Mapping Risky Driving Behavior in Urban Road Networks

Georgia Lagou[1*], Eleni Mantouka[1], Emmanouil Barmpounakis[2], Eleni I. Vlahogianni[1]

[1] National Technical University of Athens, 5 Iroon Polytechniou str., 15773, Athens, Greece
[2] Ecole Polytechnique Fédérale de Lausanne (EPFL)
*e-mail: georgia2293@gmail.com

Abstract

Increased urbanization and widespread use of private vehicles increases the need for mobility and the quality of driving behavior in terms of aggressiveness and accident risk. An important tool in this direction is the spatial mapping and analysis of driving incidents on road hazard maps. This paper attempts to leverage smartphone based driving risk metrics to study the spatial distribution of aggressiveness and risky driving. To this end, DBSCAN and the BI-Directional-Extension-based frequent closed sequence mining (BIDE) algorithm are trained and evaluated on more than 100,000 driving trips of more than 200 drivers in the metropolitan Athens area. The analyses result to identifying the network's dangerous points based on the collective driving behavior among different users. Finally, the spatial representation of the groups resulting from the application of the algorithms on risk maps was made to serve as a tool for informing and warning the drivers of the network. The paper ends by discussing the limitations of the proposed approach and further research directions.
Keywords: *crowdsourcing; driving behavior; road safety; clustering; hazard maps*

1. Introduction

Together with the recent urbanization, people's needs for transportation are also increased. One of the main challenges for Transportation Engineers is to manage the road system effectively in order to improve users' driving behavior to achieve safe, efficient and environmentally friendly conditions. However, when it comes to identifying driving hazards, most approaches are based on police reports after an accident has occurred, making it difficult to enable precautionary actions early enough, or expensive Naturalistic Driving Studies (NDS) with instrumented vehicles, using GPS, IMUs and cameras (Vlahogianni & Barmpounakis, 2017).

Smartphones are considered as an economical manner to crowdsource information on users' mobility patterns including the manner they drive. Literature has to show numerous efforts to monitor driving behavior using smartphones (Castignani, Derrmann, Frank, & Engel, 2015; Mantouka, Barmpounakis, & Vlahogianni, 2019; Toledo, Musicant, & Lotan, 2008; Vlahogianni & Barmpounakis, 2017). The main driving indicators that have been taken into consideration are (Handel et al., 2014): harsh Braking and Acceleration, harsh Left/Right Cornering, speeding, mobile usage etc. Smartphones overcome the limitations on other telematics approaches (e.g. OBDs) due to the low cost (the users crowdsource and not oblige them to buy a telematics device), their penetration to user population, their transparent data collection mechanisms (the users may easily resort to their data to check their behavior or even

interact with them through the proper software, usually developed to work in a smartphone) (Vlahogianni & Barmpounakis, 2017).
However, smartphones are susceptible to uncertainty of the location-based information, noisy signals and battery drain (Handel et al., 2014; Paefgen, Kehr, Zhai, & Michahelles, 2012; Vlahogianni & Barmpounakis, 2017). These issues should be properly addressed to ensure the efficiency and sustainability of smartphone-based driving behavior monitoring systems.

From a conceptual perspective the ability to have complete information on how someone is driving in terms of aggressiveness and risk can have far-reaching implications to the understanding of network wide traffic and road safety – if this information is considered collectively with the existing roadway and traffic conditions, as well as weather and incidents. The spatial distribution and further visualization accident and of extreme driving phenomena could hint to a unified problematic condition in the road network that should be alleviated taking proactive traffic and safety measures (Han et al., 2015). To this end, (Chau, Sato, Kubo, & Namatame, 2015) proposed a method for generating traffic safety maps based on differences in individual recognition of the road environment by using smartphone data from various users. In (Han et al., 2016), authors analyzed the road accidents hotspots based on natural nearest neighbor clustering algorithm. Hotspot identification through clustering has been also followed in (Anderson, 2009; Han et al., 2016; Ouni & Belloumi, 2019; H. Wang, De Backer, Lauwers, & Chang, 2019; Zeng et al., 2018).

