International Symposium “Navigating the Future of Traffic Management” June 29 - July 3 2025, Athens, Greece

A Geo-Spatial Analysis of Unsafe Traffic Events and Crash Occurrence at Urban Intersections: Insights from Telematics Data and Machine Learning

Stelios Peithis[1], Dr. Paraskevi Koliou[1], Prof. George Yannis[1]

[1] National Technical University of Athens, Department of Transportation Planning and Engineering, Athens, Greece

Introduction

Road safety remains a critical global issue, with over **1.19 million deaths** annually. In Europe, progress is uneven, despite initiatives like the EU Vision Zero aiming to halve road deaths by 2030.

Traditional crash-based analyses are retrospective and limited by **underreporting** and low **data granularity**, requiring crashes to occur before action is taken.

Smartphone-based telematics enable a proactive shift by capturing frequent unsafe driving events, like harsh braking, rapid acceleration, and speeding, as surrogate safety indicators.

This study explores how these unsafe events relate to crash occurrences at **478 intersections** in central Athens, leveraging:

- Telematics data from smartphone apps
- Police-reported crash records
- Machine learning models (e.g., XGBoost, Random Forests)

The goal is to classify intersections by crash risk and identify urban **hotspots** for **proactive safety interventions and data-driven policymaking.**

Objectives

1. **Integrate** telematics-based unsafe driving events (e.g., harsh braking, speeding) with crash data at urban intersections.
2. **Classify** intersections into risk levels (Low, Medium, High) using crash occurrences.
3. **Evaluate** the predictive power of unsafe events in identifying crash-prone locations.
4. **Apply** machine learning models (Random Forest, XGBoost) for crash risk classification.
5. **Compare** feature importance to determine which unsafe behaviours are most predictive.
6. **Inform** urban planning and traffic safety strategies using proactive, data-driven insights.

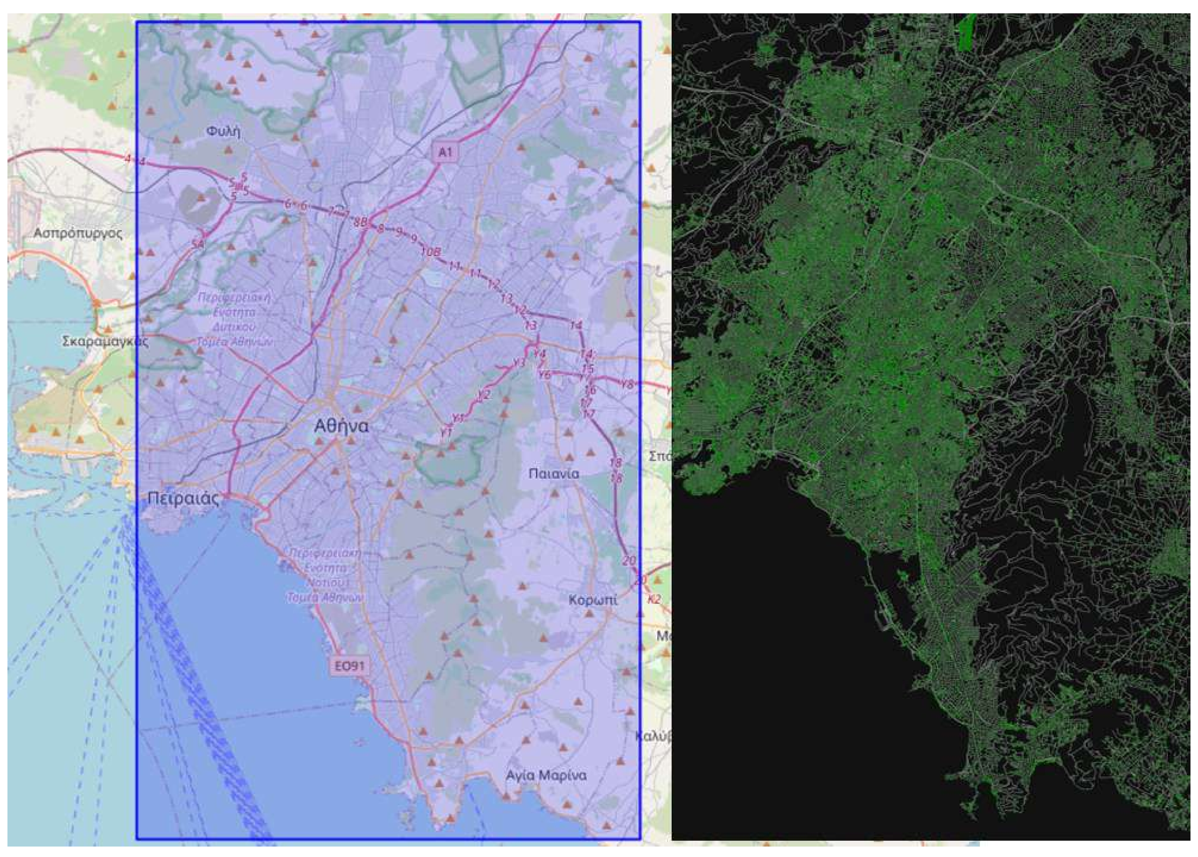

Figure 1: Study network of Athens, used for OSM data.

