Analysing Harsh Events with Spatiotemporal Machine Learning Techniques Using Mobile Data

Armira Kontaxi[1*[0000-0002-0935-0889]], Haris Sideris[2[NA]], Dimitris Oikonomopoulos[2[0000-0003-4832-8032]], George Yannis[1[0000-0002-7017-0756]]

[1]Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece
*akontaxi@mail.ntua.gr
[2]OSeven Single Member Private Company, Athens, Greece
hsideris@oseven.io
[2]OSeven Single Member Private Company, Athens, Greece
doikonomopoulos@oseven.io
[1]Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece
geyannis@central.ntua.gr

Abstract. The proliferation of smartphone-based telematics has enabled scalable, high-resolution monitoring of driving behaviour across space and time. In that framework, the aim of this study is to investigate harsh driving events using large-scale smartphone telematics collected during a five-month naturalistic experiment that included a 30-day gamified competition embedded within baseline operation. From a cohort of 95 drivers that participated in the competition, per-second GPS traces were used to map severity-weighted hotspots, and trip-level metadata supported machine-learning analysis of different phase effects. An Extreme Gradient Boosting (XGBoost) machine learning classifier was trained with an 80/20 stratified split (13,247 training, 3,312 test trips) and minority-class upsampling. Feature importance was dominated by distance, followed by hour of day and speeding; experiment phase showed a smaller but measurable contribution, and mobile usage was minimal. Spatial visualization revealed pronounced hotspots along urban cores and major corridors. These findings suggest that smartphone-based feedback and gamification, such as those provided by the telematics app of OSeven, promote safer driving, while they can also be operationalized within a risk-modelling workflow to prioritize locations, times, and drivers for targeted interventions.

Keywords: Road Safety, Smartphone Data, Harsh Events, Machine Learning Models.

1 Introduction

Driver behaviour monitoring has gained prominence in transportation research [1], but collecting accurate, real-time data at scale remains difficult and costly. The ubiquity of smartphones offers a practical alternative: app-based sensing enables low-cost data

collection, trip analytics, and real-time feedback that can reduce crashes and injuries [2]. In addition, smartphone-supported naturalistic driving studies unobtrusively capture behaviour in everyday conditions, yielding more realistic insights than laboratory or instrumented-test settings [3].

Smartphones, beyond passively collecting driving data, enable real-time driver feedback and gamified incentives that can strengthen safety culture. Evidence on feedback/gamification is largely positive, with effectiveness contingent on feedback modality, timing, and incentive design [4,5].

Methodologically, studies employ statistical, machine learning, and deep learning approaches tailored to the outcome of interest. Harsh events are commonly used as surrogate safety measures, with prior work applying generalized mixed-effects models [6] and spatial ML (e.g., Spatial Zero-Inflated models, Spatial Random Forest [7]) to relate harsh events to driving-behavior indicators and exposure. However, integrated spatiotemporal modeling that explicitly examines feedback/ gamification effects remains limited.

Addressing this gap, the present study uses large-scale naturalistic driving data from smartphone sensors to evaluate the impact of feedback/gamification on harsh events. XGBoost models are employed to assess whether, and to what extent, gamification delivered via a smartphone application reduces the incidence of harsh events during driving.

2 Methodology

2.1 Experiment design

In order to achieve the research objective, an innovative smartphone application developed by OSeven (www.oseven.io) for the purpose of the "O7Insurance" (implemented under the National Recovery and Resilience Plan Greece 2.0) research project was exploited aiming to record, analyse and improve driver behavior. "O7Insurance" introduces a novel approach to vehicle insurance management by enabling drivers to handle all aspects of their coverage through a mobile application, supported by OSeven's telematics technology.

Within the framework of the "O7Insurance" project, OSeven has also developed a seamless integration platform for collecting and transferring raw data and recognizing the driving behaviour metrics via Machine Learning (ML) algorithms. After the end of each trip, the application transmits all data recorded to the central database of the OSeven backend platform via an appropriate communication channel, such as a Wi-Fi network or cellular network (upon user's selection) e.g. 3G/4G (online options). The data collected is highly disaggregated in terms of space and time. Also, the data provided by OSeven for the current analysis was in a completely anonymized format, in full compliance with GDPR and personal data protection policies the company has in place.

A variety of different metadata are eventually calculated, as shown in **Fig. 1**. Exposure indicators include total distance, driving duration, road-network type (via GPS/map matching, e.g., OSM), time of day, weather (linked from external providers), and self-reported trip purpose. Driving-behaviour indicators capture speeding (duration

and degree of exceedance), counts and severity of harsh events, specific manoeuvres such as harsh braking/acceleration, and distraction from mobile phone use. For fair comparisons across trips and drivers, all indicators can be normalized by distance or time. For further details on data processing and metadata derivation, the reader is referred to earlier studies from the same research project [8, 9].

