Review

A comprehensive review on data-driven driver behaviour scoring in vehicles: technologies, challenges and future directions

Vaibhav Shirole[1] · Aniket K. Shahade[1] · Priyanka V. Deshmukh[1]

Received: 23 October 2024 / Accepted: 3 March 2025
Published online: 13 March 2025

Abstract
The automotive industry is transforming because of the incorporation of cutting-edge technologies like Big Data, deep learning (DL), machine learning (ML), and the Internet of Things (IoT). This is especially true regarding improving driver behaviour analysis and vehicle performance. Road safety and car diagnostics are greatly aided by real-time monitoring and predictive analytics, made possible by connected vehicles' massive data generation via onboard sensors, IoT devices, and telematics. Previous studies have mainly focused on individual technologies and lacked comprehensive discussions on the integration of generative AI. This paper covers the literature on driving behaviour analysis, generative AI, predictive maintenance and profiling. It emphasizes how well ML and DL models categorize driver behaviour, spot dangerous driving habits, and forecast when a car requires maintenance. Furthermore, the function of Generative AI is examined in terms of giving drivers personalized and dynamic feedback, enhancing overall driving performance, safety, and fuel efficiency. The paper also addresses data privacy challenges, real-time monitoring, and combining various data sources. Emerging trends in hybrid AI models and large language models are discussed as promising directions for improving predictive maintenance systems and optimizing vehicle performance.

Keywords Artificial intelligence · Data-driven methods · Driver behaviour scoring · Generative AI

1 Introduction

The automobile industry is witnessing radical changes, highly steered by technologies such as machine learning (ML), deep learning (DL), Big Data, and the Internet of Things (IoT) [1]. In the present century, connected vehicles are equipped with advanced sensor systems that can gather a considerable amount of data about the Driver's behaviour, the performance of the car, and the environment. This convergence of technology offers immense research prospects in exploring driver behaviour in a way that can provide critical inputs toward road safety and vehicle diagnostics, thereby leading to an overall improvement in the driving experience [2]. The risks of accidents and wear and tear parts of a vehicle could be predictively mitigated by using analysis of driver behaviour data, which is a significant aspect of innovation in the automotive sector [3].

With the integration of IoT devices, Onboard diagnostics (OBD) dongles, and high-tech telematics, data collection is constant, real-time, and thorough enough to provide comprehensive datasets for profiling drivers [4]. These datasets determine the key behavioural patterns in proper driving and vehicle longevity. Predictive analytics and machine learning algorithms have made Driver profiling accurate and personalized, providing actual feedback on target behaviours, such

✉ Aniket K. Shahade, aniket.shahade@sitpune.edu.in; Vaibhav Shirole, vaibhavshirole16@gmail.com; Priyanka V. Deshmukh, priyanka.deshmukh@sitpune.edu.in | [1]Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India.

as excessive braking or accelerating, speeding, and many more [5, 6]. Figure 1 shows how a basic process is followed for driver behaviour scoring. Consequently, this is followed by timely interventions that change drivers to safer practices and avoid dangerous situations.

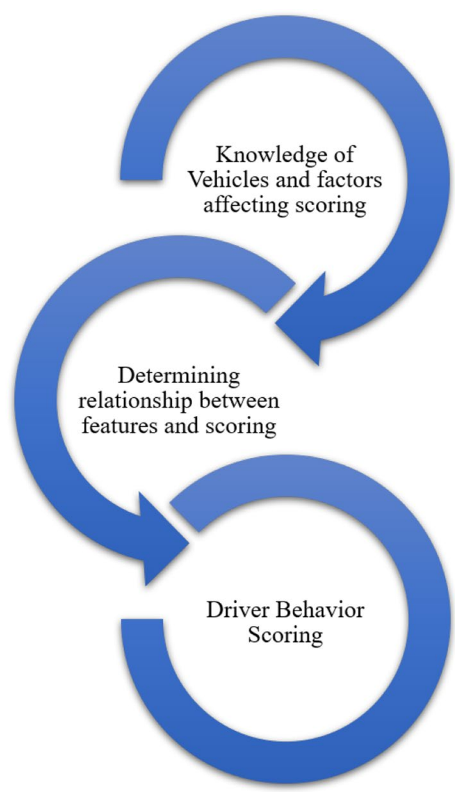
Driver behaviour analysis is catching much more than traditional usages: laws and mitigation of environmental impacts. Figure 2 depicts various applications of driver behaviour scoring systems. The system scores individual drivers to optimize operations and minimize risks associated with accidents. Insurance companies use the data gathered through telematics to evaluate drivers' conduct. Depending on the driving pattern, personal premiums are offered to promote safer road use [7]. Road safety monitoring systems can detect and counter risky behaviour through real-time driver data, avoiding accidents. Indeed, eco-driving for fuel efficiency is another significant application. Therefore, by drawing feedback from the latter AI-based model, drivers may select energy-saving driving styles, lowering fuel consumption and emissions [8]. Predictive maintenance focuses on driving data to predict incidents about potential vehicle problems, thus involving interventions before costly repairs become necessary. Personalized driving feedback through AI gives recommendations based on individual needs to improve driving skills, thereby increasing the safety and overall performance of the Driver on the road [9].

This field has been further transformed by introducing machine learning, artificial intelligence, and predictive maintenance. This technology enables vehicle safety by detecting potential failures before they happen. Predictive models can analyze petabytes of data gathered from vehicles to determine anomalies that cannot be detected during regular inspections, thus lowering the chances of breakdowns and long downtimes and making a car more reliable overall [10]. It not only makes it safer but also reduces costs related to maintenance and prolongs the vehicle's life.

However, this only relies on predictive analytics and may omit more complex factors relevant to equipment breakdown or hazardous driving conditions. A balance to this would be found in Generative AI: contextualizing insights into the analytics process offers the platform a broad base to synthesize vast amounts of complex data into human-readable reports full of detailed observations and practical advice [11]. This brings a data-driven analysis and natural language processing to generate a dynamic system monitoring and suggestions for improving driver behaviour and vehicle performance [12].

The culmination of these technologies, as shown in Fig. 3, has led to intelligent transportation systems that can monitor and influence driver behaviour through feedback and suggestions. The focus is now on proactive vehicle management and personalized coach driver training, which results in improved road safety, optimized vehicle maintenance, and enhanced overall transport productivity. As the technologies develop, they will play a more and

Fig. 1 Driver behavior scoring process



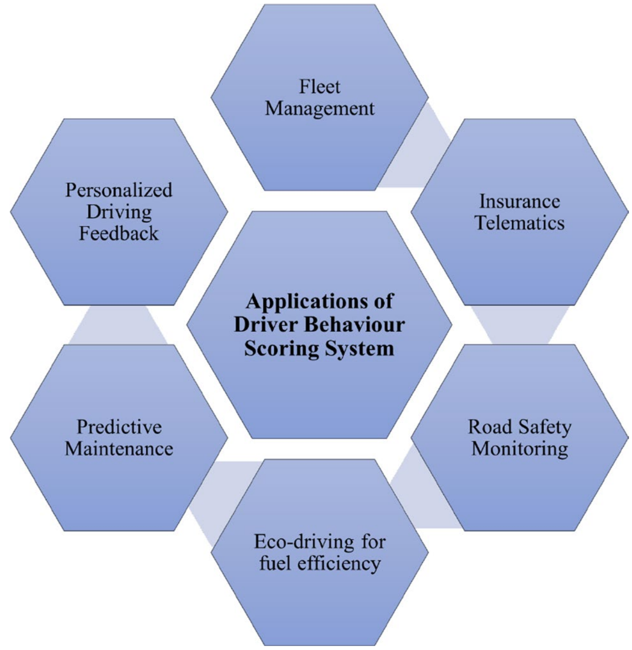


Fig. 2 Applications of driver behaviour scoring systems

more instrumental role in shaping the automotive sector's future. This study aims to explore the technologies in AI for driver behaviour analysis while addressing the challenges and discussing the emerging trends in the field. The main contributions include:

1. An in-depth review of ML, DL and GenAI technologies in improving driver behaviour analysis and optimizing vehicle diagnostics.
2. Identifying current challenges and using hybrid AI models by leveraging LLMs within the automotive industry.
3. Review and utility of public datasets available for driver behaviour analysis.

