Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2026) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2026 Toulouse 6-10 July 2026

Identifying Dangerous Driving Behaviour through Big Data and Machine Learning Techniques

Hector Kamvoussioras, Thodoris Garefalakis, Eva Michelaraki*, George Yannis

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

Abstract

Predicting risky driving behaviour in real-time is crucial for road safety, allowing for early intervention to prevent crashes. Using the Safety Tolerance Zone (STZ) to classify driving performance into three levels (i.e. normal, dangerous and avoidable accident), The aim of this study was to identify dangerous behaviour and explore the key factors influencing each level. To achieve this objective, data from a driving simulator experiment were exploited and four classification models: Ridge Classifier (RC), Support Vector Machines (SVM), Random Forests (RF) and eXtreme Gradient Boosting (XGBoost) were developed. To better understand the contribution of individual features, SHAP analysis was conducted. Through the systematic feature selection process, the most relevant variables were identified and organized into two distinct Groups, A and B. Results revealed that RF and XGBoost models consistently achieved the best performance, outperforming the other techniques across all three safety levels, reaching 95% in prediction accuracy. It was also demonstrated that time to collision was the most influential factor in Group A, whereas speed emerged as the dominant predictor in Group B, with higher values strongly associated with risky behaviour. This research highlights the risk indicators, offering valuable insights for shaping safety interventions and ultimately enhancing road safety.

Keywords: Driving behaviour; Road safety; Support Vector Machines; Random Forests; Machine Learning Techniques.

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

1. Introduction

Road safety is an important area that affects the state, society, economy and many other areas. Europe has made significant efforts to reduce the number of traffic crashes. The significance of modern automobile technology and transportation automation to improving road safety has received special attention in recent decades.

According to Whittingham (2004), human mistakes are responsible for 94% of major crashes. Therefore, it is crucial to investigate new technology in order to improve road safety by reducing human error. The goal of using Advanced Driver-Assistance Systems (ADAS) is to either prevent or lessen the severity of unavoidable collisions (Masello et al, 2023). These systems use a range of sensors, including cameras, radar and ultrasonic technologies, to collect vital information about the vehicle's environment in order to proactively prevent collisions or lessen their effects. Even though ADAS is a huge advancement in improving road safety, there are still issues with using these technologies to accurately forecast and reduce crash risks in real time. The customization of interventions is a major obstacle. Even with technological developments, current interventions frequently lack customisation to fit the unique driving behaviour profiles of each individual.

In order to build precise models that can recognize unsafe driving patterns and establish the foundation for in-vehicle interventions, prior published research focuses on understanding how different factors influence hazardous driving. There is a lack of personalization of interventions and a clear connection between targeted interventions and real-time driving behaviour, despite a number of in-vehicle and post-driving interventions being recommended (Osman et al., 2019; Bakhshi & Ahmed, 2022). To foresee possible risks and create tailored remedies in real time, machine learning prediction models may evaluate enormous volumes of data, such as driver behaviour, environmental variables and road conditions (Xie & Zhu, 2019; Mozaffari et al., 2020).

Because it offers a safe and regulated setting for the thorough investigation of driver behaviour, cognitive functions and decision-making under on-road driving circumstances, simulator driving studies have emerged as a key approach in transportation research (Ali et al., 2020; Voinea et al., 2023). Understanding the intricate mechanics of human interaction with cars and the road has been made much easier thanks to these synthetic settings, which mimic real driving situations. They get over the difficulties of researching crash precursors and near-crash conditions, which are frequently uncommon and challenging to see in practical situations. Researchers can examine a variety of issues, including driver distraction, reactions to different traffic conditions and the effect of developing technology on road safety, by immersing participants in virtual driving experiences. Simulator driving experiments, as opposed to real-world testing or naturalistic driving research, provide for exact control over experimental design and the gathering of high-quality driver performance data. The examination of critical safety concerns and the development of solutions to enhance road safety and driver education have completely transformed the field of transportation research through this methodology.

The overarching objective is to create a comprehensive framework for defining, developing, testing and validating a Safety Tolerance Zone (STZ), designed to reduce risky driving behaviour in real-time conditions. This framework functions within a dynamic system that continuously monitors driving performance, considering both driver background factors and risk indicators related to driving task and environmental conditions. At the core of this concept is the STZ, which categorizes driving behaviour into three distinct levels: normal, dangerous and avoidable accident.

More specifically, the normal driving level describes the portion of the STZ when there is no indication that a collision scenario is likely to occur at that moment, given the status of the world. The STZ is conceptually to be understood as a time-space window where the human operator's self-regulated vehicle control can be qualified as “normal driving” as long as a moment-to-moment registration of the current state-of-the-world does not detect the potential for a crash course to start developing. The STZ phase where the possibility for the onset of a collision scenario is observed is referred to as the second level. The STZ conceptual status shifts from “normal driving" to “dangerous phase” in the event that the objective state-of-the-word changes. More specifically, the latter indicates that the possibility of a crash path forming has been initiated, making the human operator's self-regulated vehicle control less safe. Reduced driver competency, increased work demand from outside variables or a mix of external and driver-related factors could have caused this. Lastly, the third level or “avoidable accident phase”, describes the specific STZ subzone when a collision scenario is actually beginning to take shape but the driver still has the opportunity to stop it. The STZ's conceptual status shifts from “dangerous” to “avoidable accident” phase in the event that the objective state of the world changes in this way. More precisely, this indicates that the possibility of a collision has been triggered,

making the human operator's self-regulated vehicle control even less safe. This could be caused by outside factors affecting the road traffic system, a decline in the operators' skills or both.

Real-time interventions delivered within the vehicle are designed to assist drivers during critical moments by issuing timely warnings as behaviour approaches the “dangerous” or “avoidable accident” phases. The level of intervention is calibrated to the severity of risk, with more noticeable warnings provided as the likelihood of a crash increases. Beyond these immediate responses, the system also incorporates post-trip interventions. Using data collected during the journey, it generates personalized feedback on driving style, safety-critical events and contributing risk factors. This information, accessible through mobile and web-based platforms, promotes long-term improvements in driving behaviour through interactive and gamified learning experiences.

Overall, this approach introduces an innovative method to enhance road safety by integrating advanced technologies with proactive intervention strategies. It seeks to transform driving safety by establishing a dynamic STZ and tailoring interventions to the driving context. The ultimate goal is to equip drivers with the awareness and tools needed to drive responsibly, reduce crash risks and foster a culture of safer road use.

Taking all the aforementioned arguments into consideration, the goal of this study is to categorize dangerous driving behaviour and identify the key factors influencing each level. Using the STZ to classify driving behaviour into three levels (i.e. normal, dangerous and avoidable accident), this study fills in the gaps in current research on real-time interventions and driving behaviour prediction. To achieve this objective, data from a driving simulator experiment were exploited and four classification models: Ridge Classifier (RC), Support Vector Machines (SVM), Random Forests (RF) and eXtreme Gradient Boosting (XGBoost) were developed.

The structure of the paper is as follows: it begins with an introduction, followed by a literature review focusing on driving behaviour analysis through machine learning techniques. Next, the study process is outlined, including the theoretical background of the models. The subsequent section describes the data collection and processing. Finally, the analysis results are presented, along with conclusions that highlight implications for road safety.

2. Background

Several studies and research have been carried out in order to identify the key factors influencing dangerous driving behaviour. More specifically, two primary approaches are commonly employed for such analyses: (i) simulator-based experiments and (ii) naturalistic driving studies. More recent efforts have placed emphasis on developing classification models capable of detecting risky driving patterns, thereby enhancing the effectiveness of driver assistance systems. Many studies also focus on interpreting these models and examining the significance of the variables involved, in order to better understand the conditions that lead drivers to engage in unsafe behaviour.

The classification of driving behaviour is crucial for understanding and reducing traffic crashes and their resulting injuries, ranging from minor to severe. Accurate recognition of different driving styles is considered a foundation for advancing traffic safety initiatives. For example, Xie & Zhu (2019) employed a Random Forest (RF) model within a manoeuvre-based framework, achieving an average accuracy of 70.47% in classifying driving behaviour. Likewise, Amsalu et al. (2015) used naturalistic driving data with Support Vector Machines (SVM), reaching an accuracy above 97%. Furthermore, Shi et al. (2019) proposed a robust framework for risk prediction that integrates advanced feature selection methods, risk-level labelling, strategies for handling imbalanced datasets and the application of the XGBoost classifier, which achieved an overall accuracy of 89%. Collectively, these contributions highlight the value of sophisticated classification methodologies in capturing driving behaviour and their implications for road safety.

The analysis of driving behaviour is a critical factor in ensuring road safety. Shangguan et al. (2021) proposed a methodology for assessing and predicting drivers’ real-time risk levels. Using clustering algorithms, they identified four distinct stages of risk. Alongside risk prediction, classification algorithms were also developed. Their analysis revealed that variables such as speed difference, distance from the preceding vehicle, overall speed and acceleration play a particularly significant role in predicting a driver’s risk state. Garefalakis et al. (2024) analysed different classification techniques and examines their ability to detect dangerous driving behaviour based on a large-scale field-trial and simulator experiment. Towards that end, four classification algorithms, namely SVM, RF, AdaBoost and Multi-Layer Perceptron were implemented. In the simulator experiment, RFs and MLPs emerged as the top-performing models with an accuracy of 84% and 82%, respectively, while in the naturalistic driving experiment, RF and AdaBoost maintained robust performance, with high accuracy of 75% and 77%, respectively.

