Hybrid Modelling for Risky Driving Behavior Classification: Insights from Naturalistic Driving Study

Eleni Maria Theodoraki
Research Associate
Department of Transportation Planning and Engineering
National Technical University of Athens, Athens, Greece, GR15773
Email: elemartheo@gmail.com

Thodoris Garefalakis
Research Associate
Department of Transportation Planning and Engineering
National Technical University of Athens, Athens, Greece, GR15773
Email: tgarefalakis@mail.ntua.gr

Eva Michelaraki
Research Associate
Department of Transportation Planning and Engineering
National Technical University of Athens, Athens, Greece, GR15773
Email: evamich@mail.ntua.gr

George Yannis
Professor
Department of Transportation Planning and Engineering
National Technical University of Athens, Athens, Greece, GR15773
Email: geyannis@central.ntua.gr

Word Count: 6792 words + 2 table (250 words per table) = 7,292 words

Submitted: December 13, 2024

ABSTRACT

Driver behavior significantly impacts road safety, serving as a critical factor in traffic crash risks. Human error accounts for a large proportion of crashes, emphasizing the need for targeted interventions. To address this challenge, the i-DREAMS project introduced a "Safety Tolerance Zone (STZ)" framework. This innovative framework is designed to maintain drivers within safe operational boundaries by utilizing both real-time interventions, such as in-vehicle alerts, and post-trip feedback mechanisms, including personalized reports and recommendations. This study introduces and evaluates three hybrid machine learning models—DNN-RF, CNN-LSTM, and RNN-AdaBoost—to classify risky driving behavior into three safety levels: Normal, Dangerous, and Avoidable Accident. The hybrid approach combines the strengths of deep learning and traditional machine learning techniques to enhance predictive accuracy and robustness. To achieve this, a naturalistic driving experiment was conducted in Belgium and the United Kingdom, yielding a comprehensive dataset encompassing 69 drivers, 15,389 trips, and 265,512 minutes of recorded driving data. This dataset reflects diverse driving conditions and behaviors, providing a rich basis for analysis. Among the hybrid models, the Deep Neural Network-Random Forest (DNN-RF) model demonstrated the highest accuracy, achieving approximately 97% in both datasets. Critical driving variables identified as predictors included total travel distance, average speed, harsh acceleration, and harsh braking. To further enhance the interpretability of these machine learning models, the Local Interpretable Model-agnostic Explanations (LIME) algorithm was applied. LIME provided valuable insights into regional differences: harsh acceleration and braking were found to be the most influential factors in predicting risky behaviors in Belgium, whereas trip distance and harsh acceleration were more critical in the UK dataset. These findings underscore the potential of machine learning models to offer actionable insights into the factors contributing to hazardous driving behaviors, allowing authorities and organizations to develop real-time interventions and region-specific strategies.

Keywords: road safety; driving behavior classification; hybrid classification models; Local Interpretable Model-agnostic Explanations (LIME)

INTRODUCTION

Road transport is a cornerstone of modern society, supporting both societal function and economic growth. The rapid increase in private vehicle ownership, particularly over the past few decades, has resulted from technological advancements, industrial expansion, and rising personal mobility demands. While the widespread use of automobiles has brought undeniable benefits, it has also introduced significant challenges - chief among them being the issue of road safety. According to the World Health Organization (WHO), road traffic injuries are one of the leading causes of death globally, responsible for approximately 1.19 million deaths each year (1). Beyond the loss of life, road crashes have profound social and economic consequences.

Human error is a leading factor in the vast majority of traffic crashes, with studies consistently showing that approximately 90-95% of all road crashes are linked to driver-related behaviors. These behaviors include speeding, violating traffic rules, distracted driving, fatigue, and driving under the influence of alcohol or drugs. Given the significant role human error plays in road safety incidents, addressing these behaviors has become a central focus of efforts to reduce crashes. Autonomous driving technologies, which can eliminate or reduce the impact of these human factors, hold significant promise for enhancing road safety. For instance, autonomous vehicles have the potential to reduce crashes by as much as 93% by eliminating common driver errors (2). Similarly, addressing critical human errors in decision-making and attention could further reduce crash rates (3).

The rise of advanced vehicle technologies, including autonomous driving and intelligent transportation systems (ITS), has provided new tools to analyze and predict driver behavior in real-time. These technologies aim to reduce the human error factor in road safety, which has traditionally been challenging to address. Recent studies highlight the potential of machine learning (ML) and deep learning (DL) algorithms to identify dangerous driving patterns and predict crash risks by analyzing naturalistic driving data. This opens up possibilities for real-time interventions to prevent crashes and enhance overall road safety.

Several studies have already explored the use of ML techniques to predict crashes and assess driver behavior. For example, Wang et al. (2020) (4) examined the correlation between various behavioral and environmental factors on driving risks, while Peppes et al. (2021) emphasized the role of ITS in enabling autonomous vehicles to prevent crashes by anticipating driver errors. Additionally, Shi et al. (2019) (5) developed a framework for risk assessment using unsupervised learning to predict driving risks based on naturalistic driving data. Moreover, recent advancements in machine learning algorithms, such as Random Forests, Long Short-Term Memory (LSTM) networks, and Deep Neural Networks (DNN), have demonstrated high accuracy in predicting driver behavior based on various factors such as vehicle speed, distance from other vehicles, and sudden braking or acceleration events. For instance, the combination of DNN and Random Forest models has proven particularly effective in classifying driver behavior into different risk levels, as demonstrated in the research conducted by Yang et al. (2021) (6) within the i-DREAMS project.

