# Predicting Driver Behaviour in a Cross-Country Naturalistic Driving Study Using Machine Learning Techniques

Giannis Roukos, Thodoris Garefalakis[0000-0003-0151-7864], Eva Michelaraki*[0000-0002-7167-4630], George Yannis[0000-0002-2196-2335]

National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou str., 15773, Athens, Greece
* evamich@mail.ntua.gr

**Abstract.** As traffic incidents continue to pose serious public health risks, identifying and predicting unsafe driving behaviour has become increasingly important. This paper is an outcome of the European Union's Horizon 2020 i-DREAMS project and applies the Safety Tolerance Zone (STZ) concept by classifying trip-level driving behaviour into three safety levels using naturalistic driving data. The aim of this work was to develop machine learning models in order to classify driver behaviour into three safety levels. To achieve this objective, a naturalistic driving experiment was conducted and data from Belgium and the United Kingdom were collected and analyzed. Variable importance was assessed using the Random Forest algorithm, resulting in the selection of nine key features. Four classification models were trained and evaluated through confusion matrices and standard performance metrics. SHapley Additive exPlanations (SHAP) values were used to interpret the contribution of each input variable, leading to the identification of CatBoost and LightGBM as the most effective models. Further SHAP analysis revealed that average speed was the most influential predictor across all safety levels, while harsh driving events played a crucial role in identifying dangerous behaviour. The findings highlighted the potential of interpretable machine learning models for real-time safety monitoring and suggest incorporating speed management and harsh event detection into future in-vehicle safety systems.

**Keywords:** driving behaviour, road safety, real-time crash prediction, machine learning, classification models.

## 1. Introduction

Road safety remains one of the most pressing public health concerns globally. Despite long-term declines in fatalities, road traffic crashes still account for over 1.19 million deaths each year worldwide [1]. In the European Union (EU), the number of road deaths has decreased by 60% since 2000, however, progress in recent years has plateaued [2]. Human error is widely recognized as the primary contributing factor in approximately 90-95% of road crashes, underscoring the critical need for targeted interventions that address risky driver behaviour [3].

Efforts to enhance traffic safety are increasingly supported by technological advancements such as Advanced Driver Assistance Systems (ADAS), which aim to detect, predict, and mitigate potentially hazardous driving in real time. These systems often rely on data-driven models trained to identify patterns of unsafe behaviour from large-scale sensor and behavioral data [4]. In this context, machine learning (ML) methods offer promising capabilities for classifying driver risk profiles with high accuracy and adaptability.

This paper is a standalone applied study developed using i-DREAMS field-trial data. The objective is to develop and evaluate interpretable machine-learning models for classifying trip-level driving behaviour into three STZ safety levels using naturalistic driving data from Belgium and the United Kingdom, and to examine the most influential predictors across safety levels. The contribution of this paper is applied. It presents an interpretable workflow for trip-level STZ classification using naturalistic driving data collected under the i-DREAMS protocol and reports comparative results across two national datasets (i.e., Belgium and the United Kingdom). The workflow includes leakage-aware variable handling (excluding label-defining variables), feature selection, class balancing, model training (i.e., Random Forest, CatBoost, LightGBM, and Multi-Layer Perceptron) and evaluation under a consistent procedure per country, and SHAP-based interpretation to identify influential predictors across safety levels [5, 6]. This workflow is differentiated by its trip-level operationalization of STZ classification using harmonized field-trial variables and by explicitly reducing label leakage through the exclusion of label-defining inputs. Average speed and harsh driving events emerged as influential features associated with unsafe driving behaviour [7], supporting STZ-based safety monitoring applications.

## 2. Background

Understanding and predicting driving behaviour has become an increasingly critical domain in road safety research, particularly in light of the growing availability of naturalistic driving data and advancements in machine learning. Driver behaviour, influenced by complex interactions between human, vehicle, and environmental factors, plays a central role in crash causation. Studies show that behavioral factors such as speeding, harsh acceleration, distraction, fatigue, and stress are among the primary contributors to unsafe driving [8].

