset of naturalistic trajectories on multilane freeways. For this reason, it is most consistent and reasonable to adopt a freeway scenario rather than urban or suburban roads. The choice of a 14 km four-lane segment reflects a trade-off between realism and computational efficiency: it provides sufficient distance for vehicles to accelerate, cruise, and interact so that stable traffic states and realistic conflicts can develop without being dominated by boundary effects. Several on-ramps and off-ramps are embedded along the segment to mimic real freeway operations and generate complex merging/diverging interactions that are important sources of conflicts in mixed CAV–HDV traffic. We have added a concise explanation of these design choices in Section 4.1 of the revised manuscript. Below is the revised version:

Revised (Page 6 Line 17—23, Page 7 Line 1—2):

This choice is consistent with the calibration of 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).

Comment 2:

The paper's research scenario is limited to four-lane highways and does not cover more common urban roads, suburban roads, etc. The authors should consider whether adding pedestrians, non-motorized vehicles, and other elements to the model would significantly impact its performance. Incorporating these elements would enhance the model's robustness in real-world traffic environments. If the authors are unable to include them, this limitation should be explicitly stated in the paper.

Response to Comment 2:

We appreciate this important comment. As clarified in the revised Section 4.1, the present framework is built on an EIDM car-following model calibrated using the highD dataset, which

contains naturalistic trajectories on multilane freeways (See Page 6 Line 16—19). We therefore intentionally choose a freeway segment as the primary simulation environment, since this is the most consistent setting for applying the calibrated model and analysing mixed CAV–HDV interactions at higher speeds.

Our focus in this paper is on segment-level vehicle–vehicle conflict prediction from a macroscopic perspective, rather than on microscopic interactions with pedestrians or non-motorized vehicles. Incorporating these road users would require additional behaviour models and data, and would effectively lead to a different research problem. For example, recent studies on pedestrian–vehicle or motorcycle–pedestrian interactions use dedicated microscopic and reinforcement-learning frameworks tailored to those interactions, rather than freeway segment-level risk prediction (Nasernejad et al., 2021; Lanzaro et al., 2022).

[Lanzaro, G., Sayed, T., & Alsaleh, R. (2022). Can motorcyclist behavior in traffic conflicts be modeled? A deep reinforcement learning approach for motorcycle-pedestrian interactions. *Transportmetrica B: transport dynamics*, *10*(1), 396-420.\

Nasernejad, P., Sayed, T., & Alsaleh, R. (2021). Modeling pedestrian behavior in pedestrian-vehicle near misses: A continuous Gaussian Process Inverse Reinforcement Learning (GP-IRL) approach. *Accident Analysis & Prevention*, 161, 106355.]

At the same time, we agree that additional contextual factors (e.g., weather, pavement friction, POIs) are relevant for real-world conflict risk. These variables are more aligned with our macroscopic setting but cannot be fully and reliably represented in the current simulation. We therefore focus on core traffic-flow variables—time of day, flow, speed, and occupancy—as inputs, which we regard as a simple yet effective choice for freeway mixed-traffic scenarios.

Following the reviewer's suggestion, we have made this limitation explicit in the Conclusion. The revised text now 1) states that the model is calibrated and evaluated only on a four-lane freeway with motorized traffic, and that direct transferability to urban or suburban networks with pedestrians and non-motorized vehicles is limited, and 2) refines the practical implications to focus on freeway and urban-expressway safety management, which better matches the simulation setting.Below is the revised version:

Revised (Page 25 Line 25, Page 26 Line 1—21):

The proposed framework has several practical implications. It can be embedded as a safety prediction component in CAV cloud management systems for freeway corridors and urban expressways, integrated into freeway traffic management centers and ramp control or variable speed limit systems to support mixed CAV–HDV operations, and used within regional expressway operation platforms to provide real-time conflict or crash risk warnings at bottlenecks and merging/diverging areas, thereby enhancing the safety management and visualization of freeway networks. The limitations of this study are summarized as follows: 1) The model is calibrated and evaluated in a microscopic simulation of a four-lane freeway segment with motorized traffic only. Although the simulation is grounded in highD trajectory data, we do not yet validate MS-STGNet on large-scale field observations of mixed CAV–HDV traffic, and the direct transferability of the results to urban or suburban road networks with signalised intersections, pedestrians, and non-motorised vehicles is therefore limited. 2) The current experiments focus on a single 14 km corridor with specific demand patterns; additional facilities and more diverse demand scenarios would further test the generalizability of the

framework. 3) The predefined manifold similarity matrix remains static over time, preventing the model from capturing previously unseen traffic state transitions unless it is retrained. 4) The proposed framework currently focuses on binary conflict/non-conflict prediction. Although the sigmoid activation in the output layer produces continuous risk 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.

Comment 3:

MS-STGNet does not explicitly specify which traffic features contribute most significantly to conflict prediction. I suggest the authors add this information to Section 6.

Response to Comment 3:

We thank the reviewer for this insightful comment. All models in our study, including MS-STGNet, take as inputs macroscopic traffic features (flow, mean speed, occupancy) together with the CAV penetration rate, which are processed jointly within the spatiotemporal graph network. We note that Section 6.5 already compares model performance across five CAV penetration scenarios and highlights penetration rate as a key scenario variable in mixed CAV–HDV traffic (See Page 18 Line 21—24); in the revised version we make this role more explicit and connect it to the feature-oriented discussion in Section 6.10.

Section 6.10 and Appendix C have been expanded to analyse how traffic volume and resulting speed-dispersion patterns affect conflict risk and model behaviour. Using 500 hours of simulation with low/medium/high volume levels, we show that higher volumes produce more pronounced speed oscillations along the segment and substantially higher empirical conflict rates. The supplementary trajectory plots (Figures C1–C3) highlight that HDVs and heavy vehicles generate larger and more frequent speed fluctuations, while increasing CAV penetration smooths these perturbations. Combined with the segment-level risk profiles in Fig. 9, these results indicate that CAV penetration rate, traffic volume, and speed separation are among the most influential traffic features for conflict prediction, and that MS-STGNet aligns its predicted risk with these structures more reliably than STGCN and STGAT.

We also note that deep spatiotemporal graph networks do not natively provide scalar feature-importance scores as in tree-based models; a full attribution analysis (e.g., SHAP or GNN-specific explainability) is thus left for future work. Below is the revised version:

Revised:

In Section 6.10 (Page 23 Line 28—29, Page 24 Line 1—17):

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

In Appendix C (Page 28 Line 7—13, Page 29):

Appendix C. Supplementary vehicle position–speed trajectories

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

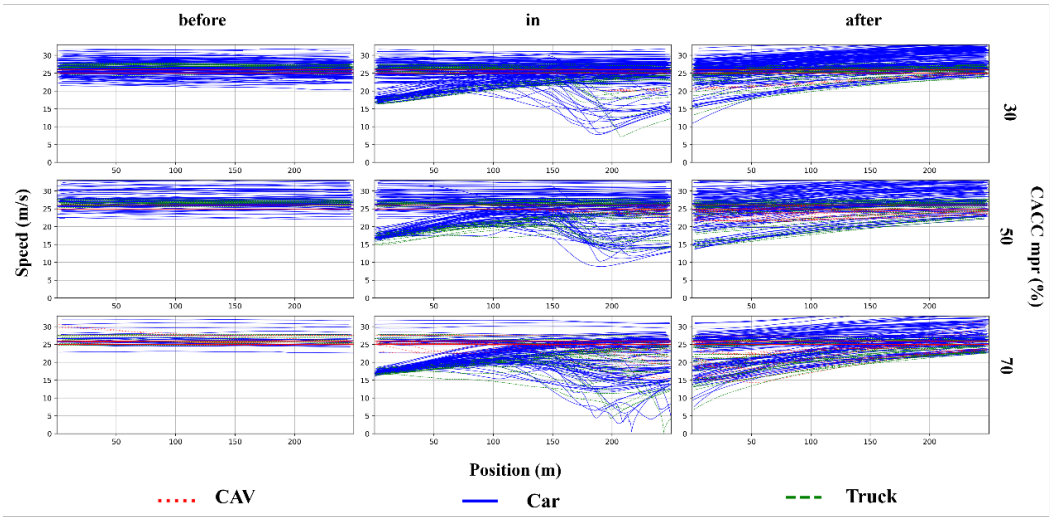

Fig. C1. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 3000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

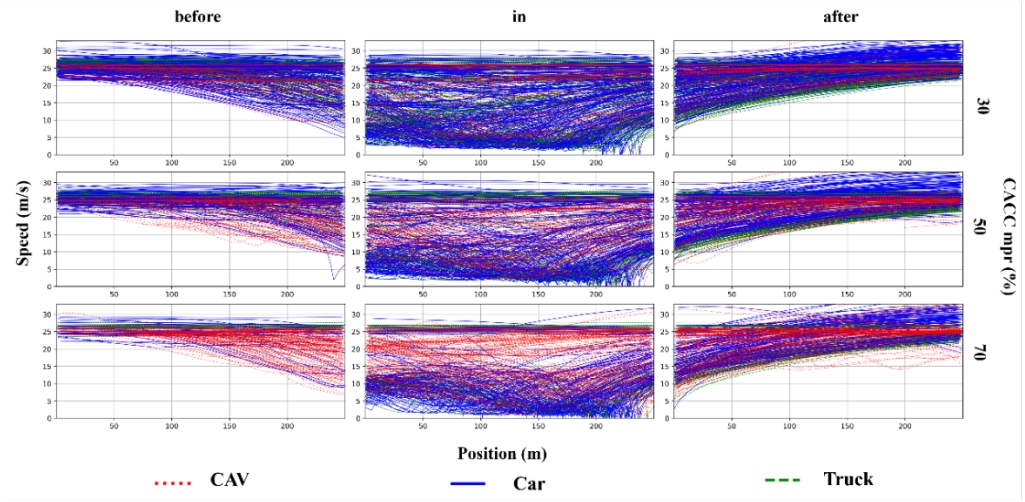

Fig. C2. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 6000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

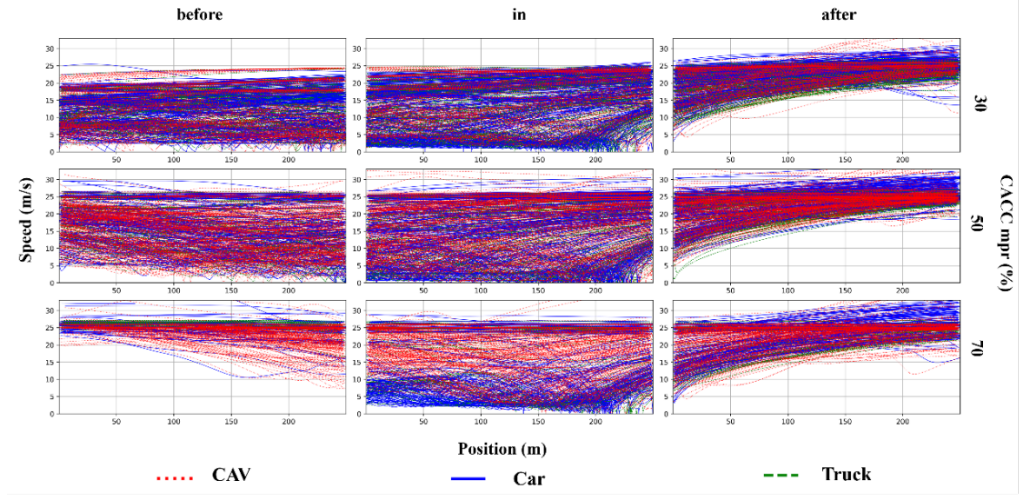

Fig. C3. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 9000

vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

Comment 4:

The finding of insufficient dialogue with existing research and it is suggested that the authors strengthen the discussion of the current study's results with the existing literature after obtaining the results in Section 6.5.

Response to Comment 4:

We appreciate this constructive suggestion and agree that the empirical findings should be more clearly related to existing work. In the revised manuscript, we have strengthened the end of Section 6.5 in two respects: 1) We explicitly note that our results are consistent with previous studies showing that spatiotemporal graph models such as STGCN and STGAT generally outperform traditional machine-learning and sequence models for traffic prediction tasks. 2) We clarify that MS-STGNet extends these STGNN approaches by embedding a manifold-based state-similarity prior into the adaptive graph learning process. We also connect our findings to recent manifold-learning studies in traffic flow and safety analysis, emphasizing that our framework complements this line of work by demonstrating how manifold-informed similarity can be integrated into a deep spatiotemporal graph network for mixed CAV–HDV conflict prediction. We hope this enhanced discussion in Section 6.5 addresses the reviewer's concern about insufficient dialogue with existing research. Below is the revised version:

Revised (Page 18 Line 31—43):

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.

Comment 5:

The authors' current research focuses solely on dichotomous predictions of conflict presence or absence, without addressing graded predictions of conflict severity (e.g., minor scrape risk versus severe collision risk). This is also a limitation of the study.