However, while these machine learning models offer powerful predictive capabilities, they are often referred to as "black boxes" with limited transparency regarding how specific factors contribute to their predictions. This lack of interpretability poses a challenge, particularly in safety-critical domains like road safety, where understanding how and why a model makes certain predictions is crucial. Interpretability methods allow for greater transparency by providing insights into which variables (e.g., vehicle speed, sudden acceleration, or proximity to other vehicles) influence model decisions. This is vital for ensuring trust, accountability, and safety in applying these models to real-world road safety scenarios. Despite the growing body of research on machine learning and road safety, the application of interpretability methods remains limited. Few studies have incorporated such techniques into predictive models, leaving a significant gap in ensuring transparency in machine learning models used for road safety. This lack of interpretability is a concern in real-world, safety-critical applications like autonomous driving, where understanding the factors that drive predictions is essential for trust and safety.

This research aims to address these gaps by (1) integrating interpretability methods into machine learning models for predicting dangerous driving behaviors, thereby providing greater transparency into how these models operate, and (2) applying these models to datasets from multiple countries (e.g., Belgium and the UK) to assess the generalizability of the findings.

The paper is structured as follows : after the introduction, an extensive literature review is conducted on driving behavior analysis using deep learning techniques. This is followed by the description of the research methodology, which includes the data collection process as well as the theoretical background of the models. Finally, the results of the analysis are presented, in order to draw conclusions, related to road safety.

METHODS

Data Collection

As part of the i-DREAMS research project, a naturalistic driving experiment was conducted with participants from Belgium and the UK. For the Belgian cohort, the study included 43 drivers, resulting in a large dataset consisting of 7163 trips and 147337 minutes of driving data. In the UK, 26 drivers were involved, resulting in 8226 trips and 118175 minutes of recorded driving time. The experiment was designed to collect comprehensive data on driving behavior and road environments, facilitating an in-depth analysis of dangerous driving behaviors.

The experiment was conducted over a four-month period, divided into four distinct phases. Phase 1, lasting four weeks, served as a baseline with no interventions. In Phase 2, in-vehicle real-time warnings were introduced through adaptive Advanced Driver Assistance Systems (ADAS), also for a four-week period. During Phase 3, drivers received performance feedback via a mobile phone app, while Phase 4 extended this approach by introducing gamification elements to encourage safer driving behavior. Throughout all phases, the focus was on monitoring real-time driving behavior and evaluating the effectiveness of both real-time interventions (ADAS warnings) and post-driving feedback mechanisms. **Figure 1** provides an overview of the different phases of the experimental design of the i-DREAMS on-road study.

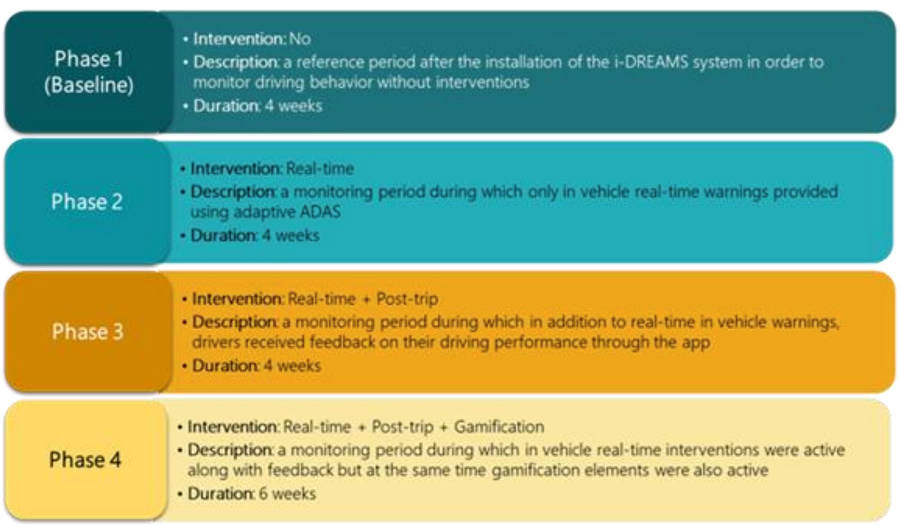


Figure 1 Overview of the different phases of the experimental design

A Data collection employed several cutting-edge technologies, including an OBD-II device installed in each vehicle to capture hundreds of driving parameters. The Mobileye system, integrated with mobile networks, further facilitated data gathering without user interaction. To categorize driving behavior, each 30-second interval of the trip was assigned to one of three safety levels: Normal, Dangerous, or Avoidable

Accident. These levels were determined based on intervention thresholds from the literature and the classification of variables such as speed and headway distances.

Definition of "Safety Tolerance Zone"

Before the development of classification algorithms, it was necessary to categorize the driving data into one of three levels of the "Safety Tolerance Zone" These categories were critical in structuring the dataset for the machine learning models. The Safety Tolerance Zone levels - Normal, Dangerous, and Avoidable Accident - were established based on real-time intervention thresholds derived from the international literature and validated within the i-DREAMS project. This categorization process was essential for understanding how driving behavior relates to road safety risks, with each intervention level reflecting different levels of driving risk.

To harmonize the classification with standards from the literature, the intervention levels were mapped using two primary indicators : speed and headway distance (the distance between the driver's vehicle and the preceding vehicle). This mapping process was consistent for both the Belgian and UK datasets. It was expected that normal driving behaviors would form the majority class, while dangerous behaviors and avoidable accidents would naturally be minority classes, given the focus on safety-critical driving events.

