

## Article

# Modeling and Sustainability Implications of Harsh Driving Events: A Predictive Machine Learning Approach

Antonios Kostopoulos, Thodoris Garefalakis , Eva Michelaraki , Christos Katrakazas  and George Yannis 

Department of Transportation Planning and Engineering, National Technical University of Athens, 5 IroonPolytechniou Str., 157 73 Athens, Greece; ackostopoulos@gmail.com (A.K.); evamich@mail.ntua.gr (E.M.); ckatrakazas@mail.ntua.gr (C.K.); geyannis@central.ntua.gr (G.Y.)

\* Correspondence: tgarefalakis@mail.ntua.gr

**Abstract:** Human behavior significantly contributes to severe road injuries, underscoring a critical road safety challenge. This study addresses the complex task of predicting dangerous driving behaviors through a comprehensive analysis of over 356,000 trips, enhancing existing knowledge in the field and promoting sustainability and road safety. The research uses advanced machine learning algorithms (e.g., Random Forest, Gradient Boosting, Extreme Gradient Boosting, Multilayer Perceptron, and K-Nearest Neighbors) to categorize driving behaviors into ‘Dangerous’ and ‘Non-Dangerous’. Feature selection techniques are applied to enhance the understanding of influential driving behaviors, while k-means clustering establishes reliable safety thresholds. Findings indicate that Gradient Boosting and Multilayer Perceptron excel, achieving recall rates of approximately 67% to 68% for both harsh acceleration and braking events. This study identifies critical thresholds for harsh events: (a) 48.82 harsh accelerations and (b) 45.40 harsh brakings per 100 km, providing new benchmarks for assessing driving risks. The application of machine learning algorithms, feature selection, and k-means clustering offers a promising approach for improving road safety and reducing socio-economic costs through sustainable practices. By adopting these techniques and the identified thresholds for harsh events, authorities and organizations can develop effective strategies to detect and mitigate dangerous driving behaviors.

**Keywords:** road traffic safety; naturalistic driving experiment; driving behavior analysis; driving behavior; harsh events; machine learning



**Citation:** Kostopoulos, A.; Garefalakis, T.; Michelaraki, E.; Katrakazas, C.; Yannis, G. Modeling and Sustainability Implications of Harsh Driving Events: A Predictive Machine Learning Approach. *Sustainability* **2024**, *16*, 6151. <https://doi.org/10.3390/su16146151>

Academic Editor: Victoria Gitelman

Received: 31 May 2024

Revised: 12 July 2024

Accepted: 17 July 2024

Published: 18 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Road safety is a major concern for the institutions of the European Union and its national components. According to research conducted by the World Health Organization [1], approximately 1.19 million human deaths per year are related to road crashes, making it the leading cause of fatalities for individuals aged 5–29 years. Consequently, the European Union has implemented substantial measures to eliminate road fatalities by adopting Vision Zero, which aims to eradicate fatal road crashes by 2050. This initiative aligns with sustainable practices by promoting safer transportation systems and reducing the socio-economic costs associated with road crashes. As outlined by the EU Road Safety Policy Framework for 2021–2030 [2], the mid-term goal of the Vision Zero project is to reduce fatalities in road crashes by 50% during the decade 2021–2030, based on the Safe System approach and using contextual data.

Human behavior is a major factor in causing severe road injuries beyond the roadway geometric design, traffic volume, and other risk indicators. However, the mechanisms through which these behaviors lead to crashes remain largely unclear, posing a considerable challenge in the realm of road safety analytics. Thus, recent studies have focused on driving behavior analysis by developing appropriate machine and deep learning models to enhance road safety and optimally detect and quantify the relationships between different features.

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 [3]. Speed has a direct impact on the frequency and severity of road crashes. A 1% increase in average speed leads to an increase of approximately 2% in the incidence of mild injuries from road crashes, 3% in the incidence of serious injuries, and 4% in fatalities [4]. 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 [5]. Therefore, to prevent and classify risky driving behavior, it is important to highlight the spatiotemporal dimension in which accelerations and brakings take place and examine their categorization into 'Dangerous' and 'Non-Dangerous' classes.

This study aims to bridge these gaps through the utilization of advanced machine learning algorithms that predict dangerous driving behaviors. It leverages a comprehensive dataset of over 356,000 trips, analyzing patterns of harsh accelerations and braking events to elucidate their predictive significance for road safety. The establishment of specific thresholds for 'Dangerous' and 'Non-Dangerous' driving behaviors furnishes actionable insights for traffic safety interventions. Employing a variety of analysis techniques, this study evaluates the impact of driving behavior features on the occurrence and severity of harsh events through feature selection and regression models. Subsequent developments include the use of classification algorithms, such as Random Forest, Gradient Boosting, Extreme Gradient Boosting, Multilayer Perceptron, and K-nearest Neighbors, to categorize the dependent variables into predetermined safety levels. The k-means clustering method establishes optimal thresholds for distinguishing between safe and risky driving behavior, identifying 48.82 harsh accelerations and 45.40 harsh brakings per 100 km as key thresholds. The results indicate that Gradient Boosting and Multilayer Perceptron algorithms outperform the others, achieving recall rates of approximately 67% to 68% for both harsh acceleration and harsh braking events.

This research offers an innovative approach to predicting harsh events, with the key contribution being the distinction between dangerous and non-dangerous driving behavior based on harsh events. The predictive nature of this study is emphasized by its ability to anticipate dangerous driving behaviors before they occur. Unlike traditional data-fitting models, these predictive models utilize real-time data to estimate the likelihood of harsh events, enabling proactive interventions. This capability is particularly important for enhancing Advanced Driver Assistance Systems (ADAS), where early warnings about potentially dangerous driving can significantly improve driver safety and overall road safety. This integrated approach enhances predictive accuracy and provides valuable insights for developing effective strategies to mitigate dangerous driving behaviors, thereby contributing to the overarching goal of improving road safety.

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. Following this, a detailed description of the research methodology is provided, including the theoretical basis of the models, as well as the data collection and processing procedures. The study results are then presented, leading to the formation of conclusions related to road safety.

## 2. Literature Review

In recent years, with the advancement of Intelligent Transport Systems (ITS), research has focused on the analysis of driving behavior to make safety systems capable of predicting and improving dangerous driving behavior. Given this context, several studies have applied various machine and deep learning techniques to create accurate predictive models. For instance, Papadimitriou et al. (2019) [6] proposed a methodology to quantify the correlation between dangerous driving behavior and mobile phone use through naturalistic driving data analysis. In this study, a model of binary logistic regression was deployed with an accuracy of 70%, revealing a large correlation between harsh events and mobile phone use. The study by Yang et al. (2021) [7] aimed to evaluate driving performance in real-time

by focusing on detecting the optimal number and thresholds for driving safety levels. To implement the research, three different clustering techniques were applied to the dataset; the K-means algorithm, hierarchical clustering, and Gaussian Mixture Models (GMM) combined with the Expectation-Maximization algorithm (EM), resulting in four optimal safety levels and a Support Vector Machine classification algorithm with 97.9% overall accuracy. The study of Yarlagadda et al. (2021) [8] proposed a framework using k-means clustering to identify aggressive driving patterns in heavy passenger vehicle drivers and revealed 13.5% of accelerations and 34.7% of braking maneuvers as aggressive [8]. The study establishes thresholds for aggressive maneuvers and demonstrates the heterogeneity in drivers' aggressive behavior through k-means clustering, emphasizing its potential for developing personalized driver assistance systems. Furthermore, Ali et al. (2021) [8] established the imperative goal of enhancing traffic safety on freeways by proposing an analysis protocol to distinguish between normal and risky driving in both clear and rainy weather conditions. The study employs a k-means cluster analysis, to classify driving patterns into normal and risky conditions, and results showed that risky driving patterns in rainy conditions start, on average, one second earlier and extend for three seconds longer compared to clear conditions.

Moreover, Zhang et al. (2016) [9] intended to classify driving behaviors by using only low-level sensors, such as data collected from the diagnostic outlet of the car (OBD) and smartphone sensors. Results showed that using the combined dataset of both smartphone and OBD sensors and applying an SVM algorithm, classification accuracy was 86.67%. Ghandour et al. (2021) [10] applied a methodology to classify behavior based on different psychological states of the driver in real driving conditions. This approach was comprised of several machine learning algorithms, such as Random Forests, Artificial Neural Networks, and Gradient Boosting, with the alternative behaviors being classified into three separate psychological levels. As a result, Gradient Boosting was found to be the optimal method to identify and predict the levels.