Response to Comment 5:

We are grateful for this thoughtful comment and agree that modelling graded conflict severity would further enhance practical relevance. In the current study, we formulate the problem as binary conflict/non-conflict prediction because our simulation provides reliable labels for conflict occurrence, but not well-validated categorical labels for different severity levels.

We would like to clarify, however, that the output of MS-STGNet is a continuous risk score in [0,1] obtained via a sigmoid activation. The threshold of 0.5 is used only for computing classification metrics (See Page 16 Line 2—5); the raw probabilities are used directly to construct the segment-level risk profiles in Section 6.10 and can already be interpreted as a continuous measure of conflict risk intensity.

Nevertheless, we fully acknowledge that we do not explicitly design or evaluate a multi-level severity model in this work. In the revised Conclusion, we have (i) added this point to the list of limitations, stating that the framework currently focuses on binary prediction, and (ii) explicitly highlighted, in the future work paragraph, the extension from binary conflict detection to graded or ordinal conflict severity prediction by combining the continuous risk scores with appropriate severity labels (e.g., minor, moderate, severe). Below is the revised version:

Revised (Page 25 Line 25, Page 26 Line 1—21):

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

Comments from Reviewer #5:

Comment 1:

The evaluation is conducted using simulated traffic datasets generated with SUMO. Although the authors claim to have calibrated their simulation models (EIDM and CACC) using real-world datasets, the conflict data analyzed are synthetic. It is recommended to validate the proposed MS-STGNet model with a real-world traffic dataset (such as highD, NGSIM, etc.) where traffic conflicts can be extracted or inferred.

Response to Comment 1:

We sincerely appreciate this important comment. At the early stage of this study, our initial intention was indeed to develop and validate MS-STGNet directly on real-world mixed CAV–HDV data. However, after a thorough review of existing datasets, we found that currently available data sources cannot simultaneously meet the two core requirements of our problem: 1) truly mixed CAV–HDV traffic, and 2) long, spatially continuous freeway segments with macroscopic measurements (flow, speed, occupancy) suitable for segment-level conflict prediction over continuous time series.

On the one hand, classical trajectory datasets such as NGSIM and highD contain only human-driven vehicles and therefore do not match our target mixed-traffic scenario. Nevertheless, to ensure that our simulation is not based on purely theoretical assumptions, we calibrated the human-driven car-following model directly on highD freeway trajectories, so that HDV behaviour in the simulation reflects empirically observed acceleration, deceleration, and headway patterns rather than arbitrary parameter choices.

On the other hand, recent autonomous-vehicle datasets such as the Lyft Level 5 AV Dataset (Houston et al., 2021), nuScenes (Caesar et al., 2020), and the Waymo Open Dataset (Sun et al., 2020) do provide mixed traffic with AVs/CAVs, but their structure is not well suited to our macroscopic conflict-prediction task. As summarized in Table 1 of Hu et al. (2022), these AV datasets are organized into short trajectory segments: Waymo comprises 1,000 segments with a temporal resolution of 0.1 s and a typical segment length of 20 s; Lyft Level 5 contains 366 segments at 0.2 s resolution and 25–45 s duration; and nuScenes includes 1,000 segments with 0.5 s resolution and 20 s duration. These segments are collected from the viewpoint of individual AVs and are neither spatially contiguous along a single freeway facility nor temporally continuous over long periods. Hu et al. (2022) explicitly note that substantial preprocessing and reconstruction are required even to obtain usable car-following trajectories from these segment-based recordings, and that the resulting data remain fragmented in space and time for macroscopic analyses.

Table 1. Overview of three AV trajectory dataset.

Dataset	Number of segments	Resolution (s)	Length of each segment (s)
Waymo	1000	0.1	20
Lyft	366	0.2	25–45
nuScenes	1000	0.5	20

[Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... & Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer*

vision and pattern recognition (pp. 11621-11631).

Houston, J., Zuidhof, G., Bergamini, L., Ye, Y., Chen, L., Jain, A., ... & Ondruska, P. (2021, October). One thousand and one hours: Self-driving motion prediction dataset. In Conference on Robot Learning (pp. 409-418). PMLR.

Hu, X., Zheng, Z., Chen, D., Zhang, X., & Sun, J. (2022). Processing, assessing, and enhancing the Waymo autonomous vehicle open dataset for driving behavior research. *Transportation Research Part C: Emerging Technologies*, *134*, 103490.

Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., ... & Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2446-2454).]

Similarly, Zhang et al. (2025) show that such AV trajectory datasets are particularly suitable for microscopic behaviour analysis and data-driven stochastic fundamental-diagram modelling, but they are inherently sparse in space and time and therefore not directly aligned with macroscopic, segment-level modelling of traffic states over extended freeway sections. In our study, however, the prediction target is whether a given freeway segment will experience a conflict within a continuous time series, based on macroscopic indicators (flow, speed, occupancy) monitored along a 14 km stretch. This requires long, contiguous observations along one facility, which current AV datasets do not provide.

[Zhang, X., Yang, K., Sun, J., & Sun, J. (2025). Stochastic fundamental diagram modeling of mixed traffic flow: A data-driven approach. *Transportation Research Part C: Emerging Technologies*, *179*, 105279.]

These limitations are consistent with the broader challenges identified in recent reviews of machine-learning-based crash prediction. For example, Ali et al. (2024) point out that empirical safety studies for mixed CAV–HDV environments are still scarce, that most real-time crash and conflict prediction models are developed for conventional freeway or urban networks without CAVs, and that many studies necessarily rely on simulation or indirectly inferred data due to the lack of suitable field datasets. Against this background, we adopted a hybrid strategy that combines empirical calibration with large-scale simulation.

[Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis & Prevention*, *194*, 107378.]

Given these data limitations, we adopted a hybrid "real-trajectory calibration + simulation-based testing" strategy. At the microscopic level, we construct mixed CAV–HDV traffic with multiple penetration rates (10%, 30%, 50%, 70%, and 90%) and generate trajectories at 0.2s resolution. Using these trajectories, we compute widely used surrogate safety measures TTC, DRAC, and DDR—to identify both longitudinal and lateral conflicts. At the macroscopic level, we aggregate the simulated data along a 14 km four-lane freeway with ramps and extract continuous time series of flow, speed, and occupancy for each segment, paired with the conflict/non-conflict labels derived from TTC/DRAC/DDR. This design allows us to study conflict prediction as a segment-level time-series classification problem under a wide range of demand levels and CAV penetration rates, while keeping driver behavior anchored in real freeway observations and defining conflicts via physically interpretable criteria.

We fully acknowledge that this hybrid validation cannot replace large-scale field testing on real mixed CAV–HDV networks. In the revised Conclusion, we now explicitly state that 1)

the current evaluation is conducted in a calibrated freeway simulation, 2) the direct transferability to urban or suburban networks is therefore limited, and 3) the lack of suitable real-world mixed-traffic datasets is a key limitation of the present work. We also clarify that, once continuous macroscopic observations of mixed CAV–HDV traffic over long freeway segments become available, we plan to retrain and evaluate MS-STGNet on those data and to systematically compare its performance with other state-of-the-art safety models in real-world environments. Below is the revised version:

Revised (Page 25 Line 25, Page 26 Line 1—21):

The proposed framework has several practical implications. It can be embedded as a safety prediction component in CAV cloud management systems for freeway corridors and urban expressways, integrated into freeway traffic management centers and ramp control or variable speed limit systems to support mixed CAV–HDV operations, and used within regional expressway operation platforms to provide real-time conflict or crash risk warnings at bottlenecks and merging/diverging areas, thereby enhancing the safety management and visualization of freeway networks. The limitations of this study are summarized as follows: 1) The model is calibrated and evaluated in a microscopic simulation of a four-lane freeway segment with motorized traffic only. Although the simulation is grounded in highD trajectory data, we do not yet validate MS-STGNet on large-scale field observations of mixed CAV–HDV traffic, and the direct transferability of the results to urban or suburban road networks with signalized intersections, pedestrians, and non-motorized vehicles is therefore limited. 2) The current experiments focus on a single 14 km corridor with specific demand patterns; additional facilities and more diverse demand scenarios would further test the generalizability of the framework. 3) The predefined manifold similarity matrix remains static over time, preventing the model from capturing previously unseen traffic state transitions unless it is retrained. 4) The proposed framework currently focuses on binary conflict/non-conflict prediction. Although the sigmoid activation in the output layer produces continuous risk 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.

Comment 2:

It is advisable to test the model in different traffic conditions (low, medium, and high congestion), different road geometries (e.g., merging ramps), or even an urban environment.

Response to Comment 2:

We thank the reviewer for this constructive suggestion. In the revised manuscript, we clarify more explicitly how varying traffic conditions and geometries are already incorporated, and we delimit the current scope to freeway environments.

Section 6.1 explicitly states that the simulation covers 500 hours with hourly demand randomly sampled between 2,500 and 10,000 veh/h and stratified into low-, medium-, and high-volume ranges (See Page 14 Line 10—16). Section 6.10 and Appendix C have been expanded with additional analyses: the supplementary position–speed trajectories (Figures C.1–C.3) show that increasing volume leads to stronger speed oscillations and higher conflict occurrence, supporting our use of traffic-state variations as predictors.

Section 4.1 has been clarified to emphasize that the four-lane freeway includes multiple on-ramps with 250 m acceleration lanes (See Page 6 Line 22—23, Page 7 Line 1—2). The case study in Section 6.10 (Fig. 9) is conducted on one such merging segment, explicitly distinguishing pre-merging, merging, and post-merging subsegments.

We agree that extending the framework to urban or suburban networks would further enhance its scope. However, the present study is built on an EIDM model calibrated with highD freeway trajectories, making a freeway corridor the most consistent evaluation setting (See Page 6 Line 16—19). Extending MS-STGNet to urban networks would require additional behaviour models and data sources and is therefore left as future work. This limitation and planned extension are now explicitly stated in the Conclusion. Below is the revised version:

Revised:

In Section 6.10 (Page 23 Line 28—29, Page 24 Line 1—17):

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 C.1–C.3) 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 in the proposed framework: MS-STGNet is particularly effective at aligning its predicted risk with these underlying speed dispersion structures, while

STGCN and STGAT tend to generate spurious conflict probabilities in disturbance zones.

In Appendix C (Page 28 Line 7—13, Page 29):

Appendix C. Supplementary vehicle position–speed trajectories

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

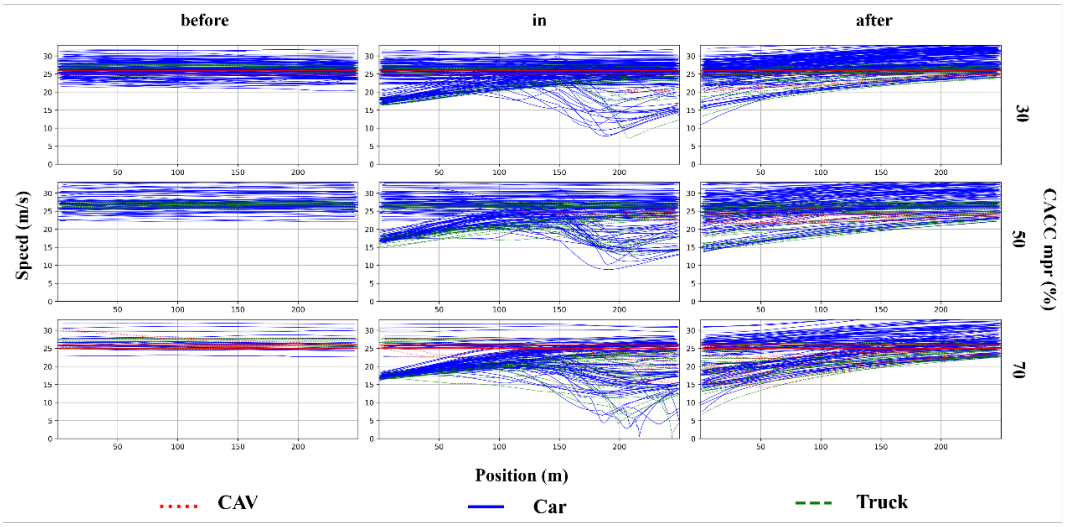

Fig. C1. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 3000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

