PACTRANS Region 10

STARLab
Smart Transportation Applications & Research

ite 2026
Western District Meeting
Bellevue,Washington • June 28-July 1, 2026

Network-Scale Safety Screening Using Connected Vehicle Trajectories:

Extraction of Surrogate Safety Measures and Crash Modeling for Signalized Intersections

Mehrdad Nasri, Jingyi He, Muhammad Monjurul Karim, Shucheng Zhang, Yinhai Wang

Department of Civil & Environmental Engineering
University of Washington

07/01/2026

Agenda

We leverage connected vehicle trajectories to extract surrogate safety measures and model crash patterns—enabling proactive, data-driven safety screening across the network.

01 | Background

Safety challenges and the need for proactive, network-scale assessment.

02 | Introduction & Motivation

Why surrogate safety measures and connected vehicle data matter.

03 | Methodology

SSM extraction, data integration, and modeling with statistical, ML, and graph-based methods.

04 | Conclusion

Key findings and implications for proactive safety interventions.

Background

Traffic safety remains a major public concern.

Limitations of traditional (reactive) approaches

01 Reactive Approach Doesn't Work → **02** Rare Crashes Lead to Sparse Data → **03** Crash-Involved Drivers Are Not Representative

01 Reactive Approach Doesn't Work

- Actions are typically taken only after crashes have occurred.
- This delay reduces the opportunity to prevent future crashes.

02 Rare Crashes Lead to Sparse Data

- Severe crash types (e.g., fatalities) are relatively infrequent.
- Limited observations restrict our ability to build robust models using crash data alone.

03 Crash-Involved Drivers Are Not Representative

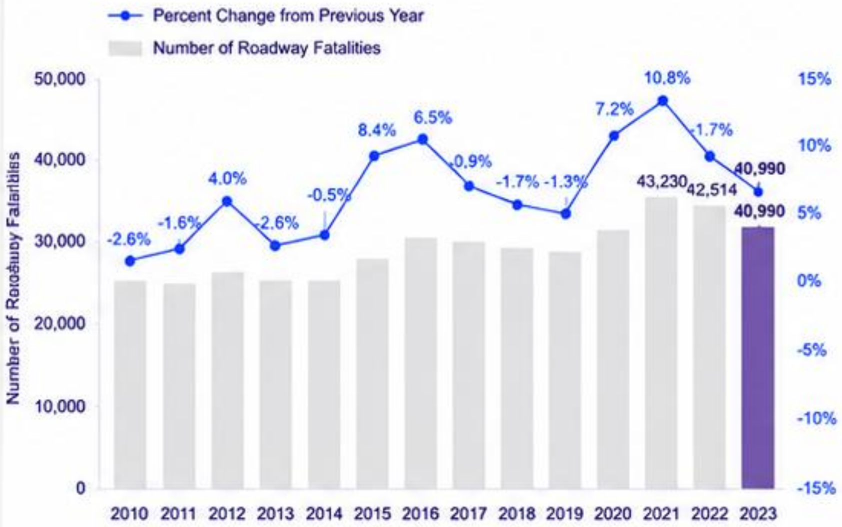
- Drivers in crashes may not represent the broader driving population.
- This can introduce bias in risk estimation if unaddressed.

How can we identify dangerous intersections *before* crashes occur?

Identifying locations with high risk **in advance** requires looking beyond crash data and using additional information.

An estimated 40,990[1] people died on America's roads in 2023

After significant increases in lives lost in 2020 and 2021, the nation started to see a decline in roadway deaths in 2022 and 2023.



Source: 2025 progress report on the National Roadway Safety Strategy, USDOT

[1] Preliminary estimate

Introduction

1. PROBLEM

Crashes are only the tip of the iceberg.

