ORIGINAL PAPER

Unfolding the dynamics of driving behavior: a machine learning analysis from Germany and Belgium

Stella Roussou[1*] , Eva Michelaraki[1], Christos Katrakazas[1], Amir Pooyan Afghari[2], Christelle Al Haddad[3], Md Rakibul Alam[3], Constantinos Antoniou[3], Eleonora Papadimitriou[2], Tom Brijs[4] and George Yannis[1]

Abstract

The i-DREAMS project focuses on establishing a framework known as the 'Safety Tolerance Zone (STZ)' to ensure drivers operate within safe boundaries. This study compares Long-Short-Term-Memory Networks and shallow Neural Networks to assess participants' safety levels during i-DREAMS on-road trials. Thirty German drivers' trips and Forty-Three Belgian drivers were analyzed using these methods, revealing factors contributing to risky behavior. Results indicate i-DREAMS interventions significantly enhance driving behavior, with Neural Networks displaying superior performance among the algorithms considered.

Keywords On-road field trials, Driving behavior, Long-short-term-memory network (LSTM), Neural network, Machine learning

1 Introduction

In the modern era, road safety stands as a critical global concern, given the alarming statistics of approximately 1.3 million annual fatalities resulting from road accidents, along with millions more enduring non-fatal injuries [20]. This persistent challenge has spurred significant research efforts and technological innovations aimed at curbing these alarming figures and creating safer road environments. In response to the alarming global road safety crisis, extensive research efforts and technological advancements have been dedicated to creating safer road environments.

Intelligent Transportation Systems (ITS) have emerged as a transformative force, utilizing advanced technologies to enhance road safety significantly. ITS have revolutionized road safety measures by harnessing the power of real-time data analysis and advanced technologies. These systems, encompassing various innovative solutions such as smart traffic management, predictive analytics, and vehicle-to-vehicle communication, have drastically reduced accidents and improved overall traffic management [11]. By enabling proactive measures and optimizing traffic flow, ITS has played a pivotal role in minimizing risks on the roads, thus making significant strides toward achieving global road safety goals.

The integration of sophisticated machine learning techniques, such as Long Short-Term Memory (LSTM) networks and Neural Networks (NN), has further propelled road safety efforts. LSTM networks, known for their ability to comprehend complex temporal patterns, and Neural Networks, with their deep learning capabilities, have been instrumental in predicting and preventing

*Correspondence:
Stella Roussou
s_roussou@mail.ntua.gr
[1] Department of Transportation Planning and Engineering, National Technical University of Athens, 5 Heroon Polytechniou str, Athens GR-15773, Greece
[2] Safety and Security Science Group, Delft University of Technology, Jaffalaan 5, Delft 2628 BX, The Netherlands
[3] Chair of Transportation Systems Engineering, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany
[4] School for Transportation Sciences, UHasselt, Transportation Research Institute (IMOB), Agoralaan, Diepenbeek 3590, Belgium

Roussou *et al. European Transport Research Review*   (2024) 16:40

Page 2 of 13

road accidents. Researchers have harnessed LSTM networks to analyse sequential driving data, accurately anticipating hazardous driving behaviours and enabling timely interventions [5]. Additionally, Neural Networks, through their capacity to process vast datasets, have been deployed to identify intricate patterns within driving behaviour data, enhancing the precision of safety predictions [21].

By leveraging the predictive power of LSTM networks and Neural Networks, road safety predictions have reached unprecedented accuracy levels. The fusion of these advanced technologies with real-time data from Intelligent Transportation Systems holds the promise of creating comprehensive, proactive safety measures, ultimately making our roads safer for everyone [6, 21].

The i-DREAMS project, funded by the European Commission Horizon 2020 initiative, strives to establish a 'Safety Tolerance Zone' (STZ) ensuring safe driving behaviour. Through continuous monitoring of risk factors related to task complexity (e.g., traffic and weather) and coping capacity (e.g., driver's mental state and vehicle status), i-DREAMS aims to determine appropriate STZ levels and interventions, maintaining drivers within safe limits. The STZ comprises 'Normal', 'Dangerous', and 'Avoidable Accident' levels. 'Normal' indicates a low likelihood of a crash, 'Dangerous' implies an increased possibility, and 'Avoidable Accident' suggests a high probability with time for preventive action.

Building upon i-DREAMS principles, this paper compares machine-learning techniques (LSTM and Neural Network) to identify risky driving behaviour levels. Data from 30 German drivers and 43 Belgian drivers were analysed, and the models were developed based on these principles and objectives. The paper is organized into sections exploring the project's introduction, the methodology including the aim of the study, the characteristics of the experiment, the machine learning techniques used on driving behaviour analysis and their results. Conclusions drawn shed light on the NN's and LSTM's models ability to predict risky driving behaviour and risky driving instances.

2 Background

Naturalistic driving studies (NDS) have been widely utilized in recent years to examine unsafe driving behavior [13]. There are certain traffic, driver, vehicle, and environmental factors that affect the risk of driving [16]. Furthermore, recent studies focus on identifying driving behaviors and categorizing them as risky or safe in order to improve road safety [14]. Researchers have utilized models to evaluate unsafe driving behavior based

on the driver's state [5] and specific features of the driver, such as demographics [18], in a more anthropocentric approach.

Other studies [15, 16, 21] have proposed models for identifying unsafe driving based on characteristics related to driving behavior, such as speed, time to collision, and time to headway. Overtaking behaviour of motorized vehicles by measuring the lateral distance between the bike and passing vehicle and a statistical model was developed to predict the probability of an unsafe critical maneuver and cyclists' safety perception [22].

Furthermore, the continuous development of Intelligent Transportation Systems (ITS) as well as the increasing availability of real-time data streams from in-vehicle sensors, GPS systems, and mobile devices has opened new opportunities for the application of machine learning models in real-time risk prediction and Advanced Driver Assistance Systems (ADAS). By continuously analyzing sensor data and contextual information, these models can provide timely alerts and warnings to drivers, assist in making safer driving decisions, and contribute to the prevention of crashes.

Machine learning is emerging as a powerful tool also in the field of road safety and it has become crucial to analyze the complex and heterogeneous data that are today available from new technologies [23, 24]. In recent years, classification models have been widely used to identify risky driving behavior. Several studies have explored the application of ML and DL techniques for classifying risky driving behaviors. For example [1], developed an LSTM-based model to identify driving behavior using sensor data, based on three levels of driving behavior (i.e., normal, drowsy, or aggressive) defined by the authors.

However, while ML and DL techniques show promise in classifying risky driving behaviors, several challenges persist, including data collection, preprocessing, feature selection, model generalizability, and interpretability of learned representations. Overcoming these challenges is pivotal to ensuring the reliability and applicability of ML and DL models in real-world driving scenarios.

