Proposal for Virginia Transportation Safety Index (VTSI)

Jason Cusati
Computer Science
Virginia Tech
Blacksburg, VA, USA
djjay@vt.edu
(LEADER)

Cheng-Shun Chuang
Computer Science
Virginia Tech
Alexandria, VA, USA
cchengshun@vt.edu

Victory Uhunmwangho
Computer Science
Virginia Tech
Blacksburg, VA, USA
victoryu@vt.edu

***Abstract*—This project proposes the development of a dynamic, real-time Safety Index for urban intersections that integrates multimodal traffic datasets to provide actionable risk assessments. Unlike traditional crash-based approaches, which rely on multi-year historical data, the proposed framework fuses crash records, vehicle counts, vulnerable road user (VRU) volumes, speed distributions, and sensor-detected safety events into three interpretable indices: a VRU Index, a Vehicle Index, and a Combined Index, each scaled from 0–100. These indices are refreshed in 15-minute intervals to capture temporal fluctuations in risk, with computation supported by a FastAPI–PostgreSQL backend and interactive visualizations delivered through a React frontend. The system is designed to serve stakeholders such as transportation agencies, autonomous vehicle platforms, and delivery fleets by enabling real-time, exposure-adjusted safety scoring that informs routing, infrastructure investment, and Vision Zero policy goals. Through this approach, the project bridges the gap between static crash datasets and the live, population-specific decision support required for modern mobility systems.**

I. Introduction

A. Motivation

Roadway safety analysis is a critical task for transportation planning and engineering. Crash data collected by agencies such as VDOT provide a foundation for constructing quantitative indicators of roadway risk. However, the ability to act on safety risks in real time remains limited. While technology companies such as Google, Apple, and INRIX provide traffic flow visualizations, these services focus on travel time estimation rather than proactive safety assessment. Their models often lag actual conditions by several minutes, and the underlying training data can be weeks or even months old, reducing their utility for immediate decision-making. This gap highlights the need for a system that integrates crash history, traffic exposure, and real-time dynamics into a unified safety index.

B. Problem Statement

Raw crash counts alone are insufficient for assessing roadway risk, as they do not account for exposure, severity, or temporal context. Existing commercial tools provide aggregate congestion information but fail to generate actionable insights about safety risk at specific intersections or road segments. Furthermore, archived safety scoring methods often rely on multi-year data, producing results that are stale and less useful for real-time safety management [1], [2]. A data-driven approach is needed that combines crash severity, vehicle and vulnerable road user volumes, and dynamic speed variance into an interpretable 0–100 index, refreshed in 15-minute intervals.

C. Significance

Developing such a Safety Index has direct implications for both public agencies and specialized driver populations. For VDOT, an exposure-adjusted and severity-weighted index provides a transparent and evidence-based tool to prioritize interventions, support Vision Zero goals, and engage the public with clear safety metrics. For technology adopters such as autonomous vehicles, advanced driver assistance systems (ADAS), and large delivery fleets, the index can provide real-time safety context, enabling routing decisions that avoid high-risk intersections during periods of heightened danger. By delivering actionable, fresh, and population-specific insights, the proposed system bridges the gap between static crash datasets and the real-time decision engines required for modern mobility systems.

II. Literature Review: Traffic Safety, Real-Time Analysis, and Machine Learning in ITS

A. Traffic Safety Assessment

Traditional traffic safety assessment methods rely heavily on multi-year crash data to identify high-risk sites. For example, state and federal agencies typically require 3–5 years of crash records for hotspot analysis and project evaluation [3]. This retrospective approach often results in delays: safety concerns are only addressed after enough collisions have occurred. Researchers have identified this as a major limitation, leading to a “wait-for-crashes” paradigm that is slow to adapt to emerging hazards [4].

Recent studies call for a shift to proactive safety management, leveraging surrogate safety measures (e.g., hard braking, time-to-collision) and live traffic data to assess risk in real time. Hossain *et al.* describe how real-time monitoring of traffic conditions can detect precursors to crashes and enable interventions before collisions occur [4]. This vision is especially relevant for supporting Vision Zero goals, which require timely and actionable insights rather than archived reports.