To mitigate these risks, recent research has focused on classifying driver behaviour using supervised machine learning algorithms trained on real-world driving data [9]. These approaches have been shown to outperform traditional statistical models in terms of classification accuracy and scalability [10]. However, many of these models operate as "black boxes," making them difficult to interpret and less suitable for safety-critical applications where transparency and explainability are essential [11].

The i-DREAMS project (intelligent Driver and Road Environment Assessment and Monitoring System) was designed to address these challenges by developing a dynamic driver monitoring framework that continuously assesses driving risk in real time. Central to this framework is the concept of the Safety Tolerance Zone (STZ), which classifies each driving moment into one of three levels: normal, dangerous, or avoidable accident [12]. This classification allows for adaptive feedback and intervention

strategies tailored to the driver's current state and context. The STZ classification draws upon a range of data sources, including vehicle telemetry (e.g., speed, acceleration), physiological signals (e.g., heart rate variability), and contextual variables (e.g., road type, time of day). Prior research within the i-DREAMS framework has demonstrated the feasibility of integrating these data streams to derive a robust understanding of driver risk [13, 14]. Yet relatively few studies have applied interpretable machine learning models to assess feature contributions and cross-country variations in driving behaviour patterns [6, 14]. This study builds upon previous work by leveraging data collected from naturalistic driving experiments in Belgium and the United Kingdom. Using feature selection, data balancing techniques, and interpretable ensemble classifiers, the research aims to identify critical predictors of unsafe driving while maintaining transparency through SHAP-based explanations. As such, the study contributes to a growing body of literature on interpretable AI in transportation and offers practical insights for the future of in-vehicle safety systems.

## 3. Data Description

### 3.1 Study Design and Data Collection

This study was conducted within the framework of the Horizon 2020 i-DREAMS project, aiming to define and monitor a Safety Tolerance Zone (STZ) based on naturalistic driving data. The data were collected during the third phase of the i-DREAMS field trials, which included instrumented vehicles and wearable sensors that captured both behavioral and physiological indicators. The current analysis focuses on Belgium and the United Kingdom, which were selected because they offered sufficient sample size and complete availability of the trip-level variables required for the modelling pipeline under the i-DREAMS data collection protocol. In total, the dataset comprises 813 trips (21,412 driving minutes, 42 drivers) from Belgium and 3,317 trips (58,458 driving minutes, 54 drivers) from the United Kingdom.

Each trip is labelled using the STZ framework, which classifies driving behaviour into three safety levels: Level 0 (normal), Level 1 (dangerous), and Level 2 (avoidable accident). These classifications were derived using domain-specific thresholds and i-DREAMS algorithms incorporating variables such as time gaps, harsh events, and physiological state. In this study, the trip was used as the unit of analysis, and variables that directly contributed to STZ labelling were excluded from modelling to avoid bias. Data sources included vehicle telemetry from OBU and Mobileye systems (e.g., speed, distance, harsh braking/acceleration, speed-limit compliance), physiological data from the Empatica E4 device (heart rate and inter-beat interval, IBI), and contextual features such as time of day, trip duration, road type, and driver ID. Following feature selection, the variables retained for modelling are presented in Table 1.

**Table 1.** Description of Selected Features for Modeling.

| Feature | Description |
|---|---|
| ME_Car_speed_mean | Mean vehicle speed during the trip |
| IBI_value_mean | Mean inter-beat interval (physiological stress indicator) |
| DEM_evt_ha_lvl_L_sum | Total harsh acceleration events at low severity |
| DEM_evt_ha_lvl_L_mean | Average severity of harsh acceleration events at low level |
| DEM_evt_hb_lvl_L_mean | Average severity of harsh braking events at low level |
| DrivingEvents_Map_evt_ha_mean | Mean harsh acceleration mapped across trip segments |
| DrivingEvents_Map_evt_hb_mean | Mean harsh braking mapped across trip segments |
| DrivingEvents_Map_evt_hc_mean | Mean harsh cornering mapped across trip segments |
| DEM_evt_hc_lvl_L_sum | Total harsh cornering events at low severity |

### 3.2 Model Interpretability with SHAP

This study employed supervised machine-learning models to classify trip-level driving behaviour into three predefined safety levels. Four algorithms were evaluated: Random Forest, LightGBM, CatBoost, and Multi-Layer Perceptron (MLP). These models were selected because they can handle tabular data, non-linear relationships, and class imbalance under a consistent comparative setup.