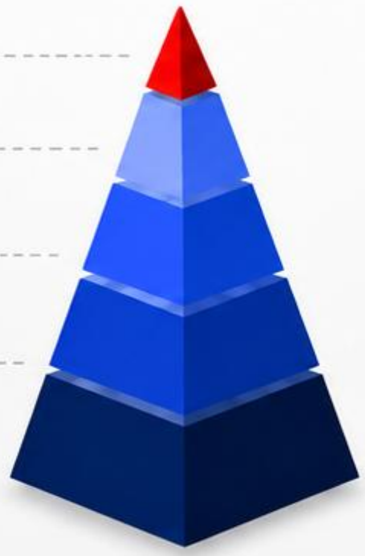
Crashes (Fatal & Serious)

Serious Traffic Conflicts

Light Traffic Conflicts

Normal Interactions

Undisturbed Passages

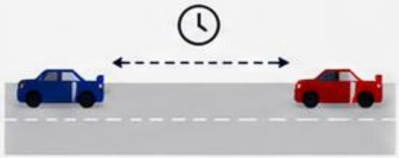


Relying on crash data alone is reactive, sparse, and incomplete.

2. APPROACH

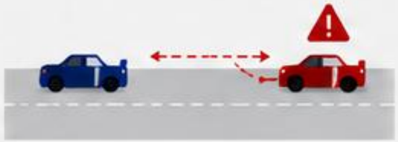
Surrogate Safety Measures (SSMs) quantify different conflict mechanisms.

TTC
Time-to-Collision
Time remaining before a potential collision.



PET
Post-Encroachment Time
Time after one vehicle clears the conflict point.

DRAC
Deceleration Rate to Avoid Crash
Required deceleration to avoid a collision.



Aggregate SSM events at each intersection and link them with crash records.

3. CHALLENGE

?

How can we **extract realistic SSMs** from **large-scale** connected vehicle **trajectories** and use them to **explain crash patterns** at signalized intersections?

Knowledge Gap & Research Questions

We need to move from **incomplete understanding** to **actionable insights**.

Knowledge Gap

Gap 1
Limited use of connected vehicle data for network-wide intersection safety analysis.

Gap 2
Most studies focus on individual SSMs rather than comparing multiple measures.

Gap 3
Relationships between SSMs, crash types, and crash severity remain unclear.

Research Questions

01 Which SSM (DRAC, TTC, PET) is most **strongly associated** with crash frequency?

02 Do SSMs explain different **crash types** and **injury severities** differently?

03 How do statistical, machine learning, and graph-based models differ in capturing **crash–SSM** relationships?

Our goal: Evaluate the ability of **surrogate safety measures** to explain crash patterns across types and severities using **multiple modeling approaches**.

Data Sources & Integration

We integrate crash records, roadway network, and connected-vehicle trajectories from the **City of Tucson** to construct a comprehensive intersection-level dataset.