B. Real-Time Traffic Analysis

Real-time traffic analysis provides the foundation for proactive safety by monitoring and responding to live traffic streams. Recent work shows how machine learning models trained on traffic flow and weather data can predict imminent crash risk. Zhang *et al.* (2022) demonstrated statewide crash risk prediction using random forests, SVM, and XGBoost on live traffic feeds combined with crash records, showing strong potential for real-time deployment [5]. Yeole *et al.* (2021) applied LSTM networks to predict crash risk at intersections, outperforming traditional regression approaches.

For emergency responders, real-time analytics are critical. Connected vehicle (CV) technologies can provide situational awareness and warnings to approaching drivers, reducing secondary collisions. Seattle's recent pilot (dCAAMP) integrates live 911 dispatch data with CV systems, creating geofenced "emergency zones" that alert nearby vehicles and navigation apps in real time [6]. Such systems directly address the gap in providing responders with immediate safety context.

C. Machine Learning in Intelligent Transportation Systems

Machine learning (ML) has rapidly become central to modern Intelligent Transportation Systems (ITS). A comprehensive review of 191 studies (2018–2023) shows ML's effectiveness in leveraging heterogeneous data sources for safety prediction [7]. Ensemble methods and deep learning models can capture complex spatiotemporal patterns in traffic streams that classical methods often miss.

Zhao *et al.* [8] introduced an AdaBoost-based risk predictor in vehicular ad-hoc networks (VANETs) with SMOTE oversampling, achieving an AUC of 0.77 in real time. More recently, Hou *et al.* [9] developed a deep transfer learning model for predicting traffic conflicts (near crashes) using trajectory data, enabling proactive warnings before actual collisions occur. Reinforcement learning has also been applied: Gong *et al.* [10] trained adaptive traffic signal controllers to minimize both delay and conflicts, demonstrating ML's ability to balance efficiency and safety in real time.

ML also addresses data imbalance through synthetic data generation. GAN-based augmentation has been shown to improve detection of rare crash events, reducing false negatives [7]. Combined with connected vehicle streams, these ML models pave the way for ubiquitous, low-latency safety scoring systems that inform drivers, responders, and infrastructure controllers in real time.

D. Summary

The literature indicates a clear paradigm shift: from retrospective, crash-based safety scoring to proactive, real-time, ML-driven systems. This transition addresses the limitations of stale insights and equips both connected vehicles and emergency responders with immediate risk information. Future work lies in scaling these models to statewide networks, fusing multiple live data streams, and ensuring robustness against imbalanced and sparse data environments.

III. Literature Review: Flow-First Strategies to Reduce Incidents

A. Safety index merging methods

The development of a Safety Index for urban unsignalized intersections can be informed by prior research in systemic safety analysis, Empirical Bayes evaluation, severity weighting, and network screening. Four key references provide complementary insights into this process.

Montella et al. [11] develop and validate a procedure to rank unsignalized *urban* intersections using a composite Safety Index (SI) derived from road safety inspections. The SI combines two components: (i) *Exposure*—the extent to which road users are exposed to hazards; and (ii) a *Risk Index*—scores for safety issues that increase crash likelihood. The SI includes separate vehicle and pedestrian components and is computed from a structured checklist of detailed and general safety issues assessed in the field. Validation against Empirical Bayes (EB) safety estimates and potential-for-improvement (PFI) on a large urban sample showed strong agreement, supporting the SI's utility for proactive screening and prioritization.

Persaud and Lyon [12] highlight the importance of the Empirical Bayes (EB) approach in before–after safety studies. EB adjusts observed crash frequencies with expected crash counts derived from Safety Performance Functions (SPFs), mitigating regression-to-the-mean bias. In a Safety Index context, EB-adjusted frequencies offer a statistically rigorous baseline for benchmarking or validating index outputs.

Schultz [1] emphasizes incorporating crash severity via weights (e.g., fatal and injury crashes weighted higher than property-damage-only). Embedding severity weighting into a Safety Index ensures locations with rarer but more severe outcomes are prioritized appropriately.

The Maricopa Association of Governments (MAG) [2] describe practical network screening using measures such as crash rate, excess crash frequency, and equivalent property damage only (EPDO). These measures, paired with exposure (e.g., traffic volumes) and site characteristics, provide scalable screening inputs to a composite Safety Index.