2 Research methodology

A literature review is an essential component of academic research. In the present paper, a systematic approach is undertaken to ensure an accurate and in-depth review of the literature on Driver Behaviour. A systematic literature review (SLR) framework, as shown in Fig. 4, was applied to identify, evaluate, and synthesize all the relevant studies.

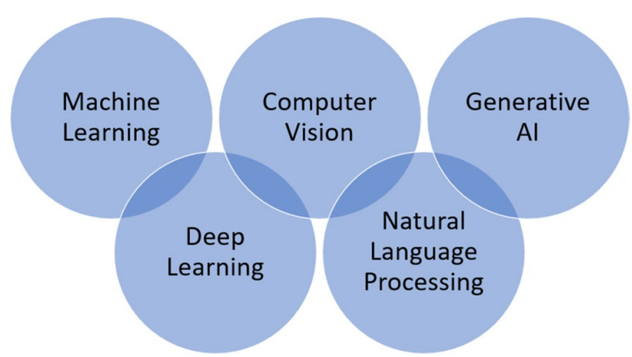


Fig. 3 Emerging AI techniques in driver behaviour

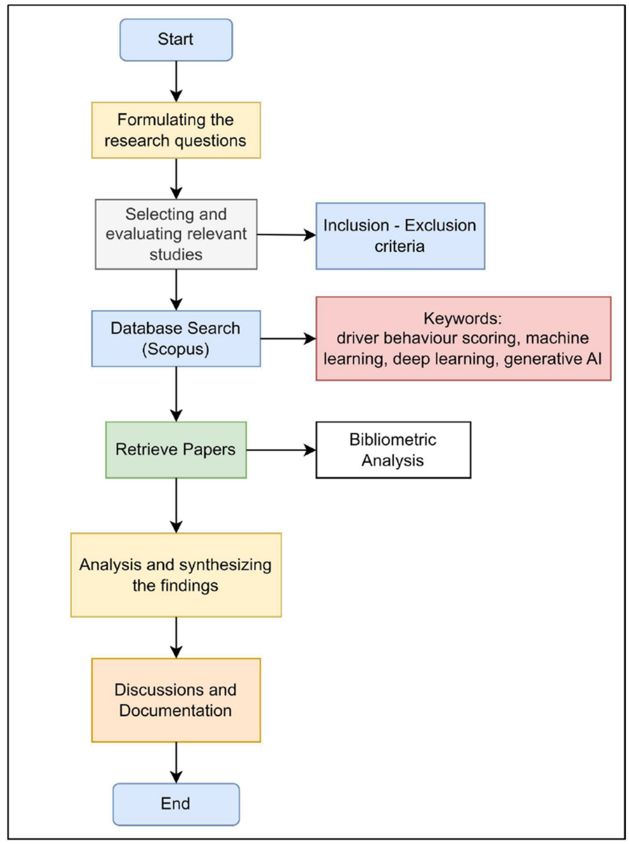


Fig. 4 Research methodology framework

2.1 Research questions

The review begins by formulating well-crafted research questions to answer core aspects of driver behaviour scoring systems. The key research questions identified for this review were:

Q1. What are the main objectives and categories of driving behaviours found in earlier studies on driver behaviour score systems?

Q2. How are various forms of driving behaviour examined and classified in driver behaviour scoring studies?

Q3. Which datasets, data sources, and essential elements are used in studying driver behaviour scoring systems?

Q4. What kinds of machine learning models are used, and what is the effectiveness of these models in classifying driver behaviours?

Q5. How are advancements in Generative AI influencing the research in driver behaviour reporting?

2.2 Search strategy and information sources

Literature with relevant articles is extracted using a systematic search strategy after research questions are identified with clarity. As such, this research covers many databases under IEEE Xplore, Springer, Scopus, and Google Scholar so that related keywords can help retrieve presumably relevant articles. The keywords used for these searches were “driver behaviour scoring," “machine learning for driver analysis," "predictive maintenance using driver data," and “generative AI for driver behaviour reporting.” However, boolean operators were used to condition the search to retrieve the relevant scope. The documents were restricted to reviews, articles, preprint and book chapters (Fig. 5).

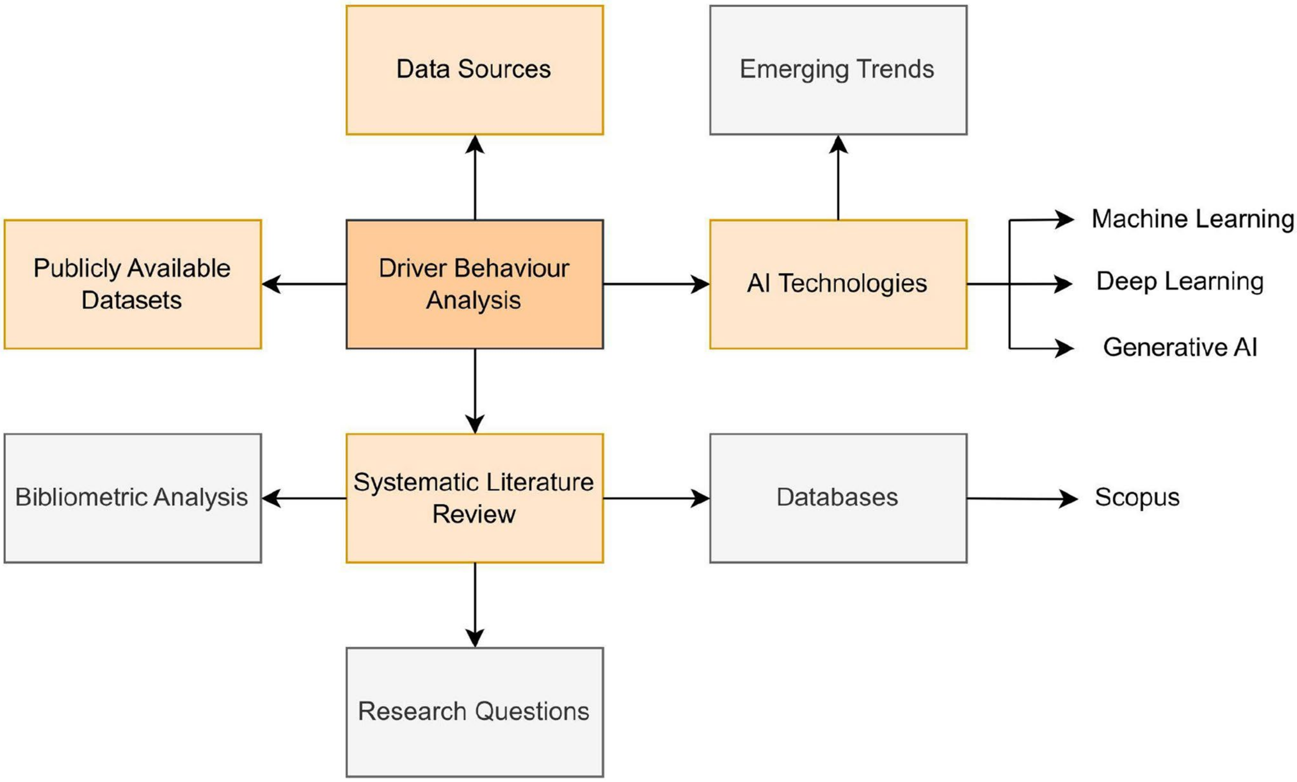


Fig. 5 Taxonomy tree for analytics

2.3 Inclusion and exclusion criteria

There should be strict inclusion and exclusion criteria to ensure only the most relevant of high-quality papers are used. Papers should be evaluated regarding their relevance to inclusion in how they concentrate on applications in machine learning, in terms of datasets, and how they integrate Artificial Intelligence (AI) for reporting. Table 1 presents a concise view of the criteria set used for the systematic literature review.

2.4 Final selection and preparation of detailed review

After filtering and data extraction, the final set of articles is curated for in-depth analysis. These articles represent the most relevant research studies directly addressing the research questions. This refined literature set forms the review's core, ensuring a focused discussion on current methodologies, challenges, and future directions in driver behaviour scoring systems.

