Mapping Risk: Leveraging Telematics and Machine Learning to Analyze Crash Risks at Urban Intersections

Paraskevi Koliou
Transportation Engineer, Research Associate

Together with:
Stelios Peithis, George Yannis

Department of Transportation Planning and Engineering
National Technical University of Athens

ICTR 2025

12th International Congress on Transportation Research

16-18 October 2025, Thessaloniki, Greece

HIT- HELLENIC INSTITUTE OF TRANSPORT

HELLENIC INSTITUTE OF TRANSPORTATION ENGINEERS

Introduction

Why does it matter?

- Road crashes are a significant public health issue, with over 1.19 million annual fatalities worldwide.

- Crash risk is shaped by driver behaviour (speeding, braking, acceleration), environmental conditions, and road infrastructure deficiencies.

- Urban areas concentrate a large share of crashes due to high traffic density, complex intersections, and mixed road-user behaviour.

- Intersections represent only 10–20% of the total road network yet contribute up to 50% of all urban crashes.

- Traditional safety analysis methods rely on historical crash counts, which are reactive and fail to capture real-time risk factors.

State-of-the-art

- The emergence of vehicular telematics, smartphone-based sensors, and the Internet of Things (IoT) allows for continuous monitoring of driver behaviour and traffic dynamics.

- Telematics data, such as harsh braking, rapid acceleration, and speeding events, offer fine-grained proxies for unsafe driving patterns.

- When combined with machine learning, these behavioural datasets enable predictive safety modelling, helping identify hazardous intersections before severe crashes occur.

Objectives

- **Quantify** the association between **aggressive driving indicators**, such as harsh braking, rapid acceleration, and speeding ratios, and **the frequency of crashes at intersections**.
- **Develop** a predictive modelling framework that leverages machine learning algorithms (including XGBoost, Random Forest, Gradient Boosting, and Logistic Regression) **to estimate the likelihood of crash occurrence based on telematics-derived metrics**.
- **Evaluate** the performance of these models using rigorous statistical measures (accuracy, precision, recall, and F1-score) to **determine their reliability and predictive strength**, particularly for identifying high-risk (unsafe) intersections.
- **Apply** geospatial analytics to **map and visualise spatial clusters of hazardous locations**, thereby providing actionable insights for **urban traffic safety management**.
- **Support** data-driven policy-making by demonstrating how real-time behavioural data can be used as an **early-warning tool for proactive crash prevention** and **targeted safety interventions.**

Methodology (1/2)

1. Study Area & Data Sources

- Location: **Central Athens, Greece**.
- The analysis includes **439 urban intersections** across the central municipalities of Athens, where traffic flow and driver behaviour are most heterogeneous.
- **Data inputs**:
  - **Crash data** from Greek authorities for 2019, covering all intersection-related incidents. Includes date, time, and street names, geocoded to precise coordinates.
  - **Telematics data**: 2,614 trips, 257 drivers, 8 municipalities.
  - **Road networks**: OpenStreetMap (OSM).

2. Data Processing

• Crash data were **geocoded** using **OpenStreetMap and the Overpass API** to match textual street names with **geographical coordinates**. Both datasets were **standardised under a unified coordinate system (EPSG: 2100**) for spatial compatibility.

•A **35-meter buffer zone** was created around each intersection to capture nearby telematics events.

•Through **spatial joining**, each telematics record was linked to its nearest intersection, enabling aggregation of key behavioural indicators:

- **Harsh Braking Ratio (per trip)**
- **Harsh Acceleration Ratio (per trip)**
- **Speeding Ratio (per trip)**

•Intersections were then classified based on crash history:

- **Safe (Class 0):** ≤1 crash in 2019
- **Unsafe (Class 1):** >1 crash in 2019

Methodology (1/2)

3. Modelling Approach

•The dataset was divided into **training (80%)** and **testing (20%)** subsets using **stratified sampling** to preserve class balance.

