Extended Abstract

Title: *Leveraging Smartphone Telematics for Urban Traffic Safety: A Data-Driven Analysis of Unsafe Driving Events and Crash Risk*

Authors: Paraskevi Koliou, George Yannis

Affiliation: Department of Transportation Planning and Engineering, National Technical University of Athens, Greece

Background and Objectives:
The surge in smartphone-based telematics has opened new avenues in traffic safety research, enabling granular, real-time monitoring of driving behaviour. This study investigates the relationship between **unsafe driving events**, particularly harsh braking and acceleration, and **crash occurrences** in urban environments. By integrating data from **smartphone applications**, **traffic control systems**, and **digital maps**, the research aims to identify high-risk urban junctions, understand key behavioural and infrastructural risk factors, and inform targeted safety interventions.

Data and Methodology:
Data were collected via the OSeven smartphone application from over 300 drivers operating across **Mesogeion and Vouliagmenis Avenues** in Athens, capturing more than **10,000 harsh acceleration and braking events**. This behavioural data was enriched with **traffic metrics** (e.g., speed, occupancy, flow) from **26 sensor loops**, and **road infrastructure characteristics** (e.g., lane counts, entrance/exit presence) from Google Maps.
The study employed a **multi-method framework**, including:

- **Clustering Analysis**: K-Means, DBSCAN, Hierarchical Clustering, and Gaussian Mixture Models (GMM) to identify patterns of unsafe driving.
- **Machine Learning**: A Random Forest Regressor and SHAP values for feature importance and crash risk prediction.
- **Statistical Analysis**: Generalized Linear Models (GLM) and descriptive statistics to understand correlations.
- **Dimensionality Reduction**: Principal Component Analysis (PCA) for model simplification and insight clarity.
- **Spatial Analysis**: Local Moran's I and Geary's C to identify high-risk clusters and outlier locations.

To address class imbalance in crash prediction, **SMOTE** (Synthetic Minority Over-Sampling Technique) was applied.

Results:

- **Clustering** revealed three dominant junction profiles, with one cluster (Cluster 1) showing **high frequencies of harsh braking and high crash incidence**, indicating urgent need for intervention.
- **Random Forest models** identified **mean speed difference**, **braking probability (Prob_Brk)**, and **braking frequency (Mod_Freq_Brk)** as the **strongest predictors** of crash likelihood—accounting for over **85%** of model importance collectively.
- **Spatial analysis** using Moran's I and Geary's C highlighted specific junctions (e.g., JM1, JM14, JM17) as **hotspots of unsafe events**, while also revealing **spatial outliers** with abnormal patterns.

- **Multicollinearity analysis** flagged high VIF scores for braking-related features (>75), suggesting strong internal correlation and justifying the use of PCA to ensure robust modeling.

Conclusions and Implications:
This research demonstrates the **value of smartphone-derived telematics** in enhancing urban traffic safety. By combining advanced statistical and machine learning techniques, it presents a **scalable, real-time solution** for identifying high-risk locations and understanding the behavioral underpinnings of road crashes.
Key takeaways include:

- **Driver behavior**—especially harsh braking—is a **more significant crash predictor** than road design or traffic volume alone.
- Clustering and spatial analysis enable a **proactive approach** to safety interventions, rather than waiting for crash reports.
- The methodologies used can guide the **design of smarter infrastructure**, **driver feedback systems**, and **targeted enforcement** strategies.

Future work should incorporate **additional data sources** (e.g., weather, lighting conditions), **expand spatial scope**, and experiment with **deep learning** methods for more nuanced behavioral insights.