Department of Transportation Planning and Engineering
School of Civil Engineering
National Technical University of Athens (NTUA)

Unsafe Traffic Events and Crash Occurrences: The Importance of Exploring Their Relationship Using Smartphone App Data

Paraskevi Koliou*, Virginia Petraki*, Apostolos Ziakopoulos* and George Yannis*

* Department of Transportation Planning and Engineering
National Technical University of Athens (NTUA), Greece

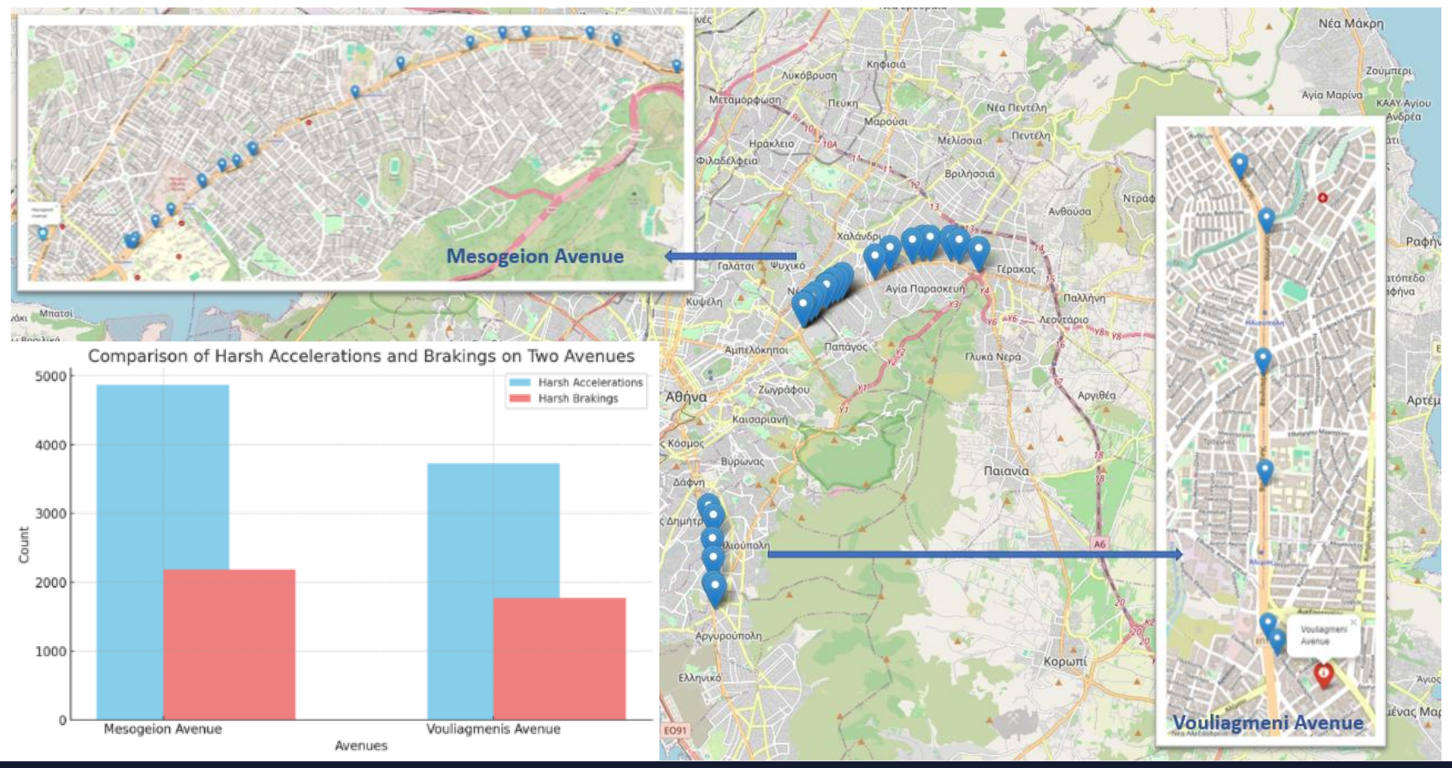
TRBAM-25-03425

Research Significance

1. **Road crashes** are a significant public health issue, with over **1.35 million** **annual fatalities worldwide**
2. Current road safety measures show **slow progress**, necessitating **new** approaches for **crash prediction and prevention.**
3. **Unsafe traffic events**, such as harsh accelerations and braking, occur **more frequently** and are **easily obtainable** using smartphone app data.
4. **Leveraging real-time data** from smartphone sensors offers a **proactive approach to traffic safety analysis and intervention.**

