Graph Attention Networks for Modeling Multi-Agent Decision Dynamics at Unsignalized Intersections

Insan Arafat Jahan
Department of Civil, Environmental and Construction Engineering, University of Central Florida
Orlando, Florida, 32816
Email: in448736@ucf.edu
ORCID: 0009-0003-5254-6021

Word Count: 3187 words + 3 table(s) × 250 = 3937 words

Submission Date: May 8, 2026

This work has not been peer-reviewed. More experiments are being conducted toward findings substantial enough to yield a journal submission.

ABSTRACT

Gap acceptance at unsignalized intersections is a safety-critical decision in which a driver on a stop-controlled approach must judge whether the available time headway in a conflicting traffic stream is sufficient to cross safely. Conventional models treat this as a single-vehicle problem, estimating a critical gap threshold from ego-vehicle kinematics alone—an approach that ignores the multi-agent reality of intersection negotiation. This paper reframes gap acceptance as a multi-agent interaction problem and applies a Spatiotemporal Graph Attention Network (ST-GAT) to model it. Using the INTERACTION Dataset v1.2 across four unsignalized intersection scenarios, we extract 24,627 labeled gap acceptance events and construct per-event scene graphs where nodes represent agents encoded with ego-relative spatiotemporal features and edges connect agents within a 10-meter spatial threshold. ST-GAT achieves AUC values between 0.793 and 0.911 across scenarios, outperforming GCN in all four and remaining competitive with logistic regression. Critically, attention weight analysis reveals three behavioral patterns consistent across scenarios: proximity drives attention allocation, non-motorized road users are disproportionately attended during gap rejections relative to acceptances, and accepted gaps are associated with faster-moving attended agents. These findings generalize across two intersection geometries and three class balance conditions, demonstrating that graph attention mechanisms can surface interpretable behavioral structure in human gap acceptance decisions that single-agent and non-attentive graph models cannot.

Keywords: Traffic Safety, Surrogate Safety, Gap Acceptance, Graph Neural Networks, Deep Learning

INTRODUCTION

Intersections account for a disproportionate share of traffic casualties in the United States, with unsignalized, stop-controlled locations responsible for nearly two-thirds of all intersection fatalities (*1*). At the center of this safety problem is gap acceptance: the decision a driver makes, in real time, about whether the available time headway in a conflicting traffic stream is large enough to cross safely. A misjudged gap acceptance is one of the most common precursors to severe intersection crashes, making accurate modeling of this decision both a fundamental behavioral science problem and a pressing engineering priority.

Classical gap acceptance models treat this as a single-vehicle problem. A driver observes an ego gap, compares it against some internal critical threshold, and accepts or rejects. This framing has produced a rich body of work built around logit models, survival analysis, and more recently machine learning classifiers trained on ego-vehicle kinematics. These approaches are tractable and interpretable, but they miss something essential: drivers at real intersections do not decide in isolation. They read pedestrians approaching the crosswalk, anticipate the speed of a vehicle still 8 meters away, and adjust their crossing decision based on the full configuration of agents in the scene, not just the gap in the primary conflict stream.

This paper reframes gap acceptance as a multi-agent interaction problem and addresses it with a Spatiotemporal Graph Attention Network (ST-GAT). Scene graphs are constructed for each decision event, with agents as nodes and spatial proximity as edges, allowing the model to jointly reason over all agents present at the decision moment. Critically, the attention mechanism produces interpretable weights that expose which agents the model attends to under accepted versus rejected gaps, enabling behavioral analysis that purely predictive models cannot support.

Using the INTERACTION Dataset v1.2 across four unsignalized intersection scenarios totaling 24,627 labeled events, we evaluate ST-GAT against logistic regression and GCN baselines and analyze attention patterns across intersection types and class balance conditions. The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the dataset and preprocessing pipeline, Section 4 presents the methodology, Section 5 reports results, and Section 6 concludes with limitations and future directions.

RELATED WORK

Gap acceptance at unsignalized intersections has been studied for decades through the lens of single-vehicle decision theory. Brilon et al. (*2*) established maximum likelihood estimation as the standard method for critical gap inference, framing the problem as a driver comparing an observed headway against a latent threshold. Subsequent work extended this framework with machine learning: Nagalla et al. (*3*) showed that random forests and support vector machines outperform logit models on multi-class gap acceptance data, while Arathi et al. (*4*) demonstrated that an ANN achieves 96.2% classification accuracy with gap duration as the dominant predictor. Li et al. (*5*) further contextualized these results using naturalistic driving data, reporting critical gaps of 5.25 s for right-turn and 6.19 s for left-turn movements. Despite their accuracy, all of these approaches model the ego vehicle in isolation, treating the gap acceptance decision as a function of gap size alone rather than as an outcome of multi-agent scene configuration.