3 Bibliometric analysis

Matplotlib and Seaborn are used for the representation and visualization of analysis.

Figure 6 indicates the number of publications proliferates in 2024 to nearly 40 and then declines throughout the remaining years. This suggests that, like advancements in predictive vehicle maintenance and safety systems, there has also been a growing demand for academic work and industry interest in driver behaviour analysis and related AI technologies in recent years.

Table 1 Criteria set for inclusion and exclusion

Criteria type	Inclusion	Exclusion
Publication year	2020–2024	Before 2020
Databases	Peer-reviewed journals and conference proceedings from Scopus, IEEE Xplore, Google Scholar and Springer	Non-academic sources, unverified databases
Study focus	Research on driver behaviour scoring systems, ML-based driver behaviour analysis, and the application of Generative AI in reporting	Studies unrelated to driver behaviour analysis, machine learning models, or AI-based reporting systems
Technology scope	Utilization of Machine Learning, Deep Learning, and Generative AI techniques for driver behaviour scoring	Studies focusing on manual or non-AI methods for driver scoring or unrelated AI applications
Data sources	Studies use mobile phones, OBD-II, Dongle, Global Positioning System (GPS), or vehicle sensor data for driver behaviour analysis	Studies using non-automotive or irrelevant datasets (e.g., medical, financial, or unrelated industries)
Application context	Focus on automotive, fleet management, and transportation industries	Studies outside of automotive, such as industrial or non-relevant sectors
Type of article	Original research articles, technical papers, and comparative analyses	Opinion pieces, reviews without original research, editorial notes

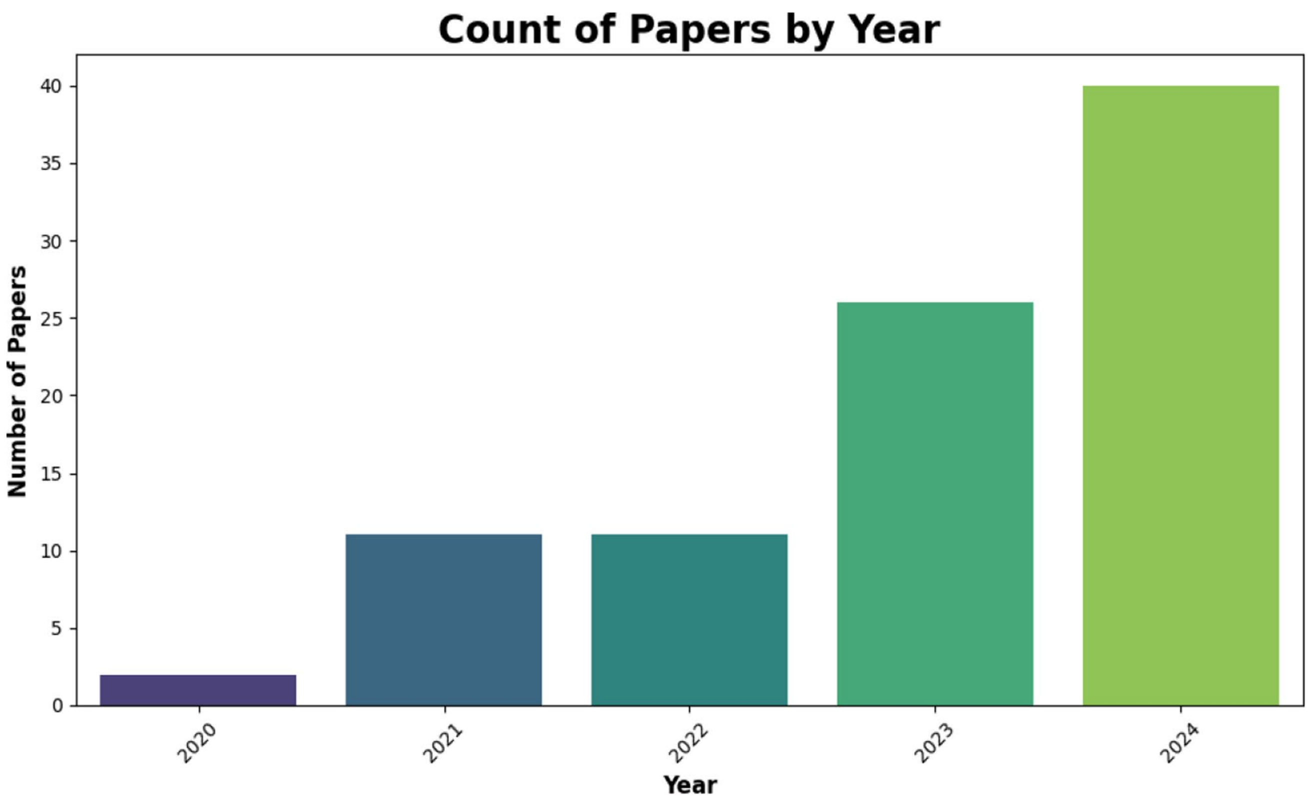
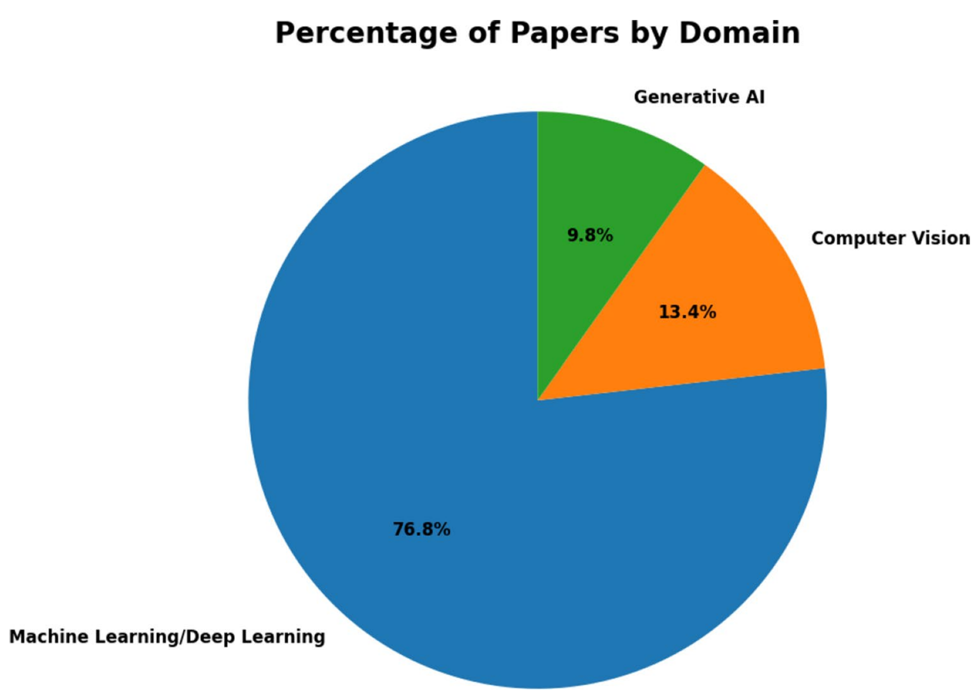


Fig. 6 Count of papers by year

Fig. 7 Percentage of papers by domain



The pie chart in Fig. 7 illustrates the distribution of focus areas in research according to pie charts, where most papers are on machine learning/deep learning, then computer vision and generative AI. This suggests that ML/DL techniques are prominent in analyzing driver behaviour with relatively fewer but equally important computer vision and generative models focusing on improving vehicle systems and predictive maintenance solutions.