In addition, Mumcuoglu et al. (2019) [11] dealt with behavior pattern recognition based on data collected from a realistic truck model simulator. For this study, a Long Short-Term Memory (LSTM) algorithm was developed to detect dangerous driving behavior in short time frames based on a dataset consisting of IMU, GPS, and Radar/LiDAR signals. Results indicate that the LSTM structure has a substantial capability to recognize dynamic relations between driving signals in short time frames and could be widely deployed in prospective analyses.

Several advanced models have demonstrated high efficiency in various applications. For example, the LSTM-ALO model has been utilized in monthly runoff forecasting, achieving superior performance due to its ability to capture long-term dependencies [12]. Similarly, the ANFIS-PSO model has shown high accuracy in system modeling, effectively combining the benefits of adaptive neuro-fuzzy inference systems with particle swarm optimization [13]. Additionally, the SVM-FFAPSO model has been leveraged for its strengths in feature selection and parameter optimization, significantly boosting predictive accuracy in environmental data analysis [14]. Furthermore, the RVM-IMRFO model is recognized for its robust prediction capabilities, making it effective in handling complex regression tasks [15]. Despite the proven effectiveness of these advanced models, they require extensive data and high computational resources, which may not be feasible given the current study's constraints.

The selected models—Random Forest, Gradient Boosting, Extreme Gradient Boosting, Multilayer Perceptron, and K-Nearest Neighbors—were chosen for their demonstrated effectiveness in similar contexts of driving behavior analysis. These models offer several advantages, including high accuracy, robustness to overfitting, and computational efficiency. Unlike more complex deep learning models, which require extensive data and computational resources, the chosen models can be effectively trained and deployed with the available dataset and computational constraints. Additionally, these models provide interpretable results, which are crucial for understanding the factors contributing to dan-

gerous driving behaviors and for developing targeted interventions. Furthermore, the key contribution of this research lies in the distinction between Dangerous and Non-Dangerous behavior, a critical aspect of road safety and essential for developing effective interventions. Therefore, simpler models were developed to retain the primary focus of this research on driving behavior distinction.

Handling imbalanced datasets in machine learning methodology has great significance in improving the accuracy of the predictive models. Due to the utilization of a real-world dataset in assessing driving behavior, data imbalance is a known issue concerning the distribution of instances across different classes (i.e., dangerous and non-dangerous behavior). The minority class most commonly consists of dangerous behavior samples, due to the fewer driving delinquent behaviors and overall dangerous conduct compared to non-dangerous ones. In this regard, Wang et al. (2021) [16] propose a framework for an automated hyperparameter optimization technique based on Bayesian optimization that aims to tune and train machine learning algorithms. The inputs of the recognition models are discrete Fourier transform coefficients of several driving features, and oversampling was best handled with the combination of the Support Vector Machine-based Synthetic Minority Oversampling Technique (SVM-SMOTE), which can significantly increase the recognition ability and minimize the designated loss function.

Another study of L. Yang et al. (2018) [17] aimed to classify driving behavior by correlating it with electroencephalography (EEG) data. For the predictive model to be optimized, the ADaptive SYNthetic (ADASYN) sampling approach was employed for adaptively generating minority data samples according to their distributions, with the highest accuracy reaching 83.50%, suggesting significant correlations between EEG data and driving behavior. Zhu et al. (2022) [18] propose a novel machine learning framework for unbalanced time series samples. The meanShift technique was applied to cluster samples and expand their volume according to sample similarity before employing three Convolutional Neural Networks (CNN) to classify driving behavior. Results indicate that using time series samples to generate data in an imbalanced classification could prove beneficial. Furthermore, in the Katrakazas et al. (2018) [19] study, researchers exploited raw speed time series data of varying duration using several imbalanced learning techniques, such as undersampling and its integration with oversampling, resulting in the best technique being the integrated combination of Repeated Edited Nearest Neighbors (REEN) with Synthetic Minority Oversampling (SMOTE-ENN) for classes that are difficult to identify. This study highlights the significance of efficient imbalanced data handling, showcasing that classification results could be enhanced by up to 40% using imbalanced learning approaches.

Most of the studies reviewed dealt with distinguishing safety levels based on socio-demographic characteristics, psychological factors, and driving characteristics such as speed and headway. This study aims to develop driver risk behavior profiles based on harsh events, which, to the authors' knowledge, is an innovative approach and can be a useful tool in the field of road safety. Most of the studies reviewed dealt with distinguishing safety levels based on socio-demographic characteristics, psychological factors, and driving characteristics such as speed, headway, etc. This study aims to develop driver risk behavior profiles based on harsh events, which, to the authors' knowledge, is an innovative approach and can be a useful tool in the field of road safety. The motivation of the current research arises from the critical need to enhance road safety through the prediction and analysis of dangerous driving behaviors. The main contribution of this study lies in its innovative approach to utilizing a comprehensive dataset of over 356,000 trips to identify and analyze patterns of harsh accelerations and braking events. By establishing specific thresholds for 'Dangerous' and 'Non-Dangerous' driving behaviors, the research provided actionable insights that are pivotal for traffic safety interventions. Moreover, the development of driver risk behavior profiles represents a novel application in the field of road safety. With the analysis of various factors, such as exposure indicators (i.e., distance, duration), mental state (i.e., distraction) and driving characteristics (i.e., speed), these profiles help identify

high-risk drivers and tailor interventions to mitigate potential dangers. This approach integrates advanced technologies like machine learning, telematics, and Big Data analytics to create comprehensive and dynamic profiles that evolve with new information. The ultimate goal is to enhance road safety through personalized strategies, early warnings and targeted education, thereby fostering a safer driving environment for all road users.

While the potential benefits of incorporating comparisons with deep learning models and hybrid machine learning models are acknowledged, the current focus is on the utilization and optimization of advanced traditional machine learning algorithms to ensure interpretability and practical applicability for immediate traffic safety interventions. Given the complexity and resource-intensiveness of deep learning models, such comparisons have been reserved for future research, as noted in the limitations section of this study. This strategic decision allows for the presentation of a clear and actionable set of findings while laying the groundwork for subsequent investigations into more complex modeling techniques.

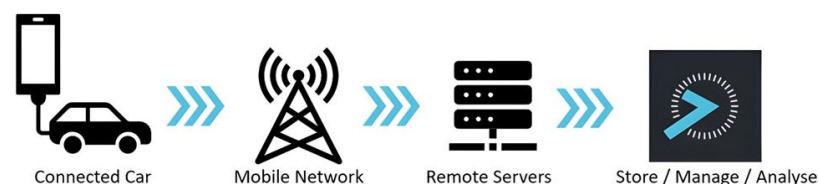
### 3. Materials and Methods

#### 3.1. Naturalistic Driving Experiment

The naturalistic driving dataset that was exploited for the analysis was collected and provided by OSeven Telematics, London, UK (<https://oseven.io/>, accessed on 15 July 2022) through a specialized and fully trip-integrated smartphone application that records and collects driving data continuously without any interference with the driving process. The application operates based on the hardware sensors of the smartphone without using any other identification equipment. Additionally, a plethora of APIs is used for data reading and temporary saving in the smartphone database before being transmitted to the company's back-end database. The collected data are specially marked spatiotemporally, and after being stored in the final database, they are converted into driving behavior and safety indexes through signal processing, machine learning algorithms, data fusion, and Big Data algorithms. More precisely, the hardware sensors that are exploited include the use of an accelerometer, gyroscope, magnetometer, and GPS, while the appointed data fusion techniques are provided by iOS and Android with nine degrees of freedom models (Yaw, Pitch, Roll), gravity, and linear acceleration, with data recording operating at a maximum frequency of 1 Hz. Data was recorded continuously throughout each trip at one-second intervals, ensuring detailed and comprehensive measurement of driving behavior during the entire trip.