Graph-based methods offer a principled framework for encoding that multi-agent structure. Veličković et al. (*6*) introduced Graph Attention Networks (GATs), which learn dynamic, neighbor-specific attention weights through masked self-attention, directly overcoming the uniform aggregation limitation of Graph Convolutional Networks. Diehl et al. (*7*) applied GNNs to highway

traffic prediction and demonstrated a 30% reduction in prediction error in high-interaction scenarios relative to non-graph baselines, motivating the use of graph structure for interaction-dense intersection scenes. Alghodhaifi and Lakshmanan (*8*) combined LSTM with GAT for trajectory prediction across vehicle-pedestrian interactions at unsignalized intersections – the closest architectural precedent to this work. Where their contribution targets trajectory forecasting, ours targets decision classification with explicit attention-based behavioral interpretation.

Recent traffic safety studies have also emphasized the importance of modeling heterogeneous risk mechanisms rather than treating safety outcomes as uniform across users or roadway contexts. Anowar et al. (*9*) used a comparative machine learning framework with interpretability analysis to examine injury severity in vehicle–non-motorist crashes, while Tahmid et al. (*10*) modeled heterogeneous motorcycle risk perception and crash exposure using structural equation modeling. Uddin et al. (*11*) further showed that segment-level crash risk and frequency can be jointly modeled through a copula-based framework while accounting for temporal instability. Although these studies address different crash-risk settings, they reinforce the broader need for behaviorally aware and interpretable models that can capture variation across road users, interaction contexts, and safety outcomes.

The INTERACTION dataset, introduced by Zhan et al. (*12*), provides drone-recorded naturalistic trajectories at unsignalized intersections with centimeter-resolution HD semantic maps and serves as the primary data source for this study. Schumann et al. (*13*) established a structured benchmark for gap acceptance prediction, identifying the asymmetric cost of false negatives in safety-critical decisions and motivating the weighted loss formulation adopted here. Broader infrastructure sensing work – including sensor fusion and ML-based road monitoring frameworks by the author (*14, 15*) – establishes the detection substrate on which behavioral prediction systems like ours operate, reinforcing the case for integrated, data-driven approaches to intersection safety.

METHODOLOGY

Problem Definition

We frame gap acceptance as a binary node classification problem over a scene graph. At each labeled decision event, the ego vehicle—the minor-road agent approaching a stop line—is embedded within a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of all agents sharing the same scene at the decision timestamp and $\mathcal{E}$ connects agents within a spatial proximity threshold. The model maps this graph to a binary label $y \in \{0, 1\}$, where $y = 1$ denotes gap acceptance (ego crosses the stop line) and $y = 0$ denotes rejection (ego remains stopped). The central hypothesis is that the decision is not purely a function of ego kinematics but is conditioned on the full multi-agent configuration, and that attention weights over $\mathcal{E}$ can surface which agents drive that conditioning.

Dataset and Preprocessing

We use the INTERACTION Dataset v1.2 (*12*), a collection of drone-recorded naturalistic vehicle trajectories at unsignalized intersections sampled at 10 Hz with HD semantic maps encoding stop line geometry. Four scenarios are used: two T-junctions (EP0, EP1) in East Palo Alto, CA, one all-way stop (MA), and one large intersection (GL), collectively spanning 24,627 labeled gap acceptance events across 5,165,005 frames. Table 1 summarizes the event counts and class distributions per scenario.

The overall pipeline is illustrated in Figure 1. Stop lines are extracted directly from the OSM-format semantic maps, and each agent frame is assigned to its nearest stop line. A 5 m

TABLE 1 Clean event statistics across four INTERACTION scenarios.

Scenario	Type	Events	Reject%	Accept%	pos_weight
EP0	T-junction	1,617	75.1	24.9	3.011
EP1	T-junction	1,823	80.9	19.1	4.278
MA	All-way stop	8,233	70.3	29.7	2.362
GL	Large intersection	12,954	90.1	9.9	9.085
Total		**24,627**			

decision zone approaching the stop line is defined following Li et al. (*5*), within which an agent must exhibit near-stopped behavior (speed < 1.5 m/s) for at least 3 consecutive frames to qualify as a gap acceptance candidate. An approach direction filter using rolling median trend over $K = 10$ frames removes agents not genuinely approaching the stop line. Events are labeled accept if the ego crosses the stop line and reject if it remains stopped. The labeling procedure is applied identically across all four scenarios.