A thorough look at the literature demonstrate that little attention has been placed in the proactive spatial analysis of near misses, as an indication of possible high accident risk areas. The driving metrics have not been introduced in the analysis of aggressive and risky network locations. To fill this gap, this paper aims to provide a spatial mapping of extreme driving behavior in order to inform and warn the drivers and decision-makers in time. In this paper, crowdsourced data are used to identify the dangerous areas prior to a potential accident. Following, spatial clustering algorithms are applied and critical areas where frequent extreme driving behavior occurs are identified. The results of the clustering algorithms are then grouped spatially and visualized in order to produce the final hazard maps.

2. Methodological Approach

The approach attempts to identify locations with similar frequency of harsh driving events and speeding occurrence based on clustering. Conceptually, harsh driving events (such as acceleration and braking) have very short duration (e.g 1-2 seconds) and emerge as a point location in the dataset. This is also because the frequency of collecting the data in the specific experiment is set to 1Hz for energy efficiency. On the other hand, speeding emerges as an event that have spatio-temporal extent, meaning it occurs in a specific location and last for several minutes. In the specific paper, these differences are treated with a different clustering policy explained below.

For the case of single point events, such as the harsh acceleration and harsh braking a DBSCAN algorithm is implemented. DBSCAN is a typical spatial clustering algorithm able to identify arbitrary-shaped clusters in any database D and at the same time to distinguish noise points (Ester, M., Kriegel, H. P., Sander, J., & Xu, 1996). DBSCAN accepts a radius value Eps(ε)

based on a user defined distance measure (Manhattan Distance, Euclidean Distance etc.) and a value MinPts for the number of minimal points that should occur within Eps radius. A cluster is created, if the total number of the neighbors around a point p is greater than MinPts. The minPts is often set to be dimensionality of the data plus one or higher. The knee in kNN distribution plot can be used to find suitable values for eps. DBSCAN separates data points into three classes: i. Core points that are at the interior of a cluster (Centre), ii. Boarder points that fall within the neighbourhood of a core point which is not a core point, and iii. Noise points, namely those points that are neither core, or boarder points.

DBSCAN does not require predefining the number of clusters in the data and can result to arbitrarily shaped clusters, is robust towards outlier detection (noise). Moreover, it requires just two points that are very insensitive to the ordering of the points in the database. On the other hand, DBSCAN does not work well on high-dimensional data in general or data sets with varying densities. DBSCAN is sensitive to the distance metric selection, as well as the selection of parameters MinPts, Eps. Consequently, the understanding of the physical aspects of the problem at hand is critical.

The case of identifying speeding clusters, a sequence pattern mining approach is implemented. The problem of sequential pattern mining was proposed by (Agrawal & Srikant, 1994), as the problem of mining interesting subsequences in a set of sequences (e.g. 0,1,1,1,1,0,0) "Given a set of sequences, where each sequence consists of a list of elements and each element consists of a set of items, and given a user specified *min_support* threshold, sequential pattern mining is to find all of the frequent subsequences, i.e., the subsequences whose occurrence frequency in the set of sequences is no less than *min_support*". The support (or absolute support) of a sequence s_a in a sequence database S is defined as the number of sequences that contain s_a, and is denoted by $\sup(s_a)$. In other words, $\sup(s_a) = \left|\{s \mid s \subseteq s_a \wedge s \in S\}\right|$. Support may be defined as a ratio relative support): $relSup(s_a) = \sup(s_a) / |S|$, that is the number of sequences containing s_a divided by the number of sequences in the database. For example, the relative support of the itemset $\langle\{b\},\{f,g\}\rangle$ is 0.5.

To provide a computationally efficient solution, we implement the BI-Directional-Extension-based frequent closed sequence mining (BIDE) algorithm (J. Wang, Han, & Li, 2007). BIDE is based on the concept of closed sequential patterns (set of sequential patterns that are not included in other sequential patterns having the same support). It is considered a pattern-growth algorithm, which avoids the curse of the candidate maintenance-and-test paradigm. BIDE implements the BackScan pruning method to check the pattern closure in a more efficient way while consuming much less memory in contrast to the previously developed closed pattern mining algorithms. It does not need to maintain the set of historical closed patterns; thus, it scales very well in the number of frequent closed patterns.