The data collected from the experiment included variables such as iDreams_Headway_Map_level_i, where i represents the intervention level ranging from -1 to 3. These variables capture the intervention levels in relation to time headway (the time gap between vehicles) and real-time speed interventions, which were critical in categorizing driving behavior into the Safety Tolerance Zone. Each intervention level was assigned a value of 0 or 1 :

- **0** : Indicates that the intervention level is not equal to i.
- **1** : Indicates that the intervention level is equal to i.

To determine the overall Safety Tolerance Zone level for each 30-second timeframe of driving data, the intervention variables were evaluated to identify the most unfavorable safety level. This ensured that the categorization of driving behavior reflected the highest risk level encountered during that interval. The levels were defined as follows :

- **Normal** : When the intervention level was -1, 0, or 1
- **Dangerous** : When the intervention level was 2
- **Avoidable Accident** : When the intervention level was 3

By classifying the data at 30-second intervals, this process captured the dynamic nature of driving behaviors and ensured that the most hazardous conditions were correctly identified, facilitating the subsequent development of classification algorithms.

Machine Learning Models

The core objective of this study was to predict driving risk levels based on real-world driving data. The classification problem was structured into three risk levels : Normal, Dangerous, and Avoidable Accident. To address this, three advanced machine learning and deep learning hybrid models were developed, which represent a relatively novel approach in road safety estimation, where hybrid models are still not widely used. These models—integrating different strengths from deep learning, machine learning, and ensemble learning—allow for more accurate and robust predictions of driving behavior compared to single-model approaches. The use of hybrid deep learning models has proven effective in various classification tasks, particularly in domains requiring both deep neural network processing and decision tree-based techniques for classification performance enhancement (7). The models developed include :

1. Deep Neural Network (DNN) - Random Forest (RF)
2. Recurrent Neural Network (RNN) - AdaBoost

3. Convolutional Neural Network (CNN) - Long Short-Term Memory (LSTM)

The application of hybrid models in road safety is relatively rare, making this study an innovative contribution to the field. Each hybrid model was chosen for its ability to handle complex driving data, allowing for nuanced predictions of risky driving behavior. To enhance model transparency, the Local Interpretable Model-agnostic Explanations (LIME) algorithm was employed, providing insights into how the models arrived at their predictions - a significant improvement in addressing the black-box nature of many machine learning algorithms.

Deep Neural Network (DNN) - Random Forest (RF)
The DNN-RF model integrates deep learning and ensemble learning techniques, combining the flexibility and power of a Deep Neural Network with the robustness of a Random Forest classifier. This hybrid approach is relatively uncommon in road safety estimation, where most studies focus on standalone models. The combination of DNN and RF allows this model to better capture both non-linear and linear relationships within complex driving data. Hybrid deep learning models that incorporate feature extraction layers from CNNs and combine them with classical machine learning methods like SVM or RF have been shown to outperform standalone models in behavioral classification tasks (8).

• **DNN Component**: DNNs are particularly effective at learning high-dimensional, non-linear relationships, such as the interactions between speed, headway distance, and acceleration, which are critical for understanding driving safety risks. The DNN processes these inputs across multiple layers, with each layer learning progressively more abstract features. It outputs a probability distribution for the three risk levels: Normal, Dangerous, and Avoidable Accident.
• **RF Component**: The RF component provides stability and enhances generalizability by combining predictions from multiple decision trees. RF is particularly effective at handling fewer complex relationships within the data, avoiding overfitting, and providing robust predictions even when faced with noisy or incomplete data.

The stacking technique combines the outputs of the DNN and RF into a secondary Random Forest model. This secondary model learns to optimize the final predictions by balancing the strengths of both DNN's probabilistic outputs and RF's categorical predictions, making the model more accurate than if either were used alone.

Recurrent Neural Network (RNN) – AdaBoost
The combination of RNN-LSTM and AdaBoost is another innovative hybrid approach in road safety estimation, where temporal patterns in driving behavior are crucial for accurate risk classification. Hybrid models such as this, which integrate sequential learning and boosting techniques, are rare in this field but offer significant advantages in handling imbalanced datasets and complex sequences of driving data.

• **RNN-LSTM Component**: The RNN component, specifically using LSTM units, is designed to handle temporal dependencies in sequential data. In the context of driving safety, this allows the model to capture how a series of actions (e.g., gradual acceleration followed by sudden braking) can evolve into risky behavior. LSTM networks are particularly adept at retaining long-term dependencies in the data, making them ideal for detecting patterns that unfold over time.
• **AdaBoost Component**: AdaBoost is an ensemble learning algorithm that improves the performance of weak learners by focusing on difficult-to-classify instances in the dataset. This is particularly valuable in scenarios like road safety, where dangerous behaviors (e.g., Avoidable Accident) are rare but critical. By applying AdaBoost to the outputs of the LSTM, the model emphasizes the instances that are hardest to classify, improving overall accuracy and reducing the chances of misclassification in the minority classes. Hybrid models applied to sensor-based driving data significantly improve classification accuracy by identifying complex sequential patterns in real-world scenarios (9).

Convolutional Neural Network (CNN) - Long Short-Term Memory (LSTM)
The CNN-LSTM model is another hybrid approach that combines spatial and temporal deep learning techniques, an approach that is still not widely adopted in road safety estimation but holds great potential for improving the classification of driving risks. Hybrid models combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) have demonstrated significant potential in time-series analysis and classification, showing improvements in model accuracy across multiple domains, including text classification (10). The CNN handles the spatial correlations between driving factors, while the LSTM captures their temporal evolution, providing a more comprehensive analysis of driving behavior. Hybrid models that combine CNNs for spatial feature extraction and LSTMs for temporal dynamics have been effectively used in other behavioral classification domains, offering robust results in tasks like driving behavior prediction (11).