It is important to note that the OSeven Telematics, London, UK platform (<https://oseven.io/>, accessed on 15 July 2022) has explicit privacy policy declarations and adheres to strong information security measures in accordance with the General Data Protection Regulation (GDPR) and relevant EU legislation. As a result, OSeven provided all data in an anonymized format, and no geolocation information for the trips (other than the relevant country) was included in the dataset.

Figure 1 illustrates the process that is carried out every time the application records a new trip.



**Figure 1.** OSeven data flow system.

For this study, the collected and processed data provided by OSeven included recordings of 356,162 different trips in an urban road network, with indexes provided for each trip exclusively amounting to 75. However, our analysis focuses on a subset of 23 key variables relevant to this study's objectives. Detailed descriptions of these variables and

their descriptive statistics can be found in Appendix A Table A1 and Table A2, respectively. The recordings occurred during the outbreak of the pandemic SARS-CoV-2 (e.g., from January 2020 to December 2020), with indexes associated with policy measures that governments took to tackle the pandemic, such as the Stringency Index and Restrictions Index, not being included in the analysis due to their irrelevant status at the moment. Overall, the Naturalistic Driving dataset consists of important risk factors for driving behavior related to traffic conditions and driver state.

To evaluate the driving quality of each trip, OSeven Telematics provides detailed score indicators, with their values ranging from 0 to 100, with 100 indicating a behaviorally perfect trip. The overall score of the user is calculated as the weighted average of the trips' scores over the last twelve (12) months, where the user was the driver, with distance as the weighting factor.

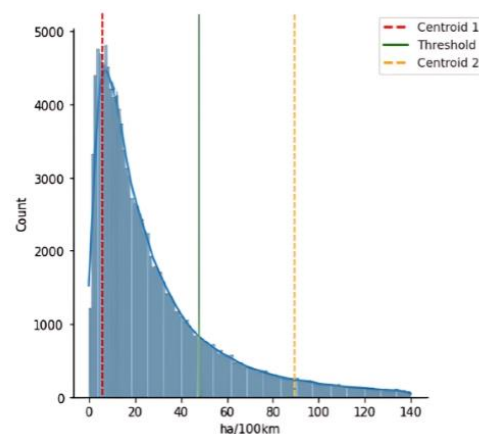
### 3.2. Definition of Driving Behavior Levels

After the examination of previous studies, the optimal methodology for determining safety levels proved to be clustering techniques [7,8,20]. Upon comparing several different techniques, the k-means clustering method was selected to find the required numerical thresholds for every dependent variable, with the given clusters representing the predefined classes. Thus, two clusters and, therefore, two centroids were selected for the computation. The subsets underwent renormalization.

The study hypothesis suggests that the k-means clustering method can effectively distinguish driving behavior into two distinct safety levels (e.g., Non-Dangerous and Dangerous) based on thresholds determined for harsh acceleration and braking events per 100 km. Additionally, the hypothesis postulates that machine learning algorithms, such as Random Forest, Gradient Boosting, Extreme Gradient Boosting, Multilayer Perceptron, and K-Nearest Neighbors, will accurately classify driving behavior into these safety levels.

#### 3.2.1. Based on Harsh Accelerations Events per 100 Km

For the set of harsh accelerations per 100 km, the k-means algorithm performed a vector quantization of the data into two clusters. The centroids of these clusters, which represent the average values of the harsh acceleration events within each cluster, were found to be 5.693 and 91.942 events per 100 km, respectively. These centroids are crucial as they indicate the central tendency of the two groups (clusters) of data points identified by the k-means algorithm. The binary distribution value threshold, which differentiates between "Non-Dangerous" and "Dangerous" driving behavior, was determined by taking the average of these centroid values. This threshold was calculated to be 48.817 harsh acceleration events per 100 km, as shown in Figure 2. Therefore, the variable data are defined as 0 (Non-Dangerous) for values below the threshold and 1 (Dangerous) for values above it.

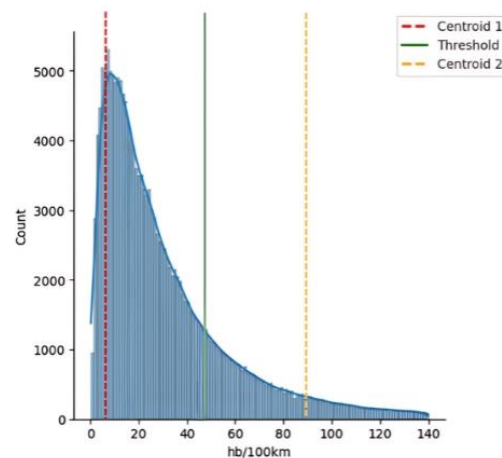


**Figure 2.** Centroids and threshold determination for the harsh acceleration events per 100 km with the K-means algorithm.

The two clusters resulting from the aforementioned process consisted of 330,395 Non-Dangerous behavior trips and 25,767 Dangerous behavior trips, with percentages of 93% and 7%, respectively.

### 3.2.2. Based on Harsh Braking Events per 100 km

Similarly, the same procedure was performed for the set of harsh braking events per 100 km. The k-means algorithm resulted in a vector quantization of the data into two clusters. The centroids of these clusters were found to be 7.975 and 82.835 events per 100 km, respectively. These centroids represent the average values of the harsh braking events within each cluster. The binary distribution value threshold, which differentiates between “Non-Dangerous” and “Dangerous” driving behavior, was determined by taking the average of these centroid values. This threshold was calculated to be 45.405 harsh braking events per 100 km, as shown in Figure 3.



**Figure 3.** Centroids and threshold determination for the harsh braking events per 100 km with the K-means algorithm.

The two clusters resulting from the aforementioned process consisted of 315,986 Non-Dangerous behavior trips and 40,176 Dangerous behavior trips, with percentages of 89% and 11%, respectively.

To conclude, the use of the k-means algorithm provided specific thresholds for harsh acceleration and harsh braking events per 100 km to distinguish between Dangerous and Non-Dangerous driving behavior. The observed thresholds of harsh events per 100 km may initially seem high; however, this could be attributed to the nature of the driving conditions, such as urban traffic density or complex road scenarios. It is crucial to highlight that the identified thresholds are not absolute but are rather contextualized within the parameters of the studied environment, including a wide range of harsh events that directly affect the outcomes. Nevertheless, the distribution of this study’s results is consistent with the literature, meaning that Non-Dangerous behavior instances are the major class while Dangerous driving behavior instances are the minority class.

The methodological framework on which the present research is based depends on the various mathematical and statistical concepts of the complex machine learning techniques applied and of the model acceptance criteria. In particular, the methodological approach is structured into a Feature Selection, a pre-processing of the provided data, and, finally, a Classification stage with all the techniques and the relevant theories applied.

### 3.3. Feature Selection

Among the 23 variables analyzed in this study, those related to harsh acceleration and harsh braking events were specifically utilized to define driving behavior as either dangerous or non-dangerous. The remaining variables were evaluated using the feature selection method to determine their significance in predicting driving behavior and to

retain the most influential ones, ensuring the highest performance of the models. Feature selection is a process aiming to identify the more effective independent variables and segregate them from the totality of variables. It is thought to be an optimization procedure for classification, as it reduces the possibility of results being deviated due to the excessive volume of input variables. According to [21], when the input variables are more abundant, the volume of the space increases so radically that the available data becomes sparse, leading to an exponential growth within dimensionality; a phenomenon called the “curse of dimensionality”. According to previous studies [22–24], a permutation importance-based feature selection is utilized.

According to Figures 4 and 5, the total distance travelled and driving duration have a major impact on the regression process. Conversely, the time of mobile usage and distance travelled during dangerous time zones (00:00–05:00) (e.g., risky\_hours in Figure 4) have the least direct effect on the regression process.