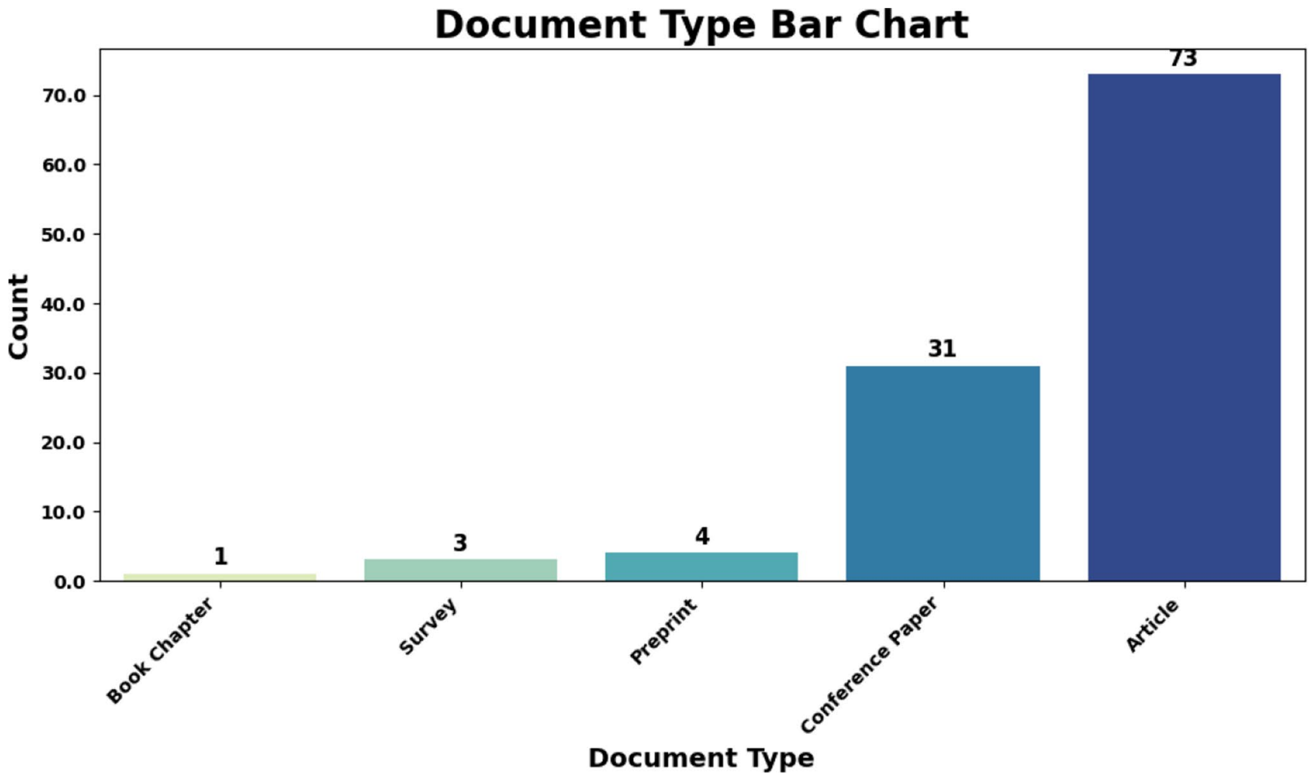


Fig. 8 Document type bar chart

The heatmap in Fig. 8 categorizes the papers by document type, mainly journal articles, followed by conference papers. This distribution highlights that most research findings are disseminated through peer-reviewed articles. At the same time, conferences are secondary outlets for sharing new insights and developments, with other document types like preprints, surveys, and book chapters being relatively rare.

4 Related work

Machine learning plays a pivotal role in developing driving behaviour scoring systems for large datasets from vehicle sensors. It extracts patterns and trends from data points like speed, acceleration, and braking. Machine learning algorithms like regression and classification models use the same to predict or identify driver behaviour anomalies. Such aspects as harsh braking, speeding, and fuel efficiency are fundamental and quantified through these machine-learning techniques. This, in turn, provides every Driver with an objective score. Relatively recent takeoffs in ML and DL technologies have experienced a transformation in their efforts to analyze and classify driver behaviour targeted towards road safety, fuel efficiency, and vehicle operation. Several researchers have used these techniques, applying different datasets and methods to produce great results.

4.1 Machine learning and deep learning

A new methodology was developed in [13] to address road accidents by incorporating smartphone motion to detect driver behaviour in advance. Their ensemble feature engineering strategy successfully used overview features Linear Regression (LR) and Random Forest Classifier (RFC) to realize the probabilistic features, which gave an efficiency rate of 99%. This study underlines the effectiveness of applying ML techniques to increase safety by recognizing the possibilities of unsafe driving.

The issue of transferring learning from one vehicle to another for the recognition of driving behaviour from a completely different kind of vehicle was dealt with in [14]. Also, they got 0.80 accuracy, which is considerably better than 0.64 for the non-transferred model, using a novel Graph Neural Network (GNN)gp-based transfer learning that they invented. This paper also mentions how well transfer learning can be used for driving behaviour recognition, specifically in multi-vehicle scenarios.

According to the paper [15], they performed a driving behaviour analysis motion-based OBD. Supervised learning-based methods were used for driver behaviour classification, including SVM, Adaboost, and random forest. The RF model had an accuracy of 100%, supporting the hypothesis that OBD data can provide timely and accurate information about driver behaviour and its classification.

The study undertaken in [16] on possible privacy issues encountered when assessing driver behaviour in hybrid electric vehicles (HEVs) came in handy. For their method, they applied Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) modelling in predicting and classifying aggressive driving. In that approach, the accuracy was 96%, with GRU providing a good solution to the problem of monitoring the Driver in real-time and keeping their privacy safe.

On the other hand, [17] includes the vehicle report, the onboard driver tracking system, and the remote detection faculties. System parameters included fuel consumption and/or exhaust emissions, and the driving style was 89% accurate for estimating the driving style. This study highlights the capability of using such technology to enhance driving style through multiple sensors.

The authors in [18] systematically reviewed driver behaviour classification research. The review encompassed various data types, data sets, features, data preprocessing, and AI techniques. The review provided a clear understanding of the factors that inhibit advancement in the study of drivers' behaviour, focusing on selecting features to optimize the performance of ML algorithms.

The research in [19] analyzed driving behaviours and patterns using the OBD-II data. They employed several models of ML to distinguish safe from dangerous driving types, ranking the performance of the nine best algorithms. The authors also noted that the models should be customized, including other data courses such as GPS and weather conditions.

The authors in [20] introduced efficiently supervised driver behaviour profiling based on ML data. They worked with the data from the naturalistic study of driving and reported obtaining 90% accuracy with the Random Forest model. This particular framework allows one to profile individual Driver risk closer to reality as it envisions and promotes safety in driving. The authors applied OBD and smartphone sensors to develop Artificial Neural Network ANN algorithms for predicting driving style and traffic conditions [21]. Their focus was primarily on dealing with topic imbalance datasets, which was appreciated when the ANN model achieved an accuracy of up to 92.09%. This technique shows combinations of different ones to facilitate detailed studies of driver behaviour.

According to [22], the research proposed a novel adaptive fuzzy recurrent neural network for macroscopic driving behaviour pattern recognition and prediction. Implementing this model of the Internet of Vehicles allowed them to predict risk-driving behaviour and assess the associated risks more accurately than any other model concerning the mean squared error. The use of fuzzy logic systems in sequential pattern recognition is one of the promising technologies for assessing and predicting drivers in real time. The paper [23] also explained the functional improvement of ML algorithms due to feature selection for the problem of driver behaviour classification. Positioning the study only on the essential features improved the classification accuracy of the study, and it was indicated that the feature was the way to go for understanding driving behaviour.

The paper used machine learning for real-time fuel consumption prediction and driving profile classification. The predictive models were reported to attain an MSE of 4.49 with XGBoost to predict fuel consumption in real time accurately. However, by including Large Language Models LLMs in this system, there would be no other improvement but the accuracy of the fuel consumption prediction, which could be used to offer feedback to drivers concerning their driving profiles. This real-time feedback can prompt better driving patterns, which reduces fuel usage and emission, as [24] has stated.

The research was done on data gathered from the OBD-II interface, and their fuel performance analytics and the essence of predictive analytics were discussed. The paper also demonstrated a strong relationship between fuel usage and parameters like throttle position and RPM using a regression model that achieved an accuracy rate of 95%. Further beneficial patterns are offered to drivers to increase fuel efficiency, decrease the cost of the operation, and execute some patterns of considerable fuel consumption relative to the conditions of the vehicle and the Driver's behaviour, which was modelled using LLMs [25].