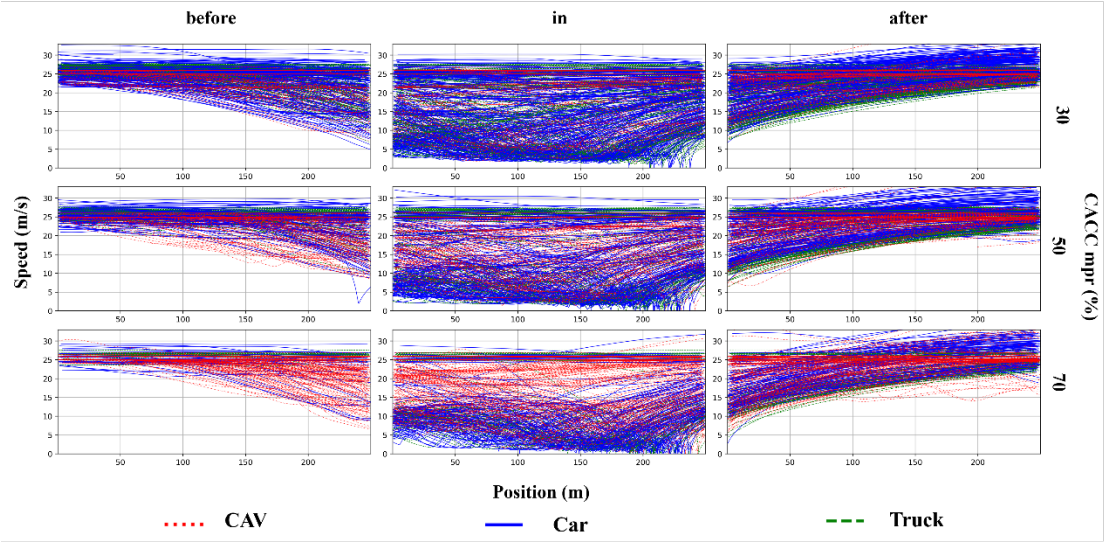

Fig. C2. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 6000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

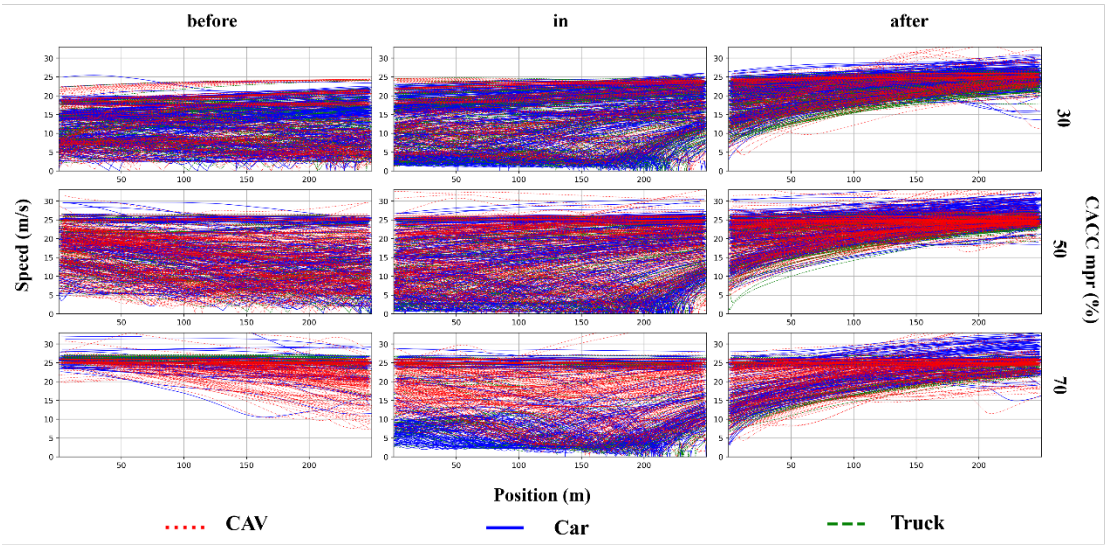

Fig. C3. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 9000 vehicles/hour. (before) pre-merging segment. (in) merging segment. (after) post-merging segment.

In Section 7 (Page 25 Line 25, Page 26 Line 1—21):

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

Comment 3:

The proposed architecture uses multiple components (convolutional residual, manifold similarity, TCN, and adaptive fusion). Processing these components could incur significant computational overhead for training and real-time inference. Could the authors provide an analysis on the computational cost of this framework?

Response to Comment 3:

We appreciate this important comment and agree that computational cost is crucial for practical deployment. In response, we have added a dedicated computational-cost comparison in Section 6.6 (Table 5), reporting for all deep models under five penetration rates: 1) GPU-MUT (peak GPU memory during training), 2) GPU-MUI (peak GPU memory during inference), and 3) the number of trainable parameters. For SVM and XGBoost, GPU-based indicators are omitted because they run on CPU and have negligible memory usage compared to deep models.

The results show that STGCN consistently has the largest parameter count and GPU memory footprint, with STGAT slightly smaller but still heavier than CNN and LSTM-CNN. By contrast, MS-STGNet uses fewer parameters than both graph-based baselines and reduces peak GPU memory by roughly 10–15% in training and 15–25% in inference across penetration rates, while still incorporating the manifold-similarity module and adaptive fusion. Compared with CNN/LSTM-CNN, MS-STGNet incurs moderately higher memory usage due to graph operations but remains in the same order of magnitude and does not introduce prohibitive overhead.

We also note that the manifold similarity matrices are computed once offline from historical data; training and inference operate on a fixed sparse manifold graph using standard spatiotemporal graph and temporal convolutions. Given this design and the hardware-agnostic indicators in Table 5, we believe that MS-STGNet achieves a reasonable balance between accuracy (Table 4) and efficiency, making it suitable for practical mixed CAV–HDV conflict prediction. We do not report wall-clock times since they depend heavily on specific hardware/software environments and are difficult to reproduce.

Revised (Page 18 Line 44—52, Page 19 Line 1—16):

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 5 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 5. 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.

Table 5
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**

Overall, these results indicate that MS-STGNet achieves superior predictive performance (as shown in Table 4) 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.

Comments from Reviewer #6:

Comments:

This manuscript presents a deep learning-based approach to stress detection in software developers using EEG signal analysis. The subject is timely and relevant, given the growing concern about mental well-being in tech-heavy occupations. The study leverages deep learning algorithms and biomedical signal processing, aligning well with current trends in AI-driven healthcare and human-centered computing.

Strengths

The paper addresses an important interdisciplinary problem combining mental health, software engineering, and AI.

Use of EEG signals provides a physiological and objective measure of stress.

The authors apply deep learning models, which are state-of-the-art for classification tasks, and present reasonable performance metrics.

Weaknesses and Suggestions

Experimental Setup: The study would benefit from a clearer description of dataset size, sampling procedures, and participant demographics.

Comparative Analysis: No strong benchmarking with baseline models or traditional ML classifiers (e.g., SVM, Random Forest). This limits the understanding of the added value of using DL.

Model Explainability: There is limited discussion of how interpretable the model's decisions are. In biomedical contexts, explainability is vital.

Writing and Structure: While generally well-organized, there are some grammatical errors and vague phrasing in the results and discussion sections that require revision.

Discussion of Limitations: The manuscript lacks a critical reflection on the generalizability of the model and the potential for bias in data collection.

Response to Reviewer #6:

We sincerely thank the reviewer for the time and effort devoted to evaluating our submission. However, we respectfully note that this specific set of comments does not appear to refer to our manuscript. Our paper focuses on traffic conflict prediction in mixed CAV–HDV freeway environments using a manifold similarity-based spatiotemporal graph neural network (MS-STGNet). The manuscript does not involve EEG data, stress detection in software developers, biomedical signal analysis, or AI-driven healthcare applications. Given this clear mismatch in topic, data, and methodology, we are unfortunately unable to provide a meaningful point-by-point response to the detailed remarks in this particular review comment, as they do not correspond to the content of our work.

ORCID Information

Zongshi Liu

Address: College of Transportation Engineering, Tongji University, Shanghai 201804, PR China

E-mail: chuochuoliu@tongji.edu.cn

ORCID: 0009-0004-4049-1956

CRediT authorship contribution statement

Zongshi Liu: Conceptualization, Methodology, Software, Writing – original draft, Data curation, Visualization, Investigation, Writing – review & editing. **Guojian Zou:** Software, Methodology. **Ting Wang:** Software, Methodology. **Meiting Tu:** Supervision, Writing – review & editing. **Hongwei Wang:** Validation, Supervision, Writing – reviewing & editing. **Ye Li:** Supervision, Writing – review & editing.

Highlights

- A manifold similarity graph neural network is proposed for traffic conflicts prediction.
- A realistic simulation environment was established to explore traffic conflicts under mixed traffic scenarios.
- The experiment results prove MS-STGNet is superior to existing competitive models.
- The manifold similarity module can effectively reduce the false alarm rate of traffic conflicts.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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

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,*]

[a]The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, 201804, PR China
[b]College of Transportation Engineering, Tongji University, Shanghai 201804, PR China
[c]Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore, 138632, Republic of Singapore

ARTICLE INFO

Keywords:
Conflict risk modeling
Manifold similarity
Mixed traffic flow
Spatiotemporal characteristics
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.

1. Introduction

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

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

* Corresponding authors: Ye Li.
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@ihpc.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

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

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.

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.

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.

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.

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

2. Related work

2.1. Mixed traffic flow modeling for traffic safety

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

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

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

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

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

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

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

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

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

2.3. Manifold learning in traffic state modelling

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

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

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

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

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

3. Preliminary

3.1. Traffic network graph

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

3.2. Embedding

Timestamps play a critical role in modeling processes. In this study, timestamp embedding specifically includes hour embedding and day-of-week embedding. Traditional timestamps are typically represented as integers, where identical numerical values are often misinterpreted as contributing equally to conflict prediction. For instance, Saturday is represented as the integer 5, and 6:00 AM is also represented as the integer 5, leading to equivalent significance in the input variables. This study addresses the issue by mapping integers to one-hot vectors, which are subsequently transformed into high-dimensional embeddings by applying two-dimensional convolutional neural networks (2-D CNNs) (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-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

study. The 2D-CNNs transformation process can be expressed as:

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

where $\mathbf{F}_{l,t}$ represents the feature output at layer l and time step t, $\mathbf{K}_{l,t} \in \mathbb{R}^{1\times1\times d}$ denotes the convolution kernel with a 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.

3.3. Problem definition

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

$$\left(\hat{Y}_{t_p+1}, \hat{Y}_{t_p+2}, \ldots, \hat{Y}_{t_p+M}\right) = f_\Theta\left(X_{t_1}, X_{t_2}, \ldots, X_{t_p}; G\right) \tag{2}$$

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

4. Simulation and data generation

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

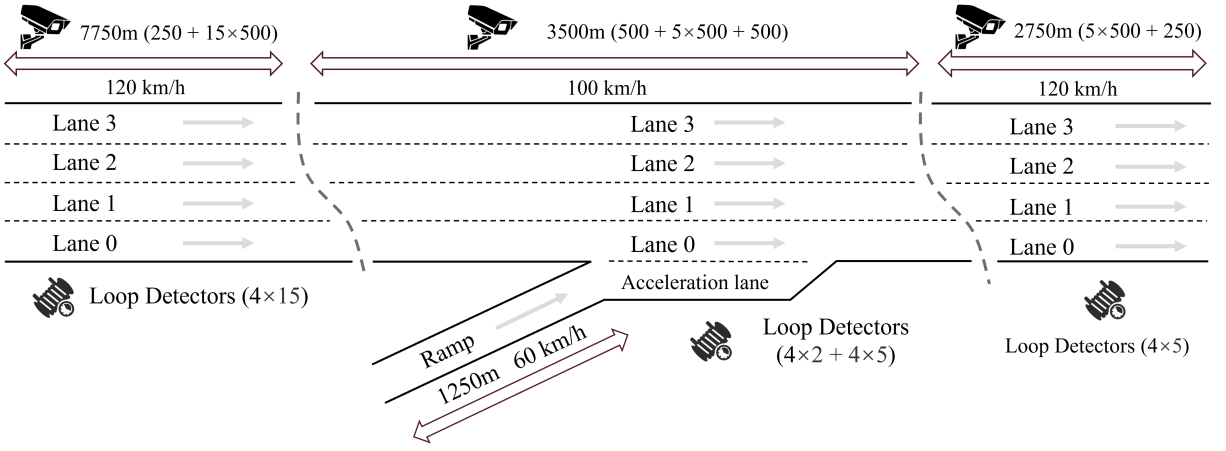

Fig. 1. The silhouette of the simulation road network.

4.1. Simulation network

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

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 Table1 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.

Table 1
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

4.4. 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).