Fig. 1. OSeven Driving Behavior Scoring

The naturalistic driving experiment spanned five months (September 2024 to early February 2025) and comprised a baseline period throughout, with a 30-day competition nested December 16, 2024–January 16, 2025, after which the study returned to baseline.

Two modes of feedback were used: Phase A (baseline) provided personalized, non-competitive feedback via the DrivingStar app, trip lists, scorecards (0–100), maps, and highlights to pinpoint unsafe behaviours (speeding, phone use, harsh braking, harsh acceleration), delivered after each trip. Phase B (competition) introduced social gamification and incentives for safe driving over 30 days; competition points equalled the sum across trips of distance travelled × a driving-behaviour factor, rewarding safer performance (see **Fig. 2**).

The analytical dataset included per-second spatial telemetry that locates incidents and grades harsh-event intensity, and a metadata layer used to quantify the competition's impact on harsh events, from a sample of 95 drivers that participated both in the baseline period as well as the competition phase.

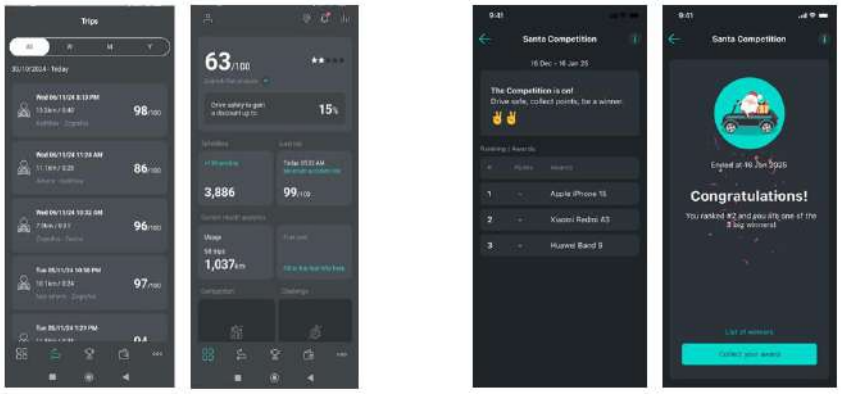

Fig. 2. Example screenshots from the application features in Phase A – Baseline (left) and Phase B – Competition (right)

2.2 Theoretical background

Extreme Gradient Boosting (XGBoost) classifier is a scalable, regularized gradient-boosted tree method that builds an additive ensemble of CART trees to minimize a penalized objective, delivering speed and accuracy [10].

Let f_k denote the k-th regression tree; the model is:

$$\hat{y} = \varphi(x_i) = \sum_{k=1}^{K} f(x_i) \tag{1}$$

with a differentiable convex training loss $l(\hat{y}_i, y_i)$ (e.g., mean squared error),

$$l(\varphi_i) = \sum_{i=1}^{I} (\hat{y}_i - y_i)^2 \tag{2}$$

and a complexity penalty per tree,

$$\Omega(\mathrm{f}) = \gamma \mathrm{T} + \frac{1}{2}\lambda \|\mathrm{c}\|^2 \tag{3}$$

where T is the number of leaves and ccc are leaf weights. The overall objective combines fit and regularization:

$$L(\varphi_i) = \sum_{i=1}^{I} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f) \tag{4}$$

Regularization (e.g., λ, γ) helps control overfitting, and tree structures are relatively robust to multicollinearity. Feature importance is summarized by Gain (loss reduction), Cover (proportion of samples affected), and Frequency (usage count). Key hyperparameters—learning rate (eta), max depth, gamma, min_child_weight, subsample/colsample, and ℓ1/ℓ2 regularization (α, λ), are typically tuned via cross-validation (often with early stopping) [11]. For classification, evaluation relies on confusion-matrix metrics (precision/recall/F1), ROC-AUC, and under strong imbalance, PR-AUC and threshold selection aligned to the desired precision–recall trade-off.

3 Results

3.1 Spatial distribution of harsh events

As a preliminary spatial exploration of safety-critical behaviour indicators, using per-second GPS telemetry, records with harsh-event intensity > 0 were mapped on an OpenStreetMap basemap in Leaflet. Clustered point markers, coloured by intensity (Low/Medium/High), were overlaid with a severity-weighted kernel heatmap (weights 1/2/3) to highlight co-location of frequency and severity.

The visualization (see **Fig. 3**) indicates pronounced hotspots in the urban core and along major arterial corridors, with secondary clusters at junctions and approach roads; peripheral areas exhibit sparser, lower-intensity activity. Because the figure reflects absolute event density rather than exposure-adjusted rates, heavily travelled corridors appear hotter and should not be interpreted as risk per kilometre. Nevertheless, the map efficiently prioritizes candidate locations for engineering review or enforcement and

motivates exposure-normalized, time-stratified, and model-based confirmatory analyses.

