Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2026) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

Road Safety and Simulation 2026 – RSS2026

Imbalanced Learning Analysis for Driving Behaviour Prediction Using Naturalistic Driving Data

Antonis Kostopoulos[a], Thodoris Garefalakis[a], Eva Michelaraki[a]* George Yannis[a]

[a]*National Technical University of Athens, Department of Transportation Planning and Engineering,*
5 Heroon Polytechniou str., 15773, Athens, Greece

Abstract

Road safety remains a major global concern, with human behaviour continuing to play a decisive role in serious traffic injuries and crashes. Although vehicle technologies and road infrastructure have improved considerably, risky driving practices such as speeding, distraction and impaired driving persist as primary contributing factors to accidents. This study investigates the prediction of hazardous driving behaviour through the analysis of more than 356,000 trips, emphasizing harsh acceleration and braking events as key indicators of driving risk. A comprehensive framework was developed to evaluate and classify driving behaviour into dangerous and non-dangerous categories. The proposed methodology integrates clustering techniques for defining safety levels, feature selection methods for identifying the most relevant variables and strategies for handling imbalanced datasets. Data obtained from a naturalistic driving experiment were analysed, with particular focus on harsh acceleration and braking patterns. Five machine learning models (Random Forest, Gradient Boosting, XGBoost, Multilayer Perceptron and K-Nearest Neighbors) were implemented to predict risky driving behaviour. The results indicated that Gradient Boosting and Multilayer Perceptron delivered the strongest predictive performance, achieving recall values of nearly 67% and 68% for harsh acceleration and braking events, respectively. Furthermore, the study identified critical safety thresholds of 48.82 harsh accelerations and 45.40 harsh braking events per 100 km. Overall, the integration of clustering methods, feature selection and machine learning models proved effective in detecting dangerous driving patterns. The proposed approach supports the development of personalized feedback systems, targeted driver training and advanced road safety strategies for authorities and transportation organizations.

Keywords: Road Traffic Safety; Naturalistic Driving Experiment; Driving Behaviour Analysis; Harsh Events; Machine Learning.

---

* Corresponding author. Tel.: +30-210-772-1265
E-mail address: evamich@mail.ntua.gr

1. Introduction

Human behaviour significantly contributes to severe road injuries, underscoring a critical road safety challenge. Despite advancements in vehicle technology and infrastructure, risky driving performance such as speeding, distraction and impaired driving remain leading causes of crashes.

Beyond the roadway geometric design, traffic volume, and other risk indicators, human behavior is the major factor in provoking severe road injuries. Although there are statistically significant relationships between different features, the exact impact of every single traffic and human behavior characteristic on the volume and severity of road injuries has yet to be found. Thus, focusing on road safety enhancement and optimally detecting and quantifying the relations between different features, recent studies focused on driving behavior analysis by developing appropriate machine and deep learning models.

It should be noted that excessive speeding and low-speed conditions and, therefore, acceleration-deceleration relationships are key factors in approximately 30% of fatal crashes (Kröyer, 2015). Speed has a direct impact on the frequency and severity of road crashes. A 1% increase in average speed leads to an increase of about 2% in the incidence of mild injuries from road crashes, 3% in the incidence of serious injuries, and 4% in fatalities (Nilsson, 2004). Harsh events, specifically acceleration and braking, play a pivotal role as key indicators in the evaluation of driving risk, particularly when assessing the degree of driving aggressiveness (Bonsall et al., 2005). Therefore, to prevent and classify risky driving behavior, it is important to highlight the spatio-temporal dimension in which accelerations and brakings take place and examine their categorization into 'Dangerous' and 'Non-Dangerous' classes.

This study tackles the complex challenge of predicting risky driving behaviour through an extensive analysis of more than 356,000 trips, focusing on patterns of harsh acceleration and braking events to highlight their predictive value for road safety. A comprehensive framework for assessing and classifying driving behaviour as either dangerous or non-dangerous was proposed. The framework incorporates defining safety levels using clustering algorithms, selecting the most relevant features and addressing issues related to imbalanced datasets.

This paper is structured as follows: Firstly, after an overview of the field of road safety, an extensive literature review is conducted on the analysis of driving behavior using emerging techniques. A detailed description of the research methodology, which also includes the theoretical basis of the models, follows, as well as the data collection and processing procedures. The results of the study are then presented, enabling the emergence of road safety-related conclusions.