Despite numerous research endeavours on analysing driver behaviour through ma-chine learning algorithms, there are currently no comparable studies in this specific domain that investigate both machine learning (ML) and deep learning (DL) algorithms [12].In summary, ML and DL models represent powerful tools for comprehending, forecasting, and mitigating risky driving behavior. Their advanced algorithms, coupled with extensive data, hold transformative potential for road safety efforts.

Given the context of the paper, exploring the application and effectiveness of Neural Networks (NN) and Long Short-Term Memory (LSTM) models

within the framework of these challenges and opportunities can provide a nuanced understanding of their impact on driving behaviors in different cultural contexts. Continued research and collaboration in this field are imperative to fully harness the benefits of advanced algorithms in enhancing driving safety, particularly in the unique driving contexts of Germany and Belgium.

3 Data description

Within the i-DREAMS project, a naturalistic driving experiment was carried out involving 30 drivers from Germany and a large database of 5,344 trips and 84,434 min was created. As for the Belgian drivers the database consisted of a varied number of drivers across the different phases of the experiment, with 39 drivers remaining consistent throughout the phases, 7163 trips and 147,337 min. The on-road trial experiment was carried out in four phases:

> Phase 1: monitoring − 30 German car drivers, 1,397 trips (23,617 min) and 39 Belgian car drivers, 1,173 trips (23,725 min).
> Phase 2: real-time interventions − 30 German car drivers, 1,322 trips (19,469 min) and 43 Belgian car drivers, 1,549 trips (31,414 min).
> Phase 3: real-time & post-trip interventions − 30 German car drivers, 1,129 trips (17,704 min) and 51 Belgian car drivers, 1,973 trips (40,121 min).
> Phase 4: real-time. post-trip interventions & gamification − 30 German car drivers, 1,496 trips (23,644 min) and 49 Belgian car drivers, 2,468 trips (52,077 min).

The on-road experiment was planned and executed based on established principles found in the relevant literature, emphasizing the evaluation of interventions aimed at aiding drivers in adhering to safe driving practices [7, 9, 17]. The experiment consisted of four distinct phases. The Phase 1, represents the monitoring phase, where no interventions were implemented and lasted 4 weeks. Phase 2 refers to the in-vehicle interventions, by providing real-time warnings using adaptive ADAS and had a duration of 4 weeks. Following, in Phase 3, which lasted 4 weeks, drivers received feedback on their driving performance through the app and in Phase 4 the drivers received feedback likewise in Phase 3, but additionally at the same time gamification elements were also active. Phase 4 lasted 6 weeks. All four phases concentrate on observing driving behaviour and assessing the influence of real-time interventions like in-vehicle warnings, as well as interventions after the trip such as post-trip feedback and gamification, on driving behaviour. A sample of the dataset used in this study is provided at the end in the section of Appendix A.

4 Methodological overview

4.1 Neural networks (NNS)

An Artificial Neural Network (ANN) is a highly complex and powerful computational model capable of capturing non-linear relationships in data. It operates as a parallel processor, simulating the behaviour of neurons in the human brain. A multi-layer perceptron ANN, which is commonly used for classification tasks, consists of three types of layers: an input layer, an output layer, and one or more hidden layers. The input layer serves as the entry point for the network, receiving the values of the explanatory variables, which represent the input data. These variables could be various features extracted from the driving dataset, such as vehicle speed, acceleration, and headway.

The hidden layer, composed of a varying number of neurons, performs calculations by summing the weighted inputs from the explanatory variables. Each neuron in the hidden layer applies an activation function, which introduces non-linearity to the model. This non-linearity is crucial for capturing complex association patterns and representing the intricate relationships between the input features and the target variable, such as different levels of risky driving behavior. The number of neurons in the hidden layer is typically determined through experimentation, as it can impact the model's capacity to learn and generalize from the data. In some crash analysis applications, a single hidden layer is often sufficient, while more complex problems may require multiple hidden layers.

Moving to the output layer, this layer aggregates the values from the hidden neurons and produces the network's final output. In the context of risky driving behavior classification, the output layer would have a number of neurons corresponding to the different classes or levels of risk. The activation function applied in the output layer depends on the nature of the problem. For instance, in a multi-class classification scenario, the softmax activation function is commonly used to calculate the probabilities of each class. These probabilities are then used to determine the predicted class or level of risky driving behavior.

The design and architecture of the neural network, including the number of layers, neurons, and activation functions, are essential considerations in achieving accurate and effective classification of risky driving behavior. Previous studies [4, 14] have explored the application of multi-layer perceptron ANNs in similar contexts, highlighting the network's ability to capture complex patterns and associations in driving data.

Roussou *et al. European Transport Research Review*   (2024) 16:40

Page 4 of 13

4.2 Long short-term memory (LSTM) networks

Long Short-Term Memory Models (LSTMs) are a specialized form of Recurrent Neural Networks (RNNs) renowned for their ability to capture long-range dependencies [5]. LSTMs have gained widespread adoption and exhibit exceptional performance in various problem domains. The LSTM-CNN deep learning algorithm is predominantly employed for detecting abnormal driving behaviour in drivers, boasting superior recognition accuracy [8]. . Cura A. et al. (2021) [3] conducted a study where they developed LSTM and CNN-based neural network models to classify and evaluate bus driver behaviour, focusing on aspects such as deceleration, engine speed pedaling, corner turns, and lane change attempts. The CNN architecture demonstrated superior performance in identifying aggressive driving compared to the LSTM network for behavioural modeling, providing valuable insights for advancing models in this domain.

Unlike conventional RNNs, LSTMs are explicitly designed to address the challenge of long-term dependencies. They possess an inherent capacity to retain information over extended periods, making them particularly suitable for tasks involving sequential data modeling.

An LSTM comprises a series of repeating modules that form a chain-like architecture [10]. Within these hidden units, LSTMs utilize memory blocks to capture long-term dependencies present in the data. This characteristic has demonstrated remarkable effectiveness in diverse time-series tasks, including activity recognition, video captioning, and language translation.

The core component of an LSTM is the cell state, also known as the memory block. It contains one or multiple memory cells that are regulated by structures called gates. These gates control the flow of information into and out of the memory block, enabling the LSTM to selectively retain relevant sequential information and discard irrelevant information. The gates are formed using a combination of sigmoid activation functions and element-wise multiplications, allowing precise control over information flow throughout the network.

Typically, an LSTM consists of three fundamental gates:

1. Forget gate: The forget gate determines which information to retain or discard from the cell state. It employs a sigmoid layer, referred to as the "forget gate layer," to make this decision.
2. Input gate: The input gate determines which new information to incorporate into the cell state and how to update it. It comprises two components: an input gate layer, implemented using a sigmoid layer, that determines which values to update, and a hyperbolic tangent (tanh) layer that generates a vector of candidate values to potentially integrate into the state. The old cell state is then updated to the new cell state based on these components.
3. Output gate: The output gate filters and determines which information to produce as the output from a memory block at a specific time step. The output is derived from the cell state but undergoes filtering. An output gate, consisting of a sigmoid layer, determines the relevant portions of the cell state to output. The filtered cell state then passes through a tanh activation function to scale the values between -1 and 1. Finally, the result is multiplied by the output of the sigmoid gate, generating the desired output.