Methodology

1. Study Area

- Location: Central Athens, Greece
- Scope: 478 urban intersections

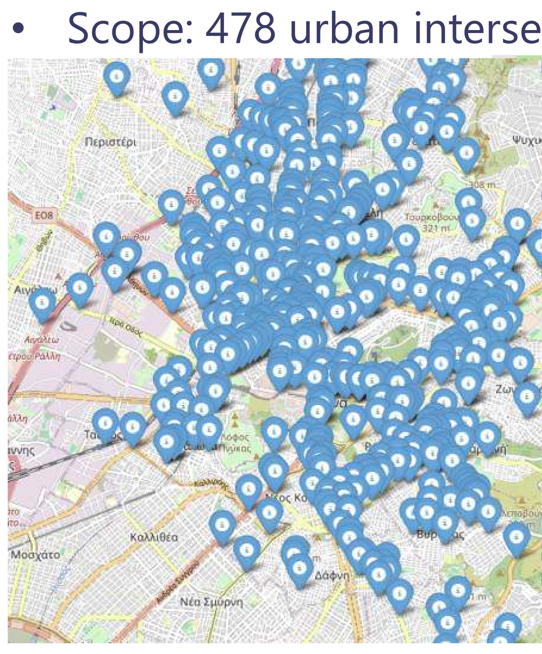

Figure 2: Plot of Crashes that occurred in intersections on a Folium Map.

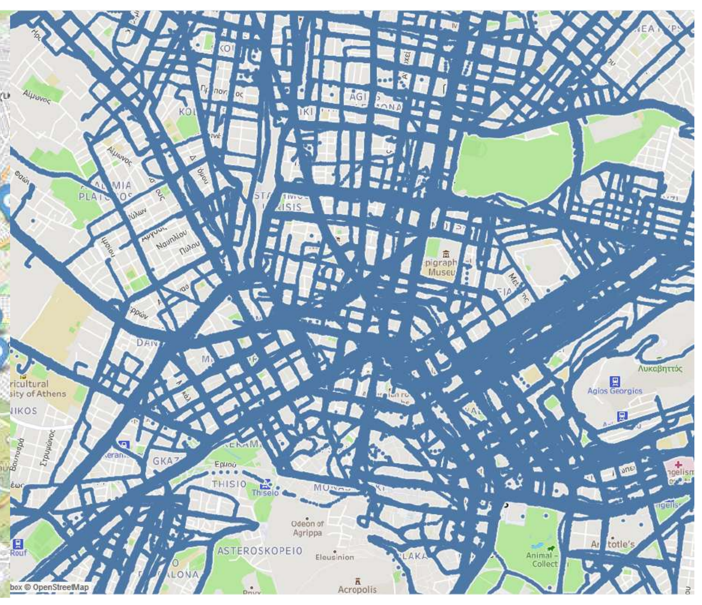

Figure 3: Telematics data collected in Athens urban roads.

2. Data Sources

- Telematics Data:
  - Collected via smartphone apps
  - 3.5M+ data points from 257 drivers (harsh braking, acceleration, speeding)
- Crash Data:
  - Greek Traffic Police records (2019)
  - Linked by intersection and street name