2. Background

Recent advances in Intelligent Transport Systems (ITS) have increased interest in analysing driving behaviour to improve the prediction and prevention of unsafe driving. Consequently, numerous studies have applied machine learning and deep learning techniques to develop reliable predictive models for driving risk assessment.

Several researchers have focused on identifying aggressive or dangerous driving patterns using clustering and classification approaches. Papadimitriou et al. (2019) investigated the relationship between mobile phone use and risky driving through naturalistic driving data, employing binary logistic regression with an accuracy of 70%. Yang et al. (2021) applied clustering techniques, including k-means, hierarchical clustering and Gaussian Mixture Models, to define driving safety levels and achieved a classification accuracy of 97.9% using Support Vector Machines. Similarly, Yarlagadda et al. (2021) used k-means clustering to identify aggressive manoeuvres among heavy passenger vehicle drivers, while Ali et al. (2021) analysed risky driving under both clear and rainy weather conditions, highlighting differences in driving behaviour between the two environments.

Other studies have explored the use of sensor and simulator data. Zhang et al. (2016) classified driving behaviour using smartphone and On-Board Diagnostics (OBD) data, achieving an accuracy of 86.67% with Support Vector Machines. Ghandour et al. (2021) examined driver psychological states using machine learning methods, identifying Gradient Boosting as the most effective model. In addition, Mumcuoglu et al. (2019) demonstrated the effectiveness of Long Short-Term Memory (LSTM) networks in detecting dangerous driving behaviour from simulator-based sensor data. A major challenge in driving behaviour analysis is the handling of imbalanced datasets, where dangerous driving events are typically underrepresented. To address this issue, previous studies have applied oversampling and hybrid learning techniques, such as SVMSMOTE, ADASYN and SMOTE-ENN, significantly improving classification performance (Wang et al., 2021). Although previous research has mainly focused on socio-demographic, psychological and speed-related factors, the present study develops driver risk profiles based specifically on harsh acceleration and braking events, providing a novel contribution to road safety analysis.

3. Materials &Methods

3.1. Naturalistic Driving Experiment

The naturalistic driving data used in this study were obtained from OSeven Telematics[†] through a dedicated smartphone application that continuously records trip-related information without interrupting the driving task. The application relies exclusively on built-in smartphone sensors and does not require any additional equipment. Various APIs are used to read and temporarily store the data on the smartphone before transferring them to the company's back-end database. The collected data are assigned spatial and temporal attributes and are subsequently transformed into driving behaviour and safety indicators using signal processing, machine learning, data fusion and Big Data techniques. The smartphone sensors used include the accelerometer, gyroscope, magnetometer and GPS.

In addition, iOS and Android data fusion methods provide nine-degree-of-freedom measurements, including yaw, pitch and roll, as well as gravity and linear acceleration. OSeven follows strict privacy and information security procedures in compliance with the General Data Protection Regulation (GDPR) and relevant European legislation. Therefore, all data were provided in a fully anonymized form, while trip geolocation details were excluded, except for the country in which each trip occurred. The dataset analysed in this study includes 356,162 trips conducted on an urban road network. For each trip, 75 indicators were available, while 23 variables, such as average speed and mobile phone use duration, were selected for the analysis. Overall, the dataset contains key driving risk factors related to traffic conditions and driver state. To assess trip quality, OSeven provides parameterized scoring indicators ranging from 0 to 100, where 100 represents an ideal driving performance. The user's overall score is calculated as the distance-weighted average of all trips made as a driver during the previous 12 months.