However, since driving behaviour analysis is often based on real-world datasets, prior research has faced challenges with class imbalance (e.g. safe vs. dangerous conditions). Relevant studies highlight that risky behaviour and crash occurrence are comparatively less frequent than safe driving and crash-free conditions. To address this imbalance, Padurariu & Breaban (2019) proposed several strategies, which can be broadly divided into two categories: preprocessing techniques (such as resampling) applied before classification and algorithm-level modifications designed to give greater weight to the minority class (e.g. Synthetic Minority Oversampling Technique - SMOTE). Building on this, Zhu et al. (2022) investigated a range of SMOTE-based methods combined with boosting approaches, including Random Oversampling and SMOTE-Adaboost.

Highly trained models often achieve strong accuracy and reliable results, yet they can be difficult to interpret. In other words, it is crucial to determine the extent to which each variable influences and contributes to the model's outcomes. Similarly, Parsa et al. (2020) sought to detect crash occurrences, using a real-time dataset of road traffic crashes, including weather, traffic, network and land-use metrics. Their findings, derived from SHapley Additive Explanations (SHAP) values, revealed that traffic-related variables were the most critical features for accurate crash detection.

Despite the widespread use of machine learning techniques in driving behaviour analysis, one persistent challenge lies in the limited interpretability of models and their results. Predictive models for driver behaviour hold substantial promise, particularly in the automotive sector, yet their opaque, "black box" nature often restricts meaningful insights into the underlying causes of behaviour.

To address this limitation, the present study not only develops machine learning models for predicting driver behaviour but also prioritizes their interpretability. With the application of SHAP values, the study highlights the key factors influencing driving behaviour and provides actionable insights for stakeholders. This approach moves beyond the black box paradigm, offering a transparent framework that delivers both accurate predictions and clear explanations of the reasoning behind them. Ultimately, the goal is to generate comprehensive knowledge of driving behaviour, thereby supporting the design of targeted interventions to improve road safety and promote driver well-being.

3. Data Overview

3.1. Driving Simulator Experiment

The main objective of this research was to examine driving behaviour with a central focus on identifying measures to reduce crash occurrences. To achieve this, a driving simulator experiment was conducted between December 2020 and January 2021 involving 36 drivers aged 29-60, with an average age of 42 years. Although the sample size is relatively modest, it is consistent with the requirements and common practice of driving simulator studies, where data collection is resource-intensive and conducted under highly controlled experimental conditions. Furthermore, each participant completed three separate driving sessions, generating a substantial number of observations that enabled the development and validation of the machine learning models. The primary objective of the study was to identify behavioural patterns and evaluate the predictive capability of the proposed models rather than to estimate population-level parameters. Consequently, the sample was designed to capture variability in driving behaviour under controlled conditions rather than to achieve statistical representativeness of the entire driving population. Nevertheless, the potential limitations regarding generalizability are acknowledged and discussed in the limitations section, while future research should seek to validate the findings using larger and more diverse driver populations.

Eligibility required more than five years of driving experience, along with good physical and mental health. Prior to participation, drivers were asked to confirm that they did not suffer from any medical condition that could impair their ability to safely operate a vehicle or complete the simulator experiment. No formal clinical assessment of mental health was conducted. The study focused on recruiting active drivers capable of safely participating in the experimental procedures.

The experiment utilised a driving simulator and implemented three distinct driving scenarios, described in detail below. In the beginning, participants completed a free driving session, designed to replicate the real conditions of a typical trip and help them become accustomed to the simulator. This familiarization drive, a common practice in simulator-based studies, was guided by the coordinator, who ensured participants were seated comfortably, addressed any questions and confirmed their readiness to proceed.

Prior to the experimental drives, participants completed a familiarization session designed to help them adapt to the simulator environment and controls. The duration of this session was not fixed; instead, participants were allowed to continue until they demonstrated adequate familiarity with the simulator and reported feeling comfortable operating the vehicle. This procedure was adopted to minimize learning effects and ensure that the recorded driving behaviour reflected the experimental conditions rather than adaptation to the simulator.

The design of the simulator experiment followed well-established principles in the literature (Papantoniou et al., 2015; Amini et al., 2021). These principles included defining outcomes, predictors and hypotheses; determining sample size and statistical power; selecting an appropriate experimental design; distributing risk scenarios among participants; setting drive durations to minimize simulator sickness; mitigating order and learning effects and accounting for potential confounding factors.

The driving simulator used in this study was specifically designed and constructed to replicate the cockpit environment of a Peugeot 206. Authentic components such as the full dashboard, functional instrument panel and driving seat were incorporated to ensure realistic driving conditions. The simulator operated on the STISIM Drive 3 software and was equipped with three 49-inch 4K-resolution screens, offering a 135° field of view. This setup provided an immersive and highly realistic environment for examining driving behaviour under controlled conditions. Figure 1 presents the driving simulator used in the experiment.

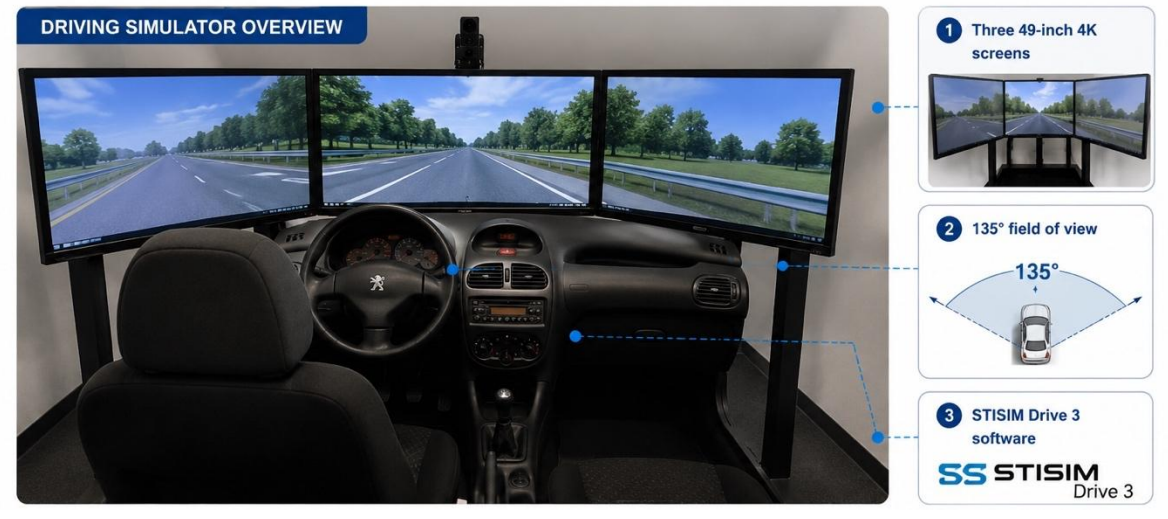


Fig. 1. Driving simulator used in the experiment

The simulator's architecture integrates multiple technologies for real-time data acquisition. A MobilEye camera, a specialized cardiogram steering wheel and dedicated simulator software worked in tandem to capture detailed driving data. Additional tools, including eye-tracking systems and video recording, further enriched the dataset by offering deeper insights into driver behaviour. A dedicated gateway enabled real-time interventions similar to those found in real vehicles. Importantly, this gateway did not store data in the cloud. Instead, it transmitted all collected and processed information via a serial interface back to the simulator computer, where it was synchronized with simulation variables and stored locally. The gateway itself did not directly alter simulation variables; the data flow remained strictly unidirectional, from the gateway to the simulator. This architecture ensured that interventions were delivered efficiently while maintaining data integrity within local storage, avoiding reliance on external cloud systems. Figure 2 illustrates the Mobileye system and its integration within the simulator environment.