•To counter **class imbalance** (few unsafe intersections), the **Synthetic Minority Oversampling Technique** (SMOTE) was applied.

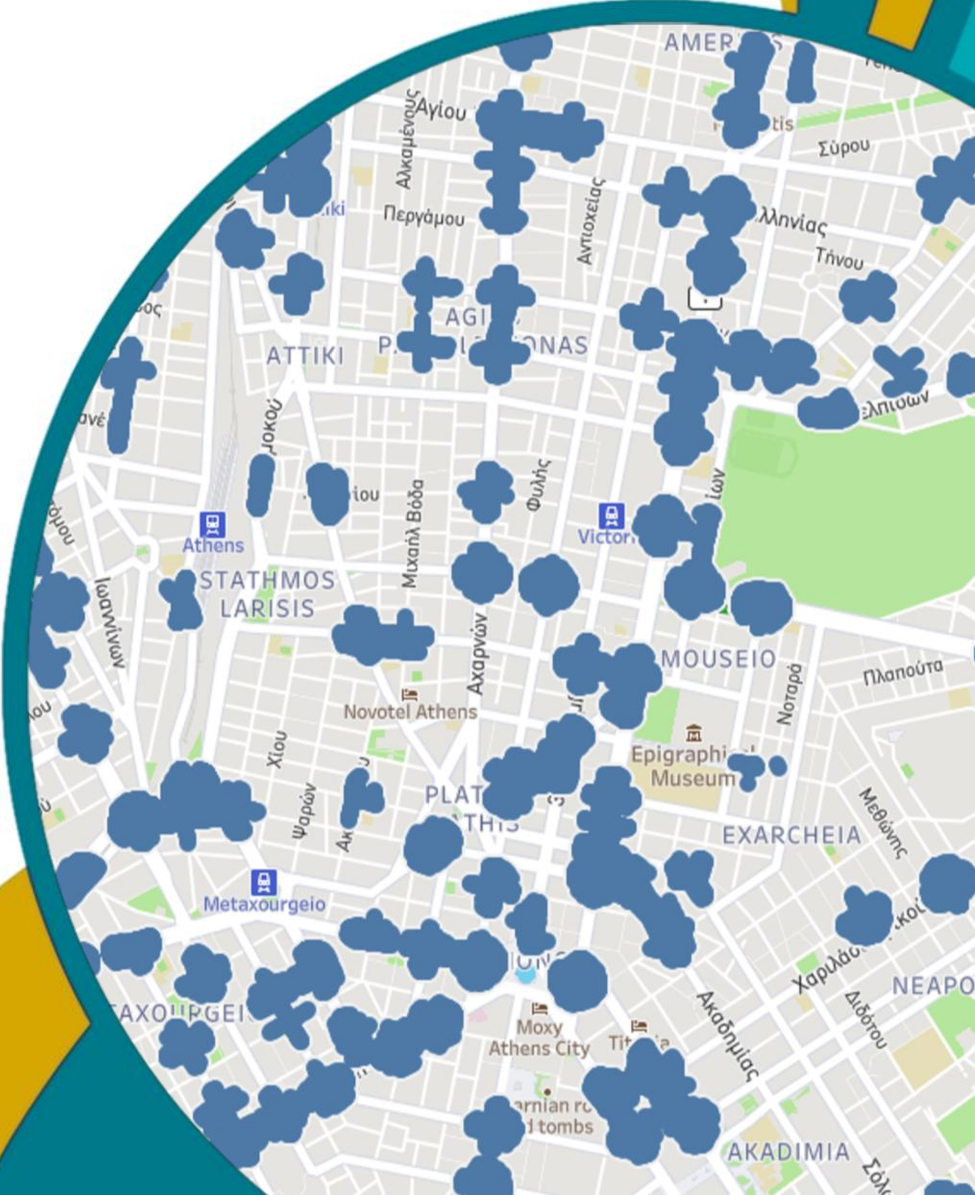
•The primary model deployed was **XGBoost (Extreme Gradient Boosting),** chosen for its ability to:

- •Handle **non-linear relationships**,
- •Manage **imbalanced data,** and
- •Deliver **high predictive accuracy.**

•Model evaluation was based on:

- •**Accuracy, Precision, Recall,** and **F1-score,** with special emphasis on **Recall** for high-risk intersections.