The Data

The dataset was collected using the OSeven smartphone application (www.oseven.io), developed for both iPhone and Android devices that recognizes driving activity without any user involvement. Using various criteria, the application starts to collect raw data from

smartphones. In order to serve the scope of this paper, driving data from more than 500 drivers in Athens, Greece is leveraged. This anonymized data consists of detailed trip information, as well as the observed driving events, namely harsh accelerations, harsh breakings and speeding that relate to aggressiveness, lack of anticipation and the degree of risky driving respectively (Mantouka et al., 2019). The events are distinguished according to whether they occur at a specific time point or whether they occur over an interval during the trip. In total, the final dataset consists of more than 6000 events. Figure 1 and 2 depict the distribution of harsh driving events and of speeding respectively.

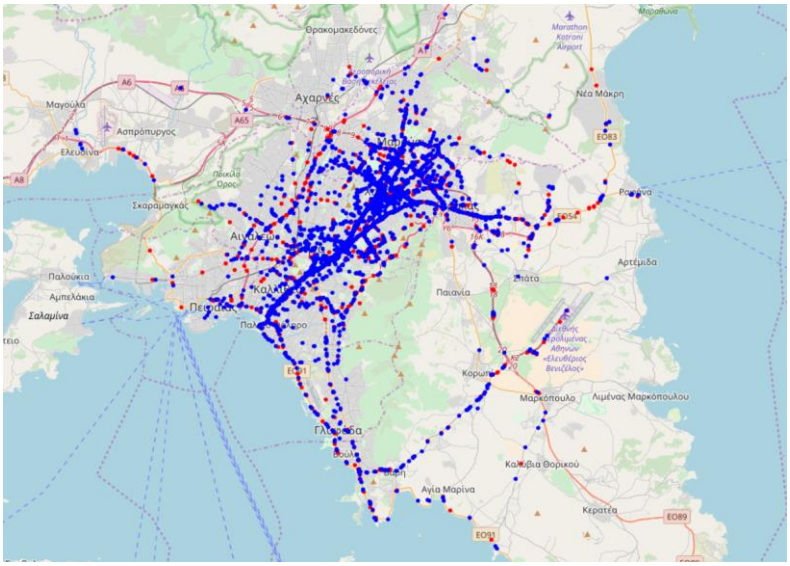

Figure 1: *Distribution of harsh events (acceleration and braking) for the entire trip dataset.*

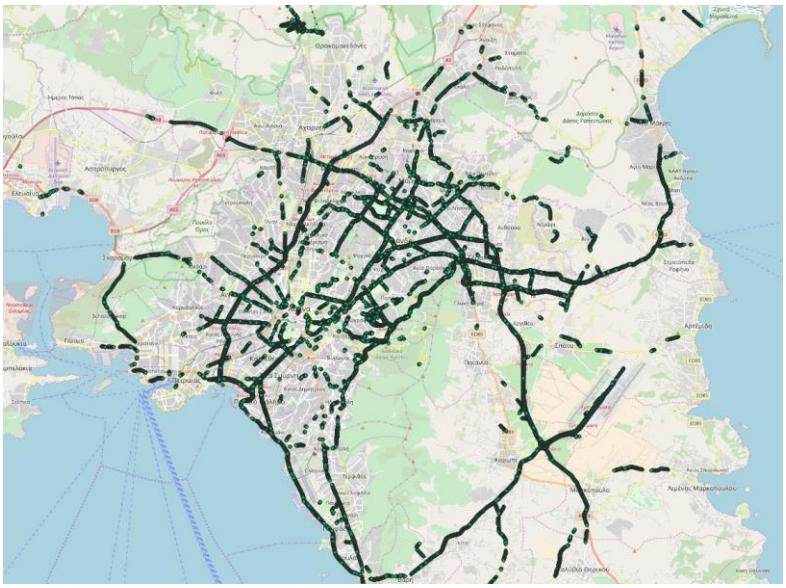

Figure 2: *Distribution of speeding events for the entire trip dataset*

3. Findings

The number of clusters that are formed using the DBSCAN algorithm relate to the parameter minPts of the algorithms as can be seen in Figure 3. In order to identify the clusters for harsh acceleration events, we choose 40 as the threshold value for the minPts per cluster, while the threshold value minPts for harsh brakings is 20 points per cluster.