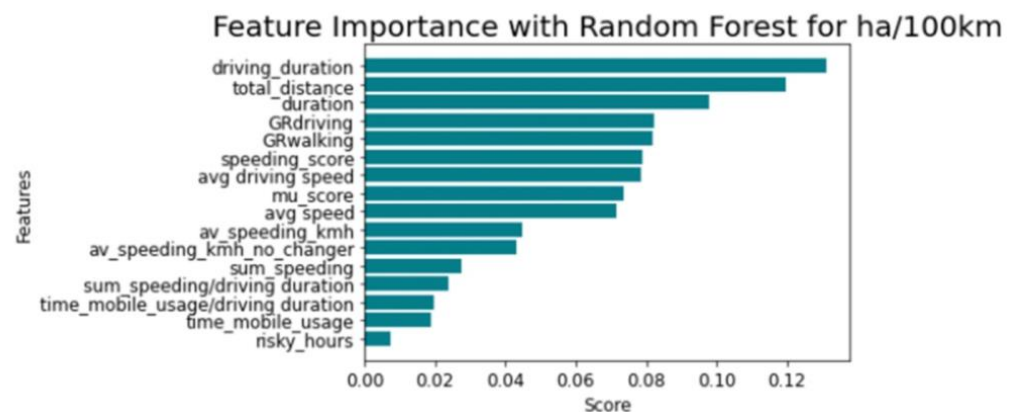


Figure 4. Permutation feature importance for harsh acceleration events per 100 km.

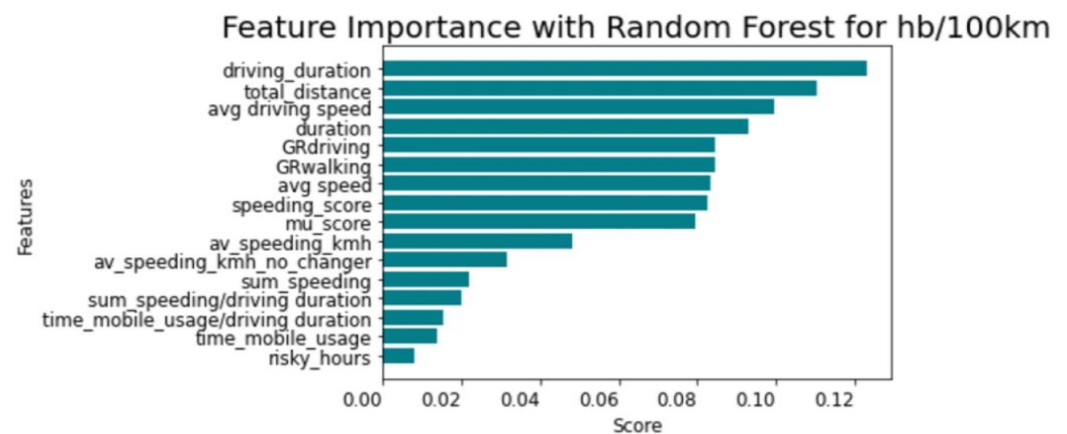


Figure 5. Permutation feature importance for harsh braking events per 100 km.

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. The following Table 1 displays some descriptive statistics for the input variables employed in the classification procedure, including mean, standard deviation, maximum, minimum, and mean values.

In this study, the use of additional noise decomposition methods was not deemed essential for real-time prediction purposes. Although feature selection helps to eliminate certain noise by identifying the most significant variables, future research will consider integrating decomposition techniques such as Empirical Mode Decomposition (EMD) or Wavelet Transform to further enhance noise management and improve predictive accuracy.



**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 (s)	769.97	967.15	61.00	23,900.00
Average driving 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%

### 3.4. Classification Process

To achieve the main objective of this study, (i.e., the identification of dangerous driving behavior), five classification models were proposed due to their strong performance and widespread use in the literature for identifying unsafe driving behavior, real-time collision prediction, and other real-world challenges. The selected algorithms were Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost, version 1.5.2), K-Nearest Neighbors (kNN), and Multilayer Perceptron (MLP). The predictive nature of these models is emphasized by their ability to anticipate dangerous driving behaviors before they occur, utilizing real-time data to forecast the likelihood of harsh events.

The dataset was divided into a training dataset and a test dataset in order to train and evaluate the classification algorithms. A training dataset follows the form  $X_{\text{training}} = (x_n, y_n)$ ,  $n = 1, N$ , where  $x_n$  is a predictor variable and  $y_n = 0.1$  is the target variable (in this case  $y$  refers to Dangerous and Non-Dangerous level).

To evaluate and compare the performance of the five models, several key metrics were employed, described by Equations (1) to (6).

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

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Sensitivity (recall)} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{f-1 score} = \frac{2 \times (\text{Precision}) \times (\text{Recall})}{(\text{Precision}) + (\text{Recall})} \quad (5)$$

$$\text{False alarm rate} = \frac{FP}{FP + TN} \quad (6)$$

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.

The proposed performance indexes are specifically selected for their ability to provide a comprehensive and nuanced evaluation of the models' effectiveness. These metrics collectively ensure that the models not only identify dangerous driving behaviors accurately but also minimize false positives and negatives. In practical terms, this means that the models are effective at correctly identifying risky driving behaviors (reducing the chance of missing dangerous incidents) and at avoiding misclassifying safe drivers as dangerous (reducing unnecessary interventions). By adopting these indexes, this study can better gauge the reliability and robustness of the predictive models, leading to more precise

and actionable insights for road safety improvements. This comprehensive evaluation framework significantly enhances the potential for developing effective interventions and policies aimed at reducing risky driving behaviors, ultimately contributing to greater road safety and sustainability.

#### 3.4.1. Random Forest (RF)

Ensemble learning approaches combine predictions from multiple weak classifiers to make a more accurate prediction than a single model. The Random Forest (RF) algorithm combines ensemble learning methods with the decision tree framework to create multiple randomly designed trees leading to powerful predictions. In particular, the RF model employs the bootstrapping and aggregation techniques. The bootstrapping technique involves training multiple decision trees in parallel using different subsets of the dataset. The final decision is made by aggregating the decisions of each decision tree. Through Grid Search, the key parameters found to be optimal were: (a) number of trees/estimators (150), (b) criterion ('entropy'), and (c) maximum depth (20).

#### 3.4.2. Gradient Boosting (GB)

Another type of ensemble method is Gradient Boosting (GB), which utilizes multiple learning algorithms to yield the best predictive ability, in the form of a boosted decision tree (BDT). The fundamental idea underlying this model is to build the new base learners to be maximally correlated with the negative gradient of the loss function, associated with the entire ensemble [25]. Grid Search optimization identified the key parameters as: (a) number of estimators (200), (b) maximum depth (6), and (c) learning rate (0.1).

#### 3.4.3. Extreme Gradient Boosting (XGBoost)

The Extreme Gradient Boosting (XGBoost) algorithm is an optimized form of the Gradient Boosting model that operates as a Newton–Raphson algorithm, using a second-order Taylor approximation, contrary to Gradient Boosting, which relies on gradient descent. More specifically, XGBoost is an implementation of gradient-enhanced decision trees, in which trees are generated sequentially with significantly higher model accuracy, in less computational training time, than standard machine learning models. The key parameters were found to be: (a) number of estimators (200), (b) maximum depth (6), and (c) learning rate (0.1).

#### 3.4.4. Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is a supplement of a feed-forward neural network. Each MLP consists of at least 3 node layers: the input layer, the hidden layer, and the output layer, where each node, except the input layer, consists of an artificial neuron using a nonlinear activation function. Each neuron of each layer is connected to the previous and next layer, with this process called neural synapses. A key performance element in classification in MLPs is the number of hidden layers, between those of input and output. In MLPs, the basic learning technique used to train all the nodes is called backpropagation. Grid Search determined that the best parameters were: (a) activation function ('tanh'), (b) alpha (0.0001), (c) hidden layer sizes (10, 30, 10), and (d) learning rate ('adaptive').

#### 3.4.5. K-Nearest Neighbors (kNN)

The K-Nearest Neighbors (kNN) algorithm is one of the most important and simple to apply to classification problems, as it does not require any primary knowledge about the distribution of the data, also overcoming the obstacle of parametric estimates of probability densities that are difficult to disambiguate [26]. For this reason, this model works most effectively on relatively small datasets. The kNN classifier is based on considering the Euclidean distance between a test data sample and training data samples by estimating their vector. More simply, the algorithm's operation classifies the data, based on their

common characteristics, as it assumes that they are in proximity. Using Grid Search, the number of neighbors was optimized to be (7).