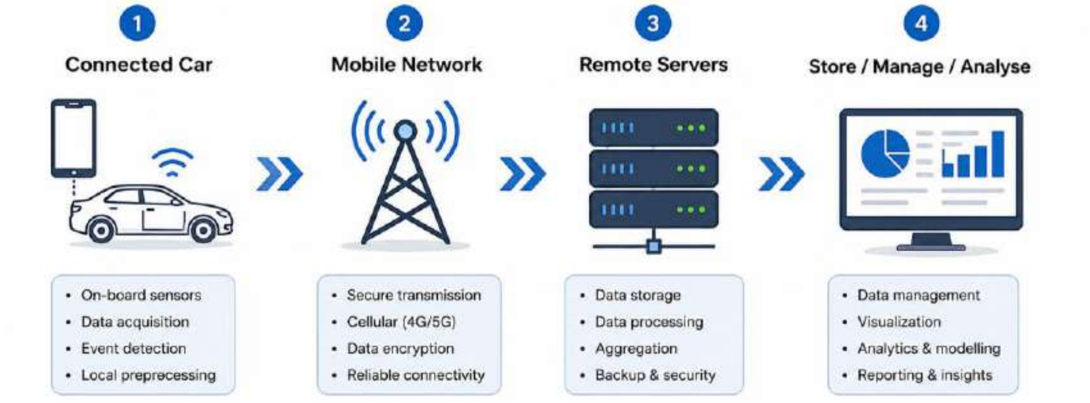

Fig 1. Overview of the data communication and management system

3.2. Definition of Safety Levels

A key and demanding stage of this study was the determination of driving behaviour levels (Garefalakis et al., 2024). Classifying driving behaviour into distinct levels requires data pre-processing and clustering into a binary structure. The K-means algorithm is a commonly used clustering method for multilevel data, known for being fast and relatively easy to implement (Kodinariya & Makwana, 2013). It groups data into predefined clusters based on the number of generated centroids, assigning each observation to the cluster with the closest mean value. This assignment

---

[†] https://oseven.io

is performed by minimizing the within-cluster sum of squares. K-means operates iteratively, continuously adjusting centroid positions until they become stable.

3.3. Feature Selection

Feature selection aims to identify the most influential independent variables and separate them from the full set of available variables. It is considered an optimization step in classification tasks, as it reduces the risk of biased or unstable results caused by an excessive number of input features. As Bellman (1966) noted, when the number of variables increases, the feature space expands rapidly and the available data become sparse, leading to the "curse of dimensionality". Following previous studies (Li et al., 2020; Roussou et al., 2024), this study applies feature selection based on permutation importance. Feature importance is used to measure how strongly the independent variables influence the dependent variable. For this purpose, feature permutation was performed and the prediction error of the model was calculated by training different regression models.

The final predictor variables were selected according to both their permutation importance scores and their interpretability in the context of driving behaviour analysis. Variables with consistently higher importance values across the examined models were retained, as they contributed most strongly to predictive performance. At the same time, particular emphasis was placed on variables with clear behavioural relevance. As a result, total travel distance and driving duration were retained as indicators of driving exposure, while average speed, speeding score and mobile use score were included because they are directly associated with driving performance and risky behaviour.

3.4. Classification Process

To achieve the aim of this study, namely the detection of dangerous driving behaviour, a methodology was developed using selected predictor variables and five classification algorithms. The machine learning models examined were Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), Multilayer Perceptron (MLP) and K-nearest Neighbors (kNN). These models were selected because of their proven effectiveness and frequent application in previous studies on risky driving detection, real-time crash prediction and other practical classification problems. The comparison also provides useful insights by examining three ensemble learning methods (RF, GB and XGBoost), one non-parametric approach suitable for datasets with limited prior assumptions (kNN) and one artificial neural network model (MLP).

For model development and assessment, the dataset was divided into training and testing subsets. The training set is expressed as Xtraining = (xn, yn), n = 1, …, N, where xn represents the predictor variables and yn = 0,1 denotes the target class, corresponding here to dangerous and non-dangerous driving behaviour. To assess and compare the performance of the five classifiers, several evaluation metrics were applied, as presented in Equations (1)-(6).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1) \qquad Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Sensitivity\ (Recall) = \frac{TP}{TP+FN} \quad (3) \qquad Specificity = \frac{TN}{TN+FP} \quad (4)$$

$$F\text{-}measure = \frac{2*(Precision)*(Recall)}{(Precision)+(Recall)} \quad (5) \qquad False\ Alarm\ Rate = \frac{FP}{FP+TN} \quad (6)$$

where: True Positive (TP) instances correspond to those that belong to class i and were correctly identified in it; True Negative (TN) instances are those that do not belong to class i and were not classified in it; False Positive (FP) cases are those that do not belong to class i but were wrongly identified as such; False Negative (FN) instances are those that do belong to class i but were not classified as such.