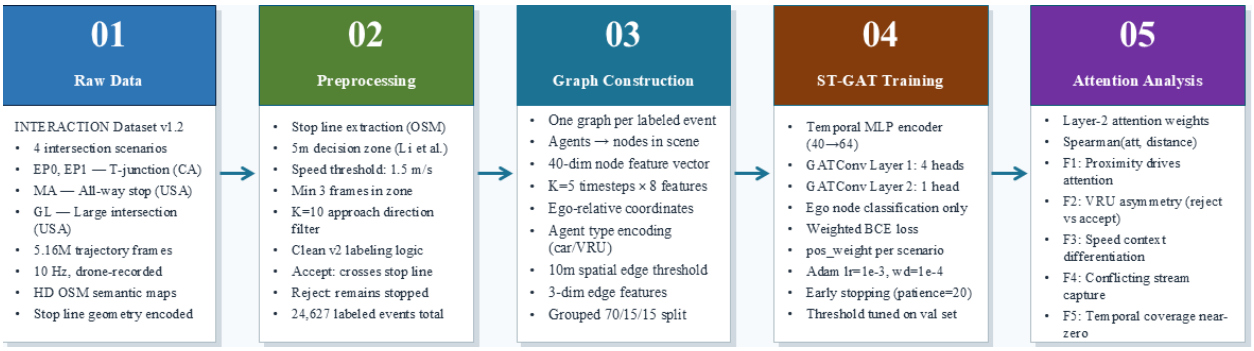

FIGURE 1 Overall methodology pipeline: from raw trajectory data through stop line extraction, event labeling, graph construction, ST-GAT training, and attention analysis.

Figure 2 illustrates stop line extraction and decision zone delineation for two representative scenarios, showing agent trajectories overlaid with extracted stop lines and labeled event positions.

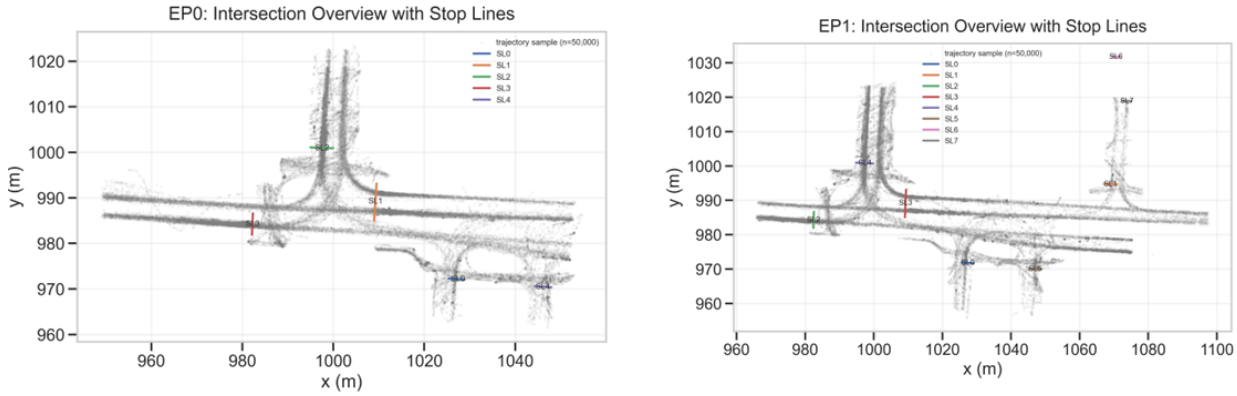

FIGURE 2 Stop line extraction and labeled gap acceptance events for EP0 (left) and EP1 (right). Agent trajectories are shown in gray; extracted stop lines are overlaid with labeled accept (green) and reject (red) decision points.