The overall methodological approach of this study offers several unique advantages that enhance its value and effectiveness. The utilization of a comprehensive naturalistic driving dataset from OSeven Telematics provides a highly detailed and realistic perspective on urban driving behavior. With over 356,162 trips recorded at one-second intervals, the dataset's high granularity allows for precise and thorough analysis of driving behaviors and harsh events. This focus on urban driving environments ensures that the developed models are specifically tailored to address the unique challenges and conditions found in urban settings, thereby improving their applicability and effectiveness. Moreover, the methodology includes robust feature selection and the application of advanced machine learning algorithms. This combination enhances the accuracy and reliability of the models in distinguishing and predicting Non-Dangerous versus Dangerous driving behaviors. The k-means clustering method further refines the analysis by establishing optimal thresholds for harsh events, ensuring a clear and actionable distinction between different safety levels. By integrating these comprehensive data collection and advanced analytical techniques, this case study not only improves the predictive power of the models but also provides valuable insights that can be directly applied to enhance road safety in urban areas.

#### 4. Results

After evaluating approaches employed in previous research [17–19], numerous re-sampling techniques, including SMOTE and SMOTE-ENN, were examined. To handle imbalanced datasets and prevent overfitting, the Synthetic Minority Oversampling Technique (SMOTE) is believed to be the most effective approach [27].

Following the oversampling process based on the SMOTE method, the development of the five classification models was performed for each category of harsh events (e.g., harsh acceleration and harsh braking events). As already mentioned, the classification of driving behavior has a binary form in two individual discrete classes: Dangerous Behavior and Non-Dangerous Behavior. To ensure optimal performance of the algorithms, a thorough optimization of hyperparameters was meticulously carried out using GridSearch before commencing with the model training.

##### 4.1. Evaluation of Classification Models for Harsh Acceleration Events

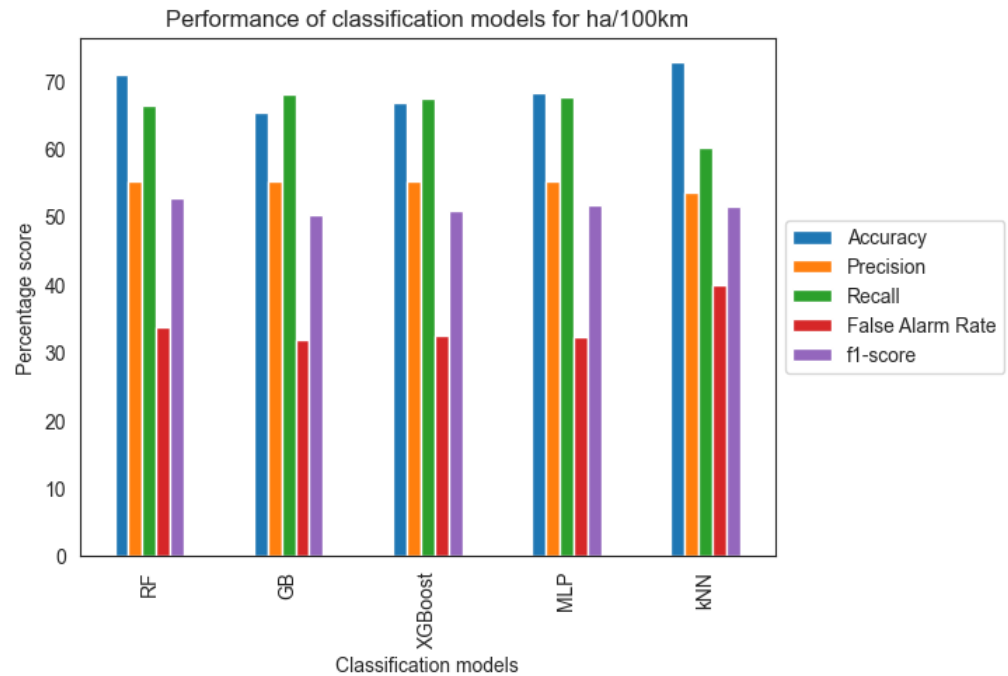
In order to properly evaluate the models, it is important to highlight that the misclassification of driving behavior that belongs to the Dangerous class is the most critical evaluation because of the potential road safety risk associated with this error. Additionally, based on the literature on data imbalance problems [28], there is a high risk that accuracy can lead to misleading conclusions (“accuracy paradox”). Based on the above, according to Table 2 and Figure 6, additional evaluation metrics such as precision, recall, false alarm rate, and f-1 score are considered

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

Classification Models	Accuracy	Precision	Recall	False Alarm Rate	F-1 Score	AUC
RF	70.83%	55.16%	66.39%	33.61%	52.64%	73.98%
GB	65.28%	55.15%	68.05%	31.95%	50.25%	75.10%
XGBoost	66.76%	55.09%	67.46%	32.54%	50.86%	74.26%
MLP	68.16%	55.26%	67.65%	32.35%	51.63%	74.67%
kNN	72.70%	53.46%	60.08%	39.92%	51.47%	64.55%

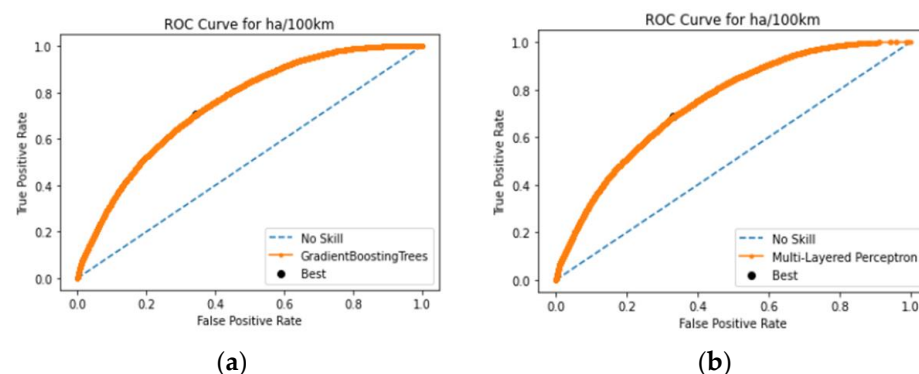
Notably, the five models achieved reasonable accuracy and recall compared to the precision metric. Based on the provided FNR scores, the models generally achieved satisfactory results in terms of capturing instances of dangerous driving behavior. A high recall indicates that the model is effective at capturing the majority of actual dangerous instances, reducing the chances of false negatives. Consequently, a lower false alarm

rate implies a smaller proportion of dangerous instances being incorrectly classified as non-dangerous. Given that misclassifying dangerous incidents as non-dangerous poses significant road safety risks compared to the other way around, it is concluded that recall, false alarm rate, and AUC score are the most important metrics evaluated.



**Figure 6.** Classification metrics for harsh acceleration events per 100 km.

The ROC curves presented in Figure 7 further illustrate the performance of the best-performing models—Gradient Boosting (GB) and Multilayer Perceptron (MLP). The ROC curve for GB (Figure 7a) demonstrates a strong ability to distinguish between dangerous and non-dangerous driving behaviors, with an area under the curve (AUC) that indicates high overall performance. Similarly, the ROC curve for MLP (Figure 7b) shows comparable effectiveness, underscoring the robustness of these models in capturing true positive rates while maintaining low false positive rates.