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

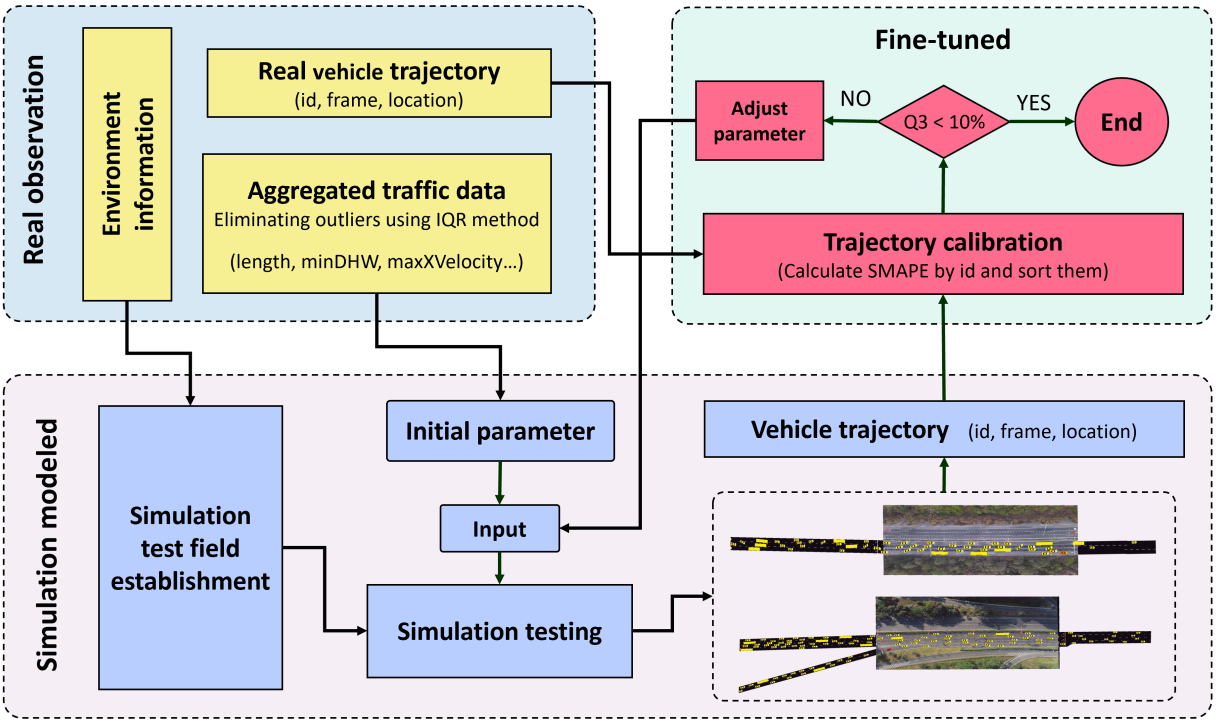

Fig. 2. The process of calibrating the EIDM model.

(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).

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

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

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

For the above formulas, where $x_i(t)$ and $x_{i-1}(t)$ are the longitudinal location of the leader and follower at timestamp 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 $i-1$. d_l and d_f denote the longitudinal distances from the subject vehicle to its immediate leader and follower, respectively, measured immediately after completing the lane-change. Correspondingly, d_l^* (d_f^*) specifies the minimum safe gap that must be maintained to the nearest leading (following) vehicle in order to satisfy the prescribed deceleration constraints. A smaller TTC / DDR or a larger DRAC indicates a more hazardous situation. Predefined thresholds are essential to detect potential traffic conflicts. A traffic conflict is identified when the TTC and DDR drop below the set threshold or the DRAC exceeds it. Referring to previous studies, this study establishes the TTC threshold for conflict identification 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 et al., 2020).

5. Methodology

5.1. Model architecture overview

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

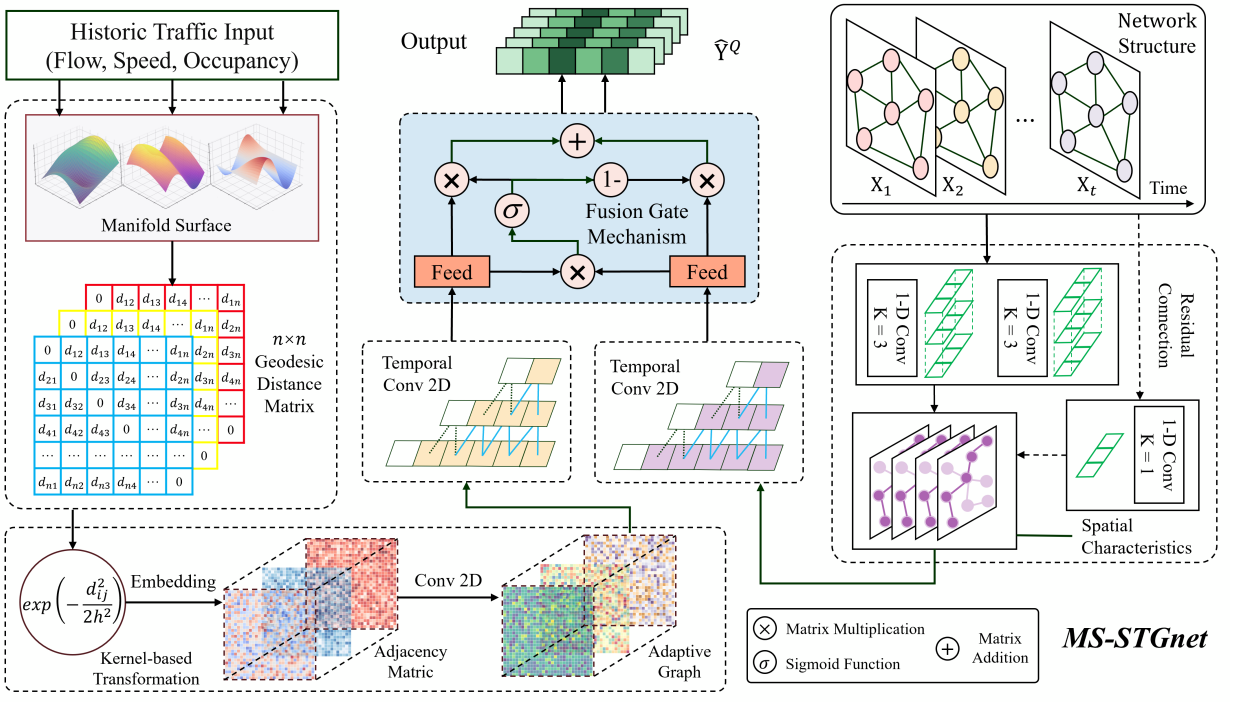

Fig. 3. Framework overview of MS-STGNet.

5.2. Residual convolutional module

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

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

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

$$\mathcal{F}_k(\mathbf{X}_t) = \sigma\Big(\mathbf{W}_k^{(1)} \circledast \sigma\big(\mathbf{W}_k^{(0)} \circledast \mathbf{X}_t + \mathbf{b}_k^{(0)}\big) + \mathbf{b}_k^{(1)}\Big) \tag{7}$$

where $\circledast$ indicates the convolution operation; $\mathbf{W}_k^{(0)}$ and $\mathbf{W}_k^{(1)}$ represent convolution kernels for the first and second 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:

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

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 $\mathbf{X} \in \mathbb{R}^{T \times W \times H \times d_x}$, and the output is $\mathbf{H}_{\text{res}} \in \mathbb{R}^{T \times W \times H \times d}$.

5.3. Manifold-similarity graph module

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

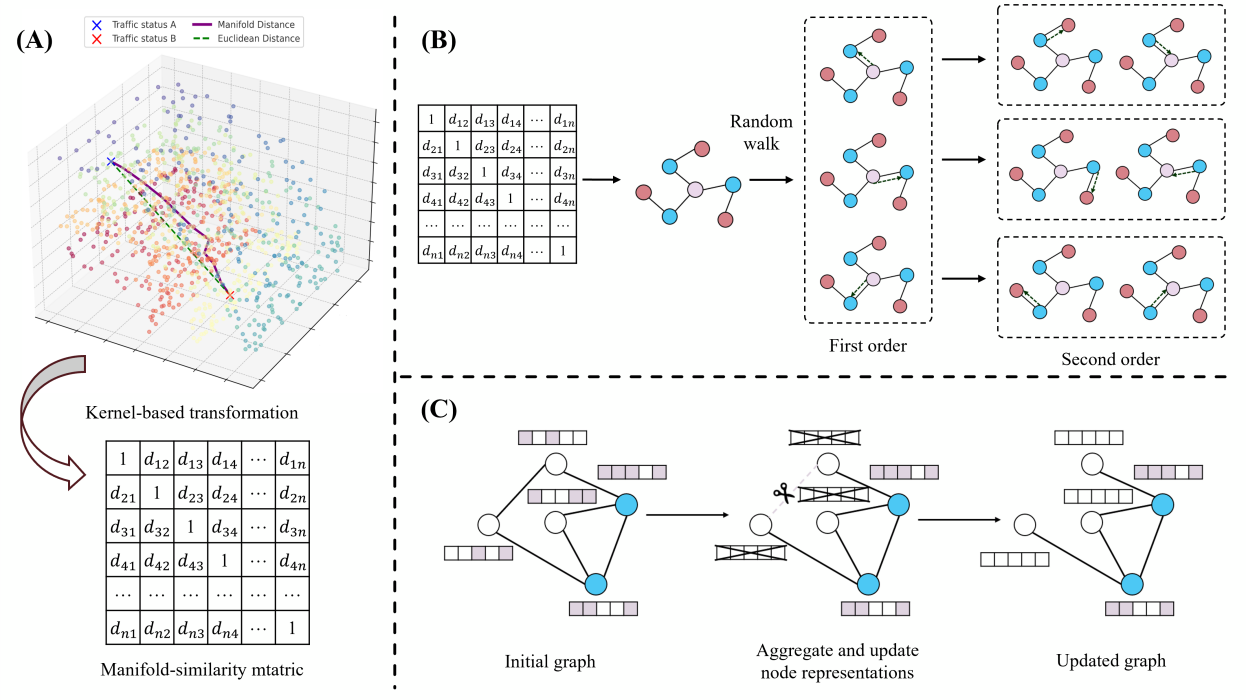

Fig. 4. Processing of self-adaptive graph based on manifold similarity. **(A)** The transformation process from manifold distance to similarity matrix. **(B)** Two-step random walk process. **(C)** Adaptive graph update process.

Specifically, we utilizes predefined similarity matrices to encode the semantic spatial dependencies between valid regions. However, solely relying on prior knowledge imposes limitations on uncovering latent spatial correlations embedded within the data. Adaptive graphs are integrated to dynamically capture global spatial relationships across

valid regions to address this problem. These adaptive graphs are initialized using predefined similarity matrices and iteratively refined during the training process. Moreover, modeling deep spatial dependencies and intricate interrelations among regions proves insufficient with single-directional and first-order graph structures. Consequently, we incorporates multi-order bidirectional random walks, enabling the aggregation and refinement of node representations by leveraging information from higher-order and bidirectional neighboring regions.

5.3.1. Manifold similarity graph

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

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

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 as an approximation of their geodesic distance:

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

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

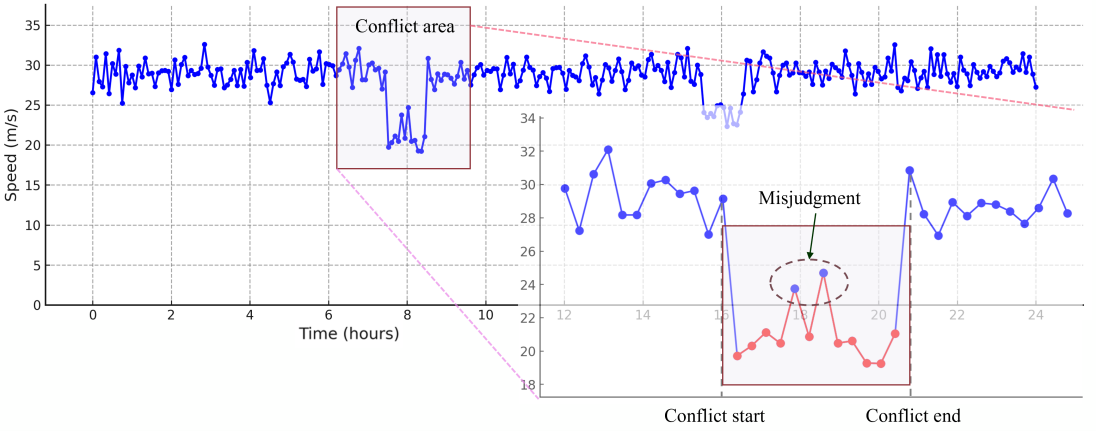

Fig. 5. Metrics for traffic conflicts based on Manifold distance and Euclidean distance.