Graph Construction

For each labeled event, a PyTorch Geometric `Data` object is constructed at the decision timestamp. Each agent in the scene constitutes a node in $\mathcal{V}$. Node features are 40-dimensional, formed by flattening $K = 5$ historical timesteps of 8 per-timestep attributes: ego-relative x and y position, velocity components v_x and v_y, heading ψ, speed, distance to stop line, and agent type encoding (car = 0, pedestrian/bicycle = 1). Edges are spatial and bidirectional, connecting all agent pairs within a 10 m Euclidean threshold at the decision timestamp. Edge features are 3-dimensional: relative distance, Δv_x, and Δv_y. Node features are z-score normalized using training set statistics only. Graphs are split by `case_id` to prevent scene-level leakage, with a 70/15/15 train/val/test ratio.

Model Architecture

ST-GAT processes each scene graph in two stages. First, a temporal MLP encodes the 40-dimensional node feature sequence into a 64-dimensional representation:

$$\mathbf{h}_i = \text{MLP}(\mathbf{x}_i) = \text{Dropout}(\text{ReLU}(\mathbf{W}_1\mathbf{x}_i + \mathbf{b}_1))\mathbf{W}_2 + \mathbf{b}_2 \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^{40}$ is the flattened spatiotemporal feature vector for node i. Second, two GATConv layers with edge features refine these representations through spatial attention. For each directed edge $(j \to i)$, the attention coefficient is computed as:

$$\alpha_{ij} = \frac{\exp\big(\text{LeakyReLU}\big(\mathbf{a}^\top[\mathbf{W}\mathbf{h}_i \,\|\, \mathbf{W}\mathbf{h}_j \,\|\, \mathbf{W}_e\mathbf{e}_{ij}]\big)\big)}{\sum_{k\in\mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^\top[\mathbf{W}\mathbf{h}_i \,\|\, \mathbf{W}\mathbf{h}_k \,\|\, \mathbf{W}_e\mathbf{e}_{ik}]))} \tag{2}$$

where $\mathbf{e}_{ij} \in \mathbb{R}^3$ is the edge feature vector and $\mathbf{a}$, $\mathbf{W}$, $\mathbf{W}_e$ are learned parameters. The first GATConv layer uses 4 attention heads with concatenation (output $\in \mathbb{R}^{256}$); the second uses a single head (output $\in \mathbb{R}^{64}$) and returns attention weights for interpretability. Classification is performed exclusively on the ego node:

$$\hat{y} = \sigma(\mathbf{W}_c \mathbf{h}_{\text{ego}}^{(2)} + b_c) \tag{3}$$

where σ is the sigmoid function and $\mathbf{h}_{\text{ego}}^{(2)}$ is the ego node's second-layer representation.

Training Configuration and Baselines

The model is trained with binary cross-entropy loss weighted by the per-scenario reject/accept ratio to address class imbalance:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} [w_+ y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \tag{4}$$

where w_+ is the pos_weight listed in Table 1. The Adam optimizer is used with learning rate 10^{-3} and weight decay 10^{-4}. Training terminates when validation weighted F1 fails to improve for 20 consecutive epochs, with a maximum of 200 epochs and batch size 16. Classification threshold is tuned via grid search on the validation set.

Two baselines are evaluated against ST-GAT across all scenarios. A logistic regression (LR) baseline uses tabular features derived from the labeled event record. For GL, a leakage-corrected feature set (LR_{fair}) replaces labeling-adjacent features with decision-time graph-derived features, justified by GL's extreme 90/10 class imbalance which caused standard features to act as near-perfect label proxies. A GCN baseline uses identical graph structure and node features as ST-GAT but replaces attention-weighted aggregation with uniform neighborhood aggregation, isolating the contribution of the attention mechanism. For EP0, an LSTM baseline trained on the full ego-vehicle approach trajectory establishes the performance ceiling of ego-trajectory-only modeling.

RESULTS

Predictive Performance

Table 2 presents the full model comparison on EP0. LSTM achieves AUC 0.976 using the complete ego-vehicle approach trajectory, establishing the ego-trajectory performance ceiling. ST-GAT outperforms GCN (0.793 vs. 0.721) despite identical graph structure and features, isolating the attention mechanism as the source of improvement.

TABLE 2 Model comparison on EP0 (T-junction). LSTM is an ego-trajectory ceiling baseline, not replicated across scenarios.