B. Synthesis & Gaps

Across engineering controls (C/AV smoothing, shockwave-aware signals, V2X intersections) and behavioral interventions (anti-tailgating and interaction-aware guidance), the unifying mechanism is *variance reduction*: shrink unexpected speed and lane-choice changes so drivers face fewer surprises. Open questions suitable for your project include: (i) how to prioritize corridors where small flow improvements most reduce rear-end risk; (ii) how to couple behaviorally informed messages with algorithmic flow control; and (iii) how to detect and preempt patterns like early braking waves and last-second exits using roadside sensing plus nudges (signing, lane guidance, advisory speeds).

IV. Dataset Description

The proposed Real-Time Safety Index will be constructed using *Smart Intersection datasets* collected from connected infrastructure and sensor systems at Virginia intersections (e.g., Glebe & Potomac). These datasets capture multimodal road user activity, vehicle kinematics, and safety events at fine temporal granularity, enabling dynamic risk monitoring.

A. Basic Safety Messages (BSM)

The *BSM dataset* contains vehicle-generated data packets compliant with SAE J2735 standards. Each record includes:

- Vehicle position (latitude/longitude).
- Speed, heading, acceleration, and yaw rate.
- Brake and stability control status.
- Vehicle size and type classification.

This stream allows identification of aggressive maneuvers, hard braking, and near-conflict trajectories in real time.

B. Personal Safety Messages (PSM)

The *PSM dataset* contains messages broadcasted from vulnerable road users (VRUs), such as pedestrians and cyclists. Each record provides:

- User position and movement type (walking, cycling).
- Device-reported activity state.
- Dynamic attributes such as acceleration and yaw rate.

By integrating PSM with BSM, the system detects potential conflicts between vehicles and VRUs.

C. Vehicle and VRU Counts

Dedicated counting datasets report aggregated flows:

- **Vehicle Count Dataset**: Records classified vehicle counts (passenger vehicles, buses, bicycles, etc.) by movement (e.g., left turn, through, right turn) and 15-minute time intervals.
- **VRU Count Dataset**: Provides pedestrian and bicycle volumes by approach and time interval.

These enable normalization of safety event rates by exposure, ensuring fair risk comparisons across intersections.

D. Speed Distribution Data

The *Speed Distribution Dataset* provides the number of vehicles by approach within defined speed bins (e.g., 0–5 mph, 6–10 mph). This distribution supports:

- Identification of speeding patterns.
- Exposure-adjusted severity modeling, since crash likelihood and outcome severity increase with higher speeds.

E. Safety Event Detection

The *Safety Event Dataset* logs events automatically detected by intersection sensors and video analytics, including:

- Intersection conflicts (IC).
- Lane change violations (LCV).
- Red-light running and near-miss events.
- Detection area and object classification (vehicle, pedestrian, cyclist).

These events are direct indicators of safety-critical interactions, forming the core of the real-time Safety Index.

F. Data Integration

- **Event-level fusion**: BSM and PSM are correlated with detected safety events to identify specific vehicle–VRU or vehicle–vehicle conflicts.
- **Exposure adjustment**: Safety event frequencies are normalized by vehicle and VRU counts, as well as speed distribution.
- **Temporal granularity**: Aggregation can be performed in near real-time (e.g., every 5–15 minutes) to generate live safety scores.

G. Data Limitations

Potential limitations include:

- Sensor blind spots and misclassification of VRUs.
- Limited adoption of PSM by pedestrians/cyclists carrying enabled devices.
- Data packet loss or latency in BSM/PSM transmission.
- Short-term variability in counts requiring smoothing for stable index trends.

Despite these challenges, the Smart Intersection dataset provides a rich, high-resolution foundation for constructing a **real-time, exposure-adjusted Safety Index** capable of capturing emerging risks at urban intersections.

V. Proposed Approaches

Integrating insights across these works, a systemic Safety Index for urban unsignalized intersections can be designed as follows:

1) **Systemic field assessment:** Use a structured road safety inspection to score safety issues and compute a *Risk Index* alongside *Exposure* measures, following Montella et al. [11].
2) **Severity weighting:** Apply severity weights to crash outcomes or to risk-issue scoring where crash history is sparse [1].
3) **Model-based benchmarking:** Where crash data exist, compare or calibrate against EB-adjusted estimates from SPFs to validate or tune the SI [12].
4) **Network scaling:** Embed the SI within a network screening workflow (e.g., EPDO, excess crash frequency) to prioritize sites region-wide [2].