•Feature importance analysis identified **Speeding Ratio, Harsh Acceleration, and Harsh Braking as the dominant predictors** of intersection crash risk.



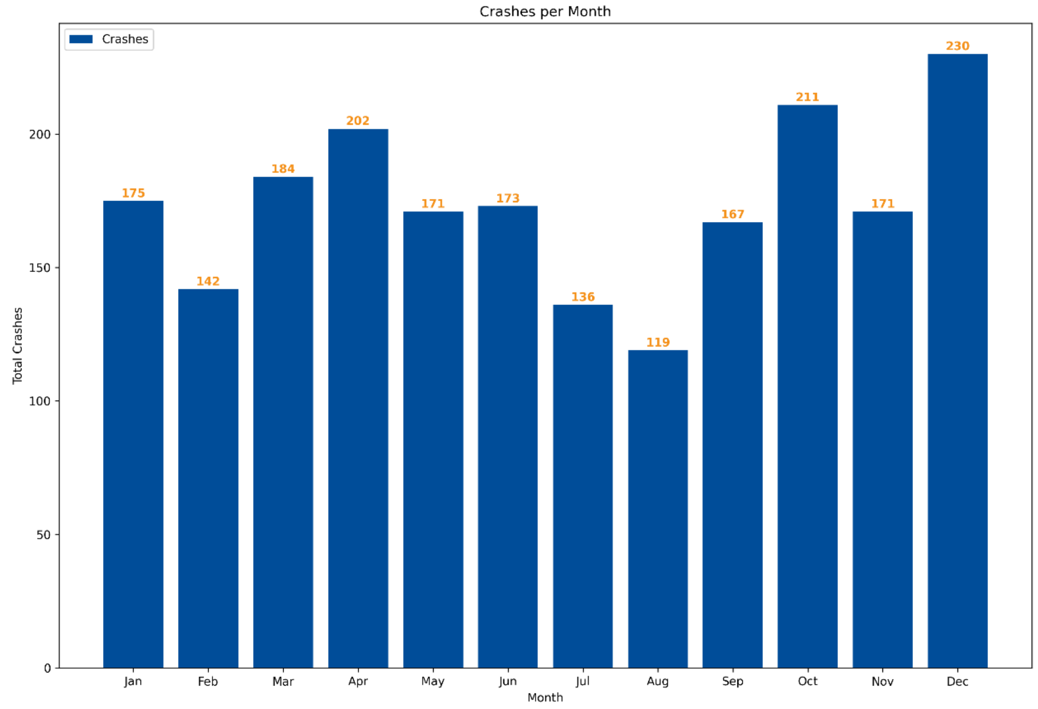
Results (1/2)