Model	AUC	Wtd F1	F1_{acc}	F1_{rej}
LR	0.800	0.774	0.563	0.839
GCN	0.721	0.737	0.477	0.817
LSTM	0.976	0.911	0.824	0.938
ST-GAT	0.793	0.754	0.545	0.831

Table 3 reports cross-scenario AUC. ST-GAT leads or matches GCN in every scenario. The largest margin appears at EP1 (0.867 vs. 0.830) where scene interaction is densest. On MA, LR and GCN close the gap – consistent with MA's more balanced 70/30 class distribution making tabular features more discriminative. GL uses LR_{fair} due to feature leakage at 90/10 imbalance; ST-GAT (0.911) outperforms GCN (0.890) and remains competitive with LR_{fair} (0.915).

TABLE 3 Cross-scenario AUC. † GL LR uses leakage-corrected features (LR_{fair}); see Section 4.

Scenario	Type	LR	GCN	ST-GAT
EP0	T-junction	0.800	0.721	**0.793**
EP1	T-junction	0.726	0.830	**0.867**
MA	All-way stop	0.805	0.813	**0.844**
GL†	Large intersection	0.915	0.890	0.911

Figure 3 shows ST-GAT AUC, weighted F1, and class-wise F1 across scenarios. Accept F1 remains the hardest metric – a direct consequence of class imbalance – and improves most on MA (0.638), the most balanced scenario. Figure 4 presents test confusion matrices; the model reliably identifies rejections across all settings while accept recall remains constrained by minority-class data scarcity.

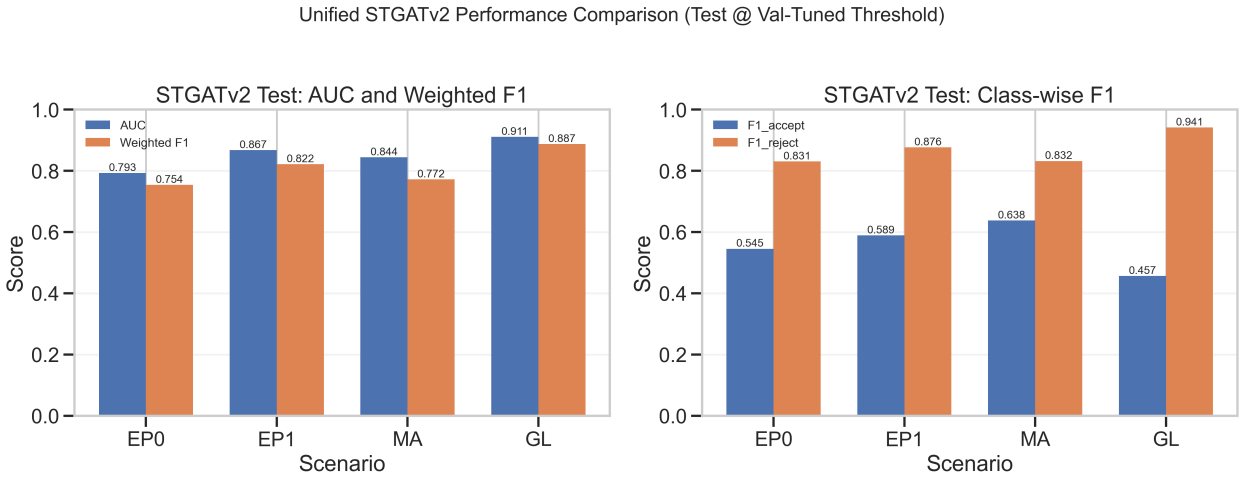

FIGURE 3 ST-GAT AUC, weighted F1, and class-wise F1 across all four scenarios.

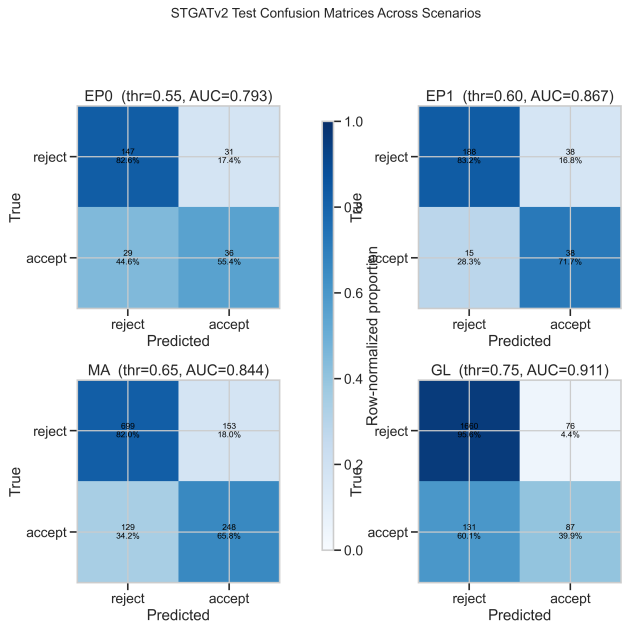

FIGURE 4 ST-GAT test confusion matrices across EP0, EP1, MA, and GL.