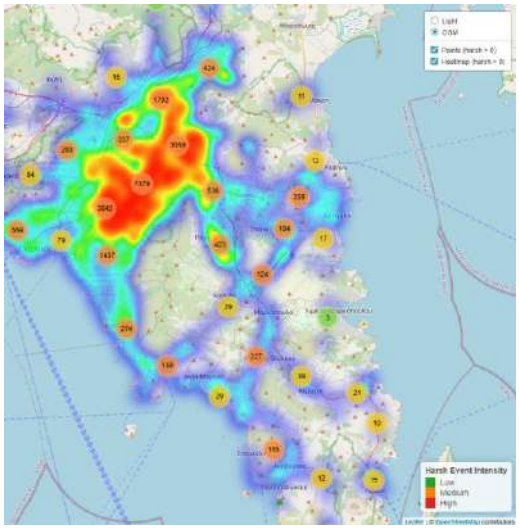

Fig. 3. Severity-weighted heatmap and clustered counts of harsh events (intensity > 0) on OSM

3.2 Competition phase and harsh events via XGBoost

Prior to model fitting, descriptive rates of harsh events per 100 km were examined across experiment phases, as shown in **Table 1**. In the pre-competition phase (n = 5,627 trips), the average rates were 14.27 harsh accelerations and 15.82 harsh brakings per 100 km. During the competition phase (n = 6,528), these averages declined to 12.32 and 13.67, representing a reduction of approximately 13–14% compared with the pre phase. In the post-competition phase (n = 4,404), rates decreased further to 9.94 for accelerations and 9.17 for brakings, corresponding to reductions of roughly 30% and 42% relative to pre. This monotonic downward trend across phases indicates that the competition period was associated with fewer harsh maneuvers, with the most pronounced improvement evident after the intervention. A possible explanation for this finding is that drivers who already improved their behavior during the competition, continued to aim for safe driving.

Table 1. Descriptive rates of harsh events per 100 km across experiment phases

Experiment phase	Average of harsh_acc_per100km	Average of harsh_brk_per100km	Count of trips
pre	14.265	15.815	5627
during	12.321	13.671	6528
after	9.939	9.169	4404

The variable of interest in the present analysis is the occurrence of harsh driving events, modelled in a binary format (yes/no). To examine the conditions under which such

events occur, detailed driving analytics collected by smartphone sensors were analysed using machine learning techniques. Gradient boosted decision trees (XGBoost) were employed to assess the relative importance of contributing variables in predicting the likelihood of a harsh event.

More precisely, XGBoost classifiers were trained on data split 80/20 into training (n = 13,247 trips) and test (n = 3,312) using stratified sampling. Because the binary harsh-event label was imbalanced, the training set was upsampled (random resampling with replacement via caret) to a 1:1 class ratio; the test set was retained unchanged. Features included speed behavior, time of day, speeding, experiment phase, and mobile usage. At the chosen operating point (threshold 0.80), the model achieves high recall (0.82) with moderate precision (0.56), indicating most true harsh events are captured while more than half of flagged cases are correct. The ROC–AUC of 0.87 suggests strong ranking ability across thresholds. However, the relatively low balanced accuracy (0.52) suggests some limitations in distinguishing between classes under imbalance conditions.

Feature importance is dominated by trip distance, with additional signal from hour of day and speeding; experiment phase contributes modestly, and mobile usage minimally (**Table 2**). More precisely, total trip distance is the most informative variable which contribute to the prediction of harsh events during a trip. Typically, longer distances raise the probability of harsh events, as they increase the exposure driving risk. Hour of day captures temporal exposure and traffic states, where risk tends to rise during rush hours (congestion, interactions, frequent stops) and late night (fatigue, low visibility). The Speeding indicator adds complementary signal to raw speed by flagging deliberate limit exceedance, which is often associated with riskier gap acceptance and shorter time-to-collision.

Table 2. Feature importance (XGBoost)

Rank	Feature	Gain	Cover	Frequency
1	Total_distance	0.5254	0.6537	0.8265
2	Hour_of_day	0.0528	0.0551	0.1034
3	Speeding	0.0428	0.0384	0.0758
4	Experiment_phase	0.0208	0.0190	0.0542
5	Mobile_usage	0.0060	0.0132	0.0091

Furthermore, as for experiment phase, this variable contributes a smaller but measurable effect: during a competition/intervention phase, behaviour may shift (e.g., temporary caution or attention), while post-phase periods can show rebound effects; the net direction depends on how incentives interacted with routes and driver habits. Mobile usage has the lowest importance in this configuration; possible reasons include sparse or noisy labels, collinearity with speed/time features, or heterogeneous “phone use” behaviours (e.g., brief glances vs. prolonged interaction) that the current features do not separate