• **CNN Component**: CNNs are typically used for spatial data, such as image processing, but in this study, they are applied to driving behavior data. The CNN is responsible for detecting local patterns and relationships between variables, such as how speed, acceleration, and braking interact in short periods. This helps the model understand how individual driving actions contribute to risk.
• **LSTM Component**: Following the CNN's feature extraction, the LSTM analyzes how these spatial features evolve over time. The LSTM processes sequences of driving data, capturing the temporal dependencies that are critical for predicting risky situations. For instance, the model can recognize that sustained high speed combined with short headway distances over time is a strong indicator of dangerous driving.
• **AdaBoost Component**: As with the RNN-LSTM model, AdaBoost is used to refine the CNN-LSTM predictions by focusing on misclassified instances. This ensures that the model handles minority classes effectively, making it better at predicting rare but critical events like Avoidable Accidents.

Multi-Class Classification and Model Evaluation
Given the class imbalance inherent in real-world driving data, where dangerous driving behaviors are far less common than normal driving, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE generates synthetic samples for the minority classes (Dangerous and Avoidable Accident), ensuring that the models are trained to handle these critical events more effectively.
Before training the models, an essential Feature Selection process was conducted to identify the most significant variables for classifying driving risk levels. The selection of features was based on their correlation and importance to the classification process. Using the Random Forest Classifier from the scikit-learn library, a feature permutation technique was employed to evaluate the impact of each variable on the models' performance. This process ensured that only the most relevant features were used, improving the models' accuracy and efficiency.

As a result, the following four variables were selected for use in the classification models :

1. **GPS_distances_sum** – Total distance traveled by the vehicle.
2. **GPS_spd_mean** – Average speed of the vehicle during the trip.
3. **DEM_evt_ha_lvl_L_mean** – Mean level of harsh acceleration events recorded during the trip.
4. **DEM_evt_hb_lvl_L_mean** – Mean level of harsh braking events recorded during the trip.

Each model was evaluated using metrics such as accuracy, precision, recall, false alarm rate, and f1-score, providing a comprehensive assessment of performance, defined by **Equation 1** to **Equation 5** :

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

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

$$Recall = \frac{TP}{TP+TN}, \quad (3)$$

$$False\ Alarm\ Rate = \frac{FP}{FP+TN}, \quad (4)$$

$$f1 - score = \frac{2x(Presicion)x(Recall)}{(Precision)+(Recall)}, \quad (5)$$

where : True Positive (TP) represents the instances which belong to class i and were correctly classified in it; True Negative (TN) represents the instances which do not belong to class i and were not classified in it; False Positive (FP) represents the instances which do not belong to class i but were incorrectly classified in it; False Negative (FN) represents the instances which belong to class i but were not classified in it.

In addition to these metrics, LIME (Local Interpretable Model-agnostic Explanations) was incorporated in this study to improve the interpretability of complex machine learning models. Given that models like DNN-RF, RNN-AdaBoost, and CNN-LSTM are often seen as "black boxes," where understanding how predictions are made can be challenging, LIME offers a way to explain the influence of individual features - such as speed, headway, and acceleration - on the classification of driving behaviors.

By creating local approximations, LIME makes it possible to understand the role of specific variables in the models' decision-making processes. This interpretability is crucial for building trust in the predictions, especially in practical applications like road safety, where the reasoning behind classifications must be clear for policy-making and real-time interventions. LIME bridges the gap between advanced predictive models and the need for transparency, making the results more accessible and actionable for stakeholders.

RESULTS

The results present the performance of three machine learning models developed to classify driving behavior into the risk categories : Normal, Dangerous, and Avoidable Accident. The models - Random Forest (RF) combined with Deep Neural Network (DNN), Convolutional Neural Network (CNN) combined with Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN) combined with AdaBoost - were applied to naturalistic driving data collected in Belgium and the UK. The evaluation of these models was based on key metrics, including accuracy, precision, recall, false positive rate (FPR), and F1-score. A comparative analysis of these metrics was conducted to assess the effectiveness of each model in predicting risky driving behaviors across both datasets.

Identification of the Safety Tolerance Zone Levels

The performance of the three machine learning models is summarized in **Table 1**, which compares the accuracy, precision, recall, false positive rate (FPR), and F1-score across both datasets (Belgium and the UK). The Random Forest (RF) combined with Deep Neural Network (DNN) model consistently achieved the highest performance in both datasets. In Belgium, the RF-DNN model achieved an accuracy of 98%, with a precision of 98%, and a recall of 93%. Similarly, in the UK dataset, the RF-DNN model achieved an accuracy of 97%, precision of 98%, and recall of 92%. The low false positive rates (FPR) of 0.96% in Belgium and 1.36% in the UK demonstrate the model's effectiveness in minimizing misclassifications of less hazardous behaviors as more dangerous.