3 Core Data Sources

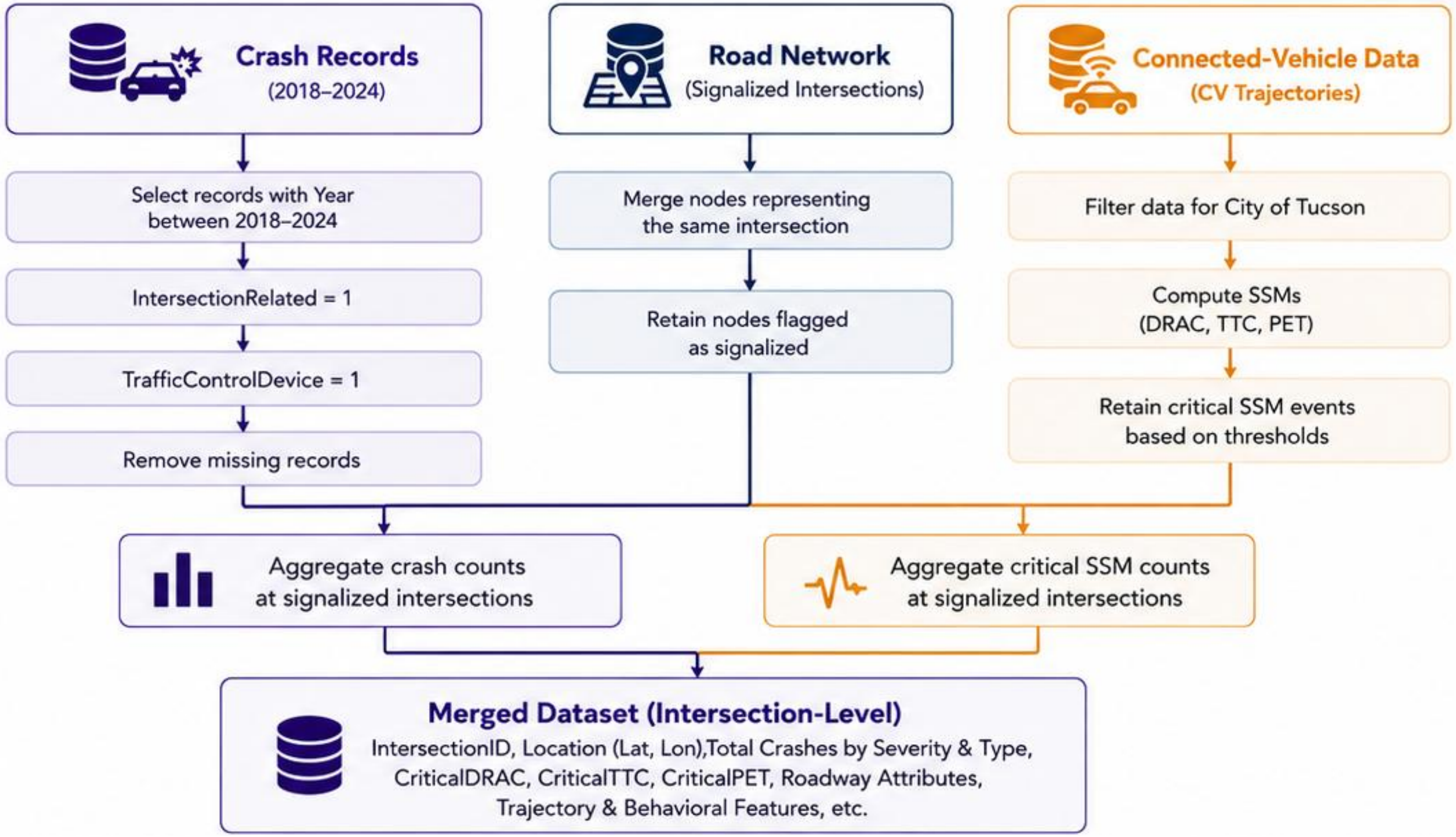
- Crash records (2018–2024)
- Road network (signalized intersections)
- Connected vehicle trajectories (CV data)

Key Processing

- Filter & clean each dataset
- Extract SSMs: DRAC, TTC, PET
- Retain critical SSM events

Integrated Output

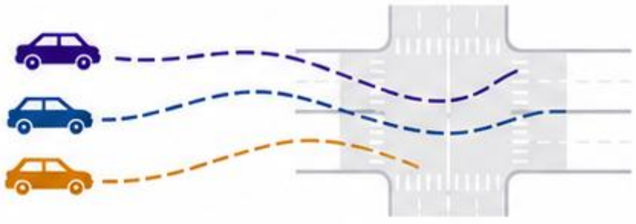
- Intersection-level merged dataset
- Crash counts + critical SSM counts
- Rich trajectory & network attributes



Framework Overview

From surrogate safety measures to crash–SSM relationship modeling.

1 Extract Surrogate Safety Measures (SSMs)

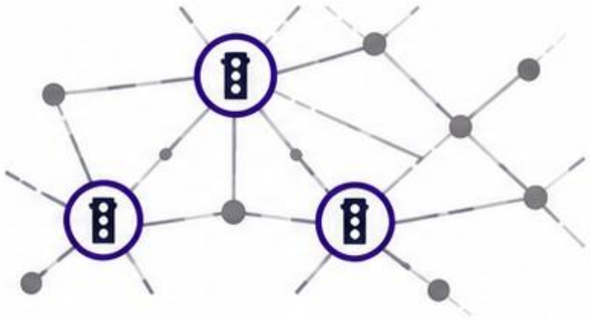


Surrogate Safety Measures

$TTC_{ij}(t) = \frac{r(t) \cdot v(t)}{\lVert v(t) \rVert^2}$	Critical TTC < 1.5 s
$PET_{ij} = t_j - t_i$	Critical PET < 1.1 s
$DRAC_{ij}(t) = \frac{\lVert v(t) \rVert^2}{r(t)}$	Critical DRAC > 3.4 m/s²