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 $\tilde{\mathbf{A}}^{\star}$ 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 $\tilde{\mathbf{A}}^{\star}$, 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:

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

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

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

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

5.4. Temporal convolutional network (TCN) layer

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

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

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 the 1D convolution kernel of the l-th TCN layer and the k-th output channel. Each residual block comprises two 1D-CONV layers, and a skip connection is introduced by adding a block's input with its output. This converts a regular

TCN block into a residual TCN block whose output is as per the given equation.

$$\begin{cases} \mathbf{H}_t^{(k)} = \mathrm{ReLU}\Big(\mathbf{f}_k^{(1)} * \mathrm{ReLU}\big(\mathbf{f}_k^{(0)} * \mathbf{H}_t^{(k)} + \mathbf{b}_{t,k}^{(0)}\big) + \mathbf{b}_{t,k}^{(1)}\Big) \\ \mathbf{H}_t^{(k)} = \mathbf{H}_t^{(k)} + \mathbf{W}_{B,k}^{(0)} * \mathbf{H}_t^{(k-1)} + \mathbf{b}_{B,k}^{(0)} \end{cases} \tag{14}$$

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 k-th residual TCN block, respectively. They are specific instances of the generic kernel $\mathbf{f}^{l,k}$ defined in Eq.(18), but we 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 learnable biases; and $*$ is the convolution operator. In this study, multiple temporal blocks based on TCN are employed to extract temporal features from the output results of the residual convolutional module and the manifold similarity graph module. These temporal blocks are designed to capture both short-term and long-term temporal dependencies, enabling the model to effectively learn time-series patterns within the input data.

$$\mathbf{H}_t = \mathrm{Stack}\big(\mathbf{H}_t^{(1)}, \mathbf{H}_t^{(2)}, \ldots, \mathbf{H}_t^{(L)}\big) \tag{15}$$

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

5.5. Adaptive channel fusion gate

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

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

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

$$\mathbf{\Phi} = \sigma\Big(\mathbf{W}_{\Phi}^{(0)} * \mathbf{H}_{\mathrm{resT}} + \mathbf{W}_{\Phi}^{(1)} * \mathbf{H}_{\mathrm{MST}} + \mathbf{b}_{\Phi}\Big) \tag{17}$$

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

$$\mathbf{H}' = \mathrm{ReLU}\Big(\mathbf{W}_H^{(1)} * \mathrm{ReLU}\big(\mathbf{W}_H^{(0)} * \mathbf{H} + \mathbf{b}_H^{(0)}\big) + \mathbf{b}_H^{(1)}\Big) \tag{18}$$

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.

5.6. Loss function

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

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

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

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

6. Experiments

6.1. Data preparation

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

Table 2
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

The loop detector data were collected at 30-second intervals and aggregated into 5-minute granularity, resulting in 6,000 time slices. Data from 5 to 20 minutes before a conflict were identified as potential precursors for predicting conflicts (Li et al., 2020; Kamel et al., 2023, 2024). Consequently, the target time step for model training was set to 1, with the most recent 4 preceding time steps (20 minutes) used as inputs for conflict prediction. During the simulation, this process generated a total of $4 \times 27 \times 6{,}000 = 648{,}000$ traffic data samples. The number of samples labeled as traffic conflicts was 24,087 (10% MPR), 24,561 (30% MPR), 19,432 (50% MPR), and 17,146 (70% MPR), and 16,580 (90% MPR). This yielded a conflict-to-non-conflict sample ratio of approximately 1:26, 1:25, 1:32, 1:36, and 1:38, highlighting the presence of significant zero-inflation in the data. This imbalance underscores the applicability of our proposed MS-STGNet model in handling rare-event scenarios effectively.

6.2. Experimental setup

PyTorch framework are utilized to construct all experiments, and the training, validation, and testing process is executed on a platform with Intel(R) Xeon(R) Gold 6336Y CPU and NVIDIA RTX 4090 GPU-24 GB card. The dataset is split into training, validation, and testing sets in a 6:2:2 ratio. During the training process, the maximum 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 to update model weights. Model performance is evaluated on the validation set after each epoch, and the weights are saved whenever a reduction in loss is observed. Additionally, an early stopping mechanism with a patience value of 10 is applied to mitigate overfitting. If the validation loss remains unchanged for 10 consecutive epochs, the training process terminates early. To reduce the impact of randomness and evaluate the stability of each method, all models are trained and evaluated five times with different random seeds orders. The detailed parameter settings in each module are summarized in Table 3.

6.3. Evaluation metrics

To evaluate the classification performance of the MS-STGNet model, we employed metrics commonly used in conflict risk analysis, including recall, accuracy, and false alarm rate (FAR) (Li et al., 2020). Additionally, the area

Table 3
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
	Padding	1/1/1
Adaptive fusion gate	Channel	64
	Fitter size	1×1
	Number of layers	1
FC	Channel	64/1
	Fitter size	1×1/1×1
	Number of layers	2
Loss function	α_t	0.5
	γ	2
	S	3
	σ	0.7
	α	0.5
	β	0.5
Other hyperparameters	Training optimizer	Adam
	Decay rate	0.9
	Batch size	32
	Learning rate	0.0005
	Dropout	0
	Epochs	200
	Patience	10

under the ROC curve (AUC) was used to assess the performance of the binary classifier; the G-mean (geometric mean) served as an indicator of a model's performance on the minority class. The descriptions are as follows:

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

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

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

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

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

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

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 incorrectly classified samples, demonstrates superior performance in predicting accident severity (Goswamy et al., 2023).
- **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.
- **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).
- **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).
- **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 (Wu et al., 2023).

6.5. Performance comparison

Table 4 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 4 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 ($p < 0.05$).

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

Table 4
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**↑1.92% ±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**↑9.09% ±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**↑2.11% ±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**↑3.01% ±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**↑2.12% ±0.017
30%	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**↑3.86% ±0.010
	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**↑3.50% ±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**↑3.58% ±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**↑2.76% ±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**↑1.84% ±0.012
50%	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**↑5.92% ±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**↑23.91% ±0.017
	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**↑6.36% ±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**↑3.85% ±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**↑5.22% ±0.011
70%	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**↑3.42% ±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**↑24.00% ±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**↑4.62% ±0.015
	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**↑7.42% ±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**↑3.86% ±0.014
90%	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**↑1.24% ±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**↑23.77% ±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**↑3.37% ±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**↑2.63% ±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**↑2.86% ±0.013

to deep learning approaches. For example, under a 30% penetration rate, the recall rates of SVM and XGBoost were 23% and 17.8% lower, respectively, than those of the proposed MS-STGNet. Additionally, their false alarm rates increased by 27.9% and 20.0%, AUC values decreased by 24.3% and 18.9%, and accuracy was reduced by 26.4% and 20.5%. These results emphasize the importance of extracting nonlinear correlations for traffic conflict prediction.

The introduction of deep learning methods significantly improved model performance. CNN and LSTM-CNN outperformed SVM and XGBoost across all metrics, demonstrating the importance of capturing spatial dependencies and temporal correlations in conflict prediction. However, deep learning methods relying on CNNs to capture spatial dependencies face a notable limitation: they cannot model spatial similarities in unconnected grid fields. This highlights the advantage of leveraging graph neural networks (GNNs), such as STGCN and STGAT, to model semantic spatial dependencies, further enhancing performance. For instance, under a 30% penetration rate, STGAT and STGCN 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% and 1.5%, and improved accuracy by 4.4% and 3.2%, respectively, compared to LSTM-CNN. These results underscore the advanced capability of utilizing the inherent graph structure of road networks to extract spatial dependencies related to conflict risks. GNNs are particularly well-suited for capturing complex relationships between road segments, 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 with spatial proximity and contextual similarity, better representing spatial dependencies.

Building on prior advancements in graph-based models, the proposed MS-STGNet model demonstrated robust performance across all penetration rate scenarios. For instance, under a 50% penetration rate, MS-STGNet outperformed the next-best models by 4.9% in recall, reduced false alarm rates by 3.3%, improved AUC by 5.3%, and increased accuracy by 3.3%. Notably, as shown in Table 4, MS-STGNet achieved a significant reduction in false alarm rates, with improvements of 23.9%, 24.0%, and 23.8% under 50%, 70%, and 90% penetration rates, respectively. This improvement can be attributed to the manifold similarity module, which reduces misjudgments in conflict-prone areas of traffic flow—a point further analyzed in subsequent sections.

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.

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 5 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 5. First, among the deep learning baselines, STGCN consistently

Table 5
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**

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 4) 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.

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 6 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:

Table 6
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**

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

6.8. Posterior probability analyses

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

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

Additionally, we observed that as market penetration increases, the probability distributions produced by the STGCN and STGAT models become increasingly skewed towards lower values, indicating a growing tendency to classify samples as belonging to the negative class. We attribute this effect to the rising ratio of non-conflict to conflict events: as penetration rates climb, the prevalence of negative instances increases, exacerbating dataset imbalance and biasing model outputs downward. Although this pattern is also evident in our MS-STGNet framework, MS-STGNet yields more balanced probability distributions across varying penetration levels, and thus across differing class proportions. Consequently, its 0.5 decision threshold more reliably separates conflict and non-conflict events, demonstrating superior stability.

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

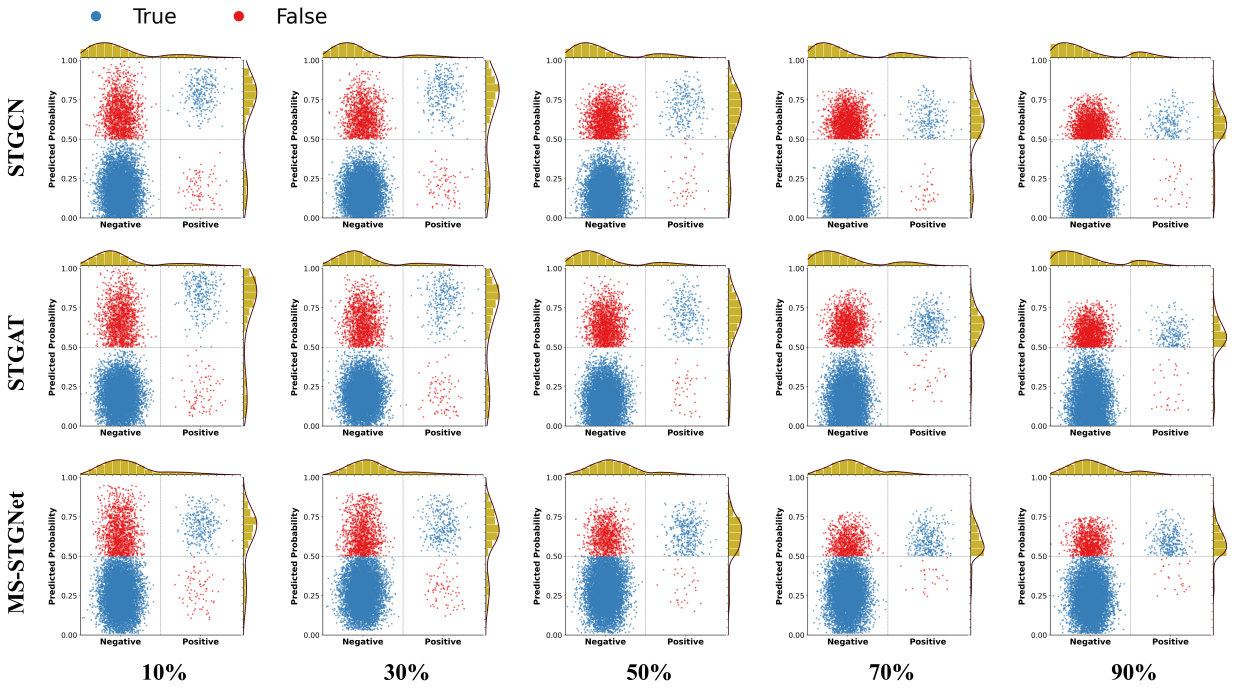

Fig. 6. Predicting probability values for STGCN, STGAT, and MS-STGNet under different market penetration rates.

performance.

6.9. Spatiotemporal heat map analysis

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

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

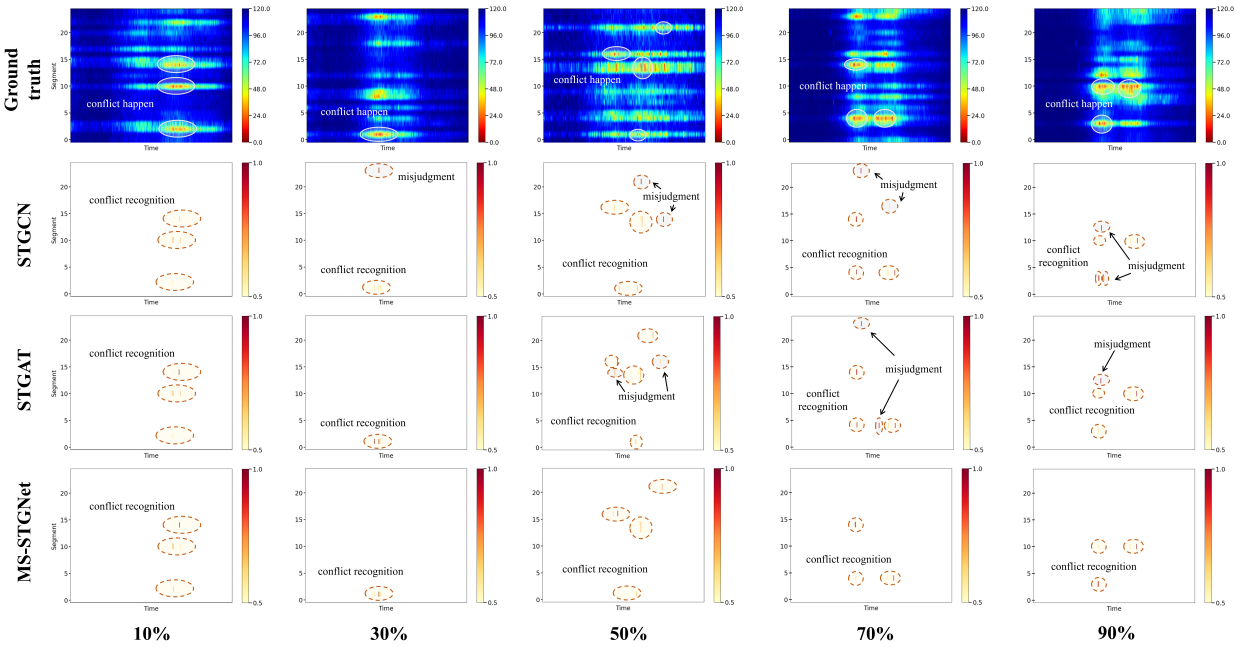

Fig. 7. Predicting probability values for STGCN, STGAT, and MS-STGNet under different market penetration rates.