A key methodological challenge in this study was the imbalanced distribution of the target classes, as dangerous driving events represented a smaller proportion of the observations compared with non-dangerous driving behaviour. This imbalance was observed in both harsh acceleration and harsh braking classifications and reflects the natural distribution of risky events in real-world driving data. To preserve class proportions and avoid bias during model development, the dataset was divided into training and testing subsets using a stratified sampling procedure. Given the imbalanced nature of the dataset, model performance was assessed using evaluation metrics specifically suited to imbalanced classification tasks. In particular, recall was considered a key performance indicator because failing to identify dangerous driving behaviour (false negatives) may have direct road safety implications. For this reason, false negative rate and F1-score were also examined alongside accuracy and precision, allowing a more balanced interpretation of model performance.

The dataset was randomly divided into training (80%) and testing (20%) subsets using stratified sampling to preserve class proportions; trips were assigned according to the available anonymized trip records, and model performance was evaluated on the testing subset without applying additional cross-validation.

Ensemble learning combines the outputs of several weak classifiers to produce more accurate predictions than an individual model. Random Forest (RF) applies this concept within a decision-tree structure by generating multiple randomly constructed trees. It uses bootstrapping to train trees on different subsets of the data and aggregation to combine their outputs into a final prediction.

Gradient Boosting (GB) is another ensemble approach that builds a sequence of boosted decision trees to improve predictive performance. Its main principle is to add new base learners that are strongly related to the negative gradient of the loss function of the overall model (Natekin & Knoll, 2013). Extreme Gradient Boosting (XGBoost) is an optimized version of Gradient Boosting. Unlike standard Gradient Boosting, which relies mainly on gradient descent, XGBoost uses a second-order Taylor approximation and operates similarly to a Newton-Raphson optimization process. It constructs trees sequentially, often achieving high accuracy with reduced computational training time compared with conventional machine learning methods.

Multilayer Perceptron (MLP) is a type of feed-forward artificial neural network. It consists of at least three layers: an input layer, one or more hidden layers and an output layer. Except for the input layer, nodes act as artificial neurons and apply nonlinear activation functions. Connections between successive layers, known as neural synapses, allow information to pass through the network. The number of hidden layers is important for classification performance, while training is mainly performed through backpropagation. K-nearest Neighbors (kNN) is a simple and widely used classification algorithm that does not require prior assumptions about the distribution of the data. It is especially suitable for relatively small datasets and avoids the complexity of estimating probability density functions (Peterson, 2009). The method classifies a test observation by calculating its Euclidean distance from training samples and assigning it to the class of the most similar nearby observations.

For all classifiers, hyperparameters were selected through an iterative tuning process based on performance on the training dataset, while model validation was conducted using the testing subset to ensure robustness and support a consistent comparison across all classification models.

4. Results

The findings showed that Gradient Boosting and Multilayer Perceptron delivered the best predictive results, achieving recall values of about 67% for harsh acceleration and 68% for harsh braking events. The study also identified key thresholds for risky driving behaviour, namely 48.82 harsh accelerations and 45.40 harsh brakings per 100 km. These values offer useful reference points for assessing driver risk. Overall, combining machine learning models with feature selection and k-means clustering proved to be a promising method for improving road safety and reducing related socio-economic impacts.

Feature selection was applied to limit the number of input variables, reduce computational requirements and enhance the predictive ability of the classification models. The selected variables were determined according to their contribution to the classification task. As shown in Figure 2, total travel distance and driving duration had the strongest influence on the regression process. In contrast, mobile phone use duration and distance travelled during high-risk night-time hours (00:00–05:00) showed the lowest direct impact.

According to the feature importance, the input variables for the classification process for both harsh acceleration and braking events, are total distance, driving duration, average driving speed, speeding score, and mobile use score. Table 1 displays some descriptive statistics for the input variables employed in the classification procedure, including mean, standard deviation, maximum, minimum, and mean values.