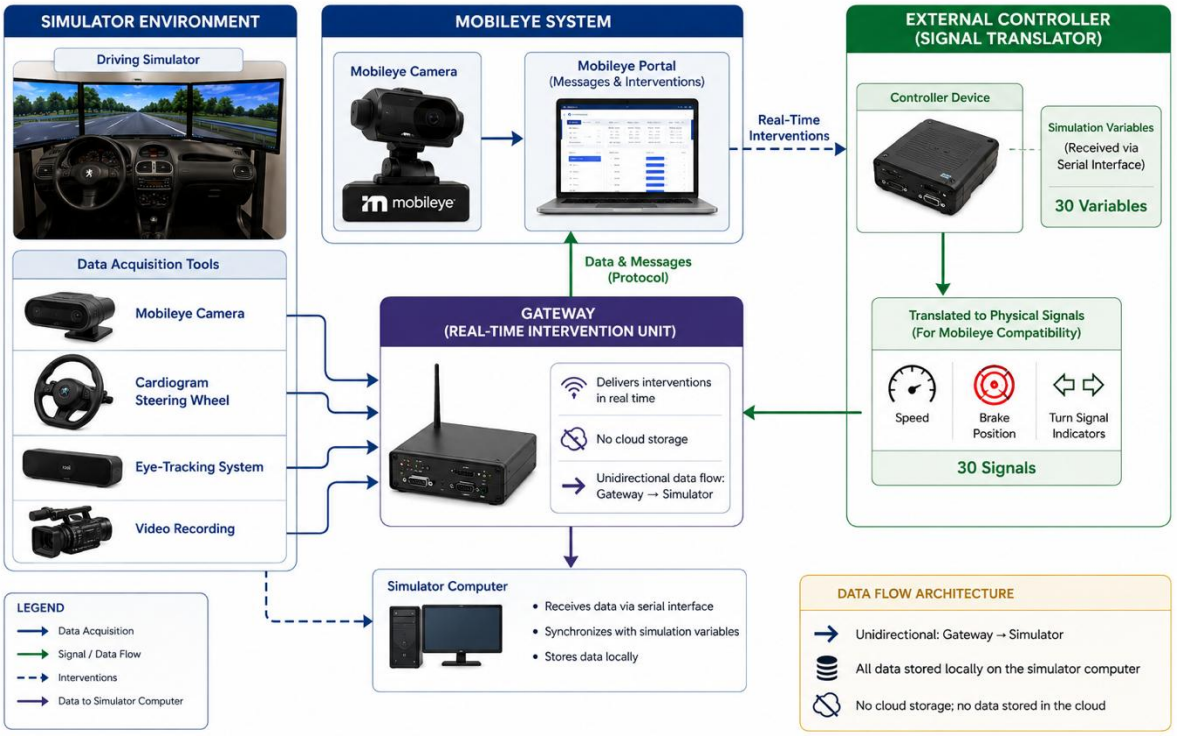


Fig. 2. Mobileye system and its integration within the simulator environment

Mobileye operates using data collected from real vehicles, including variables such as speed, brake position and usage indices, which form the basis for its interventions. These values are also accessible through the system's portal via specialized messages. To ensure compatibility with Mobileye, an external controller translates simulation variables into the required format. This controller receives simulation data via a serial interface and converts 30 variables into physical signals corresponding to vehicle speed, brake status and turn signal indicators.

The speed signal replicates the Vehicle Speed Sensor (VSS) of an actual vehicle. Typically generated by a Hall sensor or a similar device, this signal translates rotational motion into pulsed outputs. Depending on the setup, the sensor may be mounted on the gearbox's input shaft or integrated within the Anti-lock Braking System (ABS) to measure wheel rotation. By contrast, the brake and directional indicators are represented as simple digital on/off signals. For analytical purposes, all simulator data is stored locally. To ensure consistency, data from the external gateway is synchronized with the simulator's internal data through precise adjustments to the simulation loop. This allows the portal data to align seamlessly with simulation outputs at each time step. The synchronized dataset is then systematically logged, following a defined protocol and stored in JSON format to facilitate efficient analysis and retrieval.

As part of the simulator experiment, three distinct scenarios involving driving on a two-way road were implemented. Each scenario was structured into three sections, each featuring different road layouts (e.g. rural, highway), as outlined in Table 1.

Table 1. Different scenarios applied during the driving simulator experiment

Scenario	Road Section	Number of lanes	Speed Limits
A	0-6300 m	1x1	70 km/h
	6300-11300 m	2x2	90 km/h
	11300-16500 m	2x2	120 km/h
B	0-6100 m	2x2	90 km/h
	6100-12000 m	2x2	120 km/h
	12000-18200 m	1x1	70 km/h
C	0-6000 m	2x2	90 km/h
	6000-11000 m	2x2	120 km/h
	11000-17200 m	1x1	70 km/h

Each participant completed three separate driving sessions:

- Drive 1 - No interventions: A baseline monitoring drive without any in-vehicle interventions
- Drive 2 - Fixed-Threshold Interventions: A session with interventions triggered by predetermined thresholds, providing warnings during safety-critical events
- Drive 3 - Adaptive Interventions: A session where intervention triggers were adjusted based on the driver's ability to perform specific tasks, with modified conditions influencing the intervention thresholds.

The data gathered from the driving simulator experiment are presented in Table 2, organized according to the respective data collection methods. It is important to note that the "FatigueEvent" variable was measured using the Karolinska Sleepiness Scale (KSS), which is a 9-point subjective index of drowsiness. A score of 1 indicates an extremely alert state, whereas a score of 9 corresponds to a very sleepy condition in which the driver struggles to remain awake (Åkerstedt & Gillberg, 1990).

Table 2. Variables used in the driving behaviour classification model

Variable	**Description**	**Units**	**Type**
	Calculated in OM		
TTC	Time to collision with the vehicle ahead	Seconds	Numeric
Headway	Time headway to the vehicle ahead in the same lane	Seconds	Numeric
Speed	Vehicle speed	Kilometers per hour	Numeric
Distance_travelled	Distance driving	Meters	Numeric
	Data From Mobileye		
ME_ForwardCollisionWarning	Whether a forward collision warning is active	bool	Logical (False or True)
ME_LaneDepartureWarningActive	Whether Lane Departure Warning is active	bool	Logical (False or True)
	Data From Cardiogram Steering Wheel		
HandsOnEvent	Whether hands are on the steering wheel	None / both	Discrete
FatigueEvent	KSS score	1 - 9	Discrete

3.2. Data Processing

For each participant, three csv files were generated corresponding to the three driving sessions (no interventions, fixed-threshold interventions and adaptive interventions). To facilitate the data analysis, the information was segmented into 30-second intervals. Within each interval, descriptive statistics were calculated for every variable, including measures such as mean, standard deviation, minimum, maximum and median values.

The choice of 30-second aggregated values reflects the specific aims of this research. Although shorter events or micro-distractions may not be fully captured, this interval length provides a suitable balance between identifying meaningful patterns, minimizing random noise and maintaining a dataset that is both interpretable and manageable. While acknowledging the limitations inherent in aggregated statistics, this method was deemed an appropriate compromise for the study. Sensitivity analyses were also performed to reinforce the robustness of the results and to address potential concerns associated with this approach.

3.3. Definition of Driving Behaviour Safety Level

Before developing the classification algorithms, it was necessary to define three distinct safety levels of driving behaviour. Drawing on the literature, two approaches were considered: (i) clustering methods (e.g. K-means) and (ii) threshold-based methods using indicators such as speed, time to headway and time to collision. To remain consistent with previous studies, where the 'avoidable accident' category is treated as the minority class and the 'normal' category as the majority class, the final distribution across the three safety levels was determined using thresholds of the variable Speed_max.

Accordingly, the value ranges of the speed limit variable for each safety level are defined as follows:

- Normal (class: 0): Speed_max $\leqslant$ 0.8*Speed Limit
- Dangerous (class: 1): 0.8*Speed Limit < Speed_max $\leqslant$ 1.2*Speed Limit
- Avoidable Accident (class 2): Speed_max > 1.2*Speed Limit

It is important to clarify that Speed_max refers to the highest speed reached by each driver within a 30-second interval. As previously noted, the dataset was aggregated into 30-second segments, during which descriptive statistics, such as the maximum value, were computed.

3.4. Feature Selection

A crucial step in developing classification models is the feature selection process. Its primary objective is to reduce the number of input variables, thereby lowering computational cost while improving predictive performance (Chandrashekar & Sahin, 2014). Features should be selected based on their contribution to the classification process, making this step fundamental to optimizing model efficiency and accuracy.

To assess feature importance, a Ridge Classifier was applied. This linear model assigns coefficients to each feature during training, reflecting their influence on the decision-making process. Positive coefficients indicate a positive association with the target outcome, while negative coefficients represent a negative association. Figure 3 illustrates these relationships, providing a visual representation of both the magnitude and direction of each feature's effect. Features with higher absolute coefficient values are considered more influential, offering valuable insights for feature selection and supporting the interpretability of the classification model.