Identify critical events based on thresholds

2 Aggregate at Signalized Intersections



Crash Counts + **SSM Counts** (TTC, PET, DRAC)

Per signalized intersection

3 Model Crash–SSM Relationships

Statistical Model
Negative Binomial Regression

Machine Learning Models
Random Forest
XGBoost

Graph Neural Network
Diffusion GNN (APPNP)

Method Overview

Modeling the links between surrogate safety measures (SSMs) and crash outcomes at signalized intersections

1. Negative-Binomial Model

Quantify associations between SSMs and crash frequency.

Inputs (SSMs)

- **DRAC** (Critical)
- **TTC** (Critical)
- **PET** (Critical)
- ⋮

Outputs

- **Crash Frequency** (Overall & by Type)
- **Crash Severity** (Overall & by Level)

Statistical baseline with interpretable coefficients.

2. Tree-Based Ensembles

Capture nonlinear relationships and improve predictive performance.

Models

- **Random Forest**
- **XGBoost**

Outputs

- **Crash Frequency** (Overall & by Type)
- **Crash Severity** (Overall & by Level)

Nonlinear modeling with strong predictive accuracy.

3. Graph Neural Network

Model spatial dependencies across the road network.

Input

- **Connected Intersections**

Outputs

- **Crash Frequency** (Overall & by Type)
- **Crash Severity** (Overall & by Level)

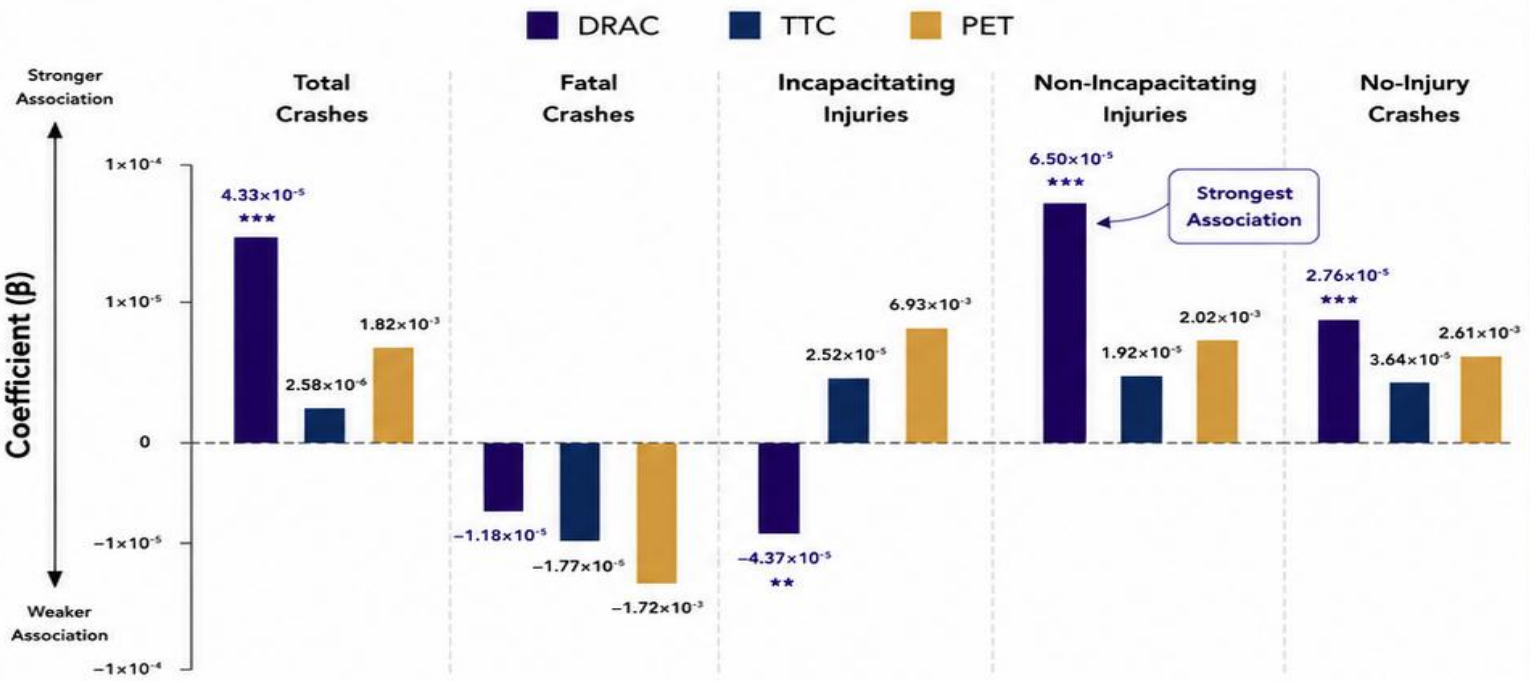
Captures spatial risk propagation beyond single locations.