4 Discussion and Conclusions

The aim of this study was to reveal the heterogeneity of potential shared cargo bike users in Budapest. The results demonstrate that the adoption potential is not uniform but shaped by a combination of perceptions, lifestyle, and demographic context.
The study integrated smartphone-based telematics with spatiotemporal visualization and machine learning to characterize harsh driving events over a five-month period that included a 30-day gamified competition. Severity-weighted heatmaps identified persistent hotspots concentrated in the urban core and on major arterial approaches. XGBoost classifiers achieved strong ranking performance (ROC–AUC ≈ 0.87) and, at a representative operating point (threshold = 0.80), high recall (0.82) with moderate precision (0.56). Feature importance was dominated by speed-related variables, with additional contributions from hour of day and speeding; the competition phase exhibited a smaller but measurable effect.

These findings suggest that smartphone-based feedback and gamification, such as those provided by the telematics app of OSeven, promote safer driving [12], while they can also be operationalized within a risk-modelling workflow to prioritize locations, times, and drivers for targeted interventions. Moreover, the combination of smartphone telematics and explainable ML can support near-real-time safety management by flagging high-risk corridors and hours, guiding targeted enforcement, feedback, or engineering audits.

At the same time, several limitations must be acknowledged. First, the observational design does not allow for causal inference regarding the impact of the competition phase, and the reported changes should be interpreted as associations rather than direct effects. From a methodological aspect, class imbalance and potential label noise constrain the achievable precision, recall trade-off. Furthermore, the results remain sensitive to threshold choice and the absence of contextual covariates, such as weather data, road geometry and traffic data.

Future research should incorporate richer contextual variables such as road classification, congestion, and weather. Furthermore, future modelling efforts should also consider probability calibration, precision–recall–AUC as a complement to ROC metrics, and advanced spatiotemporal approaches (e.g., generalized additive models, Gaussian processes, graph neural networks) [13]. Effect decomposition methods such as SHAP can clarify heterogeneity across locations, times, and drivers, and help assess the persistence of behavioural changes beyond the competition phase.

Acknowledgement

This paper was carried out within the framework of OSeven's project "O7Insurance", which is implemented under the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU, under the call “Innovation funding – Horizon 2020 (Implementation body: MIA RI)

References

1. Koesdwiady, A., Soua, R., Karray, F., Kamel, M.S.: Recent trends in driver safety monitoring systems: State of the art and challenges. IEEE Transactions on Vehicular Technology 66(6), 4550–4563 (2017).
2. Chan, T.K., Chin, C.S., Chen, H., Zhong, X.: A comprehensive review of driver behavior analysis utilizing smartphones. IEEE Transactions on Intelligent Transportation Systems 21(10), 4444–4475 (2020).
3. Ziakopoulos, A., Tselentis, D., Kontaxi, A., Yannis, G.: A critical overview of driver recording tools. Journal of Safety Research 72, 203–212 (2020).
4. Chen, W., Donmez, B.: A naturalistic driving study of feedback timing and financial incentives in promoting speed limit compliance. IEEE Transactions on Human-Machine Systems 52(1), 64–73 (2021).
5. Kontaxi, A., Ziakopoulos, A., Yannis, G.: Title of a proceedings paper. In: Proceedings of the 8th Road Safety & Simulation International Conference, pp. 1–13. Athens, Greece (2022).
6. Kontaxi, A., Ziakopoulos, A., Yannis, G.: Trip characteristics impact on the frequency of harsh events recorded via smartphone sensors. IATSS Research 45(4), 574–583 (2021a).
7. Nikolaou, D., Ziakopoulos, A., Kontaxi, A., Theofilatos, A., Yannis, G.: Spatial analysis of telematics-based surrogate safety measures. Journal of Safety Research 92, 98–108 (2025).
8. Kontaxi, A., Ziakopoulos, A., Yannis, G.: Investigation of the speeding behavior of motorcyclists through an innovative smartphone application. Traffic Injury Prevention 22(6), 460–466 (2021b).
9. Ziakopoulos, A., Kontaxi, A., Yannis, G.: Analysis of mobile phone use engagement during naturalistic driving through explainable imbalanced machine learning. Accident Analysis and Prevention 181, 106936 (2023).
10. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016).
11. XGBoost Developer Team: XGBoost documentation. Available at: http://xgboost.readthedocs.io/en/latest/index.html, last accessed 2025/06/07.
12. Kontaxi, A., Ziakopoulos, A., Yannis, G.: Exploring the impact of driver feedback on safety: A systematic review of studies in real-world driving conditions. Transportation Research Part F: Traffic Psychology and Behaviour 114, 118–140 (2025).
13. Choudhary, A., Mishra, V., Garg, R.D., Jain, S.S.: Spatio–temporal analysis of traffic crash hotspots: An application of GIS-based technique in road safety. Applied Geomatics 17, 129–146 (2025).