4.3 Performance metrics

For the classification models, the confusion matrix and several performance metrics were utilized to evaluate the model's performance. The confusion matrix provides True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) metrics. True Positive (TP) represents cases where the model correctly identified a real positive. On the other hand, False Positive (FP) indicates cases, where the model detected an anomaly that did not actually exist as Real Positive, and False Negative (FN), means there was an actual "True" that the model failed to detect as an anomaly [23, 24]. From the confusion matrix, the following well-established error metrics were calculated:

Accuracy, which measures the proportion of correctly classified observations, is defined as:

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

Precision, which quantifies the number of positive class predictions that actually belong to the positive class, is defined as follows:

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

Recall, also known as True Positive Rate, which measures the proportion of actual positive cases correctly identified by the model, is defined as follows:

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

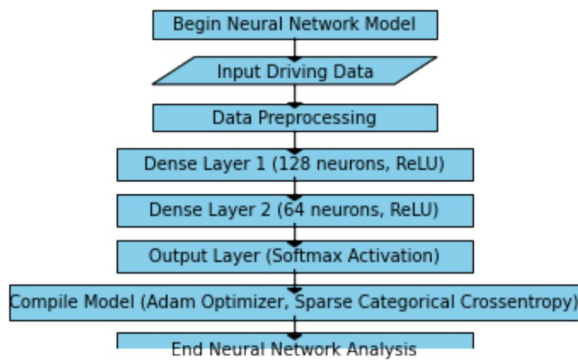F1score, which combines precision and recall into a single measure, is defined as follows:

Fig. 1 High-level algorithm description of the neural network model

Fig. 2 High-level algorithm description of the long short-term memory model

$$f1 - score = \frac{2x(Precision)x(Recall)}{(Precision) + (Recall)} \qquad (4)$$

False alarm rate, which measures the proportion of negative cases incorrectly classified as positive is defined as follows:

$$False\ Alarm\ Rate = \frac{FP}{FP + TN} \qquad (5)$$

4.4 Methodology analysis

The neural network model is structured as a multi-layer architecture, where each layer is responsible for extracting and learning different levels of patterns from the input driving data. The initial layer processes basic features, which are then passed on to subsequent layers for more complex pattern recognition. The training process involves adjusting the model weights based on historical driving data, enhancing its ability to differentiate between safe and risky driving behaviours. This model's performance is later validated using a separate set of data to determine its predictive accuracy. The following high-level description demonstrated in Fig. 1 below presents the methodology structure.

The LSTM model was structured upon the same methodology, as presented below in Fig. 2.

A snippet of the code for the neural network model and long short-term memory model has been provided in Appendix A. It is important to add that the NN model architecture consists of two dense (fully connected) layers. The first layer has 128 units, and the second has 64 units, all using the ReLU activation function. The output layer is configured with the appropriate number of units based on the classes in the target variable, utilising the softmax activation function. The
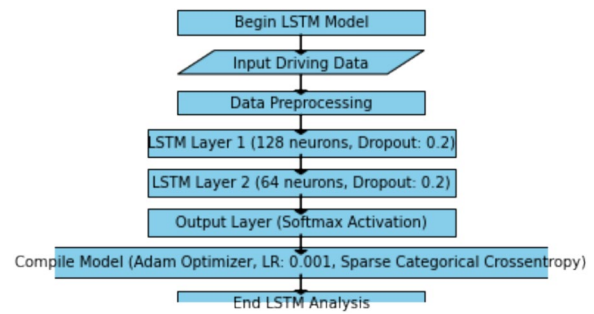
model is optimised using the Adam optimizer with a sparse categorical cross-entropy loss function. During training, the model is optimised for 100 epochs with a batch size of 32, and 10% of the training data are reserved for validation.

The hyperparameter values in the LSTM model were chosen based on established practices for working with similar datasets. The LSTM model is designed for sequential data and is composed of two LSTM layers with 128 and 64 units, respectively. Both LSTM layers use ReLU activation. The ReLU activation function was chosen for its ability to capture complex relationships. A dropout rate of 0.2 and a recurrent dropout of 0.2 were implemented to prevent overfitting. The output layer, employing softmax activation, is adapted based on the number of classes. The model is compiled with the Adam optimizer, a learning rate of 0.001, and sparse categorical cross-entropy loss. Training occurs over 100 epochs with a batch size of 64, and 10% of the training data are set aside for validation.

5 Results
5.1 Neural networks (NNs) for heading and speeding
5.1.1 German car drivers
The utilization of the Neural Networks (NNs) classification algorithms in this study serves as a valuable preparatory phase for the subsequent LSTM classification. Two feed-forward multi-layer perceptron models were applied to a subset of the German car drivers' dataset. This subset consisted of data from 30 drivers and 5,340 trips. The high accuracy achieved, exceeding 94%, highlights their effectiveness in real-time prediction of the STZ. This outcome supports the notion that real-time STZ prediction is indeed feasible. Furthermore, the low false alarm rate, with a maximum of 6%, demonstrates the models' ability to minimize incorrect predictions and reduce unnecessary alerts. The successful application of these neural

Table 1 Confusion data matrix for headway and speeding

Variable	TP	FP	FN	TN	Sum
Headway	33,378	0	1400	82	34,860
Speeding	2178	1987	63	30,632	34,860

Table 2 Assessment of classification model for headway and speeding for German car drivers

Variable	Accuracy	Precision	Recall	f1-score	False Alarm Rate
Headway	95.98%	100.00%	95.97%	97.95%	0.00%
Speeding	94.12%	52.29%	97.19%	68.00%	6.11%

network models paves the way for the implementation of LSTM-based approaches, which can leverage the temporal nature of the data to potentially enhance the precision and reliability of the STZ prediction. The upcoming subsection will delve further into the LSTM classification, building upon the foundations established by the neural network models.

After the application of the models, the identified confusion matrix was produced for the two independent variables (i.e. headway and speeding), as shown in Table 1.

From the confusion matrix, the following metrics were estimated and are depicted in Table 2.

In Fig. 3, the plot titled "Model Loss" displays the progression of the model's loss during training and validation phases across multiple epochs. The *x*-axis represents the number of training epochs, while the *y*-axis represents the corresponding loss values. The blue line represents the model's training loss at each epoch. Training loss measures how well the model is performing on the training data. As the model learns from the training data, the goal is to minimise this loss, indicating improved predictive performance. The orange line represents the validation loss at each epoch. Validation loss measures how well the model generalises to unseen data not used during training. It helps to identify if the model is overfitting (performing well on training data but poorly on new data) or underfitting (not capturing the underlying patterns).