Data Sources

- **Driving Behavior Data:** Collected from **~300 drivers in Athens** using the OSeven smartphone app (https://oseven.io), recording instances of harsh acceleration and braking, **12,500+ events.**
- **Traffic Metrics:** Obtained from the Attica Traffic Management Center, **including traffic volume, average speeds, and occupancy rates.**
- **Road Characteristics:** Extracted from Google Maps, detailing **lane configurations** and **intersection characteristics.**



Methodology

1. **Exploring** the relationship between **Unsafe Driving Events and Crash Occurrences:** Investigate how unsafe traffic events relate to crash rates.
2. **Leveraging Smartphone Data for Traffic Safety Analysis:** Utilise smartphone app data to gather detailed insights on driving behaviour, including GPS, speed, acceleration, and braking patterns.
3. **Identifying High-Risk Areas and Patterns:** Use **clustering** and **spatial analysis** methods to detect hotspots and patterns of unsafe driving behaviour.
4. **Developing Predictive Models for Crashes:** Employ advanced **machine learning techniques**, such as Gradient Boosting, to identify key predictors of crashes and create robust predictive models.
5. **Improving Road Safety Through Targeted Interventions:** **Provide actionable insights** for designing better road safety policies, improving infrastructure, and educating drivers on safer practices.
6. **Enhancing Analytical Frameworks:** Integrate **advanced clustering**, spatial, & feature importance analyses for comprehensive, data-driven understanding of **traffic safety challenges**.

Summary of Key Techniques

Method	Techniques	Equations Used
Clustering	K-Means	$\mathrm{WCSS} = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$, where C_i is the i-th cluster, x is a data point, and μ_i is the cluster centroid.
	DBSCAN	$N(p) \geq \mathrm{min\_samples}$, where $N(p)$ is the number of points in the ε-neighborhood of p.
	Hierarchical Clustering	Distance: $d_A(x_i, x_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}$, Linkage: $d_A(x_i, x_j) = \min\{d(x_i, x_j) : x_i \in A, x_j \in B\}$.
Spatial Analysis	Local Moran's I	$I_i = \frac{z_i}{m^2} \sum_{j=1}^{n} w_{ij} z_j$, where z_i and z_j are deviations from the mean, and w_{ij} is the spatial weight.
	Local Geary's C	$C_i = \frac{1}{2m^2} \sum_{j=1}^{n} w_{ij} (x_i - x_j)^2$, where x_i and x_j are feature values, and w_{ij} is the spatial weight.
Machine Learning	Random Forest	Feature Importance: $\mathrm{Importance}(X_j) = \frac{1}{T} \sum_{t=1}^{T} I_t(X_j)$, where T is the number of trees, and $I_t(X_j)$ is the importance of feature X_j in tree t.
	Gradient Boosting	Boosting minimizes: $L(y, f(x)) = \sum_{i=1}^{n} l(y_i, f(x_i))$, where l is the loss function and $f(x)$ is the prediction function.
Dimensionality Reduction	PCA	Projection: $X_{\mathrm{normalized}} = \frac{X - \mu_x}{\sigma_x}$, Eigenvector Decomposition: Data projected on components with largest eigenvalues.