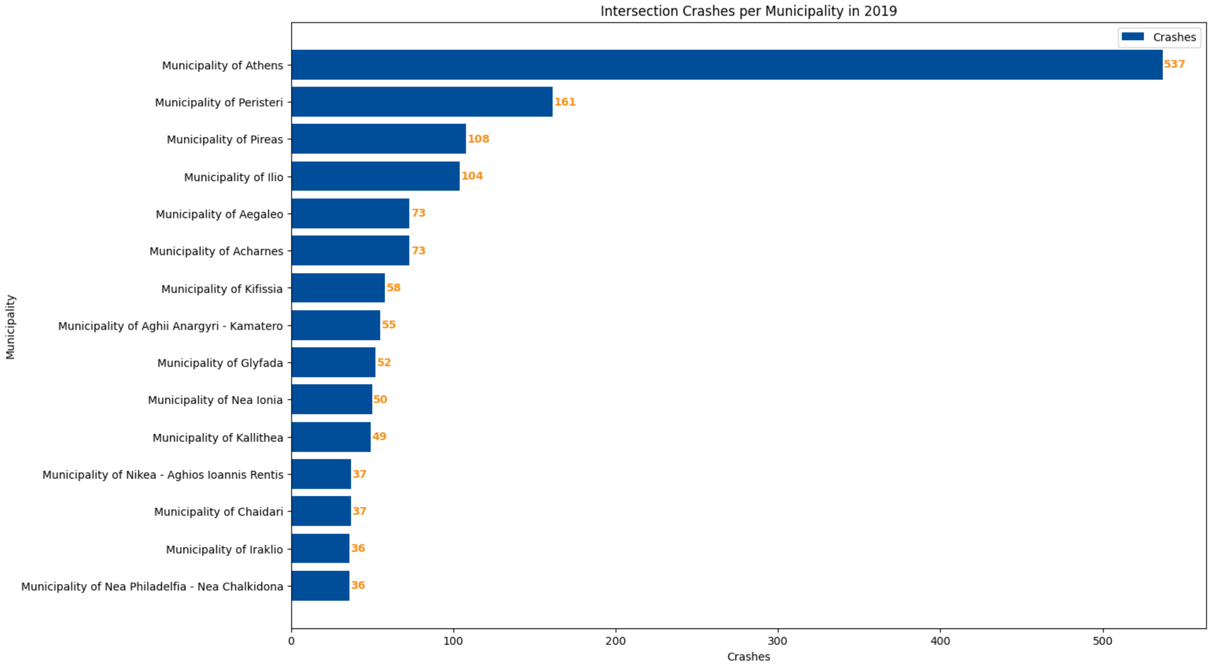
- The XGBoost classifier achieved an overall accuracy of 74%, confirming a strong predictive capacity given the data imbalance.

- Detailed performance metrics:
- Precision: 0.83 (safe class) | 0.39 (unsafe class)
- Recall: 0.84 (safe class) | 0.37 (unsafe class)
- Weighted Average F1-score: 0.74

- The model performed very well in identifying safe intersections but had lower recall for the unsafe class, a common challenge in rare-event crash modelling.

- Despite this, the framework demonstrates robust generalisation and potential for proactive risk identification when more balanced data become available.