The paper on the multivariate regression model produced the best results, registering the lowest values of mean squared error (MSE), mean absolute error (MAE) and root mean squared error (RMSE), among other models. The LLM.s may improve by building on these principles and engaging more contextual inputs in predicting fuel consumption. Fuel consumption and vehicle use can be expected through driving behaviours and vehicle characteristics, where better maintenance scheduling could be explored to lead to fewer breakdowns and less expensive maintenance [26].

The research based on autoregressive integrated moving average (ARIMA) and Ordinal Logistic Regression is used to overcome limitations in current monitoring systems referencing raw data from onboard diagnostics-II. [27]. While the system enhances user accessibility by summarising complex data into a comprehensible format, some key potential improvements include real-time actionable insights and further improvement of the prediction accuracy and communication between vehicles in the Internet of Vehicles network.

The review on Explainable Predictive Maintenance [28] followed the PRISMA technique on existing methods of explainable predictive maintenance in 2024, exploring two main XPM categories—model-agnostic and model-specific. Overall, 40 different methods were identified that would help in predictive maintenance (PDM) tasks, especially for turbofan datasets used in explainable predictive maintenance (XPM) research. It concluded that this study emphasized the need for more comparative analyses in establishing the effectiveness of various methods of XPM. and pointed out that refinement and improvement should be done on the approaches of explainable artificial intelligence (XAI) and ML to make them more applicable for Predictive Maintenance tasks.

The study [29] proposes a methodology for addressing driver fatigue monitoring based on facial recognition using deep learning algorithms for facial feature analysis: eye closure, yawning, and head movements are excellent indicators of drowsiness. The Driver's data may also concern such systems with model generalization enhancements. A systematic literature review in [30] explores the stochastic methods, statistical inference and AI techniques (ML and DL) for predictive maintenance based on big data applications. The evaluation method is challenging due to the scarcity of real datasets on automotive predictive maintenance. Similarly, another study analyses 52 studies on PdM for automotive motor vehicles [31] and discusses artificial intelligence approaches with findings, challenges and opportunities in this field.

The study in [32] introduced a novel framework for Electroencephalography **(**EEG)-based prediction of a driver's stop/go decision at a grade crossing using the functional brain network (FBN)-convolutional neural network (CNN) approach to driver decisions on stopping or going. The authors demonstrate that more robust alpha band connectivity correlates with stopping decisions and proposed that the FBN-CNN model outperformed traditional methods, such as random forest (RF) and support vector machine (SVM), in improving prediction accuracy from 76 to 90% by combining EEG data from routine and pre-decision driving stages. These studies outline the potential application of EEG and deep learning techniques in the Driver's decision monitoring.

The paper [33] examines the effects that roadside advertising has on driver attention. It is strengthened with eye-tracking data and attention to analysis through traditional statistics—Anova, multiple linear regression, and machine learning. Results indicated that roadside advertisements significantly reduce the level of attention a driver has. Results also varied and differentiated by type and size. Concerns about acceleration rates pose rather unsetting implications for road safety. The average fixation was 874.3 but extensively ranged from 80.24 to 8154.87 ms.

The study in [34] represents an embedded system for road conditions and aggressive driving behaviours with an inference time of 34 ms on Arduino Nano using minimal memory and a total accuracy of 91.9%, precision of 93.6%, and recall of 92%. With this version optimized for resource-constrained devices, real-time deployment is promising to improve road safety as it detects driving risks on edge devices under offline conditions.

The paper [35] compares LSTM networks and Neural Networks (NN) regarding driver safety assessment using i-Dreams on-road trials. The authors achieved as much as 95% accuracy with Neural Networks but did better than them in the prognosis of the risky driving behaviour. Still, LSTM models fail to understand more substantial temporal dependencies. The study concludes with the NN's promise of relating real-time information to drivers, significantly reducing road casualties. It states that further tuning of the LSTM models and investigation of contextual factors is needed to better account for accuracy and generalizability.

The research in [15] classifies driving behaviour as real-time with the OBD II protocol based on machine learning techniques like SVM, AdaBoost, and Random Forest. Parameters used for the distinction are fuel consumption, steering, and braking patterns, which distinguishes the drivers into ten classes with accuracy between 99 and 100%. Another paper [36] discusses traffic dynamics based on trajectories of vehicles by identifying adaptive cruise control behaviour using clustering methods. Real-world data might be applied to make the simulation as realistic as possible, and the way in which the advanced driving assistance system (ADAS) affects driving behaviour can be observed.

The research done in the paper [37] predicts lane-changing behaviour with machine learning, where the Extra Trees classifier has obtained the best score of traditional models and is transferable even to new datasets. It had optimal input feature selection by Recursive Feature Elimination, and it improves interpretability when using explainable AI. The authors in the paper [38] introduce spatial-temporal driving trajectory behavior anomaly detection (STDTB-AD), a deep-learning framework for detecting driving anomalies based on trajectory data. It uses a variational autoencoder to capture the spatial-temporal dependencies to achieve very high detection accuracy relative to traditional methods that only consider spatial or temporal dependencies.

The paper [39] discussed a hybrid CNN-bidirectional long short-term memory (BiLSTM)-attention model for the inference of head posse by a driver during critical tasks of the in-vehicle information system for distraction warning and vehicle takeover. Unlike traditional machine learning models, this model captures short-term and long-term dependencies, especially suited for multistep-time-series prediction tasks. Another research paper [40] presents a novel system to monitor the driving behaviour of competent truck drivers in the logistics industry of the Philippines using machine learning, IoT, and data analytics. This is demonstrated in this proposed model, which is trained on the you only look once version (YOLO) v4. It achieves high accuracy levels in object detection and driver assessment, with an average score from 75.45 to 99.89%.

The authors in the paper [41] evaluate driver distraction using real-life data and machine learning methods. Teh paper focuses more on identifying behaviours around aggressive, drowsy, and normal diving behaviours. The paper compares four classification methods. It also shows how the gradient-boosting method outperforms others in detecting driver distraction. The paper [42] focuses on Chongqing, China. It uses micro-electro-mechanical systems (MEMS) sensors and classifiers, including cubic SVM, OK k-nearest neighbors (kNN), and Gaussian NB, to assess aberrant driving behaviours. The cubic SVM reports excellent capability in identifying driver behavioural anomalies, like sudden braking that peaks at the level of F1 scores of 93%. These two studies prove the great potential of machine learning in vehicle safety-related driving behaviour analysis.

These collectively imply the strategic significance of ML and DL in enhancing understanding and improving driver behaviour. The synthesis of diverse data input, incorporation of sophisticated algorithms, and use of privacy-preserving methodologies ensure the analysis is accurate and relevant. Table 2 summarises the research that has significantly contributed to driver behaviour, profiling, and analysis with ML, DL, and other hybrid AI-related approaches. More work in this area promises to broaden the measures that promote road safety, enhance vehicle functioning and facilitate intelligent transportation systems' growth.

Generative AI improves the reporting process in driving behaviour analysis from complex data to clear, understandable, and actionable insights. Reports summarising key findings, areas of concern, and personalized recommendations will automatically be generated through highly advanced natural language processing models using generative AI. Thus, more accessible results of data analysis will be attained through AI-driven reports for a large audience, including drivers, fleet managers, and decision-makers. Recent advancements in large language models (LLMs) on many systems, such as driving behaviour evaluation, predictive maintenance, and prediction reporting, have been impressive progressive improvements in those systems. Therefore, the following papers also provide examples of how such modern technologies in predictive models and LLMs can be embedded in modernization and enhance feedback on drivers, vehicle diagnostics, and system efficiency performance.

5 Generative AI

The field of driver identification using ML, DL, and L.L.M. techniques from multiple in-car data sources: CAN-BUS, OBD-II interfaces, smartphones, GPS, and wearable technologies is studied. A unified framework incorporating LLMs for superior driver identification [43] is proposed, highlighting possibilities for hybrid models and transfer learning techniques for better adaptability and system performance.

A real-world issue of driving analytics can benefit from the rewritten prompts outlined in the paper. Using the personalization technique in context comprehension may allow the system to alter prompt feedback based on a driver's input without any misdirection in context according to the Driver's tendencies. This would enhance the diagnostics of complicated machines, particularly the ability to make predictions and provide advice on acting based on those predictions/warning signals usually issued to the drivers by non-friendly-user machines [44]. The study in [45] presented a framework with Multimodal LLMs and applied it in a real-world construction site scenario with uncrewed vehicles. It showed significant speed up in the inspection process and generated a quality report to meet regulatory standards with extensive and intensive coverage of the construction site.