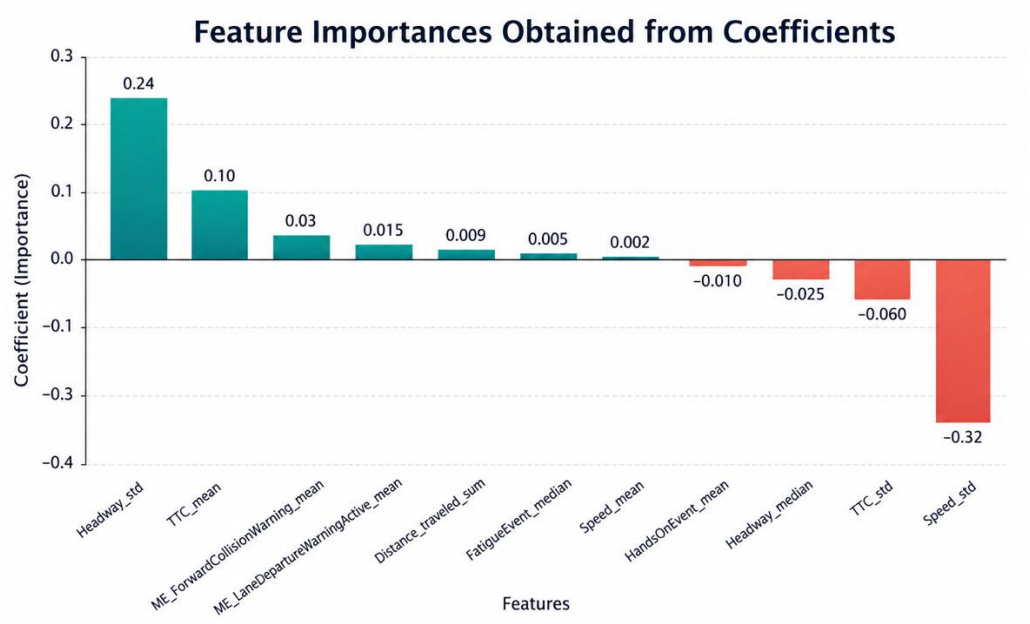


Fig. 3. Feature importance obtained from coefficients

Based on the feature selection process, the variables were divided into two groups. Group A included the most influential variables for identifying risky driving behaviour, while Group B consisted of the remaining variables, used primarily for comparative analysis and testing. To ensure statistically robust results, certain variables were retained across both groups. The segmentation of variables is summarized in Table 3.

Table 3: Independent Variables Based on Feature Selection

Group	Variables	Description
A	TTC_mean	Mean value of "TTC"
	Headway_std	Standard Deviation of "Headway"
	Speed_std	Standard Deviation of "Speed"
	ME_ForwardCollisionWarning_mean	Mean value of "ME_ForwardCollisionWarning"
B	TTC_mean	Mean value of "TTC"
	Headway_median	Median value of "Headway"
	HandsOnEvent_mean	Mean value of "HandsOnEvent_mean"
	FatigueEvent_median	Median value of "FatigueEvent"
	ME_LaneDepartureWarningActive_mean	Mean value of "ME_LaneDepartureWarningActive_mean"
	Speed_mean	Mean value of "Speed"
	Distance_travelled_sum	Sum of "Distance_travelled"

3.5. Imbalanced Learning

The imbalance inherent in many datasets poses significant challenges for predictive modelling. Most machine learning algorithms assume an approximately equal distribution of examples across classes (Garefalakis et al., 2022). When this assumption is violated, predictive performance often declines, particularly for the minority class. This issue is critical because the minority class usually carries greater importance in real-world applications, making classification errors within this group far more consequential than those in the majority class (Yen & Lee, 2006).

In the context of this study, the minority classes correspond to unsafe driving behaviour (e.g. dangerous and avoidable accident), while the majority class represents safe driving (normal). This imbalance highlights the need to apply resampling techniques during model training. Misclassifying dangerous behaviour as safe could have serious implications for road safety, underscoring the importance of addressing class imbalance. Employing robust resampling methods is therefore essential to ensure that the model is not only accurate but also reliable and practically applicable for enhancing driver safety (Abou Elassad et al., 2020).

One widely used approach is the Synthetic Minority Oversampling Technique (SMOTE), a data augmentation method designed to balance class distributions. SMOTE generates synthetic samples for the minority class by interpolating between existing observations. In particular, for each data point in the minority class, the k-nearest neighbours are identified. A neighbour is then randomly selected and the difference between the two points is multiplied by a random number between 0 and 1. This value is added to the original data point to create a new synthetic observation (Chawla et al., 2002). SMOTE addresses sample imbalance, enhancing the model's capacity to recognize minority class patterns while simultaneously mitigating the risks associated with heterogeneous distributions.

4. Methodology

This study addresses a multi-class classification problem, aiming to determine the risk level of driving behaviour across three categories: normal, dangerous and avoidable accident. To tackle this issue, four machine learning classification methods were applied: Ridge Classifier (RC), Support Vector Machines (SVM), Random Forests (RF) and eXtreme Gradient Boosting (XGBoost). These algorithms were selected for their ability to capture complex patterns and relationships within the dataset. Employing a diverse set of classifiers enables a more comprehensive understanding of driving risk assessment and allows for a robust comparison of performance across different modelling approaches.

The development of the machine learning models followed a structured process. First, the dataset was divided into two subsets: a training set, used to build and calibrate the algorithms and a testing set, used to evaluate their predictive accuracy. This segmentation ensured that model performance was assessed on unseen data, providing a rigorous and unbiased evaluation framework. To further enhance reliability, 10-fold cross-validation and GridSearch hyperparameter tuning were implemented. Cross-validation partitions the dataset into k folds (e.g. 10), iteratively training the model on k-1 folds and validating it on the remaining fold. This approach provides robust performance estimates while minimizing the risk of overfitting. GridSearch complements this by systematically exploring different hyperparameter combinations, selecting those that deliver the best results based on cross-validation outcomes.

4.1. Ridge Classifier (RC)

The Ridge Classifier (RC) is a classification variant of the Ridge Regressor. In this approach, binary targets are transformed into –1 and 1, after which the problem is treated as a regression problem and optimized using the same objective function as Ridge regression. The predicted class is then determined by the sign of the regressor's output. For multi-class classification, the task is framed as a multi-output regression problem, with the final class prediction corresponding to the output that yields the largest value.

Although it may initially appear unconventional to use a penalized least-squares loss for classification, rather than more traditional losses such as logistic or hinge loss, studies have shown that these methods often achieve comparable cross-validation performance in terms of accuracy, precision and recall. The advantage of the penalized least-squares loss used by RC lies in its compatibility with a wider range of numerical solvers, offering distinct computational efficiency benefits.

To optimize the model, GridSearch hyperparameter tuning was employed. The optimal parameters for the Ridge Classifier were identified as: (a) alpha = 0.1, (b) solver = "auto", (c) fit_intercept = true and (d) tolerance = 0.001.

4.2. Support Vector Machines (SVM)

Support Vector Machines (SVM) are supervised machine learning models widely used for both classification and regression tasks. The core idea of SVM is to construct a hyperplane in a multidimensional space that best separates the training data into predefined classes. This separation is achieved by maximizing the margin between classes while minimizing the distance of misclassified points from the decision boundary (Vapnik, 1998). With the application of kernel functions, SVMs are also capable of handling datasets that are not linearly separable.

The hyperparameters of the SVM model were optimized using the GridSearchCV tuning approach. The best-performing configuration was determined as follows: (a) kernel type = "rbf"; (b) regularization parameter C = 15; (c) kernel coefficient gamma = "scale"; and (d) decision function shape = "ovo".

4.3. Random Forests (RF)

Random Forests (RF) are supervised learning algorithms that operate by constructing an ensemble of decision trees. For each split, it randomly selects a subset of input features, which are then used to make predictions. The algorithm classifies observations into predefined classes by building multiple decision trees and aggregating their results to produce a final outcome. Unlike regression models, a single decision tree does not yield one “best” prediction for each observation; instead, it generates multiple predictions, which are then evaluated based on how well they align with the observed data (Misra et al., 2020). This ensemble approach provides a wide range of possible predictions depending on the values of input variables. As a result, classification accuracy generally improves with the inclusion of more trees, although this comes at the cost of reduced interpretability.

To optimize model performance, Grid Search was applied to the RF classifier. The best-performing set of hyperparameters was identified as: (a) number of estimators/trees = 30; (b) maximum depth = 30; and (c) split quality criterion = "gini".

4.4. Extreme Gradient Boosting (XGBoost)

EXtreme Gradient Boosting (XGBoost) is a powerful gradient-boosting algorithm that has achieved considerable success in large-scale regression and classification problems. It builds predictions by iteratively combining weak learners, typically decision trees, through reweighting, thereby producing a strong and robust model.

XGBoost is an optimized, distributed gradient-boosting library designed for efficiency, flexibility and portability. Operating within the Gradient Boosting framework, it offers parallel tree boosting (also referred to as GBDT or GBM), which enables fast and accurate solutions to a wide range of data science problems. Furthermore, the algorithm is highly scalable, capable of running in distributed environments such as Hadoop, SGE, MPI and is suitable for applications involving datasets with billions of observations.

The optimal hyperparameters obtained through Grid Search optimization for the XGBoost model were: (a) number of estimators = 25; (b) objective = "reg:linear"; (c) learning rate = 0.09; and (d) maximum depth = 30.

4.5. Evaluation Metrics

The evaluation metrics applied in the classification process serve as essential benchmarks for assessing the performance and effectiveness of machine learning models. They provide valuable insights into a model’s ability to generalize to unseen data and accurately classify instances. In this study, several key metrics were employed, as outlined in Equations (1-5). It should be noted that True Positives (TP) represent the number of instances in the positive class (event present) that are correctly classified as positive. True Negatives (TN) denote the number of instances in the negative class (no event) correctly classified as negative. False Positives (FP) are negative instances (no event) incorrectly classified as positive (event present), while False Negatives (FN) are positive instances (event present) incorrectly classified as negative (no event).