NB Model – Severity Outcome

Negative binomial regression results show the association between surrogate safety measures and crash outcomes at signalized intersections.

DRAC Shows the Strongest and Most Consistent Associations

Coefficient (β) from NB model (higher absolute value = stronger association)



*Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$*

Key Takeaways

- DRAC is statistically significant for **total crashes** and **injury-related outcomes**, especially **non-incapacitating injuries**.
- DRAC shows the **strongest association** (most consistent and highest magnitude β) across most crash outcomes.
- The explanatory power of surrogate measures **decreases** as crash severity increases, with limited association observed for **fatal crashes**.

Implication

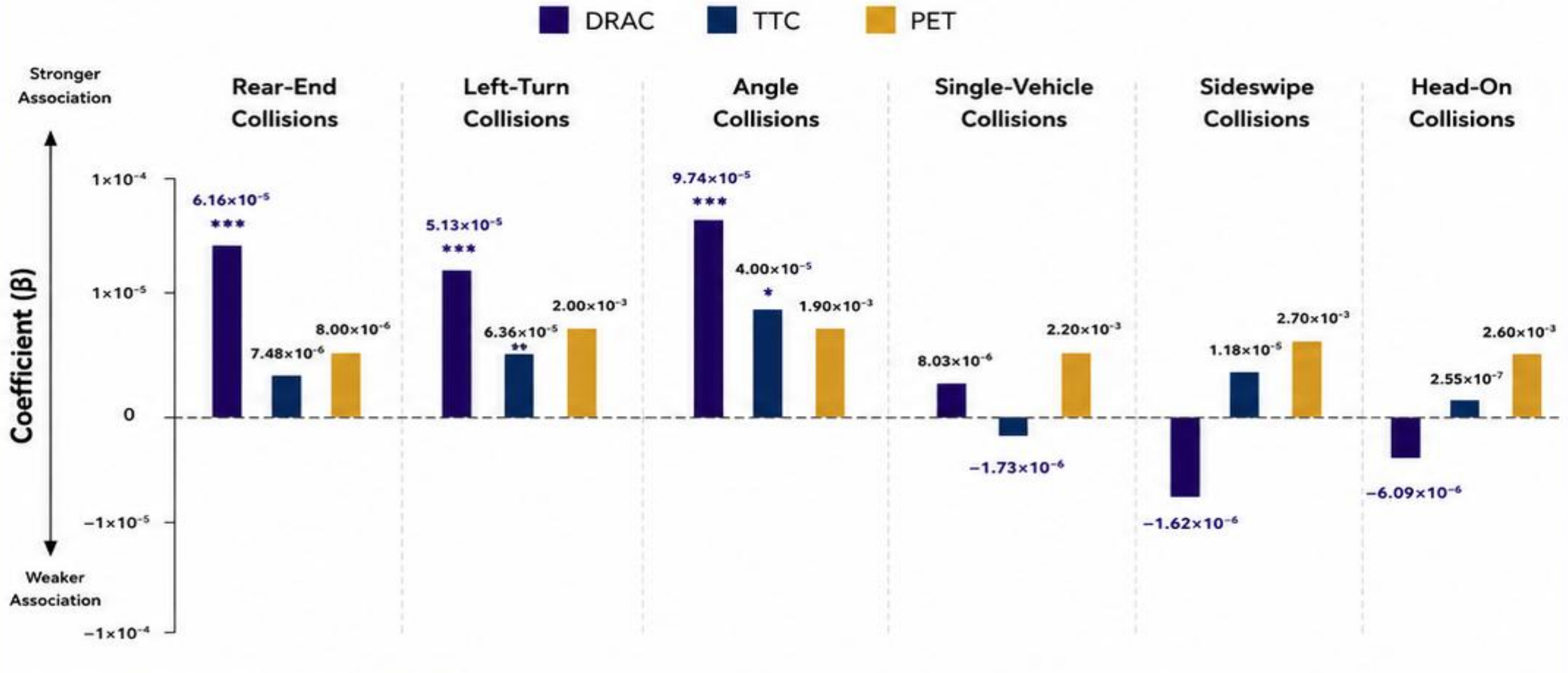
DRAC is the **most reliable** surrogate measure for identifying high-risk signalized intersections and supporting proactive, data-driven safety screening.