Recent breakthroughs with Generative AI technologies such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Large Language Models (LLMs) hold great promise for overcoming operational challenges in autonomous systems related to predictive maintenance, anomaly detection, and adaptive threat response. The comprehensive survey [46] highlights how such GenAI technologies offer new solutions. It contrasts their effect with prevailing

Table 2 Summary of the reviewed papers on machine learning and deep learning techniques in driver behaviour scoring system

Ref. no.	Algorithm/methodology used	Dataset used	Goal/output	Related gaps/limitations
[13]	Ensemble engineering [LR, RFC]	Smartphone motion data	Detect dangerous driving behaviours	Lacks generalization to various vehicle types
[14]	GNN-based transfer learning	Driving data of different vehicles	Driving behaviour recognition	Improved accuracy in diverse vehicle settings
[15]	SVM, Adaboost, RF	OBD motion-based data	Classify driver behaviour	Lacks external factors like GPS and weather conditions
[16]	LSTM, GRU	Hybrid electric vehicles data	Predict and classify aggressive driving while preserving privacy	Real-time applicability
[17, 18]	ML models [RF]	OBD and sensor data	Improve fuel consumption and emissions through driver monitoring	Requires multi-sensor integration and optimization in feature selection
[19]	Various ML models [Top 9 ranked]	OBD-II data	Classify safe and dangerous driving patterns	Customization challenges and limited external factors
[20, 21]	RF, ANN for profiling and imbalance	Smartphone sensor data and SHRP2 study	Profile individual driver risk	Lacks interpretability and broader feature set
[22]	Fuzzy RNN, IoV model	Internet-of-Vehicles data	Predict macroscopic driving behaviour	Improvement in real-time risk prediction accuracy
[25, 27]	Regression models, ARIMA and ordinal logistic regression	OBD-II data	Data analysis for driver behaviour monitoring	Integration of external data and real-time data analysis
[32]	FBN-CNNs for ECG-based predictions	EEG data from driving sessions	Predict stop/go driver decisions at grade crossings	Validation in diverse real-world scenarios
[33, 34]	ANOVA, ML and optimized ML	Driver attention data	Detect aggressive driving behaviours	Needs generalization and robust deployment scenarios
[35]	LSTMs and Neural Networks	Road trails data	Assessment of risk behaviour and driver safety	Temporal dependencies aren't fully captured with LSTMs

methods to impress advanced security measures needed for developing more adaptive and resilient autonomous systems. Another study proposed a novel prognostics framework with a conditional generative adversarial network and deep-gated recurrent unit for generating multivariate fault instances and solving the challenge of data imbalance for predicting the RUL of complex systems. With the learning of fault samples, authors claimed the accuracy improvement by 15% [47].

Another research focused on integrating Multimodal Large Language Models into an autonomous driving system by reviewing datasets, algorithms, and applications. This review has underlined the need to create new datasets and enhance current L.L.M. algorithms to handle the complexities of autonomous driving [48]. For the challenge of scarcity of abnormal data in the PdM system, a GAN model in [49] is used to generate this data in the form of acceleration signals that would help to assist on a low-frequency sensor device for early fault prediction of motors and [50] focused on increasing the effectiveness of diagnostics and maintenance in transport systems, by proposing a middleware approach integrating generative AI algorithms. The role of generative AI and E-maintenance concepts is explained in the existing categories. The research [51] makes advanced modelling of human driver behaviour using the neural policies learnt from the driving data. The authors use some of the extensions of Generative Adversarial limitation Learning, including PF-Gail for multi-agent settings, RAIL for domain-specific information and Burn-infoGAIL for latent variability, to model highway driving behaviour well.

The study in [52] utilized Drive-GPT pre-trained transformers that encourage smooth acceleration and uniform speed as efficient driving behaviours. The model was trained on 90 million data points from the telematics system of 100 vehicles; it predicts driving patterns, computes the driving score, and makes suggestions for an R-squared of 0.98 and energy efficiency improvements of 25% and [53] develops a deep latent variable model with a Wasserstein auto-encoder to accomplish the task of multi-agent prediction. This captures the interaction pattern while simultaneously ensuring kinematic constraints. It thus offers an edge over other worlds in various multi-agent decision-making scenarios. The authors of [54] synthesize danger scenarios to train autonomous-driving systems in simulations using classification inside a SUMO road network by the severity of the event and generate a high-probability collision scenario dataset, which is used then for training a conditional Wasserstein GAN to develop risk scenarios and deal with class imbalance problems. At the same time, [55] discusses how deep probabilistic generative models might be applied to modelling personalized driving behaviour. This model predicts the driving pattern, like velocity and steering angle, by incorporating driver-specific data and surrounding vehicle information coupled with the learning ability of individual styles to reproduce future driving behaviours correctly.

The paper [55] proposed a personalized prediction of driving behaviour based on probabilistic deep generative models to detect specific patterns in the velocity, acceleration, and steering angles of a driver with uncertainty and surrounding vehicle data; hence, it trains using actual driving data and infers future behaviour by applying these individual styles of driving. The research in [56] applies the STA-LSTM model optimized by the Lion Optimization Algorithm LOA to driver behaviour prediction. Optimized for predictively deciding what driving behaviour will be from 4032 GPS sensor observations in Shenzhen Province, China, this model can predict how a car would behave under variable conditions to enhance ITS safety further.

The study in [57] presents a model for predicting human driving behaviour based on Cumulative Prospect Theory, which captures more irrational or biased choices concerning the context of interacting driving scenarios and compares the performance of the CPT-based model with traditional and learning-based models for one case study of merging at roundabout results in the CPT approach with superior interpretability with fewer training data and accomplishing comparable accuracy. At the same time, the paper [43] discusses a large-scale review of identification techniques for driver detection, traditional and deep learning methods, and promising lines for future work, among which are hybrid approaches and Large Language Models that fine-tune LLMs technologies [58] and optimize performances with sophisticated techniques in memory fine-tuning as well as collaboration with multi-agents in AI applications.

The authors of the paper [59] developed a framework using Large Language Models within the autonomous vehicle environment to enhance decision-making capabilities and personalize driving experiences. Contextual understanding and reasoning capabilities of LLMs entail real-time adaptation of vehicles by the verbal commands provided by passengers. The experiments on HighwayEnv using Chain-of -thought prompting yield better decisions in a drive-by. This approach enhances safety, transparency, and the direction of autonomous driving towards adaptability, user-centricity, and more thoughtful experiences.

These studies successfully combine LLMs with predictive analytics to help provide feedback for drivers in real-time, improve fuel consumption practices, and enhance the predictive maintenance systems of vehicles. Table 3 summarises the use of GenAI for various tasks associated with this domain. Through improved accuracy, f1 scores and low error rates,

Table 3 Summary of the reviewed papers on generative AI in sectors of the automotive domain

Ref. no.	Methods/algorithms	Data used	Goal/output	Related gaps/challenges
[15, 48]	Unified framework with LLMs	Data from OBD-II, CAN_BUS, GPS, smartphones	Driver identification and hybrid models for adaptability	Lack of personalized comprehension and scarcity of hybrid models
[37]	Multimodal LLMs for uncrewed vehicles	Real-world construction site dataset	Generate quality reports for regulatory standards and inspections	Limited generalization
[38]	GANAs, VAEs, LLMs	Various PdM sources	Enhance PdM for anomaly detection and threat response	Unexplored GenAI models and lack of robustness
[39]	Conditional GAN, deep GRUs	Multivariate fault dataset	Improve remaining practical prediction accuracy	Imbalance faut data
[41]	GAN model	Low-frequency sensor data of motors	Generate synthetic data to aid fault prediction in motor systems	Data scarcity and imbalance
[43]	Extensions of PF-Gail	Highway driving data	Model highway driving behaviours to enhance multi-agent systems	Variability in highway driving, multi-agent scenario complexity
[44]	GPT-Pretrained transformer	90 million telematics data points	Predict driving patterns and compute driving scores	Needs diversity in scenarios
[45, 46]	Conditional Wasserstein GAN, Latent variable model for multi-agent	SUMO road network and multi-agent data	Generate high-risk scenarios and improve multi-agent decision-making	Kinematic constraints and class imbalance
[47, 49]	Probabilistic generative models, STA-LSTM	GPS sensor driver data	Predict driver behaviour and enhance safety	Driver-specific variability not fully captured
[50]	CPT-based model	Driving scenario dataset	Model irrational human driving behaviours and decision-making	Interpretability and fewer training data
[51]	Hybrid approaches with DL and LLMs	Driver detection datasets	Accurate and adaptable driver identification	Lack of optimized hybrid approaches
[52, 53]	LLM-based framework	HighwayEnv Data	Enhance decision-making capabilities and improve passenger experiences	Real-time adaptability

LLM-based systems would help enhance vehicle health monitoring and driver behaviour studies for safer and more effective driving.