**Figure 7.** (a) ROC curve for the GB classifier; (b) ROC curve for the MLP classifier.

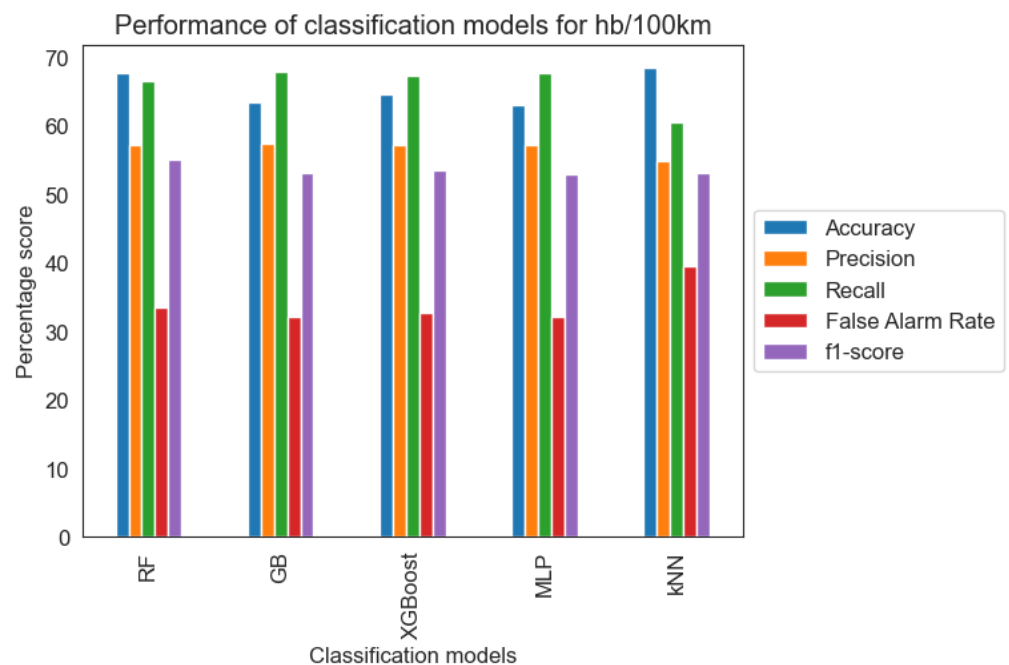
Overall, the models performed adequately in terms of predictive ability and at relatively similar performance levels compared to each other. The Gradient Boosting (GB) and Multilayer Perceptron (MLP) classifiers provide the best results in terms of recall and false alarm rate among the five models, with GB slightly outperforming MLP. Also, both models scored satisfactory AUC scores, 75.1% for the GB model and 74.67% for the MLP model, which means that the performance of the algorithm is relatively good.

#### 4.2. Evaluation of Classification Models for Harsh Braking Events

Similar to the analysis for harsh acceleration events, additional evaluation metrics were examined due to the imbalanced dataset and the hidden risk of the “accuracy paradox”. Based on Table 3 and Figure 8, the five models achieved satisfactory performance in identifying dangerous driving behavior.

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

Classification Models	Accuracy	Precision	Recall	False Alarm Rate	F-1 Score	AUC
RF	67.78%	57.20%	66.48%	33.52%	55.09%	73.62%
GB	63.36%	57.36%	67.91%	32.09%	53.13%	74.88%
XGBoost	64.53%	57.20%	67.30%	32.70%	53.60%	74.28%
MLP	62.96%	57.29%	67.80%	32.20%	52.88%	74.69%
kNN	68.45%	54.88%	60.55%	39.45%	53.19%	65.00%

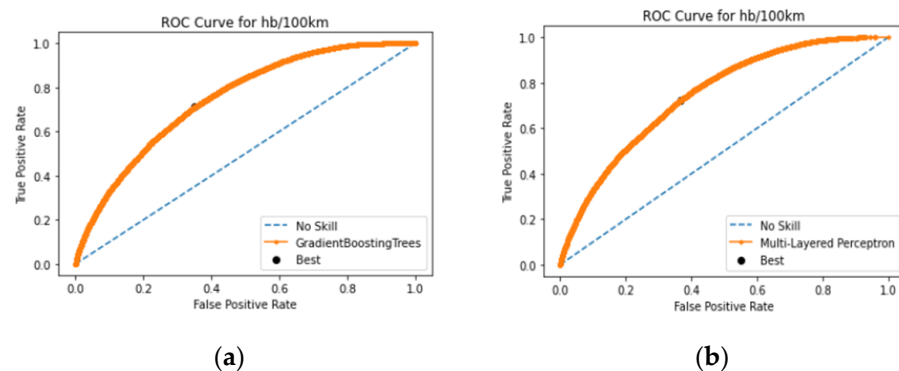


**Figure 8.** Classification metrics for harsh braking events per 100 km.

Overall, the five models demonstrated strong performance metrics similar to those observed in the analysis of harsh acceleration events. Nevertheless, the Gradient Boosting (GB) and Multilayer Perceptron (MLP) models outperformed the rest in terms of recall and false alarm rate. As mentioned in Section 4.1, these metrics are the most crucial for the hypothesis of this research. Moreover, GB and MLP achieved adequate AUC score with similar results 74.88% and 74.69%, respectively.

The ROC curves for harsh braking events, illustrated in Figure 9, provide additional insights into the performance of the best-performing models. The ROC curve for GB (Figure 9a) shows a robust performance, indicating its ability to effectively discriminate between dangerous and non-dangerous driving behaviors. The AUC score for GB reflects a strong overall performance, ensuring that the model captures a high proportion of true positives while maintaining a low rate of false positives. Similarly, the ROC curve for MLP (Figure 9b) demonstrates its competency in identifying dangerous driving behaviors associated with harsh braking. The model maintains a balance between sensitivity and specificity, evidenced by the high AUC score. This indicates that the MLP model is highly

reliable in predicting harsh braking events, thereby minimizing the risks of misclassification and enhancing road safety interventions.



**Figure 9.** (a) ROC curve for the GB classifier; (b) ROC curve for the MLP classifier.

## 5. Discussion

Regarding feature selection and feature importance, distance and total trip duration emerged as the most important factors affecting the identification of driving behavior. An increase in the distance traveled and the total duration typically worsens the driver's behavior, as signs of fatigue, sleepiness, and impaired perception appear, reaction time is affected, and psychomotor coordination is burdened. Moreover, vehicle speed was found to be an important parameter for sudden incidents and contributes significantly to accidents and dangerous driving behavior in general, confirming international literature.

Increased speed leads to a rapid decrease in the driver's perception and corresponding reaction to external factors. Finally, it is worth noting that driving during the dangerous time zone (00:00–05:00) (e.g., *risky\_hours*) is not strongly associated with harsh events compared to other factors. The data suggest that lower traffic and pedestrian volumes during these hours might mitigate the occurrence of risky behaviors typically expected due to reduced visibility and increased driver fatigue. The feature selection process used in this study is of great importance. By carefully identifying the features of driving behavior that significantly impact harsh events, the research highlights the key factors that contribute to road crashes. This knowledge strengthens the field of road safety as well as empowers researchers and safety experts to focus on specific aspects of driving behavior that require attention and intervention, streamlining efforts to mitigate risky behaviors and reduce the incidence of serious road injuries.

In addition, the use K-means clustering method, established optimal safety thresholds of harsh events to distinguish driving behavior between 'Dangerous' and 'Non-Dangerous': (a) 48.82 harsh accelerations per 100 km and (b) 45.40 harsh brakings per 100 km. By utilizing the K-means clustering method and analyzing naturalistic driving data, this framework provides a systematic approach to categorizing driving behavior on safety levels based on naturalistic driving data. Most field studies gravitate towards identifying and classifying into more than two safety levels, whereas this paper is solely focused on binary classification. Moreover, this study provides original thresholds for the distinction between dangerous and non-dangerous driving behavior as well as actionable criteria for traffic safety applications, presenting a methodological advancement over traditional heuristic-based approaches.

To handle the problem of an imbalanced distribution of minority classes, the SMOTE technique was chosen, as the variations in the data were particularly strong and the proportion of the majority class was very large. SMOTE proved to be an effective method for large and multilevel data, validating the scientific community worldwide. However, future research should consider exploring more sophisticated over-sampling techniques such as SVM-SMOTE and SMOTE-Tomek.

In the analysis of harsh driving behaviors, the data were analyzed separately for each category based on two categories of harsh incidents: (a) harsh acceleration events and (b) harsh braking events. Subsequently, five machine learning algorithms—Random Forest (RF); Gradient Boosting (GB); Extreme Gradient Boosting (XGBoost); Multilayer Perceptron (MLP); and K-Nearest Neighbors (kNN)—were meticulously developed and employed to predict these behaviors. The GB and MLP models were found to have the highest predictive ability for risky driving behaviors compared to the other algorithms for both categories of harsh events (e.g., harsh acceleration events and harsh braking events). According to Ghandour et al. (2021) [10], the gradient boosting algorithm outperforms ANN and RF classifiers, striking a significant 60% on accuracy and macro f-1 and 59% on macro precision and macro recall, in comparison to the present study's 63.4%, 53.1%, 57.4%, and 67.9%, respectively. This is due to the algorithm's ability to use clone decision trees with exemplary flexibility. While in present research MLPs are not greatly outperformed by GB, with a notable 63% in accuracy, 57.3% in precision, 67.8 in recall, and 52.9% in f-1 score. A serious decrease in performance is observed in Ghandour et al. (2021) [10] study with a 30% decrease in accuracy and macro precision, 28% in macro recall, and 27% in macro f-1 score. Nonetheless, deep learning models, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or Long Short-Term Memory (LSTM), due to their training ability and auto-corrective nature, could be investigated as a promising alternative to traditional machine learning approaches in future research, allowing for the automatic extraction of complex features and the capture of non-linear relationships for improved classification performance and generally producing better results on multilevel contextual data, which this study is associated with. Due to limitations of processing power and time, these analyses could not be performed during the time of this study.