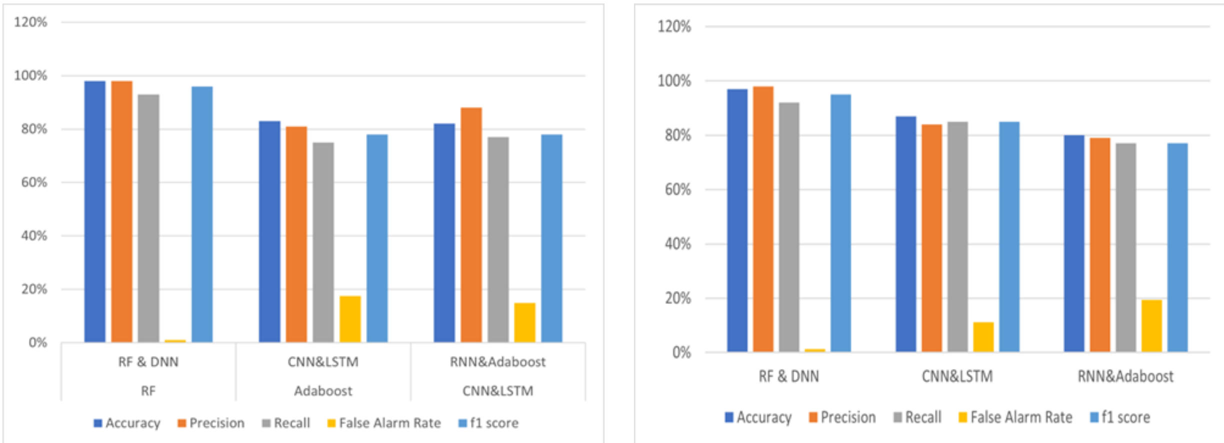
In comparison, the CNN-LSTM model demonstrated lower performance, particularly in the Belgium dataset, where its accuracy was 83%, with a recall of 75% and a higher FPR of 17.5%. This suggests that while the CNN-LSTM model is able to capture risky behaviors to some extent, it tends to misclassify more events, leading to less reliable predictions compared to the RF-DNN model. In the UK, the CNN-LSTM model showed better performance with an accuracy of 87% and recall of 85%, but the FPR remained higher at 11.11%.

The RNN-AdaBoost model also showed a reasonable performance, particularly in terms of precision, which reached 88% in the Belgium dataset. However, its overall accuracy was lower than that of the RF-DNN model, with an accuracy of 82% in Belgium and 80% in the UK. The FPR for the RNN-AdaBoost model was 14.9% in Belgium and 19.4% in the UK, indicating that it also struggled with correctly classifying some risk categories.

TABLE 1 Comparison of classification model evaluation metrics for Belgium and UK

Dataset	Model	Accuracy	Precision	Recall	FPR	f1-score
Belgium	RF & DNN	98%	98%	93%	0.96%	96%
	CNN & LSTM	83%	81%	75%	17.5%	78%
	RNN & Adaboost	82%	88%	77%	14.9%	78%
UK	RF & DNN	97%	98%	92%	1.36%	95%
	CNN & LSTM	87%	84%	85%	11.11%	85%
	RNN & Adaboost	80%	79%	77%	19.4%	77%

According to **Figure 2**, the algorithms yield high accuracy, recall, precision and f1-score which do not have a large deviation between them. A closer examination of recall underscores its importance in the context of road safety, as it reflects the model's ability to correctly identify risky driving behaviors such as dangerous or avoidable accidents. Higher recall means the model is more capable of detecting these behaviors, which is critical for preventing crashes. For instance, the RF-DNN model's recall of 93% in Belgium indicates that the vast majority of risky driving behaviors were accurately identified, ensuring its effectiveness in real-world applications where missing dangerous behavior could lead to serious consequences. On the other hand, CNN-LSTM's recall of 75% in Belgium suggests a higher likelihood of failing to detect dangerous behaviors, which reduces its practical utility in critical safety scenarios. Another key metric, the false positive rate (FPR), measures how often the model incorrectly classifies non-hazardous behaviors as hazardous. In practice, a high FPR can lead to unnecessary interventions, such as flagging safe driving behaviors as risky, which could cause unnecessary alarms and erode driver trust in the system. The RF-DNN model's FPR of 0.96% in Belgium, compared to CNN-LSTM's 17.5%, reflects the model's superior ability to avoid false alarms. This low FPR makes the RF-DNN model far more suitable for real-time applications where minimizing false positives is critical to maintaining system reliability. Precision, meanwhile, indicates how many of the model's predicted risky behaviors are actually correct. High precision ensures that the model's warnings about dangerous driving behaviors are accurate and trustworthy. Both datasets showed the RF-DNN model achieving 98% precision, meaning nearly all of its predictions were correct, thereby reducing the chance of false alarms. In contrast, the CNN-LSTM and RNN-AdaBoost models, while still performing reasonably well, demonstrated lower precision, particularly in the UK, where the RNN-AdaBoost model achieved only 79%. Finally, the F1-score balances precision and recall, offering a holistic view of model performance. The RF-DNN model's F1-score of 96% in Belgium and 95% in the UK demonstrates its strength in maintaining a balance between identifying risky behaviors and minimizing false positives. This makes it highly suitable for road safety applications where both accurate detection of risks and avoiding unnecessary warnings are crucial.



(a) (b)

Figure 2 Performance of the classification models according to the evaluation metrics (a) for Belgium and (b) for UK

Interpretability of Classification Models

To enhance the interpretability of the machine learning models, Local Interpretable Model-Agnostic Explanations (LIME) was employed. LIME was used to analyze the influence of specific driving features, such as average speed, total travel distance, harsh acceleration, and harsh braking events, on the model's predictions. As transparency in machine learning models becomes increasingly important, especially in safety-critical applications like autonomous driving, interpretability techniques such as LIME can enhance the trustworthiness of the predictions. In applications such as malware classification or driver emotion detection, interpretability tools have proven invaluable in understanding the key features driving model decisions (12). This capability is critical for stakeholder acceptance and practical deployment of AI systems in road safety. The results, as shown in **Table 2**, identify the most influential features in predicting risky driving behaviors for both Belgium and the UK.