6 Discussions

The studies on driver behaviour analysis and predictive analytics with ML and DL techniques provide deep insight into the advancements in understanding and improving driver and vehicle safety and efficiency. Several researchers employed varied datasets and methodologies to analyze driving behaviour. Studies show impressive classification accuracy and predictive capability using different machine learning models like the random forest, decision tree, logistic regression, etc., which answer the fourth question from Sect. 2.1. New trends in profiling and analyzing drivers' behaviours rely on various algorithms and methodologies that help cope with some of the principal challenges: real-time monitoring, risky behaviour prediction, and the ability to adapt to user preferences. Recent research describes the utilization of ML and DL for Driver's behaviour classification based on various data sources [60, 61], which answers the first question from Sect. 2.1. Different driving patterns can be captured from the datasets like harsh braking, harsh acceleration, cornering, overspeeding, etc., which answers the second question from Sect. 2.1. Analysis of OBD data has been proven to spot unsafe driving patterns and promote further real-time feedback, which answers the third question in Sect. 2.1. However, several challenges in the field have to be handled with this promising development. This has been highly instrumental in controlling imbalances of data, synthesis of artificial data about rare driving events, such as accidents, and improvements in the robustness of models, shown in paper [62, 63]. Generalization across different vehicle types and driving conditions is the most critical among them, possibly affecting the prediction accuracy. So far, the transfer learning approach seems the most promising solution to this challenge, even though more fine-tuning is needed for robust operation in diversified situations. Different data sources can rarely collaborate with no hitch in data processing or the choice of features, which may lessen the efficacy of ML algorithms.

Extensive studies in driving behaviour and traffic safety have been conducted through machine learning techniques and spatial analysis tools. I-Dreams presented the context-aware framework, named the "Safety Tolerance Zone" (STZ), based on applying machine learning techniques like LSTM and shallow NN to assess and improve driving behaviour in naturalistic on-road trails in Germany, Belgium, and the UK. Results indicated that real-time interventions and post-trip evaluation improved safety, and the NN performed better than the other algorithms [82]. Another study used violation data relating to traffic violations as a cause of traffic crashes, where georeferenced violation data coupled with spatial analysis methods like Inverse Distance Weighted interpolation were used to identify hotspots along expressways in Luzhou, China. The K nearest neighbours (KNN), Support Vector Machines (SVM), and CN2 Rule Inducer models were applied to classify and predict the type of violation. The analysis above shows that the model KNN, with $k=7$, Manhattan evaluation had higher prediction performance at the accuracy rate of 99%. These studies show potential in how advanced machine learning approaches and spatial analysis improve road safety by combating unsafe driving behaviours and attacking high-risk areas [83].

Generative AI is the growing theme for driver profiling with real-time analysis. However, deploying genAI models in real-time scenarios requires optimization in their inference times and demands high computational resources. To address this, lightweight architectures like TinyGANS and distilled versions of LLMs can be designed. Leveraging large language models in already-built frameworks for predictive analytics is one of the best places to refine driver behaviour and vehicle diagnostics analyses, which answers the fifth question from Sect. 2.1. Customization and refinement of LLMs for various applications, leveraging prompt-based learning techniques, can be explored in text generation tasks [64, 65].

7 Data sources and publically available datasets

Multiple data sources can be used for research and framework building for driver behaviour classification and related tasks in this domain. Several data sources mentioned in Table 4 highlight the prominent applications. These data types provide a comprehensive understanding of driving patterns and vehicle dynamics and allow for in-depth profiling

Table 4 Datasets and their applications

Data sources	Example applications	References
CAN-BUS data	Detecting aggressive driving, eco-driving analysis	[15, 37, 39]
OBD-II data	Driver behaviour analysis, Driver scoring	[12, 15, 46]
Smartphones (IMU sensors)	Detecting distracted driving, sudden acceleration/braking	[44, 45]
GPS data	Route Tracking/ Speed compliance	[48, 49]
Wearable technologies	Stress-induced behaviour detection, fatigue monitoring, driver behaviour classification	[39, 40]
Vehicle cameras	Road hazard identification, monitoring facial cues	[37, 42, 50]
In-car microphones	Identify in-car distractions, driver focus monitoring	[53]
LiDAR/radar systems	Lande detection, pedestrian proximity analysis	[43, 46]
Traffic and road conditions data	Driving adjustments based on traffic weather impact on behaviour	[40, 41, 49]
Simulators	Driver behaviour profiling, training and classification	[37, 65, 66]
Accelerometers	Harsh driving detection, fatigue monitoring and crash analysis	[53, 67–69]

and predictive analysis. Combining data from these sources makes it possible to develop more accurate models with enhanced safety and performance.

Following this, several publicly available datasets are crucial in advancing research in driver behaviour analysis. They serve as a benchmark for developing and validating ML and DL models. Notable repositories include UCI Machine Learning and Open Data Portal, which hosts driving-related datasets—Table 5 lists publicly available datasets for driver profiling and analysis.

8 Conclusion

The integration of advanced technologies such as ML, DL, and GenAI has significantly contributed to the automotive industries for driver behaviour analysis and predictive vehicle maintenance. The paper makes crucial steps forward in profiling drivers and advancing diagnostics of vehicles on account of massive data from connected cars. Accuracies are enhanced for tasks like driver behaviour classification, scoring, and profiling that draw conclusions from real-time sensor data telematics and other similar data sources. This boosted road safety while optimizing the performance of the developed frameworks. LLMs play a crucial role in personalized recommendations and feedback to ensure the safety of drivers and maximize driving efficiency. However, these systems have specific challenges like data imbalance, scarcity, and the creation of multiple data sources for btodbuildsolutions. Real-time data processing, predictions, and evaluation are other hurdles that require efficient algorithms and quality data without compromising accuracy. Explainable AI can be integrated to build trust in the AI systems. Advancements in generative AI help mitigate some challenges with the power of generating synthetic data, accurate analysis of larger datasets, and comprehensive learning from heterogeneous data.

9 Future directions

For future research, the focus is on studying different data sources for analysis of driver behaviour in order to enhance safety and accurate results. Driver behaviour profiling score can be an essential input in evaluating safety. Correlation of driver profiles and user feedback can be obtained for a reliable profiling scheme. Techniques like transfer learning and hybrid AI models can be explored, whereas models like CNN and RNN can be used for temporal analysis. Additionally, GANs can facilitate the creation of synthetic datasets, and the use of genAI and LLMs can help for personal driving insights and sophisticated feedback systems. Federated learning systems can be developed to enable decentralized training of the models to prioritize safety and privacy while implementing these solutions in the automotive domain.

Table 5 Publicly available datasets for driver profiling and analysis

Dataset name	Description	References
UAH-DriveSet	A dataset for analyzing driver behaviour and human factors collected from multiple drivers	[70, 71]
SHRP2 (Strategic Highway Research Program)	Contains driver behaviour data for various conditions to enhance safety and performance	[72]
LARA (Lane and Roadway assessment)	Captures driver interactions with roadway features focusing on lane and roadway conditions	[73]
Driving Dataset	A comprehensive dataset of driving behaviours collected via smartphones and onboard diagnostics	[74]
CAR Naturalistic Driving Dataset	Captures data from various driving environments to analyze driver behaviour	[75, 76]
Drive360	Data on vehicle dynamics in different conditions	[77]
NHTSA Fatality Analysis Reporting system	Comprehensive data on fatal crashes for analyzing risky driver behaviours	[78, 79]
NADS (National Advanced Driving Simulator)	Driving scenarios in controlled conditions	[80, 81]

Acknowledgements We extend our heartfelt gratitude to Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India, for offering invaluable support and an exceptional platform to carry out our research work. We deeply appreciate the nurturing environment that enabled us to learn, grow, and explore new horizons.