The classification models developed contribute additional insights to the existing literature and research in the field of driving behavior analysis. The classification models developed in this research not only contribute additional insights to the existing literature and research in the field of driving behavior analysis but also represent a significant advancement in the quest for improved road safety measures. By employing machine learning algorithms specifically designed for imbalanced datasets, the models demonstrate a robust ability to detect, classify, and predict dangerous driving behavior and harsh events with remarkable accuracy. Comparing and evaluating the five algorithms showed that the differences were not significant. This phenomenon supports the hypothesis that the dependent variables of harsh events were well clustered by the K-means method at two safety levels and proves that these algorithms are efficient for classifications at two-class safety levels. In the future, clustering using the Gaussian Mixture Model (GMM) can be investigated and may prove to be particularly useful due to the Gaussian distribution of the variables.

Overall, this study offers a viable alternative approach for harsh event prediction through a thorough investigation of human behavior in road safety. The insights obtained contribute significantly to the field's background knowledge and have far-reaching implications for road safety enhancement efforts. The practical applications of these findings are particularly crucial for policymakers and driving behavior analysts, who can leverage the identified thresholds for harsh driving behaviors to develop dynamic interventions. Specifically, the model's ability to differentiate between dangerous and non-dangerous driving behaviors provides a robust foundation for real-time adaptive speed management systems, significantly mitigating the risk of crashes in high-risk zones. Furthermore, the integration of these advanced machine learning techniques into existing vehicular telematics systems could enhance predictive accuracy and operational efficiency, offering traffic management authorities the capability for immediate response to potential harsh driving events. Additionally, this research can inform the development of personalized driver training programs, targeting the mitigation of risky behaviors identified through the analysis. By embedding these research findings into practical road safety measures, significant advancements can be achieved, aligning with global safety initiatives (i.e., Vision Zero).

Finally, the key novelty of this research lies in the distinction between Dangerous and Non-Dangerous behavior based on harsh events employing the K-means technique. This distinction and the context in which it is performed can extend the existing literature and provide important insights into the field of road safety, specifically in profiling drivers according to the number of harsh events.

## 6. Conclusions

This paper proposes a comprehensive framework for analyzing and classifying driving behavior as Dangerous or Non-Dangerous. The framework involves defining safety levels using clustering algorithms, selecting relevant features, and addressing imbalanced datasets. A naturalistic driving experiment collected data through OSeven Telematics' specialized smartphone application, focusing on harsh acceleration and harsh braking events. Five machine learning algorithms were developed: Random Forest (scikit-learn version 1.0.2), Gradient Boosting (scikit-learn version 1.0.2), XGBoost (version 1.5.2), Multi-layer Perceptron (scikit-learn version 1.0.2), and K-Nearest Neighbors (scikit-learn version 1.0.2), with Gradient Boosting and Multilayer Perceptron demonstrating superior predictive ability for both harsh event categories.

Although the key contribution of this research lies in the distinction of driving behavior (e.g., dangerous, and non-dangerous) based on harsh events using the k-means clustering technique, there are certain limitations that future studies need to address. Firstly, future research could explore deep learning models such as CNN, RNN, or LSTM to enhance classification accuracy and predictive power. Due to limitations in processing power and time, these models were not employed in this research. Additionally, alternative feature importance measurement methods such as the Bayesian Information Criterion or Akaike Information Criterion could provide further insights into the factors influencing dangerous driving behavior, and the potential of Gaussian Mixture Models as an alternative insightful clustering method. Moreover, this study focused solely on urban road networks, which may limit the generalizability of the findings to other types of driving environments such as rural or highway settings. Hence, future research should expand this study to include different driving environments to validate and generalize the findings. Furthermore, the dataset was collected during the SARS-CoV-2 pandemic, encompassing both lockdown and non-lockdown periods, which could have influenced driving behavior in ways that may not be typical during normal periods. Therefore, future studies should consider the temporal dynamics of driving behavior, exploring how driving patterns evolve over time and under different conditions, including external factors such as weather. By addressing these limitations, the robustness and applicability of the findings can be further enhanced, ultimately contributing to the field of road safety.

This paper makes a significant contribution to machine learning analyses of driving behavior, offering insights into data manipulation and handling. The utilization of comprehensive driving profiles, particularly focusing on the analysis of harsh accelerations and brakings, holds potential for personalized driver feedback, enhanced training programs, and advancements in the automotive industry.

**Author Contributions:** A.K., software, formal analysis, data curation, and writing—original draft preparation; T.G., formal analysis, data curation, and writing—original draft preparation; E.M., methodology, formal analysis, and writing—review and editing; C.K., conceptualization and methodology; G.Y., conceptualization, supervision, and resources. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.



**Acknowledgments:** The authors would like to thank OSeven Telematics, London, UK for providing all necessary data exploited to accomplish this study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Table A1.** Description of the data collected from the OSeven database.

Variable	Unit	Description
duration	s	Total duration of the route
total_distance	km	Total distance traveled
total_score	%	Total score
speeding_score	%	Excessive speeding score
mu_score	%	Mobile usage score during the route
hb_score	%	Score of sudden decelerations
ha_score	%	Score of sudden accelerations
risky_hours	km	Distance traveled during risky hours zone (00:00–05:00)
ha	-	Sudden accelerations on a route
hb	-	Sudden decelerations on a route
sum_speeding	s	Total duration of driving with excessive speed on a route (Speed limit + Tolerance limits)
av_speeding_kmh	km/h	Average driving speed during speed limit exceedance
time_mobile_usage	s	Total duration of mobile usage on a route
driving_duration	s	Total driving duration (excluding situations of vehicle immobility/stop, parking)
ha/100 km	-	Sudden accelerations per 100 km
hb/100 km	-	Sudden decelerations per 100 km
avg speed	km/h	Average speed of route
time_mobile_usage/driving duration	s/s	Mobile usage duration per unit of total driving duration
sum_speeding/driving duration	s/s	Driving duration with excessive speed per unit of total driving duration
av_speeding_kmh_no_changer	km/h	Average difference between driving speed and speed limit
avg driving speed	km/h	Average driving speed
GRdriving	%	Percentage daily difference in load of passenger car drivers compared to pre-pandemic conditions
GRwalking	%	Percentage daily difference in load of passenger car drivers compared to pre-pandemic conditions

**Table A2.** Descriptive statistics of the collected variables.

Variable	Mean	Standard Deviation	Variance	Min	Max
duration	962.04	1093.98	$1.197 \times 10^6$	61.0	25,549.0
total_distance	11.6	22.31	$4.979 \times 10^2$	0.5	648.68
total_score	75.66%	23.47	$5.510 \times 10^2$	10.00%	100.00%

Table A2. Cont.