In the context of multiclass classification with three risk levels of driving behaviour, several evaluation metrics were employed as key benchmarks of model performance. Precision measures the proportion of correctly predicted instances among all cases assigned to a specific risk level, reflecting the accuracy of positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

Sensitivity (Recall) calculates the proportion of correctly identified instances among all actual occurrences of a particular risk level, indicating the model's ability to detect relevant cases.

$$\text{Sensitivity (Recall)} = \frac{\text{TP}}{\text{TP+FN}} \tag{2}$$

F1-score, the harmonic mean of precision and recall, provides a balanced performance measure.

$$\text{F-measure} = \frac{2 \text{ x (Precision) x (Recall)}}{\text{(Precision)+(Recall)}} \tag{3}$$

False Alarm Rate quantifies the proportion of incorrect classifications into a particular risk level relative to all instances not belonging to that level.

$$\text{False Alarm Rate} = \frac{\text{FP}}{\text{FP+TN}} \tag{4}$$

Finally, G-means evaluates overall model performance by balancing sensitivity and specificity, ensuring reliable discrimination across all risk levels.

$$\text{G-mean} = \sqrt{\text{(Sensitivity)x(Specificity)}} \tag{5}$$

To further interpret the models, SHapley Additive Explanations (SHAP values) were calculated. This method applies game-theoretic principles to explain machine learning predictions by linking global feature importance with local interpretability. SHAP values represent the average expected marginal contribution of each variable across all possible feature combinations, thereby quantifying both their direction and magnitude of influence.

5. Results

The dataset was divided into training and testing subsets, with 70% allocated for training and 30% reserved for testing. The training data were used to instruct the machine learning models by establishing classification patterns for driving behaviour, while the testing data served to evaluate model performance. Driving instances were categorized into three classes, allowing comparisons between predicted and actual risk levels.

A key challenge, also highlighted in previous studies, is the imbalance among classes, as dangerous driving scenarios are typically underrepresented compared to safe driving conditions. Since most classification algorithms assume a relatively even distribution across classes, this imbalance often results in higher error rates, particularly for minority classes. In this context, errors are most likely to occur when predicting cases labeled as avoidable accident, given the dominance of the normal and dangerous classes. Accurate prediction across all risk levels is crucial, particularly in safety-related applications, which underscores the importance of re-sampling techniques to address imbalance and ensure fairer model outcomes.

Prior to applying the SMOTE technique, the class distribution was highly skewed, with 60% of the data labeled as normal, 36% as dangerous and only 4% as avoidable accident. After applying SMOTE to the training dataset, the distribution was balanced at approximately 33% for each class. The effectiveness of this approach was validated through sensitivity analyses and cross-validation, demonstrating that SMOTE substantially improved dataset balance and, in turn, enhanced the performance and reliability of the classification models.

5.1. Evaluation of Classification Models

Specific classification algorithms were developed to determine the driving behaviour level within each driver's 30-second interval. Four models were tested and their evaluation is summarized below.

The RC achieved moderate accuracy in predicting the normal level but performed poorly in identifying the avoidable accident and dangerous levels, indicating limited capability in recognizing risky behaviour. The SVM demonstrated reasonable accuracy across all three levels, with an overall accuracy of 73.5%. However, it also struggled with accurate predictions of the avoidable accident level. In contrast, the RF model showed strong performance across all categories, achieving 90% accuracy with minimal error rates. Its superiority was further confirmed in Group B, where accuracy increased to 95%, outperforming results from Group A.

Similarly, the XGBoost model delivered results comparable to RF, with 90% accuracy in Group A and improved performance in Group B. The comparison between the two groups highlights that incorporating additional variables enhances predictive accuracy, as Group B consistently outperformed Group A. Overall, RF and XGBoost emerged as the most effective models for reliable risk prediction, underscoring the importance of comprehensive feature inclusion.

To further identify the main contributors to model performance, both average and weighted average values were calculated. This analysis provided insights into the relative impact of each risk level on the results. The outcomes of these evaluations for both Group A and Group B are presented in Table 4.

Table 4: Performance metrics for Group A and B

	Precision	**Recall**	**F-means**	**False Alarm Rate**	**G-means**
Group A					
Ridge Classifier	43%	35%	27%	95%	19%
Support Vector Machine	59%	71%	62%	29%	77%
Random Forest	80%	88%	83%	12%	90%
XGBoost	80%	88%	83%	12%	90%
Group B					
Ridge Classifier	50%	33%	23%	67%	5%
Support Vector Machine	92%	83%	86%	17%	83%
Random Forest	92%	90%	91%	7%	95%
XGBoost	91%	93%	92%	7%	96%

Apart from the RC technique, all algorithms demonstrated strong performance, achieving high values across recall, precision, F-measure and G-means. Among them, the RF and XGBoost classifiers achieved the highest predictive capability. This superiority is further illustrated in the ROC curve (Figure 3), where both RF and XGBoost show prominent results by maintaining a high true positive rate relative to the false positive rate, underscoring their effectiveness in classification problems.