Author contributions V.S. wrote the main manuscript text, including the Introduction, Research Methodology, Bibliometric Analysis, Related Work, Data Sources and Publicly Available Datasets, Summary, Conclusion, and Future Directions. A.S. supervised the research process and provided significant input into the refinement of the Research Questions and the overall structure of the manuscript. A.S. reviewed and revised the Bibliometric Analysis, Related Work, and Conclusion sections, offering critical feedback to ensure relevance to the field. P.D. contributed to the Future Directions section by identifying emerging trends and proposing future research avenues. P.D. also reviewed the entire manuscript and provided valuable revisions to improve clarity and coherence. All authors reviewed and approved the final version of the manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Qu F, Dang N, Furht B, Nojoumian M. Comprehensive study of driver behavior monitoring systems using computer vision and machine learning techniques. J Big Data. 2024;11(1):32.
2. Herbers E, Miller M, Neurauter L, Walters J, Glaser D. Exploratory development of algorithms for determining driver attention status. Hum Factors. 2023;66(9):2191–204. https://doi.org/10.1177/00187208231198932.
3. Chai C, Li J, Islam MM, Feng R, Lu M. Hierarchical automated machine learning approach for self-optimizable driving distraction recognition based on driver's lane-keeping performance. Transp Res Record. 2024;2678(6):301–17. https://doi.org/10.1177/03611981231196152.
4. Alam W, Wang H, Pervez A, Safdar M, Jamal A, Almoshaogeh M, Al-Ahmadi HM. Analysis and prediction of risky driving behaviors using fuzzy analytical hierarchy process and machine learning techniques. Sustainability. 2024;16(11):4642.
5. Prabhu P, Athavale AA, Singh V. Development of an automated report generator using Llms and storytelling frameworks to support broadcasting in construction projects.
6. Guo B, Hansen J. Understand driving behaviors based on comprehensive grading system and unsupervised learning (No. 2024-01-2398). SAE Technical Paper; 2024.
7. Naidu KB, Vickram AS, Udhayakumar S. An early warning system of fatigue detection of drivers using novel image processing and HOG for evaluation of the driver quality using Viola-Jones over hog algorithm. AIP Conf Proc. 2024;2853(1):020141.
8. Naidu KB, Udhayakumar S. Evaluating the driver behavior and car movement and assessment of third party insurance using novel AdaBoost classifier over SVM. AIP Conf Proc. 2024;2853(1):020001.
9. Sharma PK, Chakraborty P. A review of driver gaze estimation and application in gaze behavior understanding. Eng Appl Artif Intell. 2024;133: 108117.
10. Dao MS, Zettsu K. Near-miss accident prediction on the edge: a real-time system for safer driving. In: Proceedings of the 2024 International Conference on Multimedia Retrieval. 2024; p. 1165–1169.
11. Hemalatha GV, Lakshmi SV, Karthick V. An effective framework to predict visual behavior of driver drowsiness level using improved ensemble method over XGBoost algorithm. AIP Conf Proc. 2024;2853(1):20076.
12. Jin Y. Detecting distracted drivers using transfer learning and image classification models. In: Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering, November 2023. p. 362–367.
13. Raza A, Akhtar I, Abualigah L, Zitar RA, Sharaf M, Daoud MS, Jia H. Preventing road accidents through early detection of driver behavior using smartphone motion sensor data: an ensemble feature engineering approach. IEEE Access. 2023;11:138457–71.
14. Chen S, Yao H, Qiao F, Ma Y, Wu Y, Lu J. Vehicles driving behavior recognition based on transfer learning. Expert Syst Appl. 2023;213: 119254.
15. Kumar R, Jain A. Driving behaviour analysis and classification by vehicle OBD data using machine learning. J Supercomput. 2023;79(16):18800–19.
16. Lee CH, Yang HC. A privacy-preserving learning method for analyzing hev driver's driving behaviors. IEEE Access. 2023;11:76816–26.
17. Campos-Ferreira AE, Lozoya-Santos JDJ, Tudon-Martinez JC, Mendoza RAR, Vargas-Martínez A, Morales-Menendez R, Lozano D. Vehicle and driver monitoring system using onboard and remote sensors. Sensors. 2023;23(2):814.
18. Bouhsissin S, Sael N, Benabbou F. Driver behavior classification: a systematic literature review. IEEE Access. 2023;11:14128–53.

19. Malik M, Nandal R. A framework on driving behavior and pattern using Onboard diagnostics (OBD-II) tool. Mater Today Proc. 2023;80:3762–8.
20. Abdelrahman AE, Hassanein HS, Abu-Ali N. Robust data-driven framework for driver behavior profiling using supervised machine learning. IEEE Trans Intell Transp Syst. 2020;23(4):3336–50.
21. Al-refai G, Al-refai M, Alzu'bi A. Driving style and traffic prediction with artificial neural networks using onboard diagnostics and smartphone sensors. Appl Sci. 2024;14(12):5008.
22. Li DC, Lin MYC, Chou LD. Macroscopic big data analysis and prediction of driving behavior with an adaptive fuzzy recurrent neural network on the Internet of vehicles. IEEE Access. 2022;10:47881–95.
23. Bouhsissin S, Sael N, Benabbou F, Soultana A. Enhancing machine learning algorithm performance through feature selection for driver behavior classification. Indonesian J Electr Eng Comput Sci. 2024;35(1):354–65.
24. Canal R, Riffel FK, Gracioli G. Machine learning for real-time fuel consumption prediction and driving profile classification based on ECU data. IEEE Access. 2024;12:68586–600.
25. Singh SK, Maan V, Singh AK. OBD II-based performance analysis based on fuel usage of vehicles. In: International Conference on Emerging Trends in Expert Applications & Security. Singapore: Springer Nature; 2023.
26. Rykała M, et al. Modeling vehicle fuel consumption using a low-cost OBD-II interface. Energies. 2023;16(21):7266.
27. Mahaalakshmi B, Padmapriya R, Cloudin S. Smart vehicle health monitoring system based on ARIMA and ordinal logistic regression for analysis of vehicle parameters. In: 2023 International Conference on System, Computation, Automation and Networking (ICSCAN), November 2023; IEEE. p. 1–6.
28. Cummins L, Sommers A, Ramezani SB, Mittal S, Jabour J, Seale M, Rahimi S. Explainable predictive maintenance: a survey of current methods, challenges and opportunities. IEEE Access. 2024;12:57574–575602.
29. Fan Y. Detection of driver fatigue driving behaviour based on deep learning algorithm from the perspective of traffic psychology. J Psychol Afr. 2022;32(3):303–9. https://doi.org/10.1080/14330237.2022.2075622.
30. Arena F, Collotta M, Luca L, Ruggieri M, Termine FG. Predictive maintenance in the automotive sector: a literature review. Math Comput Appl. 2021;27(1):2.
31. Ersöz OÖ, İnal AF, Aktepe A, Türker AK, Ersöz S. A systematic literature review of the predictive maintenance from transportation systems aspect. Sustainability. 2022;14(21):14536.
32. Ma S, Yan Y, Wang J, Chen D, Yang J, Liu X. Predictability of Driver's stop/go decisions at flashing-light-controlled grade crossings by coupling functional brain network and deep learning methods. Transp Res F Traffic Psychol Behav. 2024;107:115–32.
33. AlKheder S. Experimental road safety study of the actual driver reaction to the street ads using eye tracking, multiple linear regression and decision trees methods. Expert Syst Appl. 2024;252: 124222.
34. Mosleh B, Hamdan J, Sababha BH, Alqudah YA. Embedded machine learning-based road conditions and driving behavior monitoring. Int J Electr Comput Eng. 2024;14(3):2571–82.
35. Roussou S, Michelaraki E, Katrakazas C, Afghari AP, Al