The results shown in are in line with relevant literature on real-time safety evaluations [1], as well as previous project analyses utilized on simulator data [4]. Precision and f1-score metrics are probably lower due to the greater amount of 'normal' STZ level instances as compared with 'dangerous' conditions.

5.1.2 Belgium car drivers

The results from the Belgium car drivers dataset as demonstrated in the Tables 3 and 4 indicate that the models performed well, especially in the case of speeding prediction, where it achieved high accuracy and recall. For headway prediction, while the accuracy is slightly lower than that of speeding, the precision and recall values are balanced, indicating a good ability to identify true positive cases without many false positives or false negatives.

These findings suggest that the models are effective in classifying instances of headway and speeding, with the speeding model showing particularly strong performance in identifying positive cases. The absence of false positives in both cases (FP = 0) is a notable achievement, signifying a low rate of incorrectly identified positive cases.
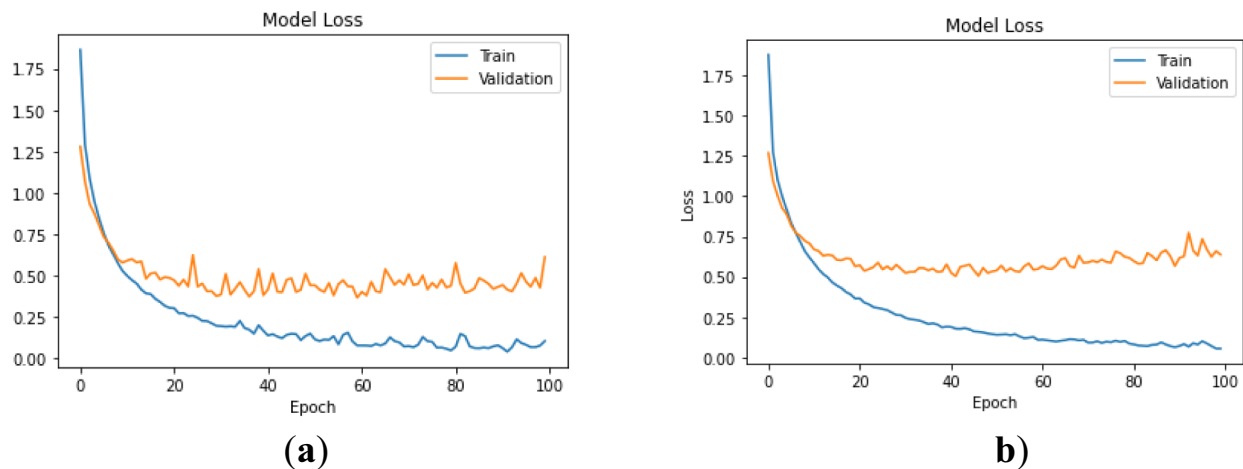

(a)

b)

Fig. 3 Model loss of the neural network of German car drivers for headway (**a**) and speeding (**b**)

Table 3 Confusion data matrix for headway and speeding

Variable	TP	FP	FN	TN	Sum
Headway	37,517	0	80	7915	45,512
Speeding	30,069	0	0	6193	36,462

Table 4 Assessment of classification model for headway and speeding for Belgian car drivers

Variable	Accuracy	Precision	Recall	f1-score	False Alarm Rate
Headway	77.19%	77.64%	77.19%	76.90%	22.81%
Speeding	83.51%	80.71%	83.51%	79.78%	16.49%

A descending trend in both training and validation loss is achieved in the neural network of Belgium car drivers for headway (a) and speeding (b) as demonstrated in Fig. 4 below.

The analysis of the results reveals a commendable performance by the models, especially in predicting instances of speeding. The model not only demonstrated high accuracy, indicating the overall correctness of its predictions, but also exhibited a high recall rate. This high recall implies that the model successfully identified the majority of actual positive cases of speeding. Similarly, the headway prediction, while having a slightly lower accuracy compared to speeding, maintained a balance between precision and recall. This equilibrium signifies the model's ability to accurately pinpoint true positive cases without generating excessive false positives or missing actual positive instances.

5.2 Long short-term memory (LSTM) for heading and speeding of level 0 and 1

5.2.1 German car drivers

Building upon the foundations laid by the previously mentioned neural network models, the subsequent subsection focuses on the application of Long Short-Term Memory (LSTM) classification for real-time prediction of the Steering Torque Zone (STZ). The LSTM approach capitalizes on the temporal nature of the data to potentially enhance the precision and reliability of the STZ prediction. The LSTM models were trained and evaluated using a subset of the German car drivers' dataset, consisting of data from 30 drivers and 5,340 trips.

The LSTM models, while showing a lower level of accuracy and precision compared to the previous neural network models, still exhibit a fair level of performance in predicting headway and speeding incidents.

For headway prediction, the model accurately identifies approximately 45.57% of instances, which is a significant improvement from random chance. The precision of 42.13% indicates that when the model predicts a positive case, it is correct 42.13% of the time. The recall of 45.57% implies that the model captures 45.57% of all actual positive cases. The F1-score of 41.11% signifies a balanced measure of precision and recall.

In the case of speeding prediction, the model performs slightly better with an accuracy of 53.14%. The precision of 49.54% indicates that nearly half of the positive predictions made by the model are accurate. The recall of 53.14% shows that the model captures 53.14% of all actual speeding cases. The F1-score of 50.81% indicates a balanced trade-off between precision and recall.
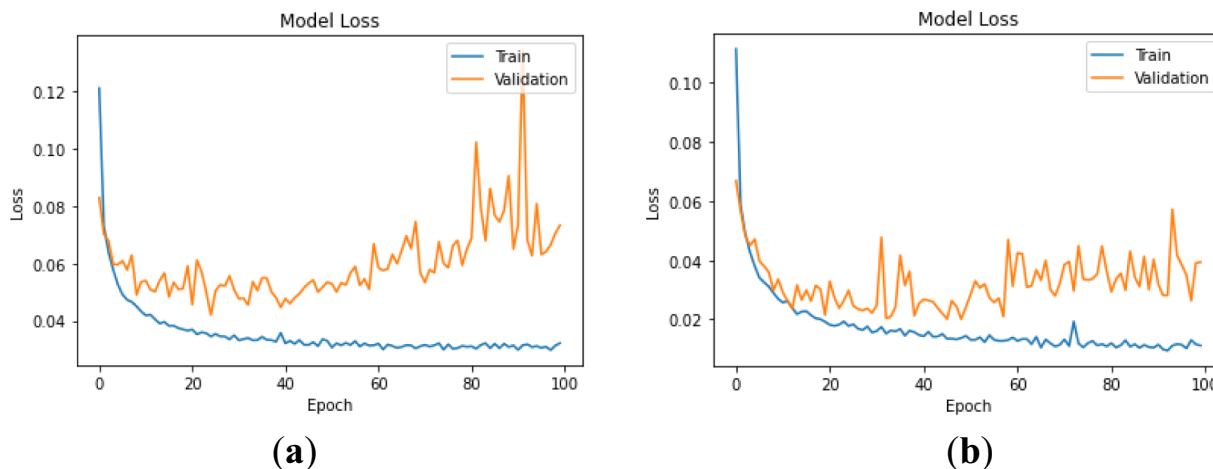