To interpret model predictions, SHapley Additive exPlanations (SHAP) was applied [5]. In this study, SHAP is used as a post-hoc interpretability tool to explain model behaviour and support transparency, not as a methodological novelty. SHAP was applied to the best-performing models to identify influential predictors across safety levels.

## 4. Results

### 4.1 Model Performance

The performance of the four machine learning models, CatBoost, LightGBM, Random Forest, and MLP, was evaluated using five key metrics: accuracy, precision, recall, F1-score, and false alarm rate (FAR). Table 2 presents the average values for each metric across all STZ levels, separately for the Belgium and UK datasets. CatBoost demonstrated the highest performance overall, particularly in the UK dataset, where it achieved an F1-score of 0.79 and the lowest FAR of 0.12. LightGBM followed closely, while Random Forest showed reasonable but lower recall and higher FAR. The MLP model consistently underperformed, indicating limitations in handling class imbalance and mixed data types. These results suggest that ensemble tree-based models, particularly gradient boosting methods, are well-suited for this classification task, offering both predictive accuracy and reliability in identifying unsafe driving behaviour.

**Table 2.** Model Performance per Country.

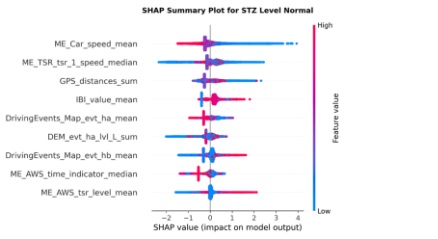
| | | | UK | | |
|---|---|---|---|---|---|
| | **Accuracy** | **Precision** | **Recall** | **f1-score** | **False Alarm Rate** |
| **Random Forest** | 77% | 78% | 73% | 75% | 14% |
| **LightGBM** | 76% | 79% | 69% | 72% | 16% |
| **CatBoost** | 75% | 75% | 71% | 73% | 15% |
| **MLP** | 68% | 69% | 61% | 64% | 21% |
| | | | **Belgium** | | |
| | **Accuracy** | **Precision** | **Recall** | **f1-score** | **False Alarm Rate** |
| **Random Forest** | 74% | 70% | 65% | 67% | 18% |
| **LightGBM** | 71% | 67% | 59% | 61% | 20% |
| **CatBoost** | 71% | 64% | 63% | 63% | 19% |
| **MLP** | 63% | 53% | 50% | 50% | 26% |

Overall, CatBoost achieved the best results across all metrics in both countries. Performance was slightly higher in the UK, possibly due to stronger signal in the physiological variables. MLP consistently underperformed, despite class balancing, suggesting that deep learning models require further optimization or additional temporal inputs for this application.

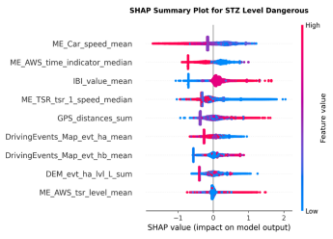
### 4.2 Model Interpretability

To gain insights into model behaviour, SHAP analysis was conducted on the CatBoost model, which outperformed all others. SHAP summary plots provide a global view of feature importance and their directional impact on the classification output. The SHAP summary plots present the most influential features for distinguishing between the three Safety Tolerance Zone (STZ) levels, Level 0 (Normal), Level 1 (Dangerous), and Level 2 (Avoidable Accident), based on their average contribution to the model output.

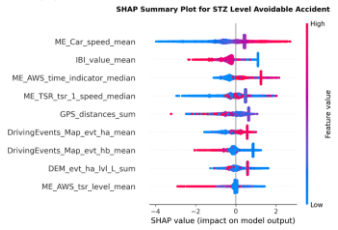
In the United Kingdom dataset (see Fig. 1), the CatBoost model drew on a balanced set of behavioral, physiological, and contextual indicators to predict STZ levels. The most influential predictors were mean vehicle speed and compliance with speed limits, highlighting the role of driving style and rule adherence in differentiating trip risk. The inter-beat interval (IBI), a physiological measure associated with cognitive workload or stress, also emerged as a key variable, particularly in distinguishing Level 2 (Avoidable Accident) trips. Other contextual features, such as trip distance and time-of-day indicators, contributed to a lesser extent, alongside indicators of harsh driving events (e.g., acceleration and braking). These results suggest that in the UK sample, the model capitalized on both external driving behaviour and internal physiological state, reflecting a multidimensional representation of risk.