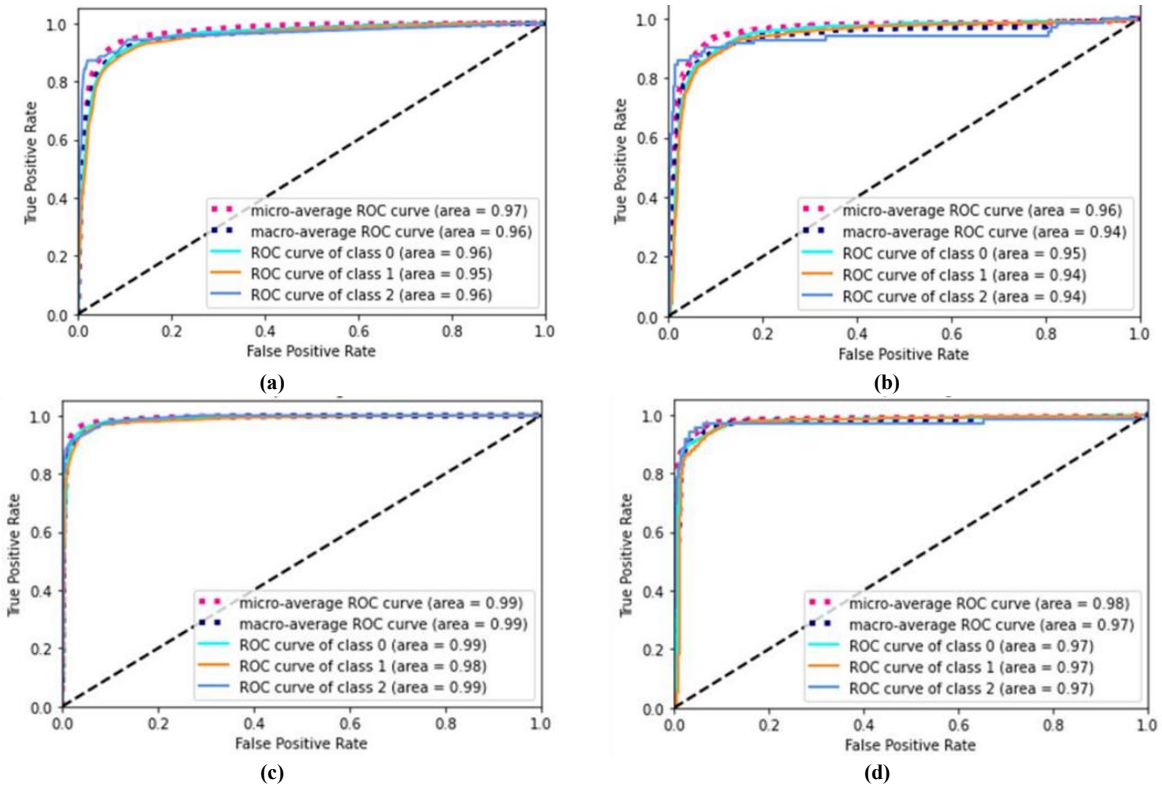


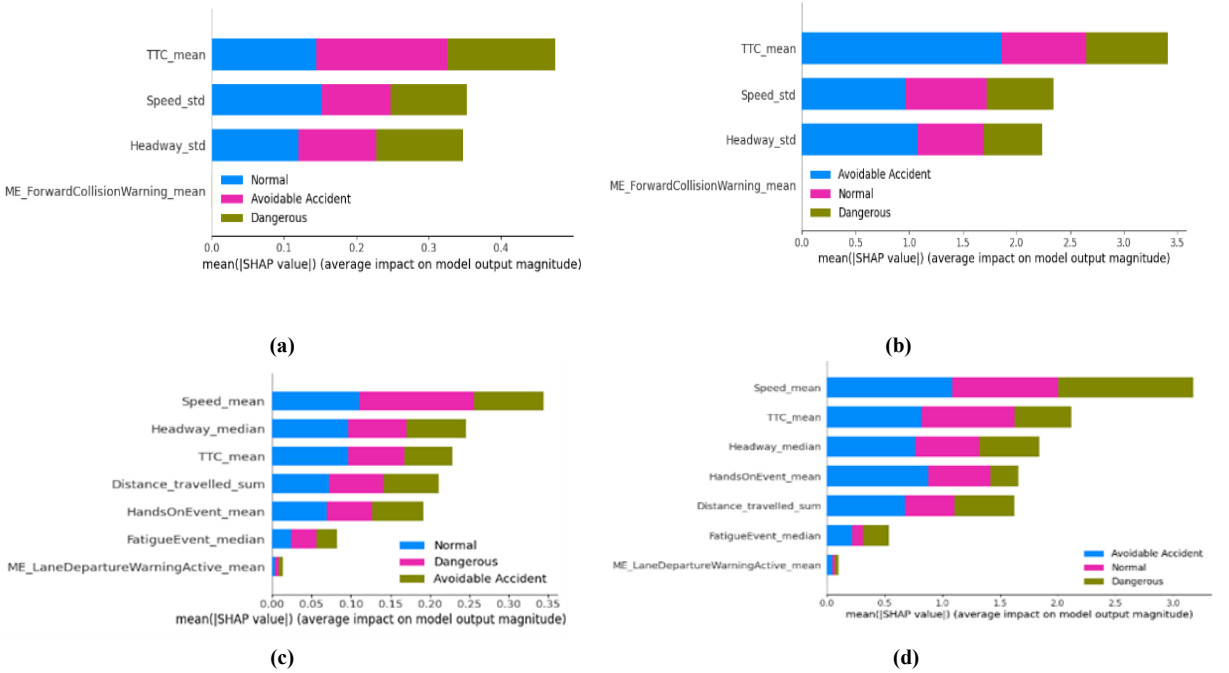
Fig. 3. ROC curves for the highest performing models (a) ROC curve for RF – Group A; (b) ROC curve for XGBoost – Group A; (c) ROC curve for RF – Group B; (d) ROC curve for XGBoost – Group B

The findings demonstrated excellent performance across both Groups A and B in classifying the three levels of driving behaviour, highlighting a strong ability to accurately distinguish between classes. Notably, Group B consistently produced superior results across the applied models. This improvement can be attributed to the broader inclusion of variables in Group B, suggesting that model accuracy is positively correlated with the amount of data incorporated. Among the tested algorithms, the RF and XGBoost models exhibited the highest capability in detecting risky driving behaviour, delivering closely aligned results with the greatest overall accuracy. Figure 3 presents the SHAP value analysis for the classification models. Higher SHAP values indicate greater feature importance, highlighting the relative contribution of each variable in predicting the three driving risk levels (normal, dangerous and avoidable accident).

Fig. 3. SHAP values (a) for RF - Group A; (b) for XGBoost - Group A; (c) for RF - Group B; (d) for XGBoost - Group B

5.2. Contribution of Each Factor

To gain deeper insights into the machine learning models, SHAP values were analysed. SHAP provides a clear interpretation of how each feature contributes to model predictions, allowing for a comprehensive understanding of



the dataset within the given problem. Figure 4 illustrates the relative contribution of each feature to the two models with the highest accuracy, offering a more detailed perspective on the influence of variables across both groups.

The distribution of SHAP values for key variables (e.g. TTC_mean, Speed_std, Headway_std, ME_ForwardCollisionWarning_mean) in relation to their impact on the model output is displayed. Each point represents a single observation, with color indicating the feature value (blue = low, red = high). Positive SHAP values shift predictions toward higher-risk classes, while negative values contribute to lower-risk classifications. The plots highlight the relative influence of each variable, showing that time-to-collision and speed are consistently strong predictors, while headway and forward collision warnings provide additional explanatory power.

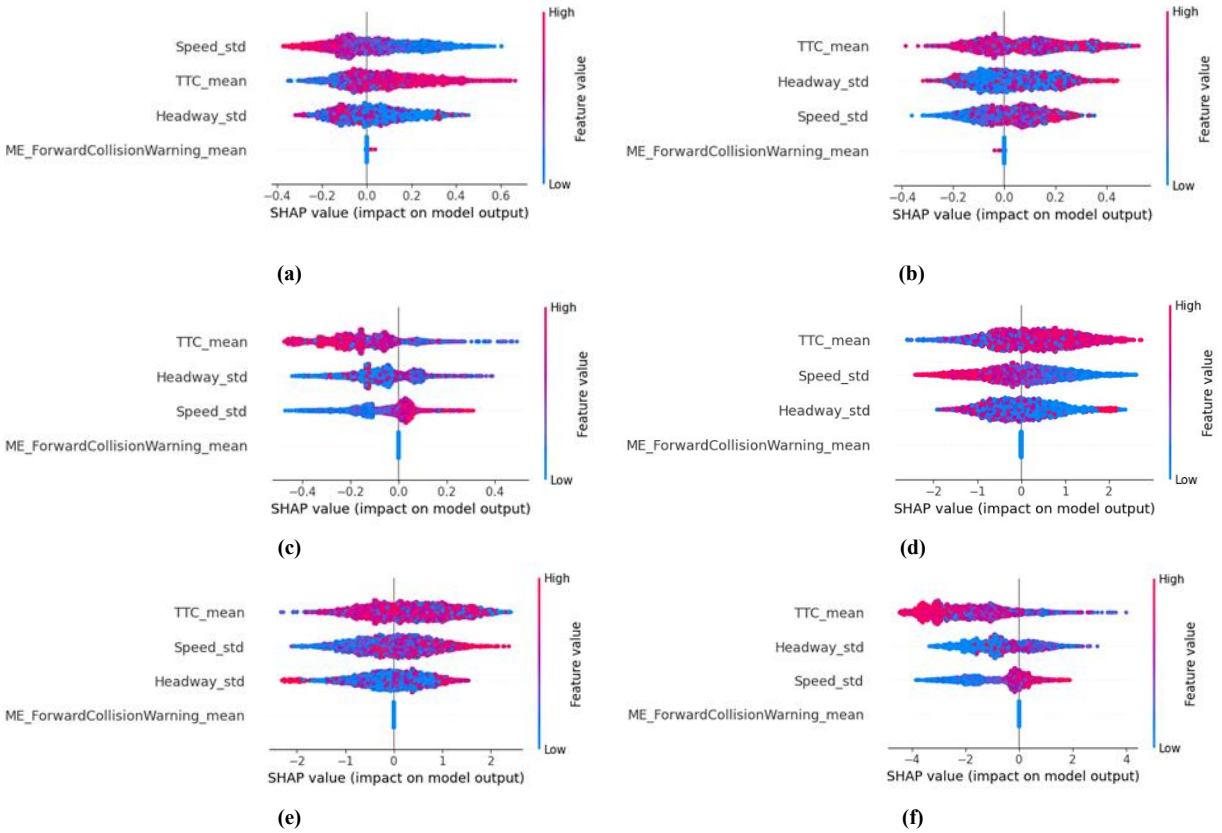


Fig. 5. SHAP values (a) for RF - Class 0 - Group A; (b) for RF - Class 1 - Group A; (c) for RF - Class 2 - Group A; (d) for XGBoost - Class 0 - Group A; (e) for XGBoost - Class 1 - Group A; (f) for XGBoost - Class 2 - Group A

For Group A, Figure 5 illustrates that a decrease in TTC_mean acts as a critical trigger for class transitions, particularly influencing the categories labeled as dangerous (1) and avoidable accident (2). Among the variables, time-to-collision was identified as the most influential factor across both models within this group. In addition, lower values of headway were associated with riskier driving behaviour, even within the normal class (0). For the dangerous (1) and avoidable accident (2) classes, headway also emerged as a key variable. In contrast, speed had a stronger effect on the normal class (0) than on the higher-risk classes. The forward collision warning variable, however, did not demonstrate a discernible impact on model predictions.

As per Group B, it was found that speed was the most influential variable across both models, with its impact on risk levels increasing proportionally with higher values. In particular, elevated speed values are strongly associated with more dangerous driving behaviour. Moreover, lower values of hands-on wheel, time-to-collision and headway contribute to shifts in driving levels, reflecting greater risk. The exposure-related variable distance travelled also plays a role, with longer distances linked to changes in safety levels. With regards to fatigue event and lane departure warning, their effect on driving behaviour was relatively minor; however, the results suggested that well-rested drivers are less likely to engage in risky driving. Taken together, the most critical determinants of risky behaviour in Group B are speed, the distance maintained between vehicles and the amount of time the driver keeps their hands on the steering wheel.