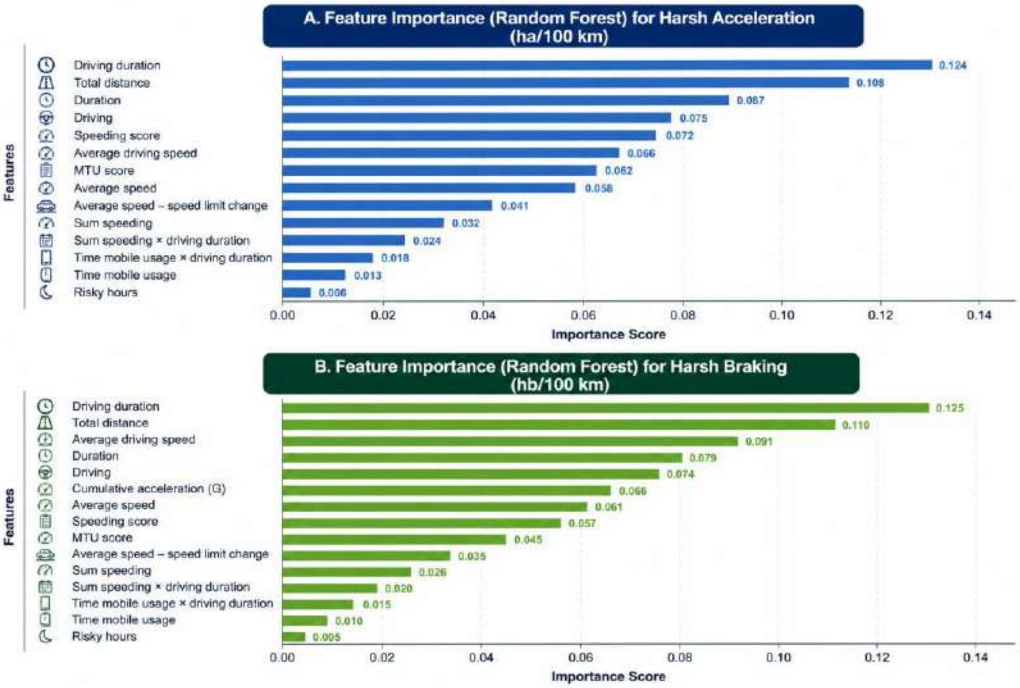

Fig. 2. Permutation feature importance for harsh acceleration/braking events per 100 km

Table 1. Descriptive statistics for input variables

Variables	Description (units)	Mean	St. Dev.	Min	Max
Total distance	Total trip distance (km)	11.60	22.31	0.50	648.68
Driving duration	Total duration of driving, i.e., duration of stops has been excluded (sec)	769.97	967.15	61.00	23900.00
Average speed	Average speed while the vehicle is in motion (km/h)	42.57	17.58	5.57	183.91
Speeding score	Excessive speed score (%)	76.52%	32.92	10.00%	100.00%
Mobile use score	Hand-held Mobile Usage Score while Driving (%)	80.53%	34.62	10.00%	100.00%

To evaluate the models appropriately, it is important to emphasize that incorrectly classifying dangerous driving behaviour is the most critical error, as it may have direct road safety implications. Moreover, previous research on imbalanced datasets (Valverde-Albacete & Peláez-Moreno, 2014) indicates that accuracy may produce misleading interpretations, a problem known as the "accuracy paradox". Therefore, as shown in Table 2, additional performance indicators, including precision, recall, false negative rate and F1-score, were also considered.

Overall, all models showed acceptable predictive performance, with relatively small differences among them. Gradient Boosting (GB) and Multilayer Perceptron (MLP) achieved the strongest results in terms of recall and false negative rate, with GB performing slightly better than MLP. Both models also produced satisfactory AUC values, reaching 75.1% for GB and 74.7% for MLP, indicating relatively good classification performance.

Table 2. Classification metrics for the developed classifiers for harsh acceleration events per 100 km

Classification Model	Accuracy	Precision	Recall	False Negative Rate	f-1 score
RF	70.83 %	55.16 %	66.39 %	33.61 %	52.64 %
GB	65.28 %	55.15 %	68.05 %	31.95 %	50.25 %
XGBoost	66.76 %	55.09 %	67.46 %	32.54 %	50.86 %
MLP	68.16 %	55.26 %	67.65 %	32.35 %	51.63 %
kNN	72.70 %	53.46 %	60.08 %	39.92 %	51.47 %

Similar to the harsh acceleration analysis, additional evaluation metrics were considered because of the dataset imbalance and the potential risk of the “accuracy paradox”. As presented in Table 3, all five models showed satisfactory ability to detect dangerous driving behaviour and achieved generally good performance results. However, Gradient Boosting (GB) and Multilayer Perceptron (MLP) performed better than the other models in terms of recall and false negative rate. As noted in Section 5.1, these indicators are particularly important for the objectives of this study. In addition, both GB and MLP obtained acceptable AUC values, with scores of 74.9% and 74.7%, respectively.