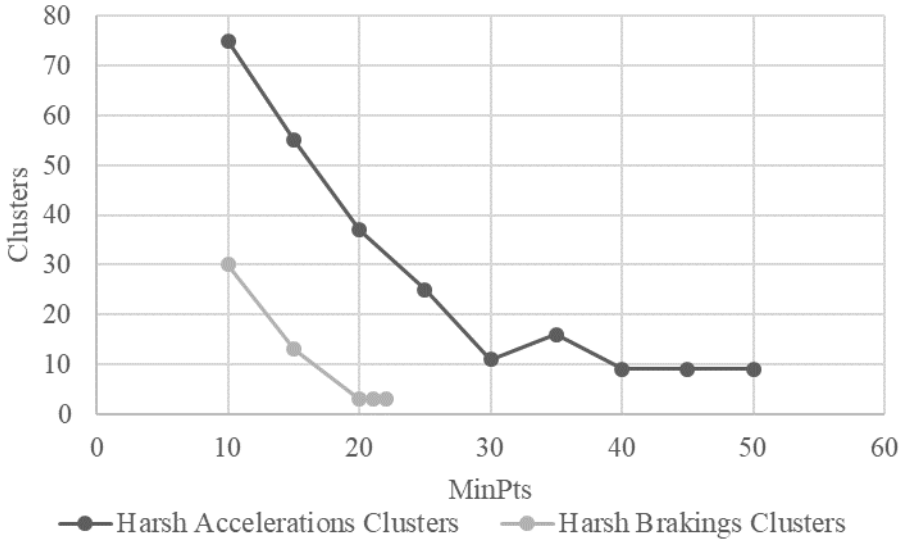

Figure 3: *Hazard Clusters versus the minimum number of harsh acceleration and brakings per cluster*

It is worth noting that as the number of these points reduces, after a certain point the algorithm will identify a single cluster until it finally detects no clusters (Number of Clusters = 0). This is expected, as, when a large threshold is set for the number of points defined by a group, there is no uniformity for the points in the database.

Critical locations based on the above analysis for the harsh accelerations and brakings can be seen Figure 4 and 5 respectively. It is interesting to note that a fairly large percentage of the points where the drivers perform either harsh accelerations or brakings have been annotated by the DBSCAN algorithms as noise. More precisely, the DBSCAN algorithm identified 4176 out of the 4894 hash acceleration events (85.39% of events) in the dataset as been "outliers" (noise). The percentage increases to 92.27% for the case of harsh brakings. These points are likely to have a greater distance than the one that is set for the formation of a cluster or if they are close enough, they do not meet the limit for the number of points that define a cluster. Evidently, this may reflect locations where there is an inconsistency in the users' driving behavior on the road network.

A thorough look at the speeds when these harsh events occurred shows that 12.53% of the total accelerations in a cluster is carried out at a speed greater than 30 km/h, whereas 17.36% of the total decelerations in a cluster is carried out at a speed greater than 30 km/h. From the speeds

above, it is clear that most sudden accelerations and braking are at a slower speed, which confirms that most of these events are near a light signal or a special sign.