VI. Safety Index Design

A. Data Sources and Key Risk Factors

To assess intersection safety, we leverage multiple data streams collected at smart intersections. The key inputs include:

- **Vulnerable Road User (VRU) Counts:** The number of pedestrians and cyclists present or crossing in each 15-minute interval. High VRU volumes mean more potential conflict points with vehicles [13].
- **Vehicle Traffic Volume:** The count of vehicles passing through the intersection. Heavier traffic increases exposure and the likelihood of conflicts [13].

- **Vehicle Speed Data:** The distribution of vehicle speeds. We derive metrics like average speed and speed variance. Higher speeds exacerbate crash severity for VRUs **fhwa˙vru˙speeds**, and high variance indicates unpredictable traffic flow.
- **Safety Event Logs:** Records of safety events such as near-misses or conflicts. VRU-involved events (e.g., pedestrian near-miss) are particularly critical inputs for the VRU index.

These inputs feed into three safety indices (VRU, Vehicle, Combined), each scaled from 0–100 for interpretability.

B. VRU Safety Index

The VRU Safety Index reflects risk to pedestrians and cyclists. Factors include:

- VRU-involved incidents (I_{VRU}),
- Vehicle traffic volume (V),
- Average vehicle speed (S),
- Speed variability (σ_S).

A possible formulation is:

$$\text{VRU\_Index} = 100 \times \left[w_1 \frac{I_{\text{VRU}}}{I_{\max}} + w_2 \frac{V}{V_{\max}} + w_3 \frac{S}{S_{\text{ref}}} + w_4 \frac{\sigma_S}{\sigma_{\max}} \right],$$

with $w_1 + \cdots + w_4 = 1$. Weights emphasize VRU incidents most heavily, consistent with Vision Zero priorities.

C. Vehicle Safety Index

The Vehicle Index captures risk of vehicle-vehicle collisions:

- VRU presence (N_{VRU}) as a secondary factor,
- Vehicle traffic volume (V),
- Speed variability (σ_S).

Formula:

$$\text{Vehicle\_Index} = 100 \times \left[u_1 \tfrac{N_{\text{VRU}}}{N_{\text{VRU,max}}} + u_2 \tfrac{V}{V_{\max}} + u_3 \tfrac{\sigma_S}{\sigma_{\max}} \right],$$

with $u_1 + u_2 + u_3 = 1$. Vehicle volume and speed variance dominate.

D. Combined Safety Index

The Combined Index aggregates the two:

$$\text{Combined\_Index} = W_{\text{VRU}} \times \text{VRU\_Index} + W_{\text{Veh}} \times \text{Vehicle\_Index},$$

with $W_{\text{VRU}} + W_{\text{Veh}} = 1$. This enables emphasis on VRU vs. vehicle risk as policy dictates.

E. Example Calculation

For a 15-minute interval at Glebe & Potomac:

- $N_{\text{VRU}} = 30$,
- $V = 253$,
- $I_{\text{VRU}} = 1$ near-miss,
- $S = 20$ mph, $\sigma_S = 10$ mph, with $S_{\text{ref}} = 25$ mph.

Assuming normalization constants $I_{\max} = 3$, $V_{\max} = 300$, $\sigma_{\max} = 20$ and weights $w_1 = 0.4, w_2 = 0.2, w_3 = 0.2, w_4 = 0.2$: The results are VRU_Index ≈ 56, Vehicle_Index ≈ 69, and Combined_Index ≈ 63.

F. Real-Time Computation and Backend Design

The system ingests continuous streams from sensors and connected vehicles, aggregates into 15-minute bins, and computes indices. A backend (e.g., FastAPI with PostgreSQL/PostGIS) handles:

- Stream processing and normalization,
- Index computation in near real-time,
- Exposing REST APIs for frontend visualization.

G. Frontend Visualization

UI components include:

- Interactive map with color-coded intersections,
- Day-vs-hour heatmaps showing 15-min interval indices,
- Trend charts with numeric variance indicators,
- Icons (pedestrian, congestion, speedometer) for dominant risk factors.

This multi-layered visualization supports both city-wide scanning and intersection-level drill-down.

VII. System Design

The system will adopt a client–server architecture with a clear separation of the backend and frontend components.