Research Results

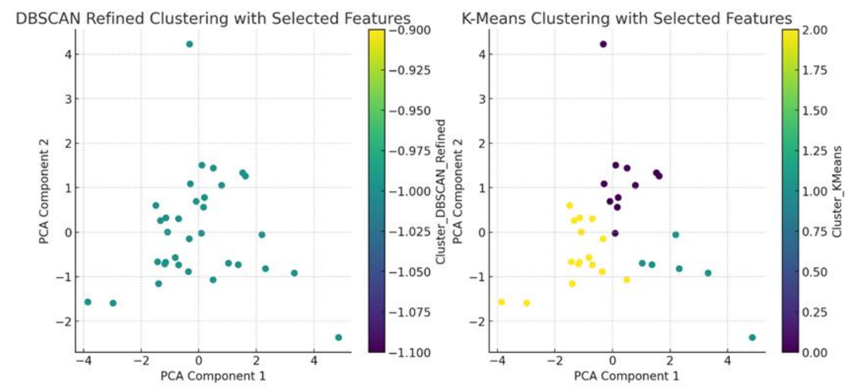


Figure 1: DBSCAN and K-Means Clustering Results

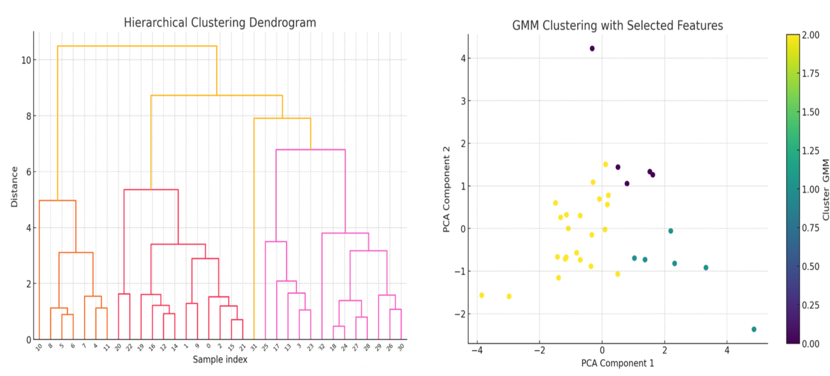


Figure 2: Hierarchical Clustering and Gaussian Mixture Models (GMM)

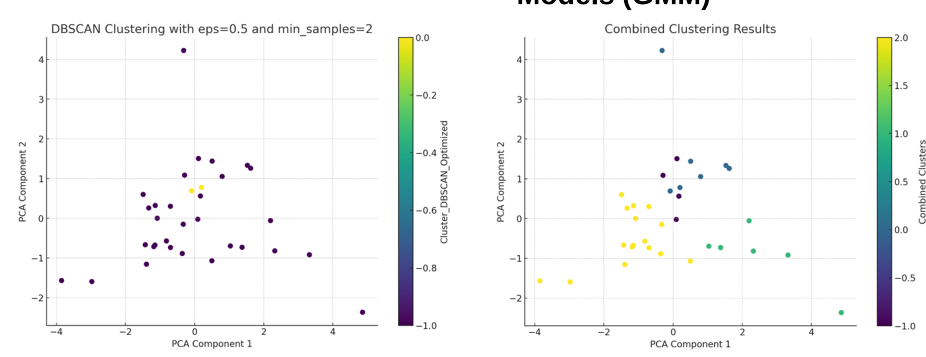
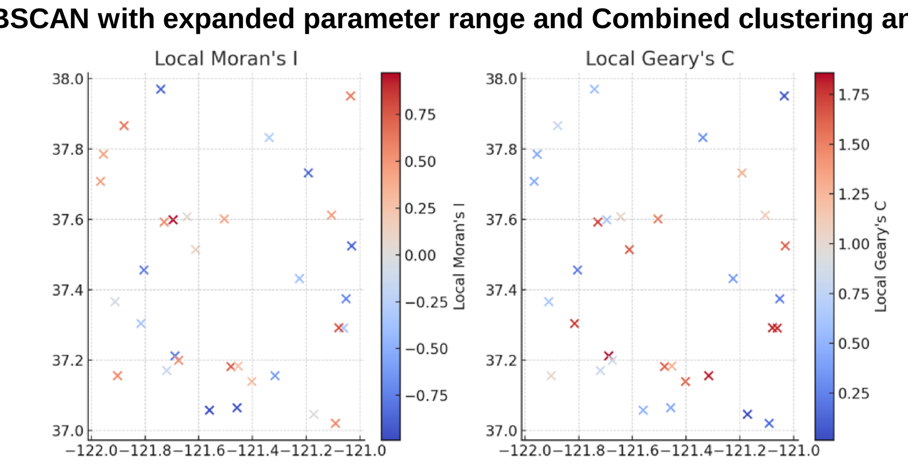


Figure 3: DBSCAN with expanded parameter range and Combined clustering analysis results



Junctions	Latitude	Longitude	UnsafeEvents	Local_Moran_I	Local_Geary_C
JM1	37.374540	-121.051114	7.618182	-0.718152	0.215783
JM14	37.212339	-121.688289	4.584000	-0.768262	1.859395
JM17	37.304242	-121.815146	3.770667	-0.338204	1.742921
JM21	37.611853	-121.105173	3.719540	0.459212	1.078684
JV9	37.065052	-121.457304	5.548628	-0.949162	0.444216

Figure 4: Significant Clusters and Outliers based on Local Moran's I and Local Geary's C values.

Conclusions

1. **Driving Behaviour:** Speed variability and **aggressive braking behaviour** (e.g., harsh braking) are **strong predictors** of crashes.
2. **Braking Metrics**: Probability of braking and frequency of harsh braking are among the **most critical factors** influencing unsafe driving events.
3. **High-Risk Areas:** Using **spatial analysis** tools specific junctions were identified as high-risk areas.
4. **Cluster Analysis:** Advanced clustering methods revealed **distinct patterns of unsafe driving events**, highlighting hotspots and spatial outliers.

Paraskevi Koliou, corresponding author
Email: evi_koliou@mail.ntua.gr