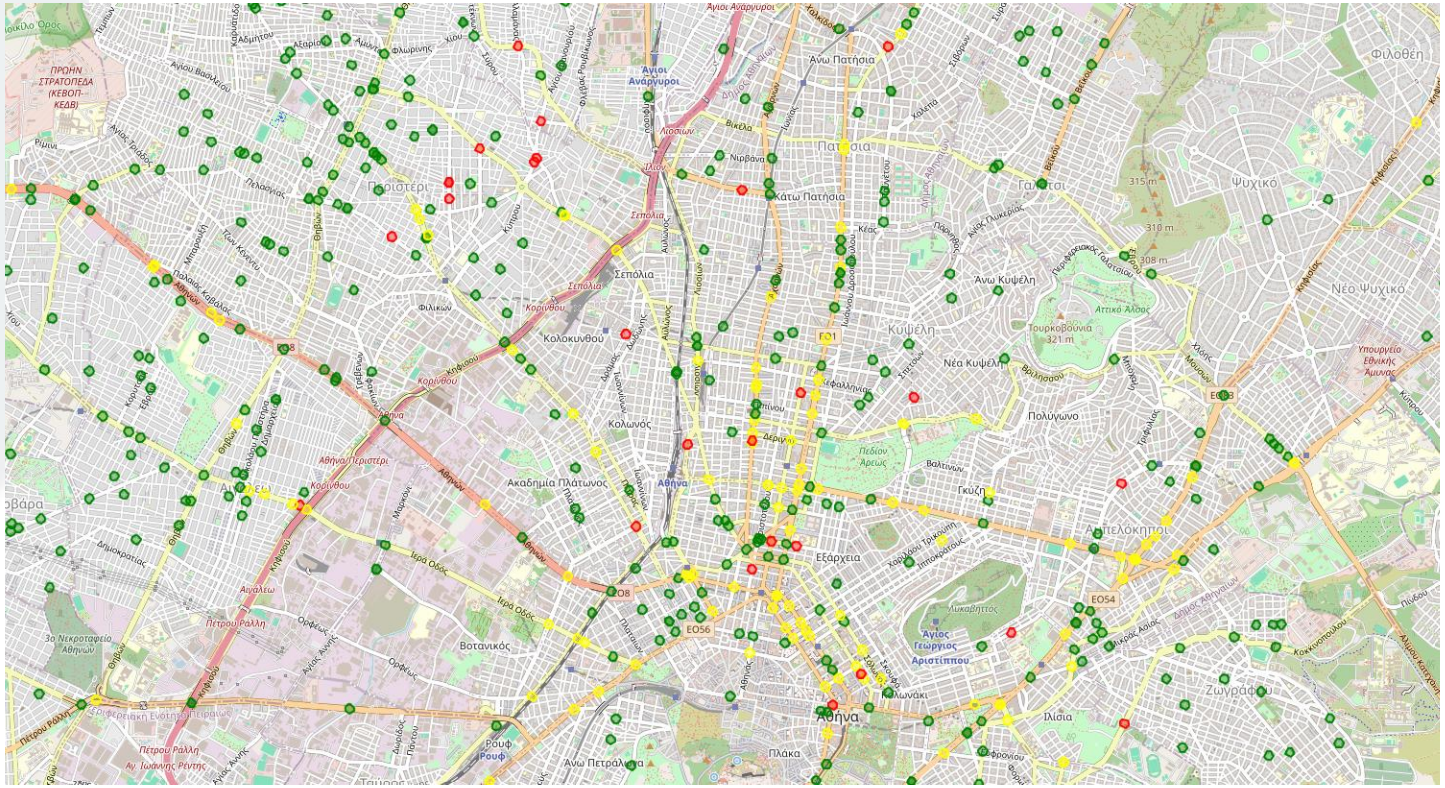
Results (2/2)

- Mapping results revealed spatial clustering of high-risk intersections within the dense urban core of Athens, where traffic complexity and mixed land uses are most intense.

- Municipality of Athens alone accounted for over half of all recorded crashes (537 incidents), illustrating the influence of urban density and intersection geometry.

- Areas exhibiting high ratios of harsh braking and speeding often coincided with intersections showing repeated crash occurrences, validating the model's capacity to identify latent risk zones.

- Some intersections with few recorded crashes but high telematics event ratios were flagged as potential "hidden danger points", demonstrating the predictive and preventive value of the approach.



Conclusions

- Unsafe driving patterns are not randomly distributed but are spatially concentrated at intersections with complex geometry, high traffic volumes, and mixed land use.
- Telematics-derived indicators provide early-warning signals of potential safety issues, even where historical crash data are limited.
- The study illustrates how machine learning can operationalise real-world behavioural data into a risk classification framework for urban safety management.
- The proposed framework enables traffic authorities and urban planners to:
  - **Identify "danger zones"** before severe crashes occur.
  - **Prioritise safety audits** and **targeted interventions** at high-risk intersections.
  - Incorporate **behavioural analytics** into **smart city systems** for continuous monitoring of traffic safety.

Mapping Risk: Leveraging Telematics and Machine Learning to Analyze Crash Risks at Urban Intersections

Paraskevi Koliou

Transportation Engineer, Research Associate

Together with:
Stelios Peithis, George Yannis

Department of Transportation Planning and Engineering
National Technical University of Athens

ICTR 2025

12th International Congress on Transportation Research

16-18 October 2025, Thessaloniki, Greece

HIT- HELLENIC INSTITUTE OF TRANSPORT

HELLENIC INSTITUTE OF TRANSPORTATION ENGINEERS