A. Backend

The backend will be implemented using **FastAPI**, chosen for its efficiency and native support for asynchronous I/O. Its responsibilities include:

- Connecting to a relational database (PostgreSQL or MySQL).
- Fetching raw crash and traffic data.
- Processing data to compute the Safety Index, severity-weighted rates, and related statistics (e.g., crash frequency, severity distribution).
- Providing RESTful APIs that deliver processed results in JSON format.

B. Database

A relational database will serve as the central repository:

- **PostgreSQL** is preferred for scalability and geospatial support (PostGIS).
- **MySQL** can be an alternative if project constraints favor it.

C. Frontend System Design

The frontend will be implemented using **React**, providing an interactive and user-friendly interface for exploring the Safety Index. To support geospatial visualization, several mapping frameworks may be considered, including the **Google Maps API**, **OpenStreetMap**, **Leaflet**, **Mapbox**, or **OpenLayers**. Each of these offers advantages in terms of performance, customization, and licensing flexibility, and the final choice can be guided by project constraints and visualization requirements.

The main responsibilities of the frontend include:

- **Interactive Mapping:** Visualize Safety Index values across intersections with dynamic, color-coded layers.

- **Heatmap Representation:** Apply heat maps or graduated coloring to indicate the severity of congestion or risk for specific object types (e.g., vehicles, pedestrians, cyclists).
- **Dashboards:** Present crash statistics, severity distributions, and traffic volume data in an accessible format.
- **Filtering Tools:** Enable users to filter results by location, time interval, and severity category for targeted analysis.

D. System Integration

The workflow is as follows:

1) Users access the React web app.
2) React calls FastAPI endpoints.
3) FastAPI queries the database, processes data, and computes Safety Index values.
4) Results are returned to React for visualization.

References

[1] G. Schultz, "Prioritizing safety funding using severity weighted risk scores," *Accident Analysis & Prevention*, 2025.

[2] Maricopa Association of Governments (MAG), "Network screening methodology for intersections," Maricopa Association of Governments, Phoenix, AZ, Tech. Rep., 2010, Accessed: 2025-09-18. [Online]. Available: https://azmag.gov/Portals/0/Documents/TSC_2010-07-20_MAG-Network-Screening-Methodology-for-Intersections-May-2010_.pdf.

[3] Federal Highway Administration, "Incorporating data-driven safety analysis in traffic impact analyses: A how-to guide," *FHWA-SA-19-026, US Department of Transportation*, 2019. [Online]. Available: https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa19026.pdf.

[4] M. Hossain, M. Abdel-Aty, M. A. Quddus, Y. Muromachi, and S. N. Sadeek, "Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements," *Accident Analysis & Prevention*, vol. 124, pp. 66–84, 2019. DOI: 10.1016/j.aap.2018.12.022.

[5] D. Wu, L. Xing, Y. Li, Y. D. Wong, J. J. Lee, and C. Dong, "A framework for real-time traffic risk prediction incorporating cost-sensitive learning and dynamic thresholds," *Accident Analysis & Prevention*, vol. 218, p. 108 087, 2025. DOI: 10.1016/j.aap.2025.108087.

[6] E. Bancroft, *Seattle keeps emergency scenes clear by alerting smart vehicles of incidents in near real-time*, Seattle DOT Blog (May 28, 2025), 2025. [Online]. Available: https://sdotblog.seattle.gov/2025/05/28/connected-vehicles-technology-pilot-program/.

[7] N. Behboudi, S. Moosavi, and R. Ramnath, "Recent advances in traffic accident analysis and prediction: A comprehensive review of machine learning techniques," *arXiv preprint arXiv:2406.13968*, 2024.

[8] H. Zhao, H. Yu, D. Li, T. Mao, and H. Zhu, "Vehicle accident risk prediction based on adaboost-so in vanets," *IEEE Access*, vol. 7, pp. 14 549–14 557, 2019. DOI: 10.1109/ACCESS.2019.2894176.

[9] Q. Hou, Y. Yang, J. Liang, X. Huo, and J. Leng, "A deep transfer learning approach for real-time traffic conflict prediction with trajectory data," *Accident Analysis & Prevention*, vol. 214, p. 107 966, 2025. DOI: 10.1016/j.aap.2025.107966.