*(a) Normal level*



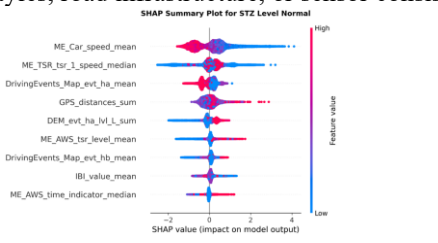

*(b) Dangerous level*



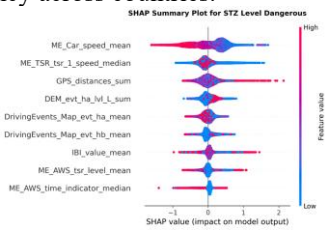

*(c) Avoidable Accident level*

**Fig. 1.** SHAP Summary Plots for the CatBoost Model – United Kingdom Dataset: (a) Normal level; (b) Dangerous level; (c) Avoidable Accident level.

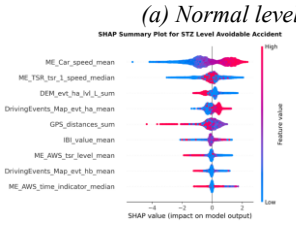
By contrast, in the Belgian dataset (see Fig. 2), the CatBoost model placed relatively greater emphasis on behavioral indicators. While mean speed and speed limit compliance again ranked highly, harsh driving events, such as hard acceleration and braking, played a more prominent role across all three STZ classes. Physiological features like IBI were retained in the model but demonstrated reduced overall impact compared to the UK case. Contextual factors (e.g., time indicators) showed more modest contributions. These differences may reflect contextual variation in driving styles, road infrastructure, or sensor consistency across countries.




*(a) Normal level*




*(b) Dangerous level*




*(c) Avoidable Accident level*

**Fig. 2.** SHAP Summary Plots for the CatBoost Model – Belgian Dataset: (a) Normal level; (b) Dangerous level; (c) Avoidable Accident level.

Overall, both datasets shared a common predictive core, but the UK model relied more on physiological and temporal signals, whereas the Belgian model placed greater weight on behavioral-event indicators.

## 5. Discussion

This study applied interpretable machine-learning models to classify trip-level driving risk under the STZ framework using naturalistic driving data from the United Kingdom and Belgium. The results show that tree-based models, particularly CatBoost, can distinguish among normal, dangerous, and avoidable-accident trips using a compact set of behavioral, physiological, and contextual features. Across both countries, mean vehicle speed and speed-limit compliance emerged as consistently important predictors, confirming their central role in characterizing unsafe driving behaviour.

Cross-country differences were also observed. In the United Kingdom, physiological indicators, particularly inter-beat interval (IBI), contributed more strongly to the identification of higher-risk trips, whereas in Belgium the model relied more on behavioral-event variables such as harsh acceleration and braking. These differences may reflect variation in road environment, traffic conditions, driver population, or signal quality across datasets. SHAP supported transparent interpretation of these class-level patterns, while the modelling pipeline reduced label leakage by excluding label-defining variables and applying a consistent procedure across countries.

Nonetheless, several limitations should be acknowledged. The dataset, although rich in sensor modalities, included a limited number of participants, and physiological data quality may vary across subjects. In addition, models were trained and validated separately for each country, future work could explore unified or transfer-learning approaches. Furthermore, model evaluation was restricted to two country datasets, therefore, transferability to other driving contexts, (e.g., non-EU countries) is not established. Future work should extend validation using more diverse data and additional driving contexts. Statistical robustness checks, such as repeated resampling, and temporal (time-based) validation were not performed within the scope of this study and are left for future work. Finally, practical deployment would require real-time implementation and testing under unseen road, traffic, and user conditions.