Table 3. Classification metrics for the developed classifiers for harsh braking events per 100 km

Classification Model	Accuracy	Precision	Recall	False Negative Rate	f-1 score
RF	67.78 %	57.20 %	66.48 %	33.52 %	55.09 %
GB	63.36 %	57.36 %	67.91 %	32.09 %	53.13 %
XGBoost	64.53 %	57.20 %	67.30 %	32.70 %	53.60 %
MLP	62.96 %	57.29 %	67.80 %	32.20 %	52.88 %
kNN	68.45 %	54.88 %	60.55 %	39.45 %	53.19

5. Discussion & Conclusions

This paper proposed a comprehensive framework for analysing and classifying driving behaviour as dangerous or non-dangerous. The framework involved defining safety levels using clustering methods, selecting the most relevant features and addressing class imbalance in the dataset. A naturalistic driving experiment was conducted using data provided by OSeven Telematics through a dedicated smartphone application that continuously records driving information without affecting the driving process. The analysis was performed separately for two types of harsh events: harsh acceleration and harsh braking. Five machine learning classifiers were developed, namely Random Forest, Gradient Boosting, XGBoost, Multilayer Perceptron and K-Nearest Neighbors.

Feature importance analysis showed that total travel distance and trip duration were the most influential variables in identifying driving behaviour. Longer trips and driving durations are associated with fatigue, reduced alertness, slower reaction times and decreased psychomotor performance, which may negatively affect driving safety. Vehicle speed was also identified as a critical factor contributing to dangerous driving behaviour and crash occurrence, confirming findings from previous studies. Higher speeds reduce drivers’ ability to perceive and respond effectively to external conditions. In contrast, driving during the high-risk nighttime period (00:00–05:00) was not strongly associated with harsh acceleration and braking events, likely due to lower traffic and pedestrian activity during these hours. The analysis was conducted separately for harsh acceleration and harsh braking events. Five machine learning models were developed for classification purposes: RF, GB, XGBoost, MLP and kNN. Among these, GB and MLP demonstrated the highest predictive capability for both categories of harsh events. Although performance differences among the models were relatively small, the findings confirmed that the K-means clustering approach effectively separated driving behaviour into two safety levels.

The study also established threshold values of 48.82 harsh accelerations and 45.40 harsh brakings per 100 km for the dangerous driving category, offering new benchmarks for driving risk assessment. Overall, this study provides a practical framework for distinguishing dangerous from non-dangerous driving behaviour based on harsh events. The proposed methodology contributes to the advancement of road safety research by improving the detection and prediction of risky driving behaviour, while also supporting the development of targeted safety interventions and driver profiling strategies.

The results showed that Gradient Boosting and Multilayer Perceptron achieved the strongest predictive performance for risky driving behaviour in both harsh event categories. Future research could further explore deep learning methods, such as CNN, RNN and LSTM models, as these approaches may capture complex patterns and nonlinear relationships more effectively. Although precision and F1-score values were relatively moderate, the models demonstrated satisfactory capability in identifying dangerous driving behaviour, while these results also indicate that further optimization is needed before large-scale real-world deployment, particularly to reduce false positive classifications and improve prediction reliability.

A limitation of this study is that the analysis was based exclusively on urban trips; therefore, the transferability of the findings to rural roads or highway environments should be considered with caution, as driving patterns and the frequency of harsh events may differ substantially across road network types. Nevertheless, future studies could examine alternative approaches to measure feature importance, such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC), both based on evaluating the likelihood function, which tends to balance the level of fit with model complexity (O'Malley & Neelon, 2014). Furthermore, clustering with Gaussian Mixture Models could have been proven truly insightful due to the nature of GMM to discover complex patterns and combat heterogeneity problems (Patel & Kushwaha, 2020). In addition, future approaches could examine data on different road networks instead of the urban road network exclusively, given the fact that the quantity and quality of harsh events would differ. Overall, this research contributes to the literature on machine learning applications in driving behaviour analysis and offers useful insights into data processing and model development. From a practical perspective, driver profiles based on harsh acceleration and braking patterns can support personalized feedback, improve driver training and contribute to innovation in the automotive sector.

References

Ali, E. M., Ahmed, M. M., & Yang, G. (2021). Normal and risky driving patterns identification in clear and rainy weather on freeway segments using vehicle kinematics trajectories and time series cluster analysis. IATSS Research, 45(1), 137–152.