[10] Y. Gong, M. Abdel-Aty, J. Yuan, and Q. Cai, "Multi-objective reinforcement learning approach for improving safety at intersections with adaptive traffic signal control," *Accident Analysis & Prevention*, vol. 144, p. 105 655, 2020. DOI: 10.1016/j.aap.2020.105655.

[11] A. Montella, C. Guida, J. Mosca, J. Lee, and M. Abdel-Aty, "Systemic approach to improve safety of urban unsignalized intersections: Development and validation of a safety index," *Accident Analysis & Prevention*, vol. 141, p. 105 523, 2020. DOI: 10.1016/j.aap.2020.105523.

[12] B. Persaud and C. Lyon, "Empirical bayes before–after safety studies: Lessons learned from two decades of experience and future directions," *Accident Analysis & Prevention*, vol. 39, no. 3, pp. 546–555, 2007. DOI: 10.1016/j.aap.2006.09.009. [Online]. Available: https://doi.org/10.1016/j.aap.2006.09.009.

[13] Federal Highway Administration, "Bicycle and pedestrian data collection in united states communities: Quantifying use, surveying users, and documenting facility extent," U.S. Department of Transportation, Washington, DC, Tech. Rep., 2016, Accessed: 2025-09-18. [Online]. Available: https://www.fhwa.dot.gov/environment/bicycle_pedestrian/publications/bikeped_data_collection/.

[14] Y. Gao and D. Levinson, "Lane changing and congestion are mutually reinforcing?" *Communications in Transportation Research*, vol. 3, p. 100 101, 2023. DOI: 10.1016/j.commtr.2023.100101.

[15] S. Mohammadi, R. Arvin, A. J. Khattak, and S. Chakraborty, "Social interactions in driving: How driver–driver interactions can lead to complex traffic dynamics," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 80, pp. 203–217, 2021. DOI: 10.1016/j.trf.2021.04.002.

[16] N. Xie and H. Wang, "Distributed adaptive traffic signal control based on shockwave theory," *Transportation Research Part C: Emerging Technologies*, vol. 173, p. 105 052, 2025. DOI: 10.1016/j.trc.2025.105052.

[17] I. Mutambik, D.-Q. Luong, Y. Zhou, B. Alotaibi, and M. S. Aziz, "Iot-enabled adaptive traffic management: A multiagent framework for urban mobility optimisation," *Sensors*, vol. 25, no. 13, p. 4126, 2025. DOI: 10.3390/s25134126.

[18] M. Khanmohamadi and M. Guerrieri, "Smart intersections and connected autonomous vehicles for sustain-

able smart cities: A brief review," *Sustainability*, vol. 17, no. 7, p. 3254, 2025. DOI: 10.3390/su17073254.

[19] S. Wang, M. Shang, M. W. Levin, and R. Stern, "A general approach to smoothing nonlinear mixed traffic via control of autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 146, p. 103 967, 2023. DOI: 10.1016/j.trc.2022.103967.

[20] M. Wu, B. Wang, Y. Yin, and J. P. Lynch, "Participatory traffic control: Leveraging connected and automated vehicles to enhance network efficiency," *Transportation Research Part C: Emerging Technologies*, vol. 166, p. 104 757, 2024. DOI: 10.1016/j.trc.2024.104757.

[21] J. W. Lee, H. Wang, K. Jang, A. Hayat, M. Bunting, and et al., "Traffic control via connected and automated vehicles: An open-road field experiment with 100 cavs," *arXiv e-prints*, 2024, Art. no. arXiv:2402.17043. DOI: 10.48550/arXiv.2402.17043. [Online]. Available: https://arxiv.org/abs/2402.17043.

[22] M. Oh, J. Shaw, and J. Dong-O'Brien, "Enhancing work zone safety: A psychological and behavioral analysis of anti-tailgating strategies using dynamic message signs and fixed warning signs," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 106, pp. 150–168, 2024. DOI: 10.1016/j.trf.2024.08.007.

[23] M. A. Arman, C. M. J. Tampère, et al., "Empirical study of lane-changing maneuvers in a weaving area based on reconstructed trajectories of floating car data," *Transportation Research Record*, 2023. DOI: 10.1177/03611981231179474.

[24] A. Shimojo et al., "How impressions of other drivers affect one's behavior when merging lanes," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 89, pp. 236–248, 2022.