## 6. Conclusions

This study developed and evaluated an interpretable workflow for trip-level STZ classification using naturalistic driving data from the United Kingdom and Belgium. CatBoost achieved the strongest overall performance, while SHAP-based interpretation showed that mean vehicle speed and speed-limit compliance were consistently important predictors, with cross-country differences in the relative contribution of physiological and behavioral variables. These findings support the applied use of transparent machine-learning models for driver risk monitoring, while indicating that model portability to other contexts should be examined in future work through broader datasets, temporal validation, and real-time deployment testing.

## References


1. World Health Organization: Global status report on road safety 2023, https://www.who.int/publications/i/item/9789240086517, last accessed 2024/03/06.
2. European Commission, Directorate-General for Mobility and Transport: Next steps towards 'Vision Zero': EU road safety policy framework 2021-2030. Publications Office (2020). https://doi.org/doi/10.2832/261629.
3. Valente, J., Ramalho, C., Vinha, P., Mora, C., Jardim, S.: Using machine learning to understand driving behavior patterns. Procedia Comput. Sci. 239, 1823–1830 (2024). https://doi.org/10.1016/j.procs.2024.06.363.
4. Lattanzi, E., Castellucci, G., Freschi, V.: Improving Machine Learning Identification of Unsafe Driver Behavior by Means of Sensor Fusion. Applied Sciences. 10, 6417 (2020). https://doi.org/10.3390/app10186417.
5. Lundberg, S.M., Lee, S.-I.: A Unified Approach to Interpreting Model Predictions. In: Guyon, I., Luxburg, U. Von, Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.) Advances in Neural Information Processing Systems. Curran Associates, Inc. (2017).
6. Cheng, C., Chen, S., Ma, Y., Khattak, A.J., Zhang, Z.: Recognition and interpretation of aggressive driving behavior for heavy-duty vehicles based on artificial neural network and SHAP. Human Factors and Ergonomics in Manufacturing & Service Industries. 34, 177–189 (2024). https://doi.org/10.1002/hfm.21019.
7. Lee, G.-S.: Machine Learning-Based Driving Style Classification Real-World Data Prediction. In: 2024 4th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC). pp. 53–57. IEEE (2024). https://doi.org/10.1109/ICAIRC64177.2024.10899816.
8. Ghandour, R., Potams, A.J., Boulkaibet, I., Neji, B., Al Barakeh, Z.: Driver Behavior Classification System Analysis Using Machine Learning Methods. Applied Sciences. 11, (2021). https://doi.org/10.3390/app112210562.
9. Feng, Y., Ye, Q., Adan, F., Marques, L., Angeloudis, P.: Driving Style Classification Using Deep Temporal Clustering with Enhanced Explainability. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). pp. 4040–4045. IEEE (2023). https://doi.org/10.1109/ITSC57777.2023.10421826.
10. Shangguan, Q., Fu, T., Wang, J., Luo, T., Fang, S.: An integrated methodology for real-time driving risk status prediction using naturalistic driving data. Accid. Anal. Prev. 156, 106122 (2021). https://doi.org/10.1016/j.aap.2021.106122.
11. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why Should I Trust You?" In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144. ACM, New York, NY, USA (2016). https://doi.org/10.1145/2939672.2939778.
12. Michelaraki, E., Katrakazas, C., Brijs, T., Yannis, G.: Modelling the Safety Tolerance Zone: Recommendations from the i-DREAMS project. In: 10th International Congress on Transportation Research. , Rhodes Island, Greece (2021).
13. Garefalakis, T., Michelaraki, E., Roussou, S., Katrakazas, C., Brijs, T., Yannis, G.: Predicting risky driving behavior with classification algorithms: results from a large-scale field-trial and simulator experiment. European Transport Research Review. 16, 65 (2024). https://doi.org/10.1186/s12544-024-00691-9.
14. Roussou, S., Garefalakis, T., Michelaraki, E., Brijs, T., Yannis, G.: Machine Learning Insights on Driving Behaviour Dynamics among Germany, Belgium, and UK Drivers. Sustainability. 16, 518 (2024). https://doi.org/10.3390/su16020518.