(a)   (b)

Fig. 4 Model loss of the neural network of Belgian car drivers for headway (**a**) and speeding (**b**)

Table 5 Assessment of classification model for headway and speeding for German car drivers

Variable	Accuracy	Precision	Recall	f1-score	False Alarm Rate
Headway	45.57%	42.13%	45.57%	41.11%	54.43%
Speeding	53.14%	49.54%	53.14%	50.81%	46.86%

Compared to the previous neural network models, these LSTM models show a lower level of accuracy and precision. However, it's crucial to note that LSTMs are particularly valuable in capturing sequential patterns and temporal dependencies in data. Despite the decrease in accuracy and precision, the LSTM models might excel in capturing nuanced patterns in the data, especially temporal ones, which could lead to more accurate predictions in specific contexts or time-dependent scenarios (Table 5).

A descending trend in both training and validation loss is achieved in the LSTM of German car drivers for headway (a) and speeding (b) as demonstrated in Fig. 5 below. The Divergence or this significant gap between training and validation loss in Fig. 5a might indicate overfitting (high training performance but poor generalisation).

5.2.2 Belgium car drivers

It is important to consider that an accuracy below 60% may not be satisfactory for a high-performance intervention system, as it could result in a relatively high number of false alarms or missed detections. However, the required level of accuracy depends on the specific use case and the associated risks. For instance, in a system aimed at detecting potential crashes or safety hazards, a higher level of accuracy may be necessary to ensure the safety of drivers and other road users.

The LSTM models for Belgium show moderate performance in predicting headway and speeding incidents. For headway prediction, the model achieves an accuracy of 58.12%, indicating it correctly classifies approximately 58.12% of the instances. The precision of 35.65% suggests that when the model predicts a positive case, it is correct 35.65% of the time. The recall of 58.12% signifies that the model captures 58.12% of all actual positive headway cases. The F1-score of 37.33% reflects a balance between precision and recall.

In the case of speeding prediction, the model performs slightly lower with an accuracy of 48.27%. The precision of 25.75% indicates that only a quarter of the positive predictions made by the model are accurate. The recall of 48.27% shows that the model captures 48.27% of all actual speeding cases. The F1-score of 32.59% indicates a trade-off between precision and recall.

The LSTM models in Belgium exhibit moderate performance, especially in identifying headway incidents. While they demonstrate a capacity to capture positive cases, there is room for improvement, particularly in reducing false positives and enhancing precision. Further refinements in model architecture, feature selection, or additional data preprocessing techniques might be necessary to enhance the accuracy and reliability of the LSTM models for both headway and speeding predictions (Table 6).

A descending trend in both training and validation loss is achieved in the LSTM of Belgium car drivers for headway (a) and speeding (b) as demonstrated in Fig. 6 below. The Divergence or this significant gap between training
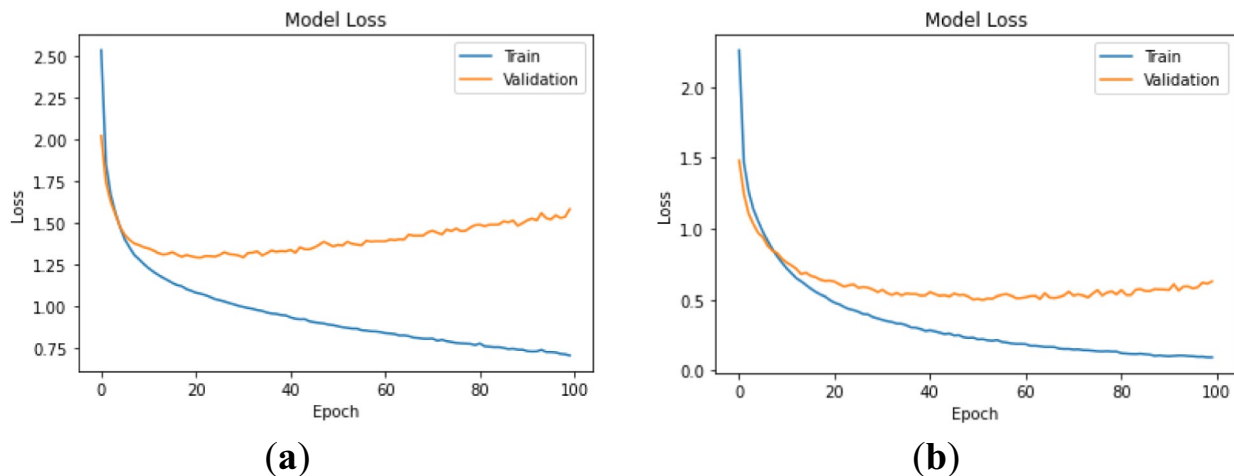

(a)   **(b)**

Fig. 5 Model loss of the LSTM model for German car drivers for headway (**a**) and speeding (**b**)

Table 6 Assessment of classification model for headway and speeding for Belgian car drivers

Variable	Accuracy	Precision	Recall	f1-score	False Alarm Rate
Headway	58.12%	35.65%	58.12%	37.33%	41.88%
Speeding	48.27%	25.75%	48.27%	32.59%	51.73%

and validation loss in Fig. 6b might indicate overfitting (high training performance but poor generalisation).

Aggregate comparisons of F1-scores between LSTM and neural networks are presented in the following graphs (Figs. 7 and 8). These figures present a comparison of F1-scores achieved by LSTM and Neural Network models across multiple epochs
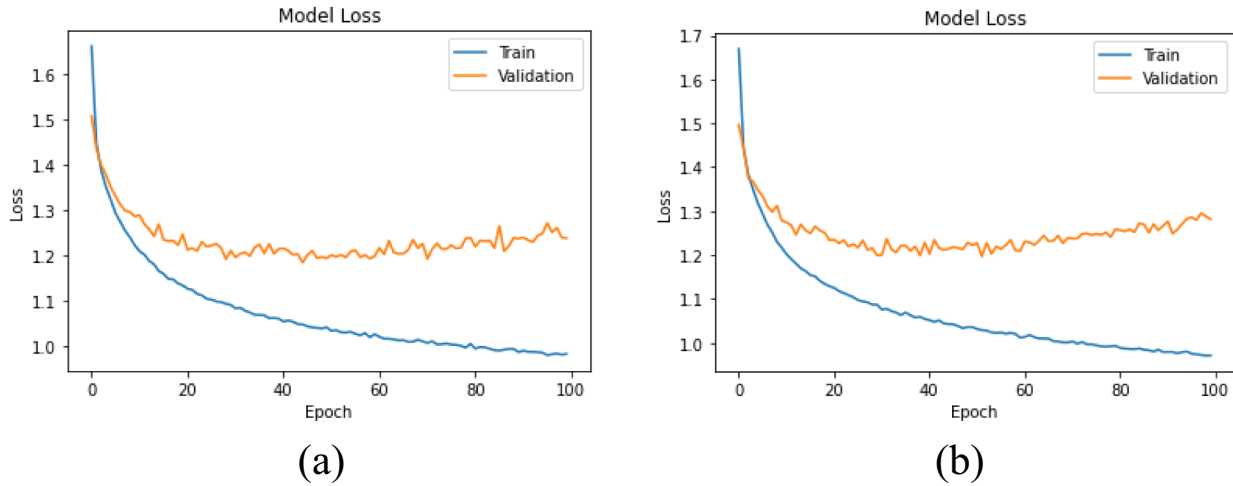