[25] C. Gutierrez-Osorio and C. A. Pedraza, "Modern data sources and techniques for analysis and forecast of road accidents: A review," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 4, pp. 432–446, 2020. DOI: 10.1016/j.jtte.2020.05.002.

Appendix A
Task Assignment for Each Member

Task	**Assigned To**
Define objectives & KPIs (delay, conflict rate, speed variance)	Victory Uhunmwangho
Draft system architecture & data flow diagram	Victory Uhunmwangho
Set repo structure, naming, meeting cadence	Victory Uhunmwangho
Ingest all CSVs; standardize timestamps/IDs	Cheng-Shun Chuang
Basic data QC (missing rates, ranges, anomalies)	Cheng-Shun Chuang
Appendix C: Data Dictionary (columns, dtypes, units)	Cheng-Shun Chuang
Engineer features (vehicle/VRU volumes, speed stats, platoons)	Cheng-Shun Chuang
Link safety events to BSM/PSM (time/space bins)	Cheng-Shun Chuang
Define train/val/test time splits	Jason Cusati
Implement baselines (exposure, severity/EPDO)	Jason Cusati
Draft & freeze Safety Index spec (factors, weights, 0–100 scaling)	Jason Cusati
Create API stubs for `/index` & `/kpi`	Cheng-Shun Chuang

Task	Assigned To
Compute TTC/PET + hard-brake flags from BSM/PSM	Jason Cusati
Validate surrogate-safety metrics on held-out day	Jason Cusati
Expand API with `/conflicts` & `/meta`	Cheng-Shun Chuang
Build dashboard v1 (map, trends, 15-min heatmap)	Victory Uhunmwangho
Draft Appendix D: Methods (index formulas, TTC/PET defs, KPI math, stats tests)	Victory Uhunmwangho
Hook dashboard to API endpoints	Cheng-Shun Chuang
Sensitivity analysis of index weights/normalizers	Jason Cusati
Ablations (drop factors; re-score)	Jason Cusati
Harden data pipeline (edge cases, missing-data strategies)	Cheng-Shun Chuang
Write Ethics/Privacy/Accessibility section	Victory Uhunmwangho
Add safety guardrails (min ped clearance, safe defaults)	Jason Cusati
Dashboard v2 (legends, tooltips, export PNG/CSV)	Victory Uhunmwangho
Integrate figures/tables into main document	Victory Uhunmwangho
Results dry-run; fix narrative gaps	Victory Uhunmwangho
Repo cleanup; README “how to reproduce”	Cheng-Shun Chuang
Final proof (numbering, captions, references)	Victory Uhunmwangho
Sanity re-compute of headline KPIs	Jason Cusati
Package submission (PDF + Appendices + code link + data dict)	Victory Uhunmwangho

APPENDIX B
SCHEDULE FOR PROJECT

TABLE II
PROJECT SCHEDULE (10 WEEKS)

Week	Task
1	Literature Review & Requirements Finalization: Review Safety Index and ITS literature; finalize project scope, datasets, and technical stack (FastAPI, PostgreSQL, React).
2	Dataset Acquisition & Preprocessing: Collect Smart Intersection datasets (BSM, PSM, vehicle counts, VRU counts, speed distribution, safety events); clean, normalize, and integrate into PostgreSQL/PostGIS.
3	Safety Index Model Design: Define VRU Index, Vehicle Index, and Combined Index formulas; establish severity weighting and exposure-adjusted normalization.
4	Model Prototyping & Validation: Implement safety index calculations; validate with benchmarks (e.g., Empirical Bayes–adjusted SPFs).
5	Backend Development: Build FastAPI services; implement ingestion, aggregation, and REST APIs for index computation.
6	Frontend Development: Develop React-based visualization; implement interactive map, dashboards, and real-time updates.
7	Backend–Frontend Integration: Connect React UI to FastAPI APIs; test real-time data flows and visualization updates.
8	System Testing & Optimization: Perform unit tests, performance tests, and debugging; address latency, blind spots, and variability smoothing.
9	Refinement & Documentation: Adjust index weighting parameters; prepare technical documentation, user guide, and API reference.
10	Finalization & Presentation: Prepare final demo and walkthrough; deliver project report, presentation, and recommendations.

APPENDIX C
ACKNOWLEDGMENT

This proposal used AI for assistance in generating content and will assist in writing codes.