NB Model – Crash Types

Negative binomial regression results show the association between surrogate safety measures and crash types at signalized intersections.

DRAC Shows the Strongest and Most Consistent Associations

Coefficient (β) from NB model (higher absolute value = stronger association)



Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Key Takeaways

- DRAC exhibits the **strongest** associations with **rear-end, left-turn,** and **angle crashes**.
- TTC shows a weaker but statistically significant relationship with **angle crashes**, while PET contributes little additional explanatory value.
- Associations are substantially weaker for **single-vehicle, sideswipe,** and **head-on** crashes.

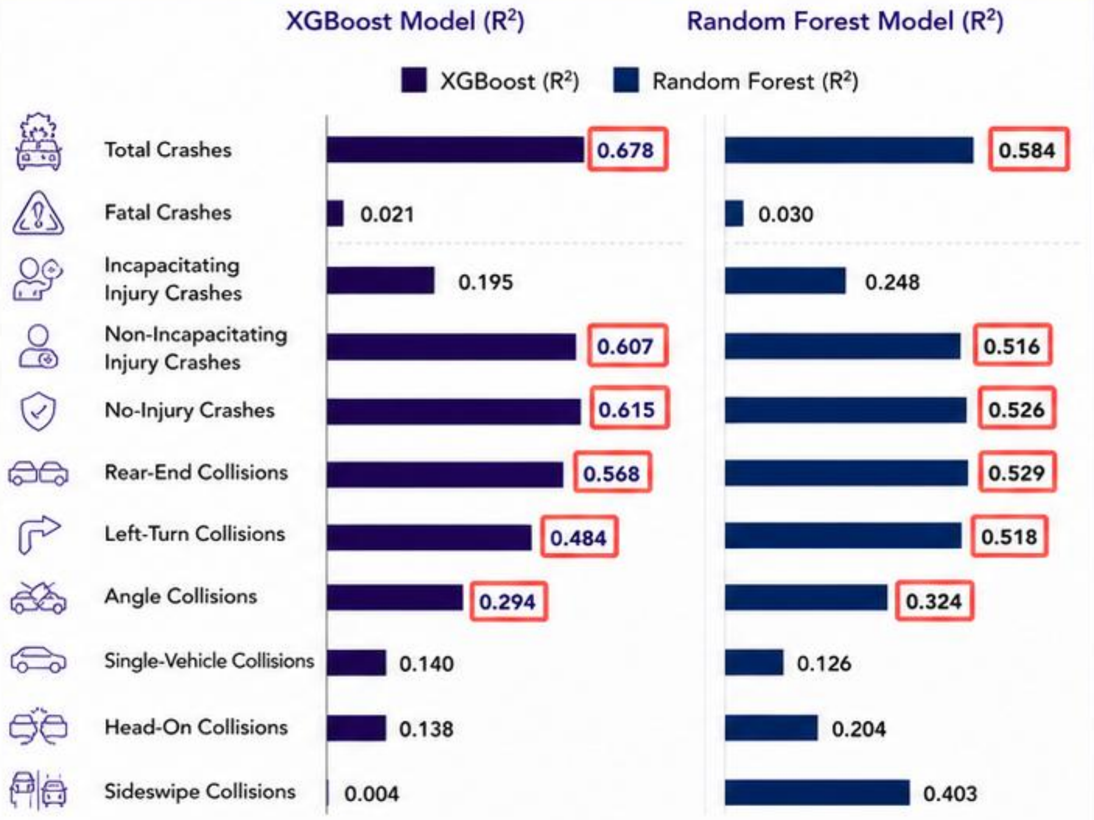
Implication

DRAC is the **most reliable** surrogate measure for identifying high-risk signalized intersections and supporting proactive, data-driven safety screening.

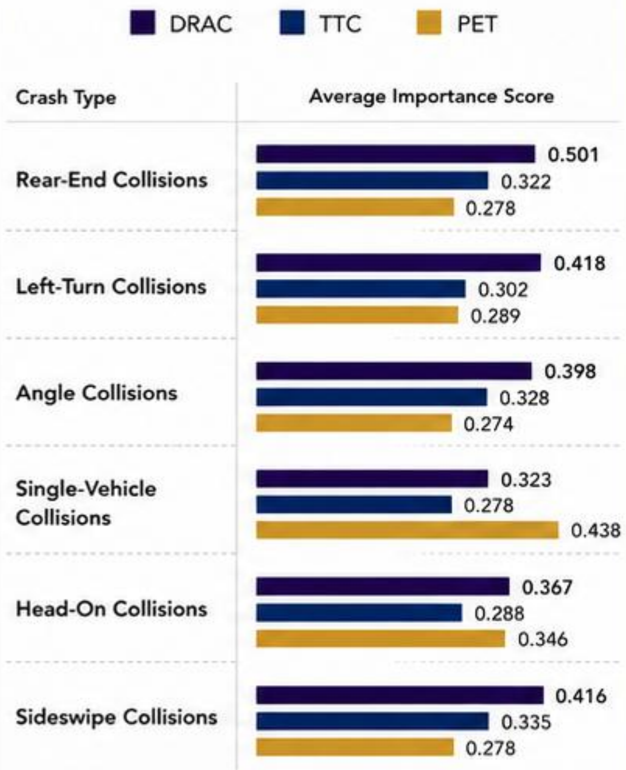
Tree Ensemble Models

Tree ensemble models (XGBoost and Random Forest) further improve predictive performance and highlight the importance of surrogate safety measures across crash outcomes.

Predictive Performance (R^2) of Tree Ensemble Models



Feature Importance (Average Across Models)



Key Takeaways

- Tree ensembles substantially improve predictive performance for high-volume crash outcomes (total, non-incapacitating, and no-injury crashes).
- DRAC consistently ranks among the most important predictors, especially for **rear-end, left-turn, and angle crashes**.
- Performance is weaker for single-vehicle, head-on, and sideswipe crashes, consistent with NB model findings.

Implication

DRAC remains the most informative surrogate in both longitudinal and crossing conflicts, while TTC provides complementary information for certain crash types.

GNN Model

Diffusion GNN analysis shows that incorporating spatial interactions further enhances the explanatory power of surrogate safety measures across crash outcomes.

Key Takeaways

- Incorporating spatial information further strengthens the influence of **DRAC**, particularly for **rear-end**, **left-turn**, and **angle crashes**.
- For **non-incapacitating** and **no-injury** crashes **DRAC** matters most and **TTC** also plays a role, implying that less-severe collisions are shaped by a blend of rapid braking and closing-gap misjudgment diffusing through adjacent intersections.
- In **angle** and **rear-end** collisions, DRAC matters most, but for **left-turn** collisions, DRAC and **TTC** produce nearly identical penalties.