Fig. 6 Model loss of the LSTM model for Belgian car drivers for headway (**a**) and speeding (**b**)
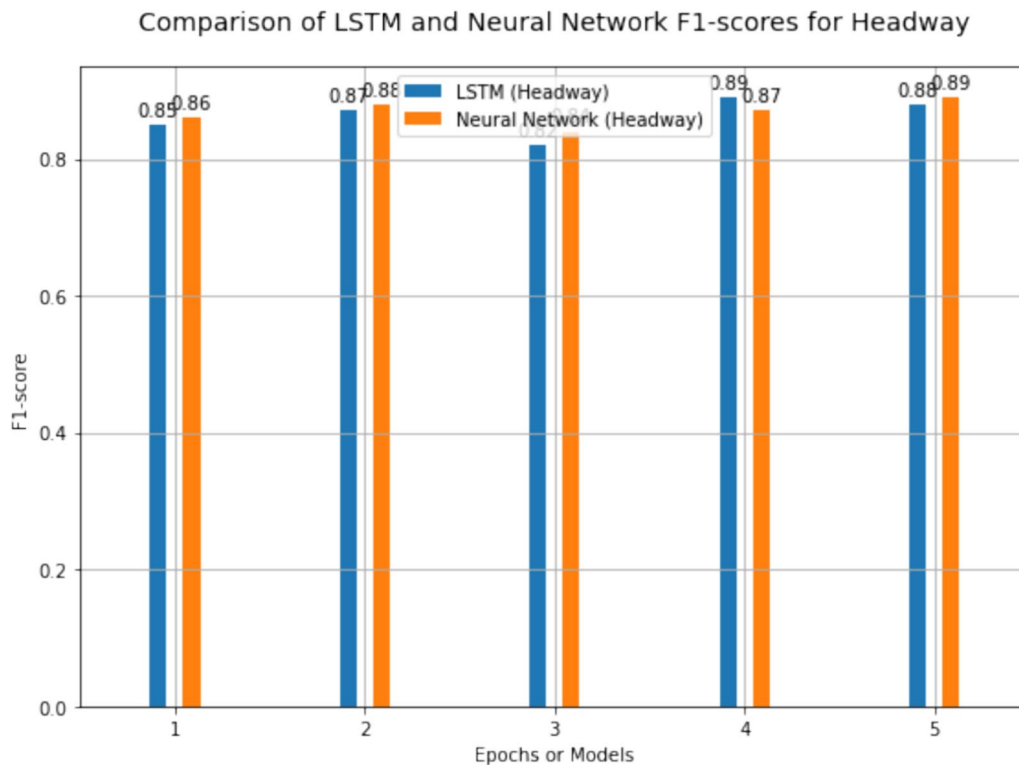

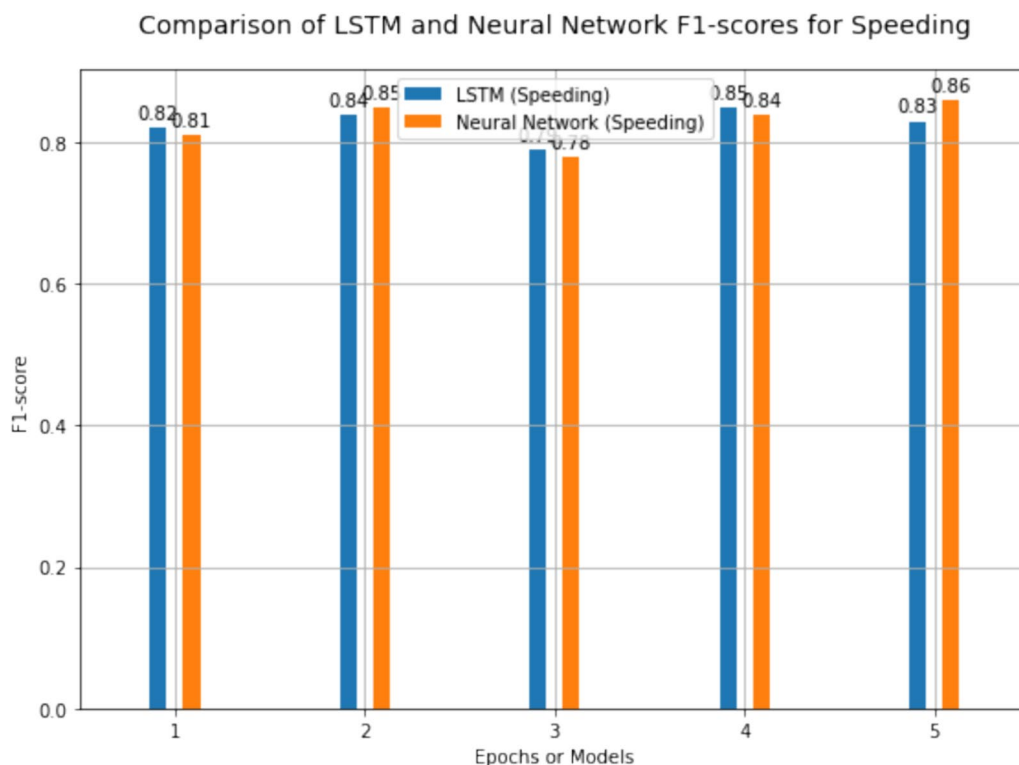**Fig. 7** Aggregate Comparison of LSTM and NN F1-scores for Headway

Fig. 8 Aggregate Comparison of LSTM and NN F1-scores for Speeding

for headway classification. Each bar represents the F1-score achieved by the respective model at a specific epoch. The LSTM model, depicted by blue bars, demonstrates varying performance across epochs, with slight fluctuations in F1-scores. In contrast, the Neural Network model, shown by orange bars, exhibits relatively stable F1-scores across epochs. This comparison provides insights into the performance consistency and potential effectiveness of each model in headway and speeding classification tasks.

6 Discussion

The objective of this study was to develop, compare, and contrast machine learning techniques for identifying risky driving behaviour. The data used in this study comprised trips from a sample of 30 German drivers and 43 Belgian drivers, and two machine learning classifiers, LSTM and a Neural Network, were developed.