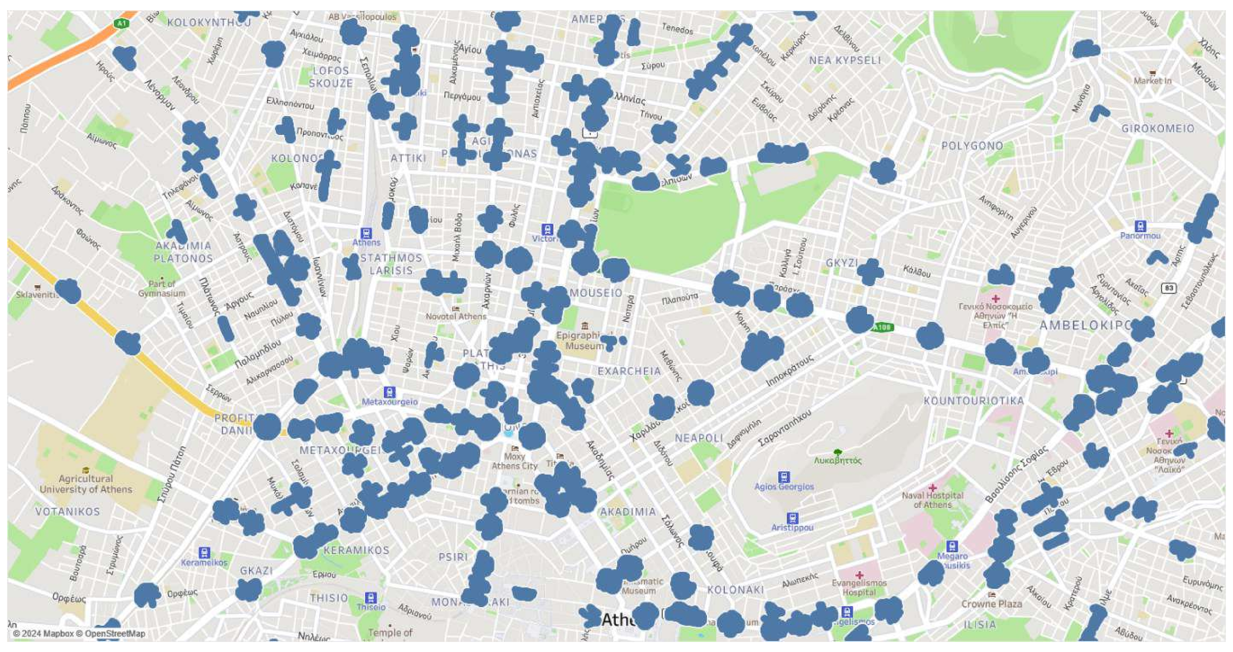

Figure 4: Telematics data assigned to intersection with centroids.

3. Data Processing

- Intersection Identification:
  - Extracted using OSMnx from OpenStreetMap
  - Intersection centroids buffered with a 50-meter radius
- Spatial Join:
  - Telematics events mapped to intersections using GeoPandas
- Aggregation:
  - Calculated unsafe driving ratios (e.g., Harsh Braking Ratio) per intersection

4. Feature Engineering

- Computed Metrics:
  - Harsh Braking Ratio, Harsh Acceleration RatioSpeeding Ratio, Number of Lanes

5. Risk Classification

- Crash risk levels:
  - Low (1 crash), Medium (2–3 crashes), High (≥4 crashes)
- Addressed class imbalance using SMOTE (Synthetic Minority Oversampling Technique)

6. Modelling Approach

- Machine Learning Models:
  - Random Forest and XGBoost
- Statistical Methods:
  - Descriptive analysis, Feature importance ranking, PCA for dimensionality reduction, DBSCAN clustering

7. Model Evaluation

- Metrics Used: Accuracy, Precision, Recall, F1-Score
- Confusion Matrix analysis to assess performance

Results

Dataset Summary

•478 intersections analyzed in central Athens
•929,543 telematics events assigned within 50m of intersection centroids
•Unsafe events quantified into:

- Harsh Braking Ratio
- Harsh Acceleration Ratio
- Speeding Ratio

Crash Risk Classification

- Intersections categorized by crash frequency:
  - Low Risk: 398
  - Medium Risk: 70
  - High Risk: 10
- Addressed imbalance using SMOTE
  - Balanced class distribution for training the model

Model Performance

Model	Accuracy	High-Risk Recall	High-Risk Precision
Random Forest	76%	50%	14%
XGBoost	80%	100%	29%

XGBoost outperformed Random Forest, especially for identifying high-risk intersections.

Feature Importance

Most predictive feature: Harsh Braking Ratio
Followed by:
Speeding Ratio
Number of Lanes (less influential)

Conclusions

1. This study demonstrates the value of telematics data in proactively identifying **crash-prone intersections** in urban environments.
2. **Harsh Braking Ratio** emerged as the **strongest predictor of crash risk**, reinforcing its role as a critical unsafe driving indicator.
3. Machine learning models, especially XGBoost, showed high potential in **accurately classifying intersections by crash risk**, particularly for high-risk locations.
4. Integrating smartphone-based unsafe driving events with crash data enables data-driven, targeted **safety interventions** - moving beyond traditional, reactive approaches.
5. These insights support urban planning, traffic safety policies, and even insurance innovations like Pay-As-You-Drive schemes.
6. The **methodology** is **scalable and adaptable** for other cities, with future improvements needed in handling imbalanced data and incorporating contextual factors like weather or traffic volume.

Acknowledgments

This project is carried out within the framework of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU (Implementation body: HFRI).

Contact Information:
Stelios Peithis, Research Associate, NTUA
Department of Transportation Planning and Engineering
Email: stelios_pithis@mail.ntua.gr