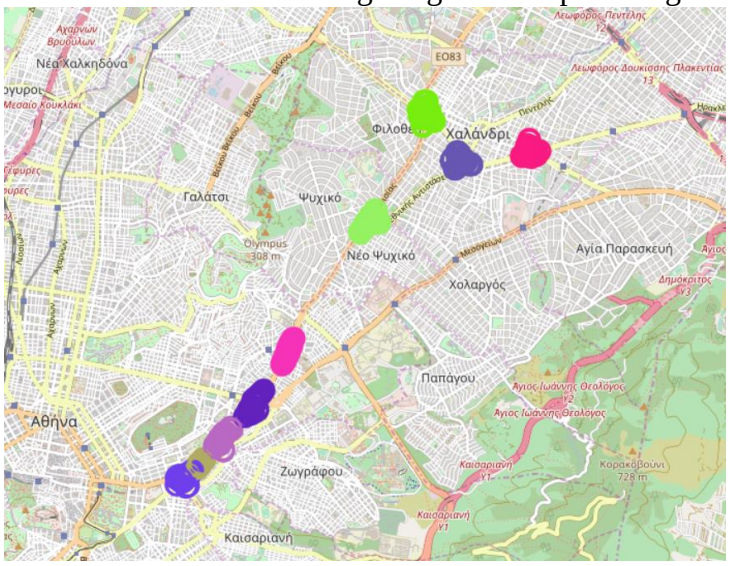

Figure 4: *Hotspot Map for Harsh Acceleration Events.*

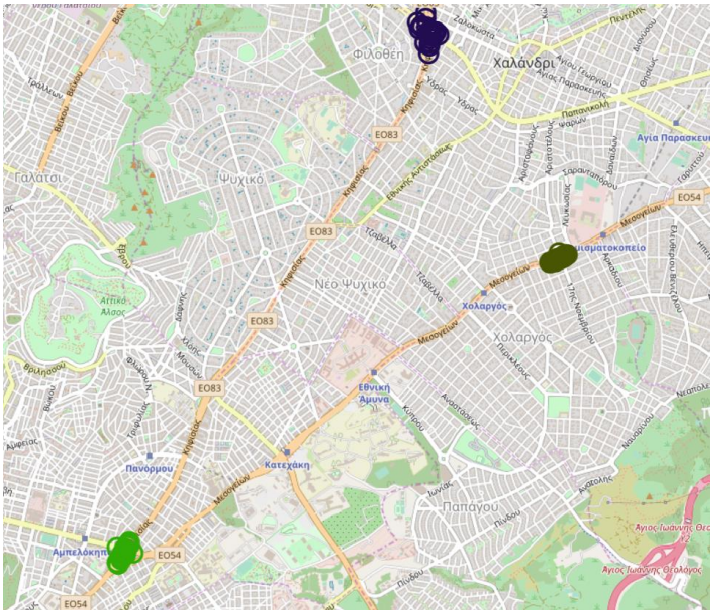

Figure 5: *Hotspot Map for Harsh Braking Events.*

The hotspots identified may shed light to the reasons that a collective harsh driving behavior may be observed. For example, Figure 6 shows the most hazardous location in terms of accelerations at the intersection of Kifissias and Alexandras Avenues. For the specific intersection, a total of 105 events from 27 different drivers have been documented and 13 of them take place at a speed greater than 30 km/h. It should be noted that in the specific hazardous cluster no harsh event is documented over the speed limit.

The cluster which is considered as most hazardous for harsh brakings is located at the intersection of Kifissias Avenue and Kapodistriou Street (Figure 7***Figure 6***). For the specific intersection, a total of 29 events from 16 different drivers have been documented and 25 of them take place at a speed greater than 30 km/h. Again, in the specific hazardous cluster, no harsh event is documented over the speed limit.