Variable	Mean	Standard Deviation	Variance	Min	Max
speeding_score	76.52%	32.92	$1.084 \times 10^3$	10.00%	100.00%
mu_score	80.53%	34.62	$1.198 \times 10^3$	10.00%	100.00%
hb_score	79.20%	21.35	$4.558 \times 10^2$	10.00%	100.00%
ha_score	81.59%	19.74	$3.897 \times 10^2$	10.00%	100.00%
risky_hours	0.37	4.01	$1.606 \times 10$	0.0	427.7
ha	0.89	1.97	3.882	0.0	121.0
hb	1.26	2.2	4.856	0.0	87.0
sum_speeding	65.63	194.31	$3.776 \times 10^4$	0.0	7697.0
av_speeding_kmh	4.0	6.03	$3.633 \times 10$	0.0	314.16
time_mobile_usage	39.11	159.27	$2.537 \times 10^4$	0.0	9901.0
driving_duration	769.97	967.15	$9.354 \times 10^5$	61.0	23,900.0
ha/100 km	11.95	27.86	$7.763 \times 10^2$	0.0	597.01
hb/100 km	16.39	29.76	$8.858 \times 10^2$	0.0	819.67
avg speed	35.13	18.89	$3.567 \times 10^2$	1.96	262.52
time_mobile_usage/driving duration	0.05	0.14	$2.079 \times 10^{-2}$	0.0	0.99
sum_speeding/driving duration	0.06	0.11	$1.133 \times 10^{-2}$	0.0	1.0
av_speeding_kmh_no_changer	9.51	11.11	$1.234 \times 10^2$	0.0	329.16
avg driving speed	42.57	17.58	$3.091 \times 10^2$	5.57	323.91
GRdriving	109.64	55.88	$3.123 \times 10^3$	0.0	241.14
GRwalking	114.35	61.94	$3.837 \times 10^3$	0.0	254.21

## References

- World Health Organization. *World Health Organization Global Status Report on Road Safety 2023*; World Health Organization: Geneva, Switzerland, 2023.
- European Commission. *Directorate-General for Mobility and Transport Next Steps towards 'Vision Zero': EU Road Safety Policy Framework 2021–2030*; Publications Office: Luxembourg, 2020.
- Kröyer, H.R.G. Is 30 Km/h a 'Safe' Speed? Injury Severity of Pedestrians Struck by a Vehicle and the Relation to Travel Speed and Age. *IATSS Res.* **2015**, *39*, 42–50. [[CrossRef](#)]
- Nilsson, G. Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety. Ph.D. Thesis, Lund University, Lund, Sweden, 2004.
- Bonsall, P.; Liu, R.; Young, W. Modelling Safety-Related Driving Behaviour—Impact of Parameter Values. *Transp. Res. Part A Policy Pract.* **2005**, *39*, 425–444. [[CrossRef](#)]
- Papadimitriou, E.; Argyropoulou, A.; Tselentis, D.I.; Yannis, G. Analysis of Driver Behaviour through Smartphone Data: The Case of Mobile Phone Use While Driving. *Saf. Sci.* **2019**, *119*, 91–97. [[CrossRef](#)]
- Yang, K.; Al Haddad, C.; Yannis, G.; Antoniou, C. Driving Behavior Safety Levels: Classification and Evaluation. In Proceedings of the 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Heraklion, Greece, 16–17 June 2021; pp. 1–6.
- Ali, E.M.; Ahmed, M.M.; Yang, G. Normal and Risky Driving Patterns Identification in Clear and Rainy Weather on Freeway Segments Using Vehicle Kinematics Trajectories and Time Series Cluster Analysis. *IATSS Res.* **2021**, *45*, 137–152. [[CrossRef](#)]
- Zhang, C.; Patel, M.; Buthpitiya, S.; Lyons, K.; Harrison, B.; Abowd, G.D. Driver Classification Based on Driving Behaviors. In Proceedings of the 21st International Conference on Intelligent User Interfaces, Sonoma, CA, USA, 7–10 March 2016; ACM: New York, NY, USA, 2016; pp. 80–84.
- Ghandour, R.; Potams, A.J.; Boulkaibet, I.; Neji, B.; Al Barakeh, Z. Driver Behavior Classification System Analysis Using Machine Learning Methods. *Appl. Sci.* **2021**, *11*, 10562. [[CrossRef](#)]

11. Mumcuoglu, M.E.; Alcan, G.; Unel, M.; Cicek, O.; Mutluergil, M.; Yilmaz, M.; Koprubasi, K. Driving Behavior Classification Using Long Short Term Memory Networks. In Proceedings of the 2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), Turin, Italy, 2–4 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
12. Yuan, X.; Chen, C.; Lei, X.; Yuan, Y.; Muhammad Adnan, R. Monthly Runoff Forecasting Based on LSTM–ALO Model. *Stochastic Environmental Research and Risk Assessment* **2018**, *32*, 2199–2212. [[CrossRef](#)]
13. Basser, H.; Karami, H.; Shamshirband, S.; Akib, S.; Amirmojahedi, M.; Ahmad, R.; Jahangirzadeh, A.; Javidnia, H. Hybrid ANFIS–PSO Approach for Predicting Optimum Parameters of a Protective Spur Dike. *Appl Soft Comput* **2015**, *30*, 642–649. [[CrossRef](#)]
14. Adnan, R.M.; Dai, H.-L.; Mostafa, R.R.; Parmar, K.S.; Heddami, S.; Kisi, O. Modeling Multistep Ahead Dissolved Oxygen Concentration Using Improved Support Vector Machines by a Hybrid Metaheuristic Algorithm. *Sustainability* **2022**, *14*, 3470. [[CrossRef](#)]
15. Adnan, R.M.; Mostafa, R.R.; Dai, H.-L.; Heddami, S.; Kuriqi, A.; Kisi, O. Pan Evaporation Estimation by Relevance Vector Machine Tuned with New Metaheuristic Algorithms Using Limited Climatic Data. *Eng. Appl. Comput. Fluid Mech.* **2023**, *17*, 2192258. [[CrossRef](#)]
16. Wang, K.; Xue, Q.; Lu, J.J. Risky Driver Recognition with Class Imbalance Data and Automated Machine Learning Framework. *Int. J. Environ. Res. Public Health* **2021**, *18*, 7534. [[CrossRef](#)]
17. Yang, L.; Ma, R.; Zhang, H.M.; Guan, W.; Jiang, S. Driving Behavior Recognition Using EEG Data from a Simulated Car-Following Experiment. *Accid. Anal. Prev.* **2018**, *116*, 30–40. [[CrossRef](#)] [[PubMed](#)]
18. Zhu, H.; Xiao, R.; Zhang, J.; Liu, J.; Li, C.; Yang, L. A Driving Behavior Risk Classification Framework via the Unbalanced Time Series Samples. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2503312. [[CrossRef](#)]
19. Katakazas, C.; Quddus, M.; Chen, W.-H. A Simulation Study of Predicting Real-Time Conflict-Prone Traffic Conditions. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 3196–3207. [[CrossRef](#)]
20. Yarlagadda, J.; Jain, P.; Pawar, D.S. Assessing Safety Critical Driving Patterns of Heavy Passenger Vehicle Drivers Using Instrumented Vehicle Data—An Unsupervised Approach. *Accid. Anal. Prev.* **2021**, *163*, 106464. [[CrossRef](#)] [[PubMed](#)]
21. Bellman, R. Dynamic Programming. *Science* **1966**, *153*, 34–37. [[CrossRef](#)] [[PubMed](#)]
22. El Haouij, N.; Poggi, J.-M.; Ghozi, R.; Sevestre-Ghalila, S.; Jaïdane, M. Random Forest-Based Approach for Physiological Functional Variable Selection for Driver’s Stress Level Classification. *Stat. Methods Appl.* **2019**, *28*, 157–185. [[CrossRef](#)]
23. Jiang, X.; Abdel-Aty, M.; Hu, J.; Lee, J. Investigating Macro-Level Hotzone Identification and Variable Importance Using Big Data: A Random Forest Models Approach. *Neurocomputing* **2016**, *181*, 53–63. [[CrossRef](#)]
24. Li, Z.; Wang, H.; Zhang, Y.; Zhao, X. Random Forest–Based Feature Selection and Detection Method for Drunk Driving Recognition. *Int. J. Distrib. Sens. Netw.* **2020**, *16*, 155014772090523. [[CrossRef](#)]
25. Natekin, A.; Knoll, A. Gradient Boosting Machines, a Tutorial. *Front. Neurobot.* **2013**, *7*, 21. [[CrossRef](#)]
26. Peterson, L. K-Nearest Neighbor. *Scholarpedia* **2009**, *4*, 1883. [[CrossRef](#)]
27. Morris, C.; Yang, J.J. Effectiveness of Resampling Methods in Coping with Imbalanced Crash Data: Crash Type Analysis and Predictive Modeling. *Accid. Anal. Prev.* **2021**, *159*, 106240. [[CrossRef](#)] [[PubMed](#)]
28. Valverde-Albacete, F.J.; Peláez-Moreno, C. 100% Classification Accuracy Considered Harmful: The Normalized Information Transfer Factor Explains the Accuracy Paradox. *PLoS ONE* **2014**, *9*, e84217. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.