Comparing the results of the LSTM model with the previous neural network models, it is evident that the LSTM model yields lower performance in terms of accuracy, precision, recall, F1-score and false-alarm rate. The LSTM model achieves a test accuracy lower than the accuracy achieved by the previous neural network models mentioned earlier. Similarly, the precision, recall, F1-score and false - alarm rate metrics also indicate lower performance compared to the previous models.

Germany's NN models demonstrate superior performance in both accuracy and precision-recall balance compared to Belgium's models and their respective LSTM counterparts. Belgium's NN models, while strong, present difficulties in achieving high precision, especially for speeding incidents. The LSTM models in both countries show potential for capturing temporal patterns, but they currently lag behind the NN models in terms of overall accuracy and precision-recall balance.

The contrasting factor between the two methods lies in the ability to capture temporal dependencies in the data. While the previous neural network models utilized a feed-forward architecture without considering the temporal aspect, the LSTM model specifically leverages the sequential nature of the data to potentially improve

prediction performance. However, in this case, the LSTM model did not outperform the previous models. This outcome suggests that the temporal dependencies present in the data may not be crucial for accurately predicting the STZ or that the LSTM model's architecture and hyperparameters need further tuning to achieve better results. Further analysis and experimentation may be required to determine the optimal approach for predicting the STZ accurately.

The results of predictive real-time analyses demonstrated that the level of STZ can be predicted with an accuracy of up to 95%. Additionally, post-trip explanatory studies highlighted the capacity of state-of-the-art econometric models to provide insights into the complex relationship between risk and the interdependence of task complexity and coping capacity.

Among the machine learning algorithms, Neural Networks proved to be the best approach for capturing complex relationships between various driving parameters and predicting the likelihood of potential risks or crashes. These algorithms were trained using the i-DREAMS data and deployed in real-time applications, such as in-vehicle systems or mobile applications, to provide immediate feedback and guidance to drivers regarding their driving behaviour. This feedback aimed to assist drivers in making informed decisions, improving their driving habits, and reducing crash risk.

The identification of safe driving behaviour through the ensemble of machine learning algorithms and i-DREAMS data has the potential to revolutionize road safety interventions. By leveraging data-driven insights and advanced analytics, this approach can contribute to creating a safer driving environment, reducing the number of crashes, and ultimately saving lives.

7 Conclusions
The findings of this study have significant implications for road safety interventions. The insights derived play a pivotal role in refining the capabilities of the STZ by providing a deeper understanding of driving behavior dynamics and improving the prediction of risky driving scenarios. By leveraging machine learning algorithms and data-driven insights, it is possible to identify safe driving behavior, provide immediate feedback to drivers, and ultimately contribute to creating a safer driving environment. While our models hold promise, further refinement is necessary to fully maximize their potential benefits.

Analyzing the long-term impact of interventions, evaluating real-time systems, and considering human factors and driver engagement are crucial areas for further investigation. Additionally, assessing the generalizability and scalability of the developed models and interventions across diverse populations, geographic locations, and vehicle types will ensure their broader impact in improving road safety.

It is important to acknowledge the limitations of this study. The dataset used in this research consisted of trips from a sample of 30 German drivers and of a varied number of Belgian car drivers across the different phases of the experiment, with 39 drivers remaining consistent throughout the phases, which may not fully represent the diversity of driving behaviors across different regions and populations. Furthermore, the performance of the LSTM model was lower compared to the Neural Network model, suggesting that further optimization and tuning may be required for improved results.

In conclusion, this study has demonstrated the potential of machine learning techniques, particularly Neural Networks, for identifying risky driving behavior and improving road safety. The development and deployment of real-time applications based on these techniques can provide drivers with immediate feedback and guidance to help them make informed decisions, improve their driving habits, and reduce crash risk.

Future research should focus on incorporating contextual information, such as weather conditions, road infrastructure, and traffic patterns, to enhance the accuracy and applicability of the models. Personalized driver modeling, considering individual characteristics, can also lead to more effective behavior change interventions. By addressing these areas, we can further advance our understanding of safe driving behavior identification, refine intervention systems, and ultimately contribute to improving road safety, reducing the number of crashes, and preventing injuries on our roads.

Appendix A

Table 7 Dataset sample

grpby_seconds	0	30	60	90	120	150	180	210
trip_uuid_	223aRz8e5o TxyFAfYVLi4S	223aRz8e5o TxyFAfYVLi4S	223aRz8e5o TxyFAfYVLi4S	223aRz8e5o TxyFAfYVLi4S	223aRz8e5o TxyFAfYVLi4S	223aRz8e5o TxyFAfYVLi4S	223aRz8e5o TxyFAfYVLi4S	223aRz8e5o TxyFAfYVLi4S
trip_start_	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00	2021-07-26 13:16:40 + 00:002021-07-26 13:16:40 + 00:00
trip_end_	2021-07-26 13:23:51 + 00:00	2021-07-26 13:23:51 + 00:00	2021-07-26 13:23:51 + 00:00	2021-07-26 13:23:51 + 00:00	2021-07-26 13:23:51 + 00:00	2021-07-26 13:23:51 + 00:00	2021-07-26 13:23:51 + 00:00	2021-07-26 13:23:51 + 00:00
iDreams_Headway_ Map_level_−1_ mean v	−9999	1	0.866667	1	1	0.566667	0	0.666667
iDreams_Headway_ Map_level_0_mean	−9999	0	0.1	0	0	0.1	0.133333	0.333333
iDreams_Headway_ Map_level_1_mean	−9999	0	0.033333	0	0	0.333333	0.8	0
iDreams_Headway_ Map_level_2_mean	−9999	0	0	0	0	0	0.066667	0
iDreams_Headway_ Map_level_3_mean	−9999	0	0	0	0	0	0	0
iDreams_Speeding_ Map_level_0_mean	−9999	0	0	0	0.0333	0.0333	0.0333	0.0333
iDreams_Speeding_ Map_level_1_mean	−9999	0	0	0	0.0333	0.0333	0.0333	0.0333

Code Snippet for the neural network model

```
#Define and compile the neural network model.

model = keras.Sequential([.

keras.layers.Dense(128, activation='relu',

input_shape=(X_train.shape [1], )),

keras.layers.Dense(64, activation='relu'),

keras.layers.Dense(len(label_encoder.classes_),
activation='softmax') # Output layer with appropriate
number of classes])

model.compile(optimizer='adam',

        loss='sparse_categorical_crossentropy',

        metrics=['accuracy'])

# Train the model.
```

```
history = model.fit(X_train, y_train, epochs = 100, batch_size = 32, validation_split = 0.1).
```