Attention Analysis

Figure 5 summarizes attention behavior across all scenarios. Five findings emerge consistently.

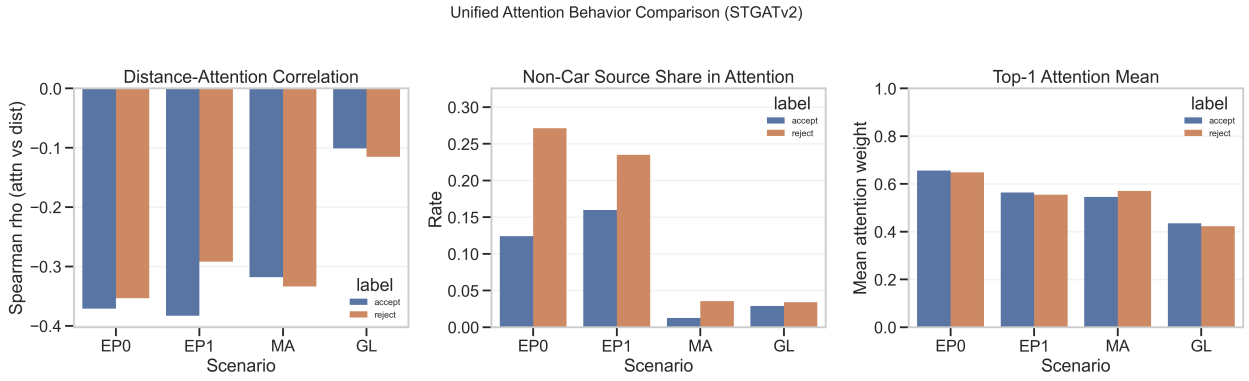

FIGURE 5 Cross-scenario attention analysis: Spearman correlation between attention weight and source distance, non-car share of top-1 attended agent by label, and top-1 attention by scenario.

Proximity drives attention: Spearman correlation between attention weight and distance is negative across all scenarios and labels (range: -0.10 to -0.38). Closer agents receive more attention in every setting.

VRU asymmetry: The top-1 attended agent is a non-motorized road user at 1.5–2.5× higher rate during rejections than acceptances. The effect is strongest at EP0 (27.0% vs. 10.8%) and consistent in direction across all four scenarios.

Speed context differentiates labels: Top-1 attended agents move faster during accepted gaps than rejected gaps across all scenarios (EP0: 2.48 vs. 1.79 m/s; MA: 2.96 vs. 1.96 m/s; GL: 3.64 vs. 1.96 m/s).

Conflicting stream capture is threshold-limited: Conflicting stream agents are rarely within 10 m at the decision moment in T-junction scenarios. The all-way stop (MA) produces more conflicting edges but no strong directional attention signal, suggesting proximity dominates over stream identity at this threshold.

Temporal coverage does not drive attention: Spearman correlation between attention weight and timestep coverage is near zero across all scenarios ($|\rho| < 0.09$), ruling out a degenerate pattern driven by observation density.

These findings replicate across two intersection geometries and three class balance conditions, confirming that ST-GAT surfaces interpretable behavioral structure inaccessible to single-agent and non-attentive graph models.

CONCLUSION

This paper reframed gap acceptance at unsignalized intersections as a multi-agent interaction problem and applied a Spatiotemporal Graph Attention Network to model it across four real-world intersection scenarios totaling 24,627 labeled events. ST-GAT consistently outperforms GCN across all scenarios, isolating attention-weighted aggregation as a meaningful contribution over uniform graph convolution. The LSTM ceiling on EP0 (AUC 0.976) establishes that ego-trajectory information is highly predictive when available; the gap between LSTM and ST-GAT reflects the trade-off between predictive dominance and multi-agent interpretability – a trade-off this work deliberately accepts in favor of behavioral transparency.

The attention analysis is the central contribution. Five findings replicate consistently across intersection geometries and class balance conditions: proximity drives attention allocation, non-motorized road users are disproportionately attended during rejections, accepted gaps are associated with faster-moving attended agents, conflicting stream capture is limited by the 10 m spatial threshold, and temporal coverage plays no role in attention assignment. These patterns are behaviorally plausible, cross-scenario consistent, and inaccessible to single-agent or non-attentive graph models.