TABLE 2 LIME results for Belgium and UK

Dataset	Feature	Value
Belgium	GPS_spd_mean	0.14
	GPS_distances_sum	0.31
	DEM_evt_ha_lvl_L_mean	2.00
UK	DEM_evt_hb_lvl_L_mean	0.56
	GPS_spd_mean	0.29
	GPS_distances_sum	0.58

For the Belgium dataset, the most significant features influencing the model's predictions were harsh acceleration (DEM_evt_ha_lvl_L_mean) and harsh braking events (DEM_evt_hb_lvl_L_mean). This suggest that aggressive driving behaviors - such as sharp speed increases and abrupt stops - play a crucial role in determining whether a driver's behavior is classified as risky. These behaviors are closely tied to real-time driving decisions and often signal heightened risk for crashes. The total travel distance also emerged as a significant predictor, although it was less dominant compared to the immediate driving behaviors (harsh acceleration and braking). The inclusion of total travel distance in the model may indicate that longer trips correlate with accumulated fatigue or prolonged exposure to risky driving conditions, both of which can elevate the risk of crashes. Although this feature played a more moderate role compared to sudden acceleration and braking, it still highlights the importance of understanding driver behavior over extended periods and how fatigue or situational factors could contribute to increased risk.

In contrast, the UK dataset revealed a different distribution of feature importance, with total travel distance (GPS_distances_sum) being the most influential factor, followed by harsh acceleration. The prominence of total travel distance in the UK dataset suggests that the length of trips plays a more critical role in predicting risky driving behaviors in this region compared to Belgium.

DISCUSSION

The current study developed hybrid classification models capable of accurately identifying dangerous driving behaviors using real-world driving data from Belgium and the UK. By employing a combination of machine learning and deep learning algorithms, the models demonstrated high precision in classifying drivers into distinct safety levels. Among the tested models, the Deep Neural Network (DNN) combined with Random Forest (RF) yielded the highest accuracy, achieving 97% in both regions. This result supports recent advancements in the application of machine learning to improve road safety, particularly for real-time driving risk prediction. For instance, Peppes et al. (2021) (13) used machine learning models to analyze vehicular data streams, showing the effectiveness of such models in identifying driving risks. However, this study goes beyond previous research by incorporating a cross-regional dataset, providing a comparative analysis between Belgium and the UK to highlight how geographical and cultural differences shape driving behaviors.

A similar approach was taken by Kwon et al. (2021), who employed a Convolutional Neural Network (CNN) combined with Long Short-Term Memory (LSTM) to classify aggressive driving behaviors, demonstrating the potential of deep learning in understanding complex driving dynamics (14). While their research focused on identifying aggressive drivers using time-series data, the present study introduces a novel hybrid modeling approach that integrates both machine learning and deep learning methods (DNN-RF). This approach not only achieved high accuracy but also enhanced model interpretability through the use of the LIME method. By classifying drivers into multiple safety levels and providing insights into specific driving behaviors that contribute to dangerous driving patterns, this study addresses a gap that has not been thoroughly explored in prior research.

A comparative analysis between Belgium and the UK revealed significant differences in the factors influencing dangerous driving behaviors. In Belgium, harsh acceleration and harsh braking emerged as the most significant predictors of risky driving, while in the UK, total trip distance and harsh acceleration played more prominent roles. These findings suggest that geographical and cultural factors influence driving behaviors. For example, the longer distances and higher speeds typical in the UK may reflect differences in road infrastructure or social norms, whereas more abrupt driving maneuvers in Belgium appear to be key risk factors. This aligns with observations from other cross-regional studies, such as the work by Shangguan et al. (2021), which emphasized the importance of tailoring safety interventions to local driving patterns (15).

The study also highlighted the importance of vehicle speed in determining crash risk, particularly in high-speed environments where reduced reaction times increase the likelihood of incidents. This finding is consistent with earlier research, such as the work by Shi et al. (2019), which demonstrated a strong correlation between speed and crash severity. Furthermore, the connection between long-distance driving and fatigue, especially in the UK, was evident in the data. Drivers engaging in longer trips were more likely to display risky behaviors, a pattern that aligns with Roshandel et al.'s (2015) research, which found that fatigue significantly impacts driver safety (16).

The inclusion of LIME in this study is a significant novelty, providing transparency into the decision-making processes of the machine learning models. One of the main challenges with deep learning models is their "black-box" nature, where the reasoning behind predictions is difficult to interpret. LIME addressed this issue by offering insights into how specific behaviors, such as harsh acceleration or braking, influenced driving safety levels. This added interpretability is critical for real-world applications, as it allows for better understanding and actionable insights into how machine learning models classify dangerous behaviors. The need for model interpretability has been highlighted in other studies, such as Liu et al. (2024), which underscores the growing need for transparency in AI applications (17).

Harsh acceleration is often linked to aggressive driving, a significant risk factor in traffic crashes, as abrupt speed increases can reduce reaction time to unforeseen obstacles or changes in road conditions, leading to a higher likelihood of crashes. Similarly, harsh braking, typically caused by driver distraction or misjudgment, poses hazards, especially in congested or high-speed conditions. These behaviors reflect diminished control over the vehicle and delayed responses to road hazards, increasing the risk of avoidable crashes. The analysis revealed regional differences: in Belgium, harsh acceleration and braking were the primary predictors of risky driving, whereas in the UK, total trip distance and harsh acceleration played more prominent roles. Longer trips in the UK may correlate with fatigue or reduced attention, contributing to risky driving behavior, while abrupt maneuvers are more prevalent in Belgium. Harsh braking had less influence in the UK compared to Belgium, suggesting differences in traffic patterns, road designs, or driving behaviors between the two regions.