Code Snipper for the LSTM model #Define the LSTM model with additional details.

```
model_lstm = Sequential().

model_lstm.add(LSTM(128, input_shape=(1, X_train.shape [1]), activation='relu', dropout = 0.2, recurrent_dropout = 0.2))

model_lstm.add(LSTM(64, input_shape=(1, X_train.shape [1]), activation='relu', drop-out = 0.2, recurrent_dropout = 0.2))

model_lstm.add(Dense(len(label_encoder.classes_), activation='softmax')).

# Compile the LSTM model with specific learning rate.

optimizer = Adam(learning_rate = 0.001).

model_lstm.compile(optimizer = optimizer,
```

Roussou *et al. European Transport Research Review*   (2024) 16:40

Page 13 of 13

```
loss='sparse_categorical_crossentropy',

metrics=['accuracy'])

# Train the LSTM model with a specified batch size.

history_lstm = model_lstm.fit(X_train_lstm,  y_train,
epochs = 100, batch_size = 64, vali-dation_split = 0.1).
```

Abbreviations

WHO	World Health Organization
NDS	Naturalistic Driving Simulator
STZ	Safety Tolerance Zone
NN	Neural Network
LSTM	Long Short Term Memory
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

Authors' contributions
Not applicable.

Availability of data and materials
Not applicable.

Declarations

Competing interests
Not applicable.

References
1. Barbosa Silva, P., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. *J Traffic Transp Eng*, *7*, 775–790. https://doi.org/10.1016/j.jtte.2020.07.004.
2. Chen, C., Zhao, X., Zhang, Y., Rong, J., & Liu, X. (2019). A graphical modeling method for individual driving behaviour and its application in driving safety analysis using GPS data. *Transportation Research part F: Traffic Psychology and Behaviour*, *63*, 118–134.
3. Cura, A., Kucuk, H., Ergen, E., & Oksuzoglu, I. B. (2021). Driver profiling using long short term memory (LSTM) and Convolutional Neural Network (CNN) methods. *Ieee Transactions on Intelligent Transportation Systems*, *22*, 6572–6582. https://doi.org/10.1109/TITS.2020.2995722.
4. Garefalakis, T., Katrakazas, C., & Yannis, G. (2022). Data-driven estimation of a driving safety tolerance zone using imbalanced machine learning. *Sensors (Basel, Switzerland)*, *22*(14), 5309.
5. Ghandour, R., Potams, A. J., Boulkaibet, I., Neji, B., & Barakeh, A. (2021). Z. Driver behaviour classification system analysis using machine learning methods. *Applied Sciences*, 11(22). https://doi.org/10.3390/app112210562.
6. Ghadiri, M., et al. (2019). Real-time prediction of driver's intention using long short-term Memory Network. *Journal of Transportation Engineering Part C: Emerging Technologies*, *27*(6), 551–566.
7. Hickman, J. S., & Geller, E. S. (2005). Self-management to increase safe driving among short-haul truck drivers. *Journal of Organizational Behaviour Management*, *23*(4), 1–20.
8. Jia, S., Hui, F., Li, S., Zhao, X., & Khattak, A. J. (2020). Long short-term memory and Convolutional Neural Network for Abnormal Driving Behaviour Recognition. *Insurance: Mathematics and Economics*, *14*, 306–312. https://doi.org/10.1049/iet-its.2019.0200.
9. Levick, N. R., & Swanson, J. (2005). An optimal solution for enhancing ambulance safety: implementing a driver performance feedback and monitoring device in ground emergency medical service vehicles. In Annual Proceedings/Association for the Advancement of Automotive Medicine (Vol. 49, p. 35). Association for the Advancement of Automotive Medicine.
10. Li, Z., Zhang, M., & Ukkusuri, S. V. (2020). *A review of Data-Driven approaches for Enhancing Traffic Safety* (Vol. 144, p. 105643). Accident Analysis & Prevention.
11. Osman, O. A., Hajij, M., Karbalaieali, S., & Ishak, S. (2019). A hierarchical machine learning classification approach for secondary task identification from observed driving behaviour data. *Accident Analysis & Prevention*, *123*, 274–281.
12. Peppes, N., Alexakis, T., Adamopoulou, E., & Demestichas, K. (2021). Driving Behaviour Analysis using machine and deep learning methods for continuous streams of Vehicular Data. *Sensors (Basel, Switzerland)*, *21*, 4704. https://doi.org/10.3390/s21144704.
13. Saleh, K., Hossny, M., & Nahavandi, S. (2017, October). Driving behaviour classification based on sensor data fusion using LSTM recurrent neural networks. In 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC) (pp. 1–6). IEEE.
14. Shangguan, Q., Fu, T., Wang, J., Luo, T., & Fang, S. (2021). *An integrated methodology for real-time driving risk status prediction using naturalistic driving data* (Vol. 156, p. 106122). Accident Analysis & Prevention.
15. 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.
16. Song, X., Yin, Y., Cao, H., Zhao, S., Li, M., & Yi, B. (2021). *The mediating effect of driver characteristics on risky driving behaviours moderated by gender, and the classification model of driver's driving risk* (Vol. 153, p. 106038). Accident Analysis & Prevention.
17. Toledo, G., & Shiftan, Y. (2016). Can feedback from in-vehicle data recorders improve driver behaviour and reduce fuel consumption? *Transportation Research Part A: Policy and Practice*, *94*, 194–204.
18. Wang, J., Lu, H., Sun, Z., Wang, T., & Wang, K. (2020). Investigating the impact of various risk factors on victims of traffic accidents. *Sustainability*, *12*(9), 3934.
19. World Health Organization (2018). Global Status Report on Road Safety 2018.
20. Wu, Y., Abdel-Aty, M., & Lee, J. (2019). Investigating the impact of Weather on driver Behaviour and Traffic Safety using Artificial neural networks. *Accident Analysis & Prevention*, *130*, 80–88.
21. Yang, K., Haddad, C., Al, Yannis, G., & Antoniou, C. (2021). Driving Behaviour Safety Levels: Classification and Evaluation. 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 1–6. https://www.researchgate.net/publication/354425668_Driving_Behavior_Safety_Levels_Classification_and_Evaluation.
22. Yaqoob, S., Cafiso, S., Morabito, G., et al. (2023). Detection of anomalies in cycling behaviour with convolutional neural network and deep learning. *Eur Transp Res Rev*, *15*, 9. https://doi.org/10.1186/s12544-023-00583-4.
23. Yaqoob, S., Cafiso, S., & Morabito, G. (2023). Deep transfer learning-based anomaly detection for cycling safety. *Journal of Safety Research*, *87*. https://doi.org/10.1016/j.jsr.2023.09.010. Pages 122–131, ISSN 0022-4375.
24. Yaqoob, S., Hussain, A., Subhan, F., Pappalardo, G., & Awais, M. (2023). Deep learning based anomaly detection for fog-assisted IoVs Network, in IEEE Access, 11, pp. 19024–19038, https://doi.org/10.1109/ACCESS.2023.3246660.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.