Diffusion GNN Analysis Results

Change in MAE (%) (Compared to NB Model)

Crash Outcome	Δ MAE (DRAC)	Δ MAE (TTC)	Δ MAE (PET)
Total Crashes	+330.0	+17.8	−0.5
Fatal Crashes	+449.6	+137.3	+20.2
Incapacitating Injury Crashes	+29.2	+10.8	+3.4
Non-Incapacitating Injury Crashes	+86.4	+40.3	+3.1
No-Injury Crashes	+110.3	+70.8	+11.8
Rear-End Collisions	+268.9	+67.8	+0.1
Left-Turn Collisions	+165.9	+172.6	−10.8
Angle Collisions	+688.4	+532.1	−1.1
Single-Vehicle Collisions	+112.1	+38.8	+8.6
Head-On Collisions	+44.2	+16.8	+2.1
Sideswipe Collisions	−22.7	−17.8	+2.1

Note: Positive values indicate increase in MAE (%) compared to NB model. Negative values indicate decrease.

Spearman Rank Correlation

Spearman rank correlations assess the monotonic relationship between surrogate safety measures (SSMs) and different crash categories.

Spearman Correlation: Surrogate Safety Measures vs. Crash Categories

	Total Crashes	Fatal Crashes	Incapacitating Injuries	Non-Incapacitating Injuries	No-Injury Crashes	Angle Collisions	Single Vehicle Collisions	Head-on Collisions	Rear-end Collisions	Left-turn Collisions	Side-swipe Collisions
Critical DRAC Events	0.72	0.33	0.52	0.69	0.70	0.49	0.28	0.30	0.74	0.75	0.46
Critical TTC Events	0.73	0.36	0.52	0.69	0.72	0.49	0.32	0.35	0.74	0.76	0.48
Critical PET Events	0.70	0.31	0.45	0.66	0.68	0.46	0.30	0.35	0.71	0.73	0.47

Key Takeaways

- All SSMs show **moderate-to-strong positive correlations** with crash outcomes.
- The strongest associations are observed for **rear-end** and **left-turn collisions** across all surrogate measures.
- **Fatal crashes** exhibit substantially weaker correlations with all surrogate measures.

Implication

Surrogate safety measures, especially DRAC and TTC, are more indicative of common conflict-related crashes rather than rare and severe outcomes.

Final remarks

1. For Severity

Associations **weaken** as crash severity increases.

Fatal crashes are the most difficult to explain.

Incapacitating injuries show moderate associations.

No-injury crashes have the strongest associations.

2. For Type of Crashes

Rear-end, left-turn, and **angle crashes** are explained most effectively.

Single-vehicle, sideswipe, and head-on crashes show **weak** or **insignificant** associations.

Crash–SSM relationships **vary** by crash type.

3. For SSMs (Across Models)

DRAC is the **most reliable** and **consistent** predictor across models and outcomes.

DRAC — Consistently shows the **strongest** relationships with crash frequency.

TTC — Provides **complementary** information, especially for left-turn and fatal outcomes.

PET — Overall contribution is **weaker** and less consistent.

Overall Takeaway

Connected vehicle–based surrogate safety measures, especially **DRAC**, provide valuable and robust information for **proactive safety screening** and **intersection risk prioritization.**

Limitation and Future Work

While our framework demonstrates strong potential, several limitations remain. Future research will focus on addressing these gaps and extending the framework.

Limitations

- Connected-vehicle data were available for only two sampled days and a 1-second sampling interval.
- Only vehicle–vehicle interactions were analyzed; pedestrian and bicycle conflicts were not included.
- Fixed thresholds were adopted for TTC, PET, and DRAC without sensitivity analysis.

Future Work

- Incorporate roadway geometry, traffic control, and weather information.
- Explore adaptive or learning-based SSM thresholds using higher-frequency trajectory data.

Thank you!

Thank you for your attention.
We welcome any questions and discussions.

Mehrdad Nasri

mehnasri@uw.edu

Department of Civil & Environmental Engineering

Jingyi He

jhe0718@uw.edu

Department of Civil & Environmental Engineering

Access Our Paper

Scan the QR code to view the full paper for more details on our methodology, results, and analysis.