The findings from this study have important implications for both policy and practical applications. From a practical perspective, the models developed could be integrated into real-time driver assistance systems (ADAS) to provide immediate feedback to drivers. Such systems could alert drivers when they engage in risky behaviors, such as harsh braking or frequent acceleration, allowing them to adjust their driving in real time. Similar approaches have been explored in previous studies, such as the work by Michelaraki et al. (2023), which demonstrated that real-time feedback through ADAS can significantly reduce crash risks by notifying drivers of dangerous behaviors as they occur (18). This study suggests that these systems could be adapted to account for regional driving patterns, such as long-distance driving in the UK or abrupt maneuvers in Belgium, further enhancing their effectiveness.

While the models showed strong performance, the study has some limitations that should be considered. One major challenge was the inherent imbalance in the dataset, with fewer samples representing dangerous driving events. Although techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) were employed to address this issue, the limited representation of extreme behaviors, such as severe crashes or near misses, remains a constraint. Other studies, such as Zhu et al. (2022), have also noted the difficulty of modeling rare but critical events in road safety research (19). Additionally, the absence of demographic and psychological data (e.g., age, gender, risk aversion) limited the scope of the analysis. Including such variables in future research could provide a more comprehensive understanding of how individual characteristics influence dangerous driving behavior (20).

Finally, expanding the geographical scope of this research to include additional regions with varying road infrastructures and traffic laws could improve the generalizability of the findings. Future studies could also explore more complex models that account for factors like urban versus rural driving, weather conditions, and time of day. Furthermore, integrating physiological data, such as heart rate or eye movements, could provide deeper insights into the impact of fatigue and stress on driving behavior (21).

In conclusion, this study offers a novel approach to classifying dangerous driving behaviors using hybrid machine learning models, integrating both deep learning and machine learning techniques. The application of the LIME algorithm adds a critical layer of interpretability, making the models more actionable and transparent. These findings have practical implications for the development of Advanced Driver Assistance Systems (ADAS) and regional safety campaigns while identifying future research directions that could further refine and enhance the current models.

CONCLUSIONS

This research developed highly effective models for classifying dangerous driving behaviors, grounded in real-world data gathered from drivers in Belgium and the UK. By employing a combination of advanced machine learning and deep learning methodologies, the models achieved impressive accuracy in identifying risky driving patterns. A key outcome of the study is the identification of specific driving behaviors—particularly harsh acceleration, harsh braking, and total driving distance—as significant indicators of driver safety. The comparative analysis between Belgium and the UK also revealed notable differences in driving habits, with speed and long-distance travel being more critical in the UK, while sudden maneuvers were more prevalent in Belgium.

The hybrid modeling approach, combining Deep Neural Networks (DNN) and Random Forest (RF), was particularly successful, consistently delivering high accuracy across both countries. The study further highlighted that, while harsh acceleration and braking were dominant risk factors in Belgium, in the UK, longer travel distances and higher speeds posed greater risks. These insights point to the need for customized safety interventions that reflect the distinct driving behaviors of different regions.

A major innovation in this research is the use of the Lime algorithm to interpret the models' decision-making processes. The ability to transparently understand how individual driving factors influence safety assessments provides crucial advantages, particularly in addressing concerns about the opaque nature of machine learning models. This transparency ensures that insights from the models are not only accurate but also actionable for stakeholders looking to enhance road safety policies and systems.

Looking ahead, there are several ways in which this research could be expanded. Increasing the size of the dataset would further enhance the reliability of the models, especially when predicting rare and critical events like severe crashes or near misses. Incorporating additional factors such as driver demographics and psychological characteristics (e.g., fatigue levels and risk tolerance) would allow for more personalized risk assessments. The inclusion of diverse driving conditions—such as urban versus rural environments and variable weather—could also improve the comprehensiveness of the models.

Applying these models in real-time settings, such as within Advanced Driver Assistance Systems (ADAS), represents a promising avenue for future research. Real-time feedback systems could offer immediate safety warnings to drivers, potentially preventing crashes before they occur. Additionally, broadening the geographic scope of future studies would help validate the models' applicability across regions with different road systems and driving cultures, offering deeper insights into how regional policies and infrastructure impact driving behaviors.

In conclusion, this study marks a significant step forward in the application of machine learning for improving road safety. By leveraging predictive modeling and real-time systems, the research opens up new possibilities for reducing crash risks and promoting safer driving practices. Its focus on interpretability, regional differences, and practical applications sets a solid foundation for future developments in the field of driver safety.

ACKNOWLEDGMENTS

The research was funded by the EU H2020 i-DREAMS project (Project Number: 814761) funded by European Commission under the MG-2-1-2018 Research and Innovation Action (RIA).

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: E.M. Theodoraki, T. Garefalakis; E. Michelaraki, G. Yannis; data collection: E.M. Theodoraki, T. Garefalakis; analysis and interpretation of results: E.M. Theodoraki, T. Garefalakis, E. Michelaraki; draft manuscript preparation: E.M. Theodoraki, T. Garefalakis, E. Michelaraki, G. Yannis. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. World Health Organization. Global Status Report on Road Safety 2023. World Health Organization. https://www.who.int/publications/i/item/9789240086517. Accessed Mar. 6, 2024.

2. Khashayarfard, M., and H. Nassiri. Studying the Simultaneous Effect of Autonomous Vehicles and Distracted Driving on Safety at Unsignalized Intersections. Journal of Advanced Transportation, Vol. 2021, 2021, pp. 1–16. https://doi.org/10.1155/2021/6677010.