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.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 separation. 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.

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

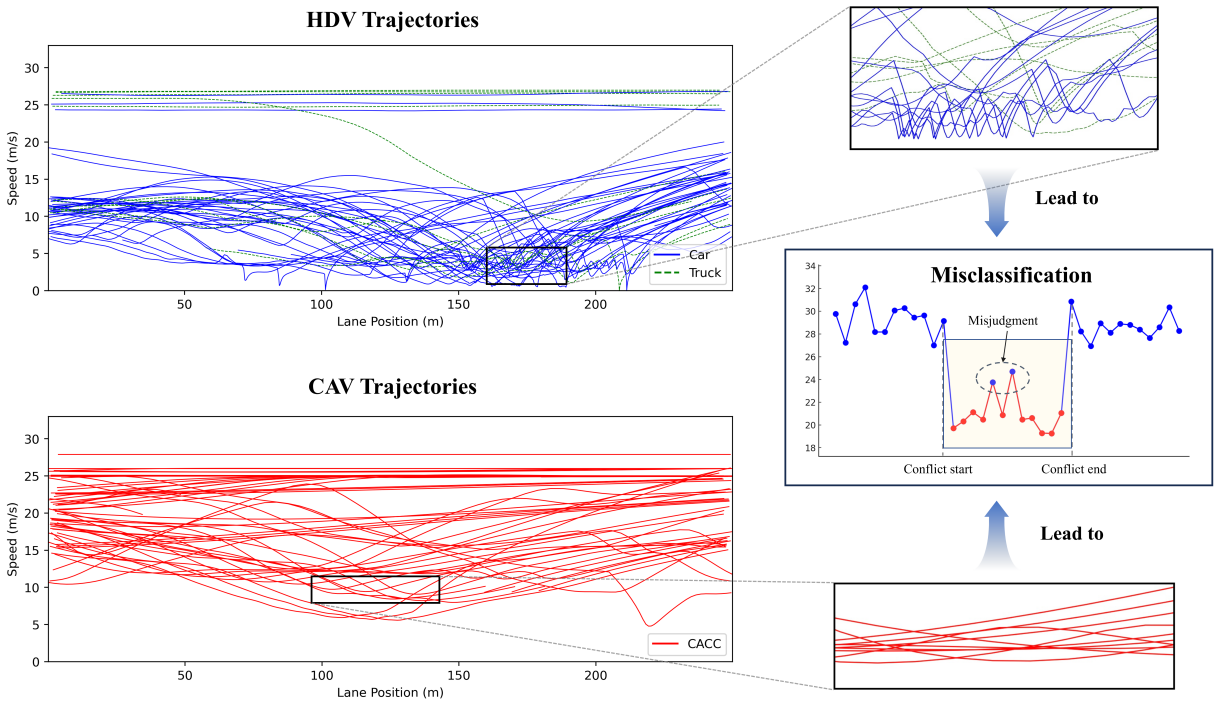

Fig. 8. Position-speed profiles of two types of vehicles.

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

7. Conclusion

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

In this study, we propose a novel Manifold Similarity-based Multi-Graph Spatiotemporal Network (MS-STGNet) for conflict prediction in mixed traffic environments. The proposed framework demonstrates strong performance in

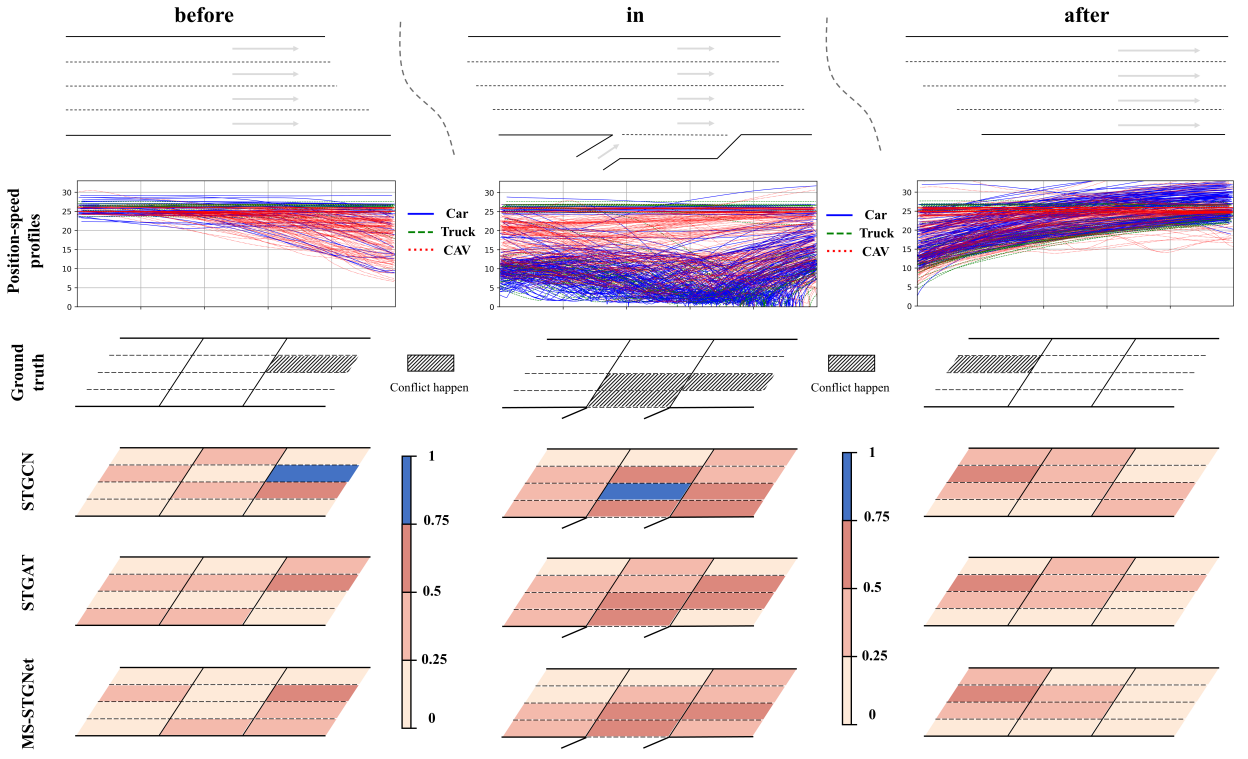

Fig. 9. Vehicle position-speed trajectories and predicting probability values for STGCN, STGAT, and MS-STGNet in **(before)** pre-merging segment, **(in)** merging segment, **(after)** post-merging segment.

scenarios where CAVs and HDVs coexist, while also exhibiting a degree of generalizability to non-mixed traffic conditions. Specifically, the model incorporates the following components: A residual convolutional network to extract geographical features in interconnected areas of the land space. A manifold similarity graph module to capture spatial semantic features in regions. A temporal convolutional network to model temporal dependencies in traffic flow data, extending spatial features into spatiotemporal representations. An adaptive fusion gate mechanism combines geographical and semantic spatiotemporal features to generate final predictions. As a main contribution of the study, we introduce a manifold similarity method to model the similarity of traffic states. Historical traffic flow data are aggregated into traffic state vectors, and manifold distances are used to calculate similarities between these vectors. The similarity matrix is then integrated into the graph network as prior knowledge, imposing a layer of physical constraints on the deep learning outcomes. This ensures that the predicted transition of traffic states from conflict to non-conflict aligns with inherent spatiotemporal patterns, which is crucial for accurately identifying different traffic states and reducing misclassifications in conflict prediction.

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

The proposed framework has several practical implications. It can be embedded as a safety prediction component

in CAV cloud management systems for freeway corridors and urban expressways, integrated into freeway traffic management centers and ramp control or variable speed limit systems to support mixed CAV–HDV operations, and used within regional expressway operation platforms to provide real-time conflict or crash risk warnings at bottlenecks and merging/diverging areas, thereby enhancing the safety management and visualization of freeway networks. The limitations of this study are summarized as follows: **1)** The model is calibrated and evaluated in a microscopic simulation of a four-lane freeway segment with motorized traffic only. Although the simulation is grounded in highD trajectory data, we do not yet validate MS-STGNet on large-scale field observations of mixed CAV–HDV traffic, and the direct transferability of the results to urban or suburban road networks with signalised intersections, pedestrians, and non-motorised vehicles is therefore limited. **2)** The current experiments focus on a single 14 km corridor with specific demand patterns; additional facilities and more diverse demand scenarios would further test the generalizability of the framework. **3)** The predefined manifold similarity matrix remains static over time, preventing the model from capturing previously unseen traffic state transitions unless it is retrained. **4)** The proposed framework currently focuses on binary conflict/non-conflict prediction. Although the sigmoid activation in the output layer produces continuous risk 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.

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).

Appendix A. Visualization of the learned manifold-similarity matrices

Appendix A presents the learned manifold-similarity matrices for flow, occupancy, and speed, denoted by $\mathbf{Matrices}^{(\mathrm{flow})}$, $\mathbf{Matrices}^{(\mathrm{occupancy})}$, and $\mathbf{Matrices}^{(\mathrm{speed})}$, 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.

$$\mathbf{Matrices}^{(\mathrm{flow})} = \begin{bmatrix} 1.000 & 0.277 & 0.268 & 0.274 & 0.745 & \cdots & 0.686 \\ 0.277 & 1.000 & 0.701 & 0.285 & 0.689 & \cdots & 0.279 \\ 0.268 & 0.701 & 1.000 & 0.707 & 0.693 & \cdots & 0.699 \\ 0.274 & 0.285 & 0.707 & 1.000 & 0.688 & \cdots & 0.759 \\ 0.745 & 0.689 & 0.693 & 0.688 & 1.000 & \cdots & 0.696 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.686 & 0.279 & 0.699 & 0.759 & 0.696 & \cdots & 1.000 \end{bmatrix}, \quad \mathbf{Matrices}^{(\mathrm{flow})} \in \mathbb{R}^{108\times108}$$

$$\mathbf{Matrices}^{(\mathrm{occupancy})} = \begin{bmatrix} 1.000 & 0.365 & 0.316 & 0.276 & 0.353 & \cdots & 0.250 \\ 0.365 & 1.000 & 0.367 & 0.327 & 0.302 & \cdots & 0.314 \\ 0.316 & 0.367 & 1.000 & 0.390 & 0.283 & \cdots & 0.358 \\ 0.276 & 0.327 & 0.390 & 1.000 & 0.237 & \cdots & 0.499 \\ 0.353 & 0.302 & 0.283 & 0.237 & 1.000 & \cdots & 0.016 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.250 & 0.314 & 0.358 & 0.499 & 0.016 & \cdots & 1.000 \end{bmatrix}, \quad \mathbf{Matrices}^{(\mathrm{occupancy})} \in \mathbb{R}^{108\times108}$$

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

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

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:

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

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

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

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

B.2. SVD-based initialization and adaptive adjacency

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 (Guo et al., 2015; Zou et al., 2024), and $\mathbf{A}^*$ can be expressed as the product of three distinct matrices, as follows:

$$\mathbf{A}^* = \mathbf{U}^*\mathbf{\Sigma}^*\mathbf{V}^{*\top} \tag{B.3}$$

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

$$\mathbf{A}^\star = \mathrm{ReLU}(\mathbf{M}_{\mathrm{lt}}\mathbf{M}_{\mathrm{rt}}) \tag{B.4}$$

where $\mathbf{M}_{\mathrm{lt}}$ and $\mathbf{M}_{\mathrm{rt}}$ are the core learnable parameter matrices, which play a crucial role in dynamically modeling the weight relationships between nodes in the graph. $\mathbf{M}_{\mathrm{lt}}$ is the left transformation matrix, designed to encode a linear transformation of the input features or spatial dependency information. It operates as a critical step in updating the representation of node relationships by applying a transformation to the input data, expressed as: $\mathbf{M}_{\mathrm{lt}} = \mathbf{W}_{\mathrm{lt}}\left(\widehat{\mathbf{U}}_\star\widehat{\mathbf{\Sigma}}_\star\right)$. Similarly, $\mathbf{M}_{\mathrm{rt}}$ is the right transformation matrix, responsible for adjusting or aggregating the information encoded in $\mathbf{M}_{\mathrm{lt}}$, expressed as: $\mathbf{M}_{\mathrm{rt}} = \mathbf{W}_{\mathrm{rt}}\left(\widehat{\mathbf{\Sigma}}_\star\widehat{\mathbf{V}}_\star^\top\right)$. The ReLU function is applied to introduce nonlinearity and ensure that the

weights remain non-negative. Subsequently, the softmax function is used to normalize the weights of each node, ensuring that their sum equals 1. This normalization guarantees a balanced distribution of information during transmission, preventing any single node from dominating the interaction:

$$\tilde{\mathbf{A}}^{\star} = \mathbf{I}_N + \mathrm{softmax}\big(\mathrm{ReLU}(\mathbf{M}_{lt}\mathbf{M}_{rt})\big) \tag{B.5}$$