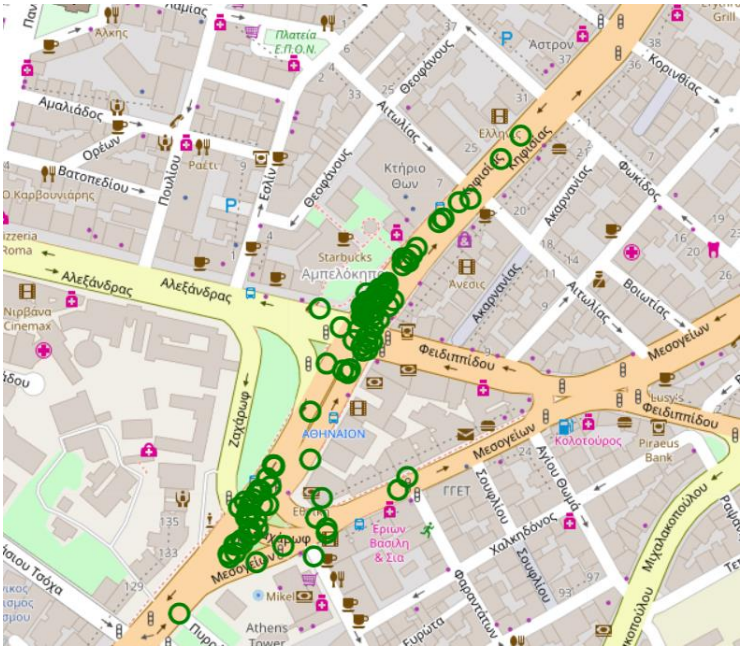

Figure 6: *Critical intersection for harsh accelerations*

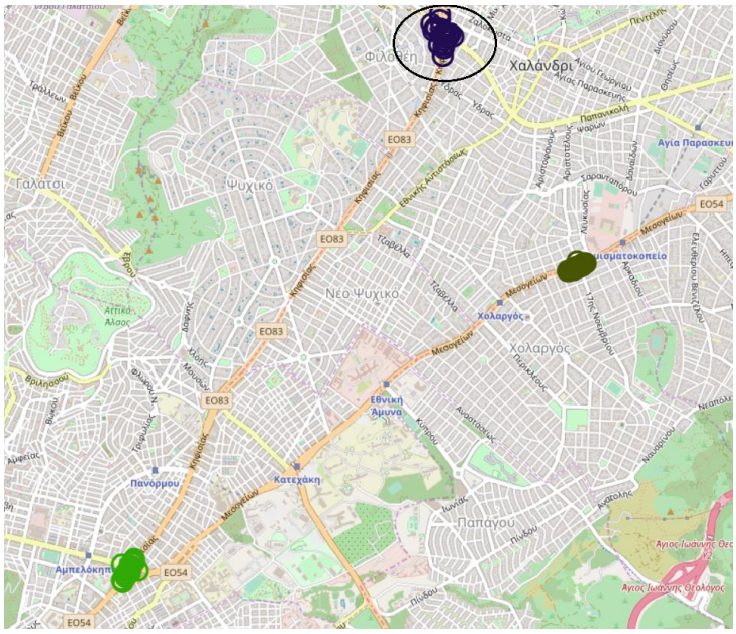

Figure 7: *Critical intersection for harsh brakings*

Evidently, in areas where a harsh acceleration occurs, it is not common for a harsh braking to occur after it. This may hint an indication of collective aggressive driving, as most of these points are highly visited parts of the Athens road network. As far as the speed of the drivers is concerned, for harsh accelerations, it is observed that most of them occur at low speeds, which relates them to the acceleration after a stop. For harsh brakings, it has been observed that they occur at higher speeds.

In Figure 8, the clusters with the most critical speeding locations are depicted and each different color represents a different cluster. From the spatial representation of the clusters related to speeding events it can be observed that, as expected, the clusters that are considered the most hazardous are found in freeways and urban motorways of the Athens metropolitan road network. Yet, in a database of 200 users, a specific part of Kifissias Avenue is presented as one of the most probable parts of the arterial for exceeding the speed limit.

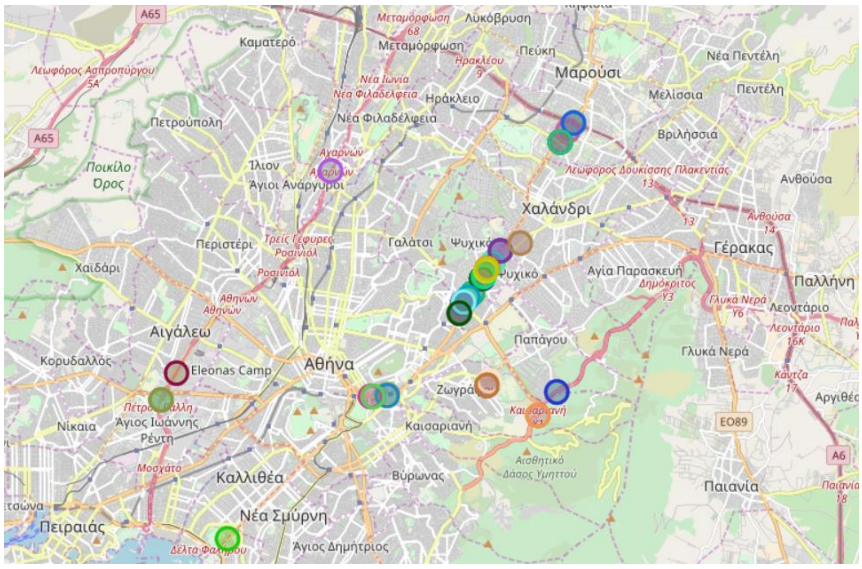

Figure 8: *Hazard map for speeding events*

4. Conclusions

Introducing new technologies, like smartphones, makes it possible to collect, store and transmit high precision data. While previous works have made significant steps when it comes to exploiting these data, there are insufficient works to their spatial representation. Thus, with this paper we intend to create a methodological approach that initially groups data based on extreme driving behavior and, then, perform spatial mapping of these on hazard maps to improve the decision making for traffic management and proactive road safety.

The methodology followed was based on clustering events that occur in specific locations on the road network, but also can extend in time. Two basic clustering algorithms were implemented, the DBSCAN to identify critical locations of harsh acceleration and braking events and the BIDE to identify the critical locations of speeding behavior.