3. Mueller, A. S., J. B. Cicchino, and D. S. Zuby. What Humanlike Errors Do Autonomous Vehicles Need to Avoid to Maximize Safety? Journal of Safety Research, Vol. 75, 2020, pp. 310–318. https://doi.org/10.1016/j.jsr.2020.10.005.

4. Wang, J., Y. Ma, X. Yang, T. Li, and H. Wei. Short-Term Traffic Prediction Considering Spatial-Temporal Characteristics of Freeway Flow. Journal of Advanced Transportation, Vol. 2021, 2021. https://doi.org/10.1155/2021/5815280.

5. Shi, X., Y. D. Wong, M. Z.-F. Li, C. Palanisamy, and C. Chai. A Feature Learning Approach Based on XGBoost for Driving Assessment and Risk Prediction. Accident Analysis & Prevention, Vol. 129, 2019, pp. 170–179. https://doi.org/10.1016/j.aap.2019.05.005.

6. Yang, K., C. Al Haddad, G. Yannis, and C. Antoniou. Driving Behavior Safety Levels: Classification and Evaluation. 2021.

7. Jaouedi, N., N. Boujnah, and M. S. Bouhlel. A New Hybrid Deep Learning Model for Human Action Recognition. Journal of King Saud University - Computer and Information Sciences, Vol. 32, No. 4, 2020, pp. 447–453. https://doi.org/10.1016/j.jksuci.2019.09.004.

8. Ahmad, H., M. U. Asghar, M. Z. Asghar, A. Khan, and A. H. Mosavi. A Hybrid Deep Learning Technique for Personality Trait Classification From Text. IEEE Access, Vol. 9, 2021, pp. 146214–146232. https://doi.org/10.1109/ACCESS.2021.3121791.

9. Savelonas, M., I. Vernikos, D. Mantzekis, E. Spyrou, A. Tsakiri, and S. Karkanis. Hybrid Representation of Sensor Data for the Classification of Driving Behaviour. Applied Sciences, Vol. 11, No. 18, 2021, p. 8574. https://doi.org/10.3390/app11188574.

10. Lee, S.-H. Text Classification of Mixed Model Based on Deep Learning. Tehnički glasnik, Vol. 17, No. 3, 2023, pp. 367–374. https://doi.org/10.31803/tg-20221228180808.

11. Singh, B., and R. Jaiswal. Impact of Hybridization of Deep Learning Models for Temporal Data Learning. 2021.

12. Sukhavasi, S. B., S. B. Sukhavasi, K. Elleithy, A. El-Sayed, and A. Elleithy. A Hybrid Model for Driver Emotion Detection Using Feature Fusion Approach. International Journal of Environmental Research and Public Health, Vol. 19, No. 5, 2022, p. 3085. https://doi.org/10.3390/ijerph19053085.

13. Peppes, N., T. Alexakis, E. Adamopoulou, and K. Demestichas. Driving Behaviour Analysis Using Machine and Deep Learning Methods for Continuous Streams of Vehicular Data. Sensors, Vol. 21, No. 14, 2021. https://doi.org/10.3390/s21144704.

14. Kwon, S. K., J. H. Seo, J. Y. Yun, and K.-D. Kim. Driving Behavior Classification and Sharing System Using CNN-LSTM Approaches and V2X Communication. Applied Sciences, Vol. 11, No. 21, 2021, p. 10420. https://doi.org/10.3390/app112110420.

15. Shangguan, Q., T. Fu, J. Wang, T. Luo, and S. Fang. An Integrated Methodology for Real-Time Driving Risk Status Prediction Using Naturalistic Driving Data. Accident Analysis & Prevention, Vol. 156, 2021, p. 106122. https://doi.org/10.1016/j.aap.2021.106122.

16. Roshandel, S., Z. Zheng, and S. Washington. Impact of Real-Time Traffic Characteristics on Freeway Crash Occurrence: Systematic Review and Meta-Analysis. Accident Analysis & Prevention, Vol. 79, 2015, pp. 198–211. https://doi.org/10.1016/j.aap.2015.03.013.

17. Liu, H., T. Wang, W. Li, X. Ye, and Q. Yuan. Lane-Change Intention Recognition Considering Oncoming Traffic: Novel Insights Revealed by Advances in Deep Learning. Accident Analysis & Prevention, Vol. 198, 2024, p. 107476. https://doi.org/10.1016/j.aap.2024.107476.

18. Michelaraki, E., M. Kallidoni, C. Katrakazas, T. Brijs, and G. Yannis. How to Define a Safety Tolerance Zone for Speed? Insights from the i-DREAMS Project. Transportation Research Procedia, Vol. 72, 2023, pp. 415–422. https://doi.org/10.1016/j.trpro.2023.11.422.

19. Zhu, S., C. Li, K. Fang, Y. Peng, Y. Jiang, and Y. Zou. An Optimized Algorithm for Dangerous Driving Behavior Identification Based on Unbalanced Data. Electronics, Vol. 11, No. 10, 2022, p. 1557. https://doi.org/10.3390/electronics11101557.

20. Song, X., Y. Yin, H. Cao, S. Zhao, M. Li, and B. Yi. The Mediating Effect of Driver Characteristics on Risky Driving Behaviors Moderated by Gender, and the Classification Model of Driver's Driving Risk. Accident Analysis & Prevention, Vol. 153, 2021, p. 106038. https://doi.org/10.1016/j.aap.2021.106038.

21. Michelaraki, E., C. Katrakazas, T. Brijs, and G. Yannis. Modelling the Safety Tolerance Zone: Recommendations from the i-DREAMS Project. 2021.