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

Appendix C. Supplementary vehicle position–speed trajectories

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

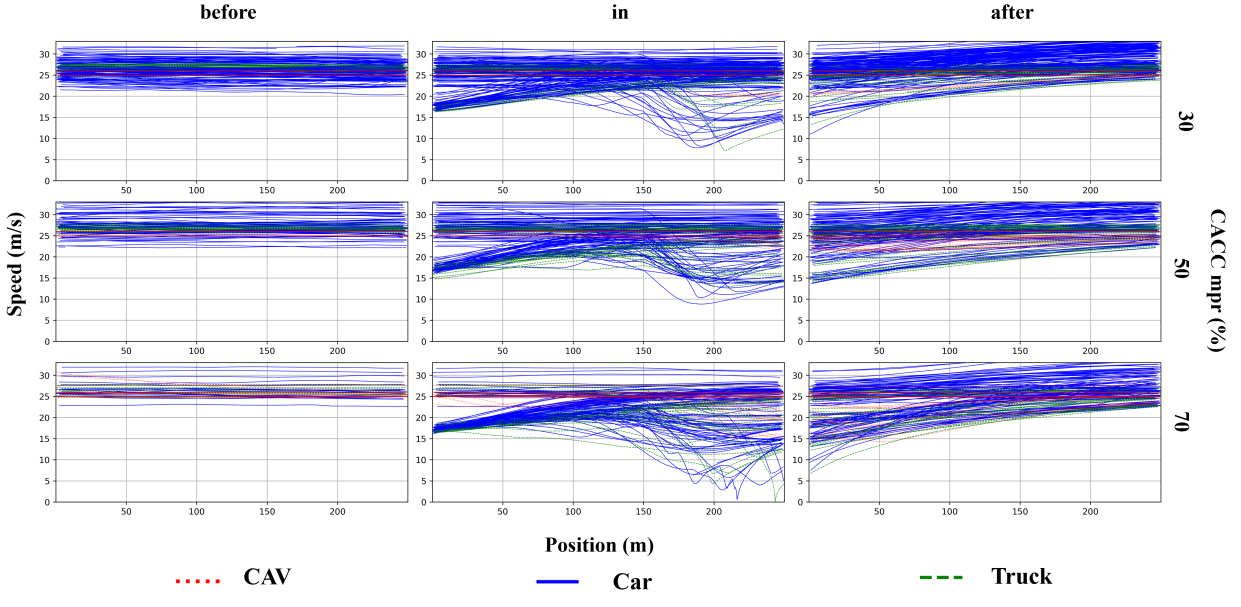

Fig. C1. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 3000 vehicles/hour. **(before)** pre-merging segment. **(in)** merging segment. **(after)** post-merging segment.

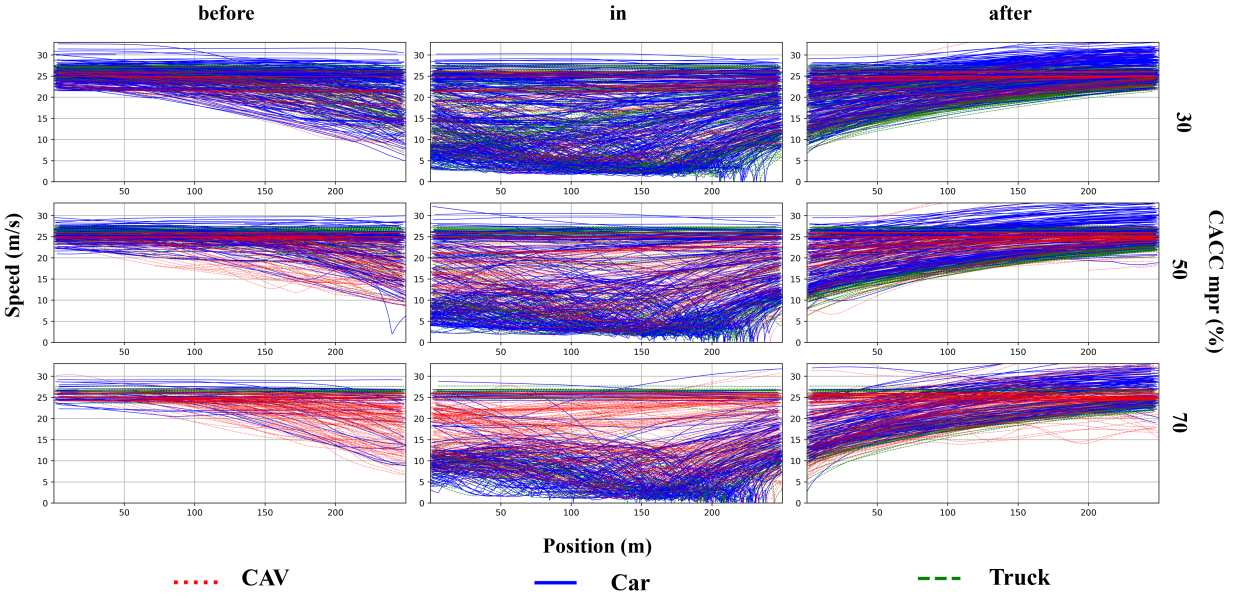

Fig. C2. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 6000 vehicles/hour. **(before)** pre-merging segment. **(in)** merging segment. **(after)** post-merging segment.

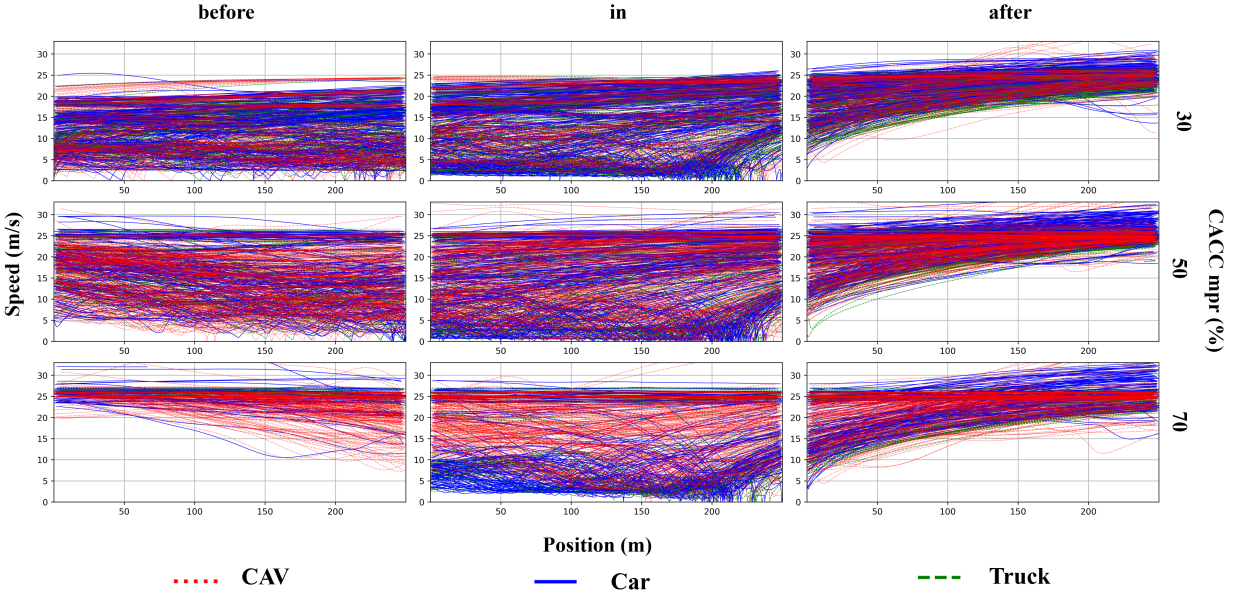

Fig. C3. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 9000 vehicles/hour. **(before)** pre-merging segment. **(in)** merging segment. **(after)** post-merging segment.

References

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

Abou Elassad, Z.E., Mousannif, H., Al Moatassime, H., 2020. A real-time crash prediction fusion framework: An imbalance-aware strategy for collision avoidance systems. Transportation research part C: emerging technologies 118, 102708.

Ahangar, M.N., Ahmed, Q.Z., Khan, F.A., Hafeez, M., 2021. A survey of autonomous vehicles: Enabling communication technologies and challenges. Sensors 21, 706.

Ali, Y., Haque, M.M., Mannering, F., 2023. Assessing traffic conflict/crash relationships with extreme value theory: Recent developments and future directions for connected and autonomous vehicle and highway safety research. Analytic methods in accident research 39, 100276.

Ali, Y., Hussain, F., Haque, M.M., 2024. Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. Accident Analysis & Prevention 194, 107378.

Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 .

Bansal, P., Kockelman, K.M., 2017. Forecasting americans' long-term adoption of connected and autonomous vehicle technologies. Transportation Research Part A: Policy and Practice 95, 49–63.

Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accident Analysis & Prevention 122, 239–254.

Basso, F., Pezoa, R., Varas, M., Villalobos, M., 2021. A deep learning approach for real-time crash prediction using vehicle-by-vehicle data. Accident Analysis & Prevention 162, 106409.

Bergel-Hayat, R., Debbarh, M., Antoniou, C., Yannis, G., 2013. Explaining the road accident risk: Weather effects. Accident Analysis & Prevention 60, 456–465.

Cai, Q., Abdel-Aty, M., Yuan, J., Lee, J., Wu, Y., 2020. Real-time crash prediction on expressways using deep generative models. Transportation research part C: emerging technologies 117, 102697.

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 first-person perspective video. IEEE Transactions on Intelligent Transportation Systems 23, 5298–5313.

Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis & Prevention 39, 657–670.

Chen, C., Fan, X., Zheng, C., Xiao, L., Cheng, M., Wang, C., 2018. Sdcae: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data, in: 2018 sixth international conference on advanced cloud and big data (CBD), IEEE. pp. 328–333.

Chen, K., Luo, Y., Zhu, M., Yang, H., 2024a. Human-like interactive lane-change modeling based on reward-guided diffusive predictor and planner. IEEE Transactions on Intelligent Transportation Systems , early access.

Chen, X., Tiu, P., Zhang, Y., Zhu, M., Zheng, X., Wang, Y., 2024b. Improving car-following control in mixed traffic: A deep reinforcement learning framework with aggregated human-driven vehicles, in: 2024 IEEE Intelligent Vehicles Symposium (IV), IEEE. pp. 627–632.

Di Vaio, M., Fiengo, G., Petrillo, A., Salvi, A., Santini, S., Tufo, M., 2019. Cooperative shock waves mitigation in mixed traffic flow environment. IEEE Transactions on Intelligent Transportation Systems 20, 4339–4353.

Fiengo, G., Lui, D.G., Petrillo, A., Santini, S., Tufo, M., 2019. Distributed robust pid control for leader tracking in uncertain connected ground vehicles with v2v communication delay. IEEE/ASME Transactions on Mechatronics 24, 1153–1165.

Fu, C., Sayed, T., 2021. Comparison of threshold determination methods for the deceleration rate to avoid a crash (drac)-based crash estimation. Accident Analysis & Prevention 153, 106051.

Galvani, M., 2019. History and future of driver assistance. IEEE Instrumentation & Measurement Magazine 22, 11–16.

Gao, X., Jiang, X., Haworth, J., Zhuang, D., Wang, S., Chen, H., Law, S., 2024. Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. Accident Analysis & Prevention 208, 107801.

Garg, M., Bouroche, M., 2023. Can connected autonomous vehicles improve mixed traffic safety without compromising efficiency in realistic scenarios? IEEE Transactions on Intelligent Transportation Systems 24, 6674–6689.

Goswamy, A., Abdel-Aty, M., Islam, Z., 2023. Factors affecting injury severity at pedestrian crossing locations with rectangular rapid flashing beacons (rrfb) using xgboost and random parameters discrete outcome models. Accident Analysis & Prevention 181, 106937.

Gu, Z., Wang, Z., Liu, Z., Saberi, M., 2022. Network traffic instability with automated driving and cooperative merging. Transportation Research Part C: Emerging Technologies 138, 103626.