Limitations: The INTERACTION dataset's fixed 3.9-second clip structure caps observable gap sizes well below the 5.25–6.19 s critical gaps reported in naturalistic studies, constraining the decision sequences the model can observe. Graph construction uses a single decision timestamp rather than evolving graph structure, and the 10 m edge threshold rarely captures conflicting stream agents in T-junction geometries.

Future work: Extending to continuous trajectory datasets such as CitySim would relax the clip constraint and enable full approach sequence modeling. Dynamic graph evolution across the approach trajectory, finer-grained edge thresholds calibrated per intersection geometry, and cross-cultural behavioral comparison across INTERACTION's international scenarios are natural extensions. The cross-scenario consistency of the attention findings also motivates direct application to real-time intersection monitoring systems where interpretable agent prioritization is operationally valuable.

ACKNOWLEDGMENTS

The authors thank the proprietors of the INTERACTION Dataset (*12*) for making the data available for this research.

AUTHOR CONTRIBUTIONS

This work, from conceptualization through experimentation and documentation, was conducted solely by the author.

FUNDING

None

REFERENCES

1. Federal Highway Administration, *About Intersection Safety*, 2024, accessed: 2026.
2. Brilon, W., R. Koenig, and R. J. Troutbeck, Useful estimation procedures for critical gaps. *Transportation Research Part A: Policy and Practice*, Vol. 33, No. 3-4, 1999, pp. 161–186.
3. Nagalla, R., P. Pothuganti, and D. S. Pawar, Analyzing gap acceptance behavior at unsignalized intersections using support vector machines, decision tree and random forests. *Procedia Computer Science*, Vol. 109, 2017, pp. 474–481.
4. Arathi, A., M. Harikrishna, and M. Mohan, Machine learning-based gap acceptance model for uncontrolled intersections under mixed traffic conditions. In *Conference of Transportation Research Group of India*, Springer, 2021, pp. 3–19.
5. Li, Y., H. Hao, R. B. Gibbons, and A. Medina, Understanding gap acceptance behavior at unsignalized intersections using naturalistic driving study data. *Transportation research record*, Vol. 2675, No. 9, 2021, pp. 1345–1358.
6. Veličković, P., G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
7. Diehl, F., T. Brunner, M. T. Le, and A. Knoll, Graph neural networks for modelling traffic participant interaction. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2019, pp. 695–701.
8. Alghodhaifi, H. and S. Lakshmanan, Holistic spatio-temporal graph attention for trajectory prediction in vehicle–pedestrian interactions. *Sensors*, Vol. 23, No. 17, 2023, p. 7361.
9. Anowar, P., Predicting Injury Severity in Vehicle-Non-Motorist Crashes: A Comparative Machine Learning Framework with Interpretability Analysis, 2026.
10. Tahmid, M. M., M. A. Raihan, S. Salam, A. Muktadir, and M. S. Hoque, Analysis of Heterogenous Motorcycle Risk Perception and Crash Exposure in Developing Country's Urban Driving Environment: Precursors and Policy Implications Using Structural Equation Modeling. *Available at SSRN 5800143*, 2025.
11. Uddin, A. S. M. N., M. Abdel-Aty, and C. Wang, Joint modeling of segment-level crash risk and frequency on arterials: A copula-based framework with temporal instability. *Available at SSRN 5624457*, 2026.
12. Zhan, W., L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle, et al., Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. *arXiv preprint arXiv:1910.03088*, 2019.
13. Schumann, J. F., J. Kober, and A. Zgonnikov, Benchmarking behavior prediction models in gap acceptance scenarios. *IEEE Transactions on Intelligent Vehicles*, Vol. 8, No. 3, 2023, pp. 2580–2591.
14. Jahan, I. A., A. S. Huq, M. K. Mahadi, I. A. Jamil, and M. Z. Shahriar, RoadSense: A Framework for Road Condition Monitoring using Sensors and Machine Learning. *IEEE Transactions on Intelligent Vehicles*, 2024, pp. 1–12.
15. Jahan, I. A., M. Abdel-Aty, and Z. Islam, Drone-Supervised Multi-Modal Sensor Fusion for Infrastructure-Based Vehicle Detection in Bird's Eye View. *IEEE Internet of Things Journal*, 2026.