Findings reveal useful insights in relation to the homogeneity of the behavior of one individual driver and many random users. In addition, the most dangerous points of the road network are identified and the hazard factors that play a significant role are further researched. Drivers' behavior is greatly affected by the presence of traffic signals and intersections, where drivers react aggressively. Moreover, although speeding critical locations are to be frequently met in high speed road network, there exist areas inside the densely populated city center where drivers may exhibit collecting speeding behavior.

The proposed approach can be used to inform drivers on their extreme behavior in certain locations. These hazard maps are an information and warning tool for each user individually, as the driver can be aware of the sections of the network where most driving events are taking place in order to make the right decisions in a timely manner. Moreover, the aggregated information on risky locations, such as speeding, can improve the proactive road safety or in real time applications as a case of Advanced Driver Assistance Systems (ADAS).

Further research should be focused on introducing the aspects of geometry and traffic conditions in the identification of critical driving hotspots. Additional variables should be also included in future research related to driving during peak hours or not, the weather conditions. The above will enhance the explanatory power of the clustering approach providing the possible causal relationships that may exist in the spatio-temporal occurrence of extreme driving events.

References

Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. *Journal of Computer Science and Technology*. https://doi.org/10.1007/BF02948845

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident; Analysis and Prevention*. https://doi.org/10.1016/j.aap.2008.12.014

Castignani, G., Derrmann, T., Frank, R., & Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. *IEEE Intelligent Transportation Systems Magazine*. https://doi.org/10.1109/MITS.2014.2328673

Chau, V., Sato, H., Kubo, M., & Namatame, A. (2015). Building Safety Road Maps Based on Difference of Judgment of Road Users by their Smartphone. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/ijacsa.2015.060902

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Kdd*. https://doi.org/10.1016/B978-044452701-1.00067-3

Han, Q., Liu, X., Zeng, L., Ye, L., Chen, D., Li, F., & Xu, Y. (2016). A large vehicle first clustering method based road section risk level estimation. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. https://doi.org/10.1109/ITSC.2016.7795684

Han, Q., Zhu, Y., Zeng, L., Ye, L., He, X., Liu, X., … Zhu, Q. (2015). A Road Hotspots

Identification Method Based on Natural Nearest Neighbor Clustering. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. https://doi.org/10.1109/ITSC.2015.97

Handel, P., Skog, I., Om, J. W., Bonawiede, F., Welch, R., Ohlsson, M., … Ohlsson, M. (2014). Insurance telematics : opportunities and challenges with the smartphone solution. *Intelligent Transportation Systems Magazine, IEEE 6.4*, *6*(4), 57–70. https://doi.org/10.1109/MITS.2014.2343262

Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identification of driving safety profiles from smartphone data using machine learning techniques. *Safety Science*, (January), 0–1. https://doi.org/10.1016/j.ssci.2019.01.025

Ouni, F., & Belloumi, M. (2019). Pattern of road traffic crash hot zones versus probable hot zones in Tunisia: A geospatial analysis. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2019.04.008

Paefgen, J., Kehr, F., Zhai, Y., & Michahelles, F. (2012). Driving behavior analysis with smartphones: insights from a controlled field study. *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, 36:1–36:8. https://doi.org/10.1145/2406367.2406412

Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C: Emerging Technologies*, *16*(3), 320–331. https://doi.org/10.1016/j.trc.2008.01.001

Vlahogianni, E. I., & Barmpounakis, E. N. (2017). Driving analytics using smartphones: Algorithms, comparisons and challenges. *Transportation Research Part C: Emerging Technologies*, *79*(June), 196–206. https://doi.org/10.1016/j.trc.2017.03.014

Wang, H., De Backer, H., Lauwers, D., & Chang, S. K. J. (2019). A spatio-temporal mapping to assess bicycle collision risks on high-risk areas (Bridges) - A case study from Taipei (Taiwan). *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2019.01.014

Wang, J., Han, J., & Li, C. (2007). Frequent Closed Sequence Mining without Candidate Maintenance. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2007.1043

Zeng, L., Hu, Y., Ye, L., Hu, X., Han, Q., Lei, J., & Zhu, Y. (2018). A new method based on PCA contribution factors for road hotspot cause analysis. In *2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI 2017 - *. https://doi.org/10.1109/UIC-ATC.2017.8397545