Guo, Q., Zhang, C., Zhang, Y., Liu, H., 2015. An efficient svd-based method for image denoising. IEEE transactions on Circuits and Systems for Video Technology 26, 868–880.

Hossain, M., Abdel-Aty, M., Quddus, M.A., Muromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. Accident Analysis & Prevention 124, 66–84.

Hossain, M., Muromachi, Y., 2012. A bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. Accident Analysis & Prevention 45, 373–381.

Hou, K., Zheng, F., Liu, X., 2024a. Enhancing mixed traffic safety assessment: A novel safety metric combined with a comprehensive behavioral modeling framework. Accident Analysis & Prevention 208, 107766.

Hou, K., Zheng, F., Liu, X., Fan, Z., 2024b. Cooperative vehicle platoon control considering longitudinal and lane-changing dynamics. Transportmetrica A: transport science 20, 2182143.

Hu, J., Huang, M.C., Yu, X., 2020. Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models. Accident Analysis & Prevention 144, 105665.

Hu, X., Sun, J., 2019. Trajectory optimization of connected and autonomous vehicles at a multilane freeway merging area. Transportation Research Part C: Emerging Technologies 101, 111–125.

Huang, Y., Bi, H., Li, Z., Mao, T., Wang, Z., 2019. Stgat: Modeling spatial-temporal interactions for human trajectory prediction, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 6272–6281.

Ivanchev, J., Eckhoff, D., Knoll, A., 2019. System-level optimization of longitudinal acceleration of autonomous vehicles in mixed traffic, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE. pp. 1968–1974.

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 different temporal resolutions. Accident Analysis & Prevention 141, 105520.

Kamel, A., Sayed, T., Fu, C., 2023. Real-time safety analysis using autonomous vehicle data: a bayesian hierarchical extreme value model. Transportmetrica B: Transport Dynamics 11, 826–846.

Kamel, A., Sayed, T., Kamel, M., 2024. Real-time combined safety-mobility assessment using self-driving vehicles collected data. Accident Analysis & Prevention 199, 107513.

Lee, G., Mallipeddi, R., Lee, M., 2012. Identification of moving vehicle trajectory using manifold learning, in: Neural Information Processing:

19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part IV 19, Springer. pp. 188–195.
Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on lstm-cnn. Accident Analysis & Prevention 135, 105371.
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 causal machine learning approach. Transportation research part C: emerging technologies 160, 104537.
Li, Y., Lu, J., Xu, K., 2017a. Crash risk prediction model of lane-change behavior on approaching intersections. Discrete Dynamics in Nature and Society 2017, 7328562.
Li, Y., Yu, R., Shahabi, C., Liu, Y., 2017b. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. arXiv preprint arXiv:1707.01926 .
Lin, J., 1991. Divergence measures based on the shannon entropy. IEEE Transactions on Information theory 37, 145–151.
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. Transportation Research Part C: Emerging Technologies 55, 444–459.
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 multi-lane freeway facilities. Transportation Research Part C: Emerging Technologies 95, 261–279.
Liu, Q., Cai, Y., Jiang, H., Lu, J., Chen, L., 2018b. Traffic state prediction using isomap manifold learning. Physica A: Statistical Mechanics and its Applications 506, 532–541.
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 highway merging area. IEEE Transactions on Vehicular Technology .
Liu, Q., Li, C., Jiang, H., Nie, S., Chen, L., 2022. Transfer learning-based highway crash risk evaluation considering manifold characteristics of traffic flow. Accident Analysis & Prevention 168, 106598.
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 information fusion under severe weather conditions. IEEE Transactions on Intelligent Transportation Systems 23, 6640–6653.
Lu, K., Ding, Z., Ge, S., 2012. Sparse-representation-based graph embedding for traffic sign recognition. IEEE Transactions on Intelligent Transportation Systems 13, 1515–1524.
Lu, Q.L., Yang, K., Antoniou, C., 2021. Crash risk analysis for the mixed traffic flow with human-driven and connected and autonomous vehicles, in: 2021 ieee international intelligent transportation systems conference (itsc), IEEE. pp. 1233–1238.
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 accuracy. Journal of Transportation Safety & Security 15, 421–443.
Ma, Y., Li, Y., Zheng, Z., Huang, H., 2024. Developing merging policies for cavs: A policy training framework combining human experience with reinforcement learning. IEEE Transactions on Intelligent Vehicles , early access.
Makridis, M., Mattas, K., Ciuffo, B., Re, F., Kriston, A., Minarini, F., Rognelund, G., 2020. Empirical study on the properties of adaptive cruise control systems and their impact on traffic flow and string stability. Transportation research record 2674, 471–484.
Martin, J.E., Rivas, T., Matías, J., Taboada, J., Argüelles, A., 2009. A bayesian network analysis of workplace accidents caused by falls from a height. Safety Science 47, 206–214.
Milanés, V., Shladover, S.E., 2014. Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. Transportation Research Part C: Emerging Technologies 48, 285–300.
Milanés, V., Shladover, S.E., Spring, J., Nowakowski, C., Kawazoe, H., Nakamura, M., 2013. Cooperative adaptive cruise control in real traffic situations. IEEE Transactions on intelligent transportation systems 15, 296–305.
Mousavi, S.M., Osman, O.A., Lord, D., Dixon, K.K., Dadashova, B., 2021. Investigating the safety and operational benefits of mixed traffic environments with different automated vehicle market penetration rates in the proximity of a driveway on an urban arterial. Accident Analysis & Prevention 152, 105982.
Organization, W.H., 2023. Global status report on road safety 2023. World Health Organization.
Papadoulis, A., Quddus, M., Imprialou, M., 2019. Evaluating the safety impact of connected and autonomous vehicles on motorways. Accident Analysis & Prevention 124, 12–22.
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 Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 3346–3351.
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 medical image classification. arXiv preprint arXiv:2212.12741 .
Saha, D., Alluri, P., Dumbaugh, E., Gan, A., 2020. Application of the poisson-tweedie distribution in analyzing crash frequency data. Accident Analysis & Prevention 137, 105456.
Salles, D., Kaufmann, S., Reuss, H.C., 2020. Extending the intelligent driver model in sumo and verifying the drive off trajectories with aerial measurements, in: SUMO Conference Proceedings, pp. 1–25.
Sameen, M.I., Pradhan, B., 2017. Severity prediction of traffic accidents with recurrent neural networks. Applied Sciences 7, 476.
Santos, K., Dias, J.P., Amado, C., 2022. A literature review of machine learning algorithms for crash injury severity prediction. Journal of safety research 80, 254–269.
Seoa, T., 2023. Understanding large-scale traffic flow using model-based and data-driven dimension reduction: with covid-19 and olympic-paralympic case study. EU Science Hub , 124.
Shirazi, M., Lord, D., 2019. Characteristics-based heuristics to select a logical distribution between the poisson-gamma and the poisson-lognormal for crash data modelling. Transportmetrica A: Transport Science 15, 1791–1803.
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.
Tan, H., Zhao, F., Zhang, W., Liu, Z., 2023. An evaluation of the safety effectiveness and cost of autonomous vehicles based on multivariable coupling. Sensors 23, 1321.
Tarko, A.P., 2012. Use of crash surrogates and exceedance statistics to estimate road safety. Accident Analysis & Prevention 45, 230–240.
Tarko, A.P., 2021. A unifying view on traffic conflicts and their connection with crashes. Accident Analysis & Prevention 158, 106187.

Theofilatos, A., Chen, C., Antoniou, C., 2019. Comparing machine learning and deep learning methods for real-time crash prediction. Transportation research record 2673, 169–178.
Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Physical review E 62, 1805.
Trirat, P., Yoon, S., Lee, J.G., 2023. Mg-tar: Multi-view graph convolutional networks for traffic accident risk prediction. IEEE Transactions on Intelligent Transportation Systems 24, 3779–3794.
Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. Accident analysis & prevention 35, 427–433.
Wang, B., Lin, Y., Guo, S., Wan, H., 2021. Gsnet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting, in: Proceedings of the AAAI conference on artificial intelligence, pp. 4402–4409.
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 perception multi-modal large model for accident analysis and prevention, in: 2024 IEEE Intelligent Vehicles Symposium (IV), IEEE. pp. 472–477.
Wang, L., Wang, K., Ma, W., Abdel-Aty, M., Li, L., 2022. Real-time safety analysis for expressways considering the heterogeneity of different segment types. Journal of safety research 80, 349–361.
Wang, Q., Wang, S., Zhuang, D., Koutsopoulos, H., Zhao, J., 2024b. Uncertainty quantification of spatiotemporal travel demand with probabilistic graph neural networks. IEEE Transactions on Intelligent Transportation Systems 25, 8770–8781.
Wang, Q., Zhang, K., Zhu, C., Zhou, Y., 2023. A multi-regional spatio-temporal network for traffic accident risk prediction. Engineering Letters 31.
Wang, T., Ngoduy, D., Li, Y., Lyu, H., Zou, G., Dantsuji, T., 2024c. Koopman theory meets graph convolutional network: Learning the complex dynamics of non-stationary highway traffic flow for spatiotemporal prediction. Chaos, Solitons & Fractals 187, 115437.
Wang, T., Ngoduy, D., Zou, G., Dantsuji, T., Liu, Z., Li, Y., 2024d. Pi-stgnet: Physics-integrated spatiotemporal graph neural network with fundamental diagram learner for highway traffic flow prediction. Expert Systems with Applications 258, 125144.
Wang, W., Jiang, X., Xia, S., Cao, Q., 2010. Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation. Safety Science 48, 1248–1262.
Wang, W., Pan, L., Liu, B., 2009. Synergetic method of traffic state recognition based on manifold learning, in: 2009 IEEE International Conference on Automation and Logistics, IEEE. pp. 587–591.
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 to road-level and minute-level traffic accident prediction. IET Intelligent Transport Systems 17, 270–284.
Yang, D., Ozbay, K., Xie, K., Yang, H., Zuo, F., Sha, D., 2021. Proactive safety monitoring: A functional approach to detect safety-related anomalies using unmanned aerial vehicle video data. Transportation research part C: emerging technologies 127, 103130.
Yang, S., Zhou, W., 2011. Anomaly detection on collective moving patterns: Manifold learning based analysis of traffic streams, in: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, IEEE. pp. 704–707.
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 and simulation method. Transportation Research Part C: Emerging Technologies 147, 103989.
Yousaf, M., Rehman, T.U., Jing, L., 2020. An extended isomap approach for nonlinear dimension reduction. SN Computer Science 1, 160.
Yu, L., Du, B., Hu, X., Sun, L., Han, L., Lv, W., 2021. Deep spatio-temporal graph convolutional network for traffic accident prediction. Neurocomputing 423, 135–147.
Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention 51, 252–259.
Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety science 63, 50–56.
Yu, R., Wang, Y., Zou, Z., Wang, L., 2020. Convolutional neural networks with refined loss functions for the real-time crash risk analysis. Transportation research part C: emerging technologies 119, 102740.
Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-time crash risk prediction using long short-term memory recurrent neural network. Transportation research record 2673, 314–326.
Zhang, C., Yan, X., Ma, L., An, M., 2014a. Crash prediction and risk evaluation based on traffic analysis zones. Mathematical Problems in Engineering 2014, 987978.
Zhang, J., Wu, K., Cheng, M., Yang, M., Cheng, Y., Li, S., 2020. Safety evaluation for connected and autonomous vehicles' exclusive lanes considering penetrate ratios and impact of trucks using surrogate safety measures. Journal of advanced transportation 2020, 5847814.
Zhang, L., Jia, Y., Niu, Z., Liao, C., 2014b. Traffic state classification based on parameter weighting and clustering method. Journal of Transportation Systems Engineering and Information Technology 14, 147–151.
Zhang, S., Abdel-Aty, M., 2022. Real-time crash potential prediction on freeways using connected vehicle data. Analytic methods in accident research 36, 100239.
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 flow topologies. IEEE Transactions on intelligent transportation systems 17, 14–26.
Zhou, J., Zhu, F., 2020. Modeling the fundamental diagram of mixed human-driven and connected automated vehicles. Transportation research part C: emerging technologies 115, 102614.
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 scenarios? Accident Analysis & Prevention 202, 107572.
Zhou, Y., Wang, M., Ahn, S., 2019. Distributed model predictive control approach for cooperative car-following with guaranteed local and string stability. Transportation research part B: methodological 128, 69–86.
Zhou, Z., Wang, Y., Xie, X., Chen, L., Liu, H., 2020. Riskoracle: A minute-level citywide traffic accident forecasting framework, in: Proceedings of the AAAI conference on artificial intelligence, pp. 1258–1265.

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 speed. Transportation research part C: emerging technologies 154, 104263.

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 individual preference for estimating travel time. IEEE Transactions on Intelligent Transportation Systems 24, 11438–11452.

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 Transactions on Intelligent Transportation Systems 25, 8221–8236.