Bellman, R. (1966). Dynamic Programming. Science, 153(3731), 34–37. https://doi.org/10.1126/science.153.3731.34

Bonsall, P., Liu, R., & Young, W. (2005). Modelling safety-related driving behaviour—impact of parameter values. Transportation Research Part A: Policy and Practice, 39(5), 425–444. https://doi.org/10.1016/j.tra.2005.02.002

Garefalakis, T., Michelaraki, E., Roussou, S., Katrakazas, C., Brijs, T., & Yannis, G. (2024). Predicting risky driving behavior with classification algorithms: results from a large-scale field-trial and simulator experiment. European Transport Research Review, 16(1), 65.

Ghandour, R., Potams, A. J., Boulkaibet, I., Neji, B., & Al Barakeh, Z. (2021). Driver Behavior Classification System Analysis Using Machine Learning Methods. Applied Sciences, 11(22). https://doi.org/10.3390/app112210562

Kröyer, H. R. G. (2015). Is 30 km/h a 'safe' speed? Injury severity of pedestrians struck by a vehicle and the relation to travel speed and age. IATSS Research, 39(1), 42–50. https://doi.org/10.1016/j.iatssr.2014.08.001

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering.

Li, Z., Wang, H., Zhang, Y., & Zhao, X. (2020). Random forest–based feature selection and detection method for drunk driving recognition. International Journal of Distributed Sensor Networks, 16(2), 155014772090523. https://doi.org/10.1177/1550147720905234

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7. https://doi.org/10.3389/fnbot.2013.00021

Nilsson, G. (2004). Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety.

Mumcuoglu, M. E., Alcan, G., Unel, M., Cicek, O., Mutluergil, M., Yilmaz, M., & Koprubasi, K. (2019). Driving Behavior Classification Using Long Short Term Memory Networks. 2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), 1–6. https://doi.org/10.23919/EETA.2019.8804534

O'Malley, A. J., & Neelon, B. H. (2014). Latent Factor and Latent Class Models to Accommodate Heterogeneity, Using Structural Equation. In Encyclopedia of Health Economics (pp. 131–140). Elsevier. https://doi.org/10.1016/B978-0-12-375678-7.00712-4

Papadimitriou, E., Argyropoulou, A., Tselentis, D. I., & Yannis, G. (2019). Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. Safety Science, 119, 91–97. https://doi.org/10.1016/j.ssci.2019.05.059

Patel, E., & Kushwaha, D. S. (2020). Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. Procedia Computer Science, 171, 158–167. https://doi.org/10.1016/j.procs.2020.04.017

Peterson, L. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883. https://doi.org/10.4249/scholarpedia.1883

Roussou, S., Michelaraki, E., Katrakazas, C., Afghari, A. P., Al Haddad, C., Alam, M. R., ... & Yannis, G. (2024). Unfolding the dynamics of driving behavior: a machine learning analysis from Germany and Belgium. European Transport Research Review, 16(1), 40.

Zhang, C., Patel, M., Buthpitiya, S., Lyons, K., Harrison, B., & Abowd, G. D. (2016). Driver Classification Based on Driving Behaviors. Proceedings of the 21st International Conference on Intelligent User Interfaces, 80–84. https://doi.org/10.1145/2856767.2856806

Valverde-Albacete, F. J., & Peláez-Moreno, C. (2014). 100% classification accuracy considered harmful: the normalized information transfer factor explains the accuracy paradox. PloS One, 9(1), e84217–e84217. https://doi.org/10.1371/journal.pone.0084217

Wang, K., Xue, Q., & Lu, J. J. (2021). Risky Driver Recognition with Class Imbalance Data and Automated Machine Learning Framework. International Journal of Environmental Research and Public Health, 18(14). https://doi.org/10.3390/ijerph18147534

Yang, K., Haddad, C. Al, Yannis, G., & Antoniou, C. (2021). Driving Behavior Safety Levels: Classification and Evaluation. 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 1–6. https://doi.org/10.1109/MT-ITS49943.2021.9529309

Yarlagadda, J., Jain, P., & Pawar, D. S. (2021). Assessing safety critical driving patterns of heavy passenger vehicle drivers using instrumented vehicle data – An unsupervised approach. Accident Analysis & Prevention, 163, 106464. https://doi.org/10.1016/j.aap.2021.106464