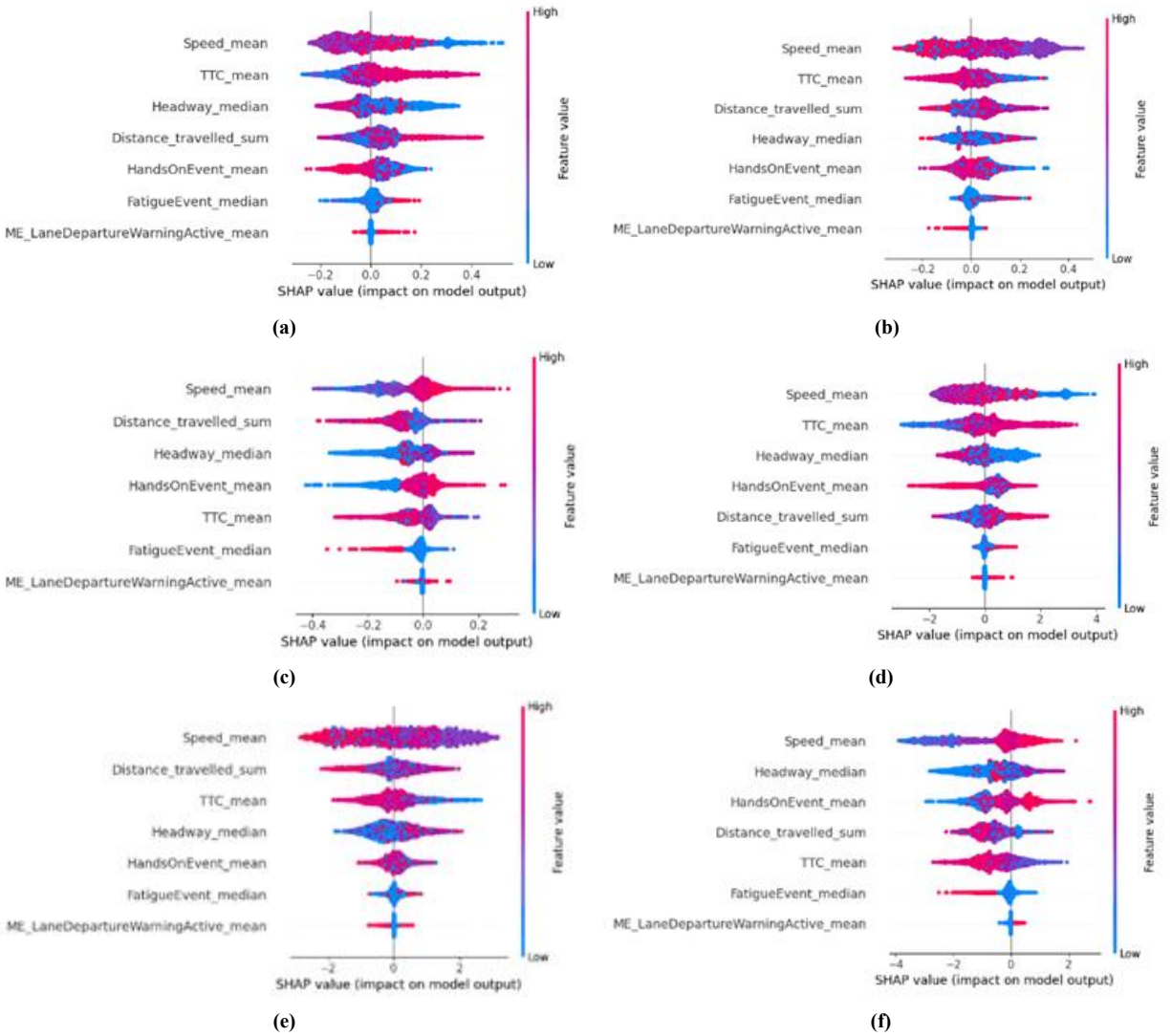


Fig. 6. SHAP values (a) for RF – Class 0 - Group B; (b) for RF – Class 1 - Group B; (c) for RF – Class 2 - Group B; (d) for XGBoost – Class 0 - Group B; (e) for XGBoost – Class 1 - Group B; (f) for XGBoost – Class 2 - Group B

6. Discussion

The analysis of driving behaviour across the three different STZ levels (normal, dangerous and avoidable accident) demonstrates the potential of machine learning algorithms to provide accurate and interpretable risk classification. Across both Groups A and B, ensemble-based models (RF and XGBoost) consistently outperformed linear models (RC and SVM), achieving superior accuracy, precision and recall. This aligns with prior studies emphasizing the effectiveness of tree-based methods in capturing nonlinear relationships and complex feature interactions in driving behaviour data (Ding et al., 2018).

The results, which compare the performance of the four classification algorithms across both feature groups, revealed significant findings regarding their predictive accuracy and overall effectiveness. In particular, in Group A, the Ridge Classifier performed poorly, with particularly low recall (35%), F-measure (27%) and G-means (19%), accompanied by a very high false alarm rate (95%). This indicates that RC was largely ineffective at distinguishing risky driving behaviour. The SVM achieved moderate performance, with recall at 71% and an F-measure of 62%, but

still fell short in comparison to ensemble methods. By contrast, both RF and XGBoost exhibited strong and nearly identical results, achieving high precision (80%), recall (88%) and F-measure (83%), while maintaining a low false alarm rate (12%) and the highest G-means (90%).

For Group B, where a broader set of features was included, overall performance improved substantially across models. The Ridge Classifier again performed poorly, with the lowest F-measure (23%) and G-means (5%), confirming its limited suitability for this task. SVM demonstrated significant improvement in Group B compared to Group A, reaching precision of 92%, recall of 83% and an F-measure of 86%. The ensemble methods (RF and XGBoost) again outperformed other models, with both achieving precision and recall above 90% and F-measure values exceeding 91%. Importantly, they also exhibited the lowest false alarm rates (7%) and the highest G-means (96%), indicating not only accuracy but also balanced performance across classes.

A key insight concerns the role of feature selection in shaping model performance. Group B, which incorporated a broader set of variables, consistently produced higher predictive accuracy than Group A. This finding suggests that integrating multiple behavioural, contextual and exposure-related factors enhances the ability of the model to distinguish between safe and risky driving patterns. It also reinforces the importance of feature comprehensiveness when developing classification frameworks for road safety research.

The SHAP value analysis further highlighted distinct variable importance patterns across the two groups. In Group A, time-to-collision emerged as the most critical predictor, with lower values strongly associated with higher-risk classifications (dangerous and avoidable accident). Headway also played a central role, particularly in differentiating between normal and risky behaviour, while speed was more strongly linked to normal driving. Results indicated that time-to-collision and headway variability are central indicators of driver safety when fewer variables are considered.

In contrast, Group B emphasized speed as the dominant predictor, with higher values correlating with increased risk. Furthermore, hands on wheel, time-to-collision and headway contributed significantly, suggesting that average speed, distance to other vehicles and sustained driver engagement with the steering wheel are strong determinants of risk. Exposure variables such as distance travelled also played a role, reinforcing that driving context and duration can influence the likelihood of risky behaviour. Although fatigue event and lane departure warning exhibited relatively limited effects, the results suggest that well-rested drivers tend to engage in less hazardous behaviour, consistent with previous findings on fatigue-related risk.

The comparison of Groups A and B revealed that expanding the feature set allows models to capture a more holistic view of driver risk, shifting emphasis from immediate situational factors (TTC, headway) toward broader behavioural indicators (average speed, hands-on control, exposure). This reflects the complexity of driving behaviour, which is shaped not only by instantaneous performance but also by accumulated behavioural patterns over time.

Overall, the results highlight three central contributions. Firstly, ensemble-based methods such as RF and XGBoost proved to be highly effective for multi-class driving risk prediction, offering superior performance compared to linear approaches. Secondly, the inclusion of a more comprehensive set of features enhanced both predictive accuracy and interpretability, underscoring the importance of capturing a wide range of behavioural and contextual variables. Thirdly, the use of SHAP values provided valuable transparency, allowing for a clearer understanding of the role and influence of key variables in model outcomes.

6.1. Limitations

Despite the rigorous methodologies employed, this study is not without limitations. One notable constraint lies in the availability and quality of data. Reliance on a relatively small sample size and a simulator-based dataset may have introduced biases, limiting the generalizability of the findings to real-world driving conditions. Future studies could address these issues by incorporating larger, more diverse datasets that reflect a broader range of drivers, environments and risk-related variables such as harsh acceleration, harsh braking or lane-changing behaviour. Expanding the feature set and adopting more advanced feature selection and evaluation strategies would also strengthen the robustness of the models. In addition, consolidating explanatory variables into a control group would streamline testing while offering a more solid framework for comparative analysis, thereby allowing for a deeper examination of their individual and combined effects.

Another limitation relates to the modelling techniques employed. The consistently weak performance of the Ridge Classifier underscores the inherent limitations of linear models in capturing the complex, nonlinear relationships that

characterize driving behaviour. While the selected algorithms performed well, particularly the ensemble methods, the inclusion of deep learning approaches such as Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs) may have enhanced predictive capabilities by capturing temporal dependencies and spatial features in the data. However, constraints in computational resources and time limited the exploration of these approaches during the research period.

Beyond these methodological limitations, other challenges should also be acknowledged. The use of a simulator, while valuable for controlled experimentation, cannot fully replicate the complexity of real-world driving conditions, where factors such as weather variability, traffic density and driver distraction can play a critical role. Furthermore, the study relied on short-term experiment, which may not adequately capture behavioural changes over extended periods of driving. Finally, the absence of cross-cultural or cross-regional validation restricts the transferability of findings, as driving behaviour and risk perception may vary significantly between different populations and traffic systems.

6.2. Future Research Directions

Looking ahead, future research should explore the integration of advanced deep learning models such as Neural Networks (Banan et al., 2020; Chen et al., 2022) and Long Short-Term Memory (LSTM) architectures (Fan et al., 2020; Roussou et al., 2024), which have demonstrated superior performance in capturing temporal and nonlinear patterns. Incorporating these methods may enhance predictive accuracy and provide more robust models for assessing driving risk. Additionally, expanding the dataset to include a broader and more diverse pool of participants, alongside richer contextual variables such as weather conditions, traffic density and road characteristics, would improve the generalizability of the findings. Longer-term data collection in real-world driving environments could also capture behavioural changes over time, offering insights that go beyond the limitations of simulator-based studies.

Another promising avenue is the inclusion of additional explanatory variables, such as harsh acceleration, harsh braking and lane-changing events, which may provide stronger indicators of risky behaviour. Furthermore, combining interpretability tools, such as SHAP, LIME (Man & Chan, 2021) and Global Sensitivity Analysis (Wen et al., 2022), could yield deeper insights into the decision-making processes of machine learning models, ensuring both accuracy and transparency. Finally, cross-regional or cross-cultural validations should be considered, as driving behaviour is often shaped by local traffic norms, infrastructure and cultural attitudes toward risk.

7. Conclusions

The aim of this research was to classify dangerous driving behaviour and identify the key factors influencing each level. Through the application of the Safety Tolerance Zone (STZ) concept, driving behaviour was classified into three phases, normal, dangerous and avoidable accident. To achieve this objective, a driving simulator experiment was conducted and a large dataset was collected and analysed.

Through a systematic feature selection process, the most relevant variables were identified and organized into two groups: Group A, consisting of TTC, headway, speed and forward collision warning, while Group B, included TTC, headway, Hands on wheel, fatigue event, lane departure warning, speed and distance travelled.

Four classification algorithms: Ridge Classifier (RC), Support Vector Machines (SVM), Random Forests (RF) and eXtreme Gradient Boosting (XGBoost) were then developed to predict driving behaviour within 30-second intervals. Evaluation across multiple metrics showed that the RF and XGBoost models consistently achieved the best performance, outperforming the other techniques across all three safety levels, reaching 95% in prediction accuracy. To better understand the contribution of individual features, SHAP value analysis was conducted. This revealed that time to collision was the most influential factor in Group A, whereas speed emerged as the dominant predictor in Group B, with higher values strongly associated with hazardous behaviour.

The findings of this study provide strong evidence of the value of predicting risky driving behaviour for road safety policy and practice. The identification of the specific factors that contribute to risky driving behaviour, such as time to collision, headway variability, speed and driver attentiveness, transportation Authorities can design more targeted and effective interventions. These insights can support the development of proactive safety programs, allowing for the timely identification of drivers at risk and the implementation of corrective measures before dangerous situations

escalate. For instance, predictive systems integrated into traffic monitoring infrastructure could alert authorities to patterns of unsafe behaviour, enabling focused enforcement and preventive campaigns in high-risk areas.

Lastly, the integration of predictive models into ADAS represents a significant opportunity for policymakers and industry stakeholders. Real-time interventions based on behavioural predictions can provide immediate feedback to drivers, fostering safer practices and reducing crash likelihood. Beyond in-vehicle applications, insurance providers and regulatory bodies could utilize these predictive tools to inform risk-based policies, promote responsible driving through incentives and ensure fairer premium structures.

Acknowledgements

The research was funded by the European Union's Horizon 2020 i-DREAMS project (Project Number: 814761) funded by European Commission under the MG-2-1-2018 Research and Innovation Action (RIA).

References

Abou Elassad, Z. E., Mousannif, H., & Al Moatassime, H. (2020). A proactive decision support system for predicting traffic crash events: A critical analysis of imbalanced class distribution. Knowledge-Based Systems, 205, 106314.

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. International journal of neuroscience, 52(1-2), 29-37.

Ali, Y., Sharma, A., Haque, M. M., Zheng, Z., & Saifuzzaman, M. (2020). The impact of the connected environment on driving behavior and safety: A driving simulator study. Accident Analysis & Prevention, 144, 105643.

Amini, R. E., Michelaraki, E., Katrakazas, C., Al Haddad, C., De Vos, B., Cuenen, A., ... & Antoniou, C. (2021). Risk scenario designs for driving simulator experiments. In 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) (pp. 1-6). IEEE.

Amsalu, S. B., Homaifar, A., Afghah, F., Ramyar, S., & Kurt, A. (2015, June). Driver behavior modeling near intersections using support vector machines based on statistical feature extraction. In 2015 IEEE intelligent vehicles symposium (IV) (pp. 1270-1275). IEEE.

Banan, A., Nasiri, A., & Taheri-Garavand, A. (2020). Deep learning-based appearance features extraction for automated carp species identification. Aquacultural Engineering, 89, 102053.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & electrical engineering, 40(1), 16-28.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

Chen, W., Sharifrazi, D., Liang, G., Band, S. S., Chau, K. W., & Mosavi, A. (2022). Accurate discharge coefficient prediction of streamlined weirs by coupling linear regression and deep convolutional gated recurrent unit. Engineering Applications of Computational Fluid Mechanics, 16(1), 965-976.

Ding, C., Cao, X. J., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. Transportation Research Part A: Policy and Practice, 110, 107-117.

Fan, Y., Xu, K., Wu, H., Zheng, Y., & Tao, B. (2020). Spatiotemporal modeling for nonlinear distributed thermal processes based on KL decomposition, MLP and LSTM network. IEEE Access, 8, 25111-25121.

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.

Garefalakis, T., Katrakazas, C., & Yannis, G. (2022). Data-driven estimation of a driving safety tolerance zone using imbalanced machine learning. Sensors, 22(14), 5309.

Bakhshi, A. K., & Ahmed, M. M. (2022). Bayesian extreme value analysis of kinematic-based surrogate measure of safety to detect crash-prone conditions in connected vehicles environment: A driving simulator experiment. Transportation research part C: emerging technologies, 136, 103539.

Man, X., & Chan, E. (2021). The best way to select features? comparing mda, lime, and shap. The Journal of Financial Data Science Winter, 3(1), 127-139.

Masello, L., Sheehan, B., Castignani, G., Shannon, D., & Murphy, F. (2023). On the impact of advanced driver assistance systems on driving distraction and risky behaviour: An empirical analysis of irish commercial drivers. Accident Analysis & Prevention, 183, 106969.

Misra, S., Li, H., & He, J. (2020). Noninvasive fracture characterization based on the classification of sonic wave travel times. Machine learning for subsurface characterization, 4, 243-287.

Mozaffari, S., Al-Jarrah, O. Y., Dianati, M., Jennings, P., & Mouzakitis, A. (2020). Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. IEEE Transactions on Intelligent Transportation Systems, 23(1), 33-47.

Osman, O. A., Hajij, M., Karbalaieali, S., & Ishak, S. (2019). A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data. Accident Analysis & Prevention, 123, 274-281.

Padurariu, C., & Breaban, M. E. (2019). Dealing with data imbalance in text classification. Procedia Computer Science, 159, 736-745.

Papantoniou, P., Papadimitriou, E., & Yannis, G. (2015). Assessment of driving simulator studies on driver distraction. Advances in transportation studies, (35).

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accident Analysis & Prevention, 136, 105405.

Roussou, S., Garefalakis, T., Michelaraki, E., Brijs, T., & Yannis, G. (2024). Machine learning insights on driving behaviour dynamics among Germany, Belgium, and UK drivers. Sustainability, 16(2), 518.

Shi, X., Wong, Y. D., Li, M. Z. F., Palanisamy, C., & Chai, C. (2019). A feature learning approach based on XGBoost for driving assessment and risk prediction. Accident Analysis & Prevention, 129, 170-179.

Shangguan, Q., Fu, T., Wang, J., Luo, T., & Fang, S. E. (2021). An integrated methodology for real-time driving risk status prediction using naturalistic driving data. Accident Analysis & Prevention, 156, 106122.

Vapnik, V. (1998). The support vector method of function estimation. In Nonlinear modeling: Advanced black-box techniques (pp. 55-85). Boston, MA: springer us.

Voinea, G. D., Boboc, R. G., Buzdugan, I. D., Antonya, C., & Yannis, G. (2023). Texting while driving: a literature review on driving simulator studies. International journal of environmental research and public health, 20(5), 4354.

Wen, X., Xie, Y., Jiang, L., Li, Y., & Ge, T. (2022). On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. Accident Analysis & Prevention, 168, 106617.

Whittingham, R. (2004). The blame machine: Why human error causes accidents. Routledge.

Xie, J., & Zhu, M. (2019). Maneuver-based driving behavior classification based on random forest. IEEE Sensors Letters, 3(11), 1-4.

Yen, S., & Lee, Y. (2006). Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. Lecture notes in control and information sciences, 344, 731.

Zhu, S., Li, C., Fang, K., Peng, Y., Jiang, Y., & Zou, Y. (2022). An optimized algorithm for dangerous driving behavior identification based on unbalanced data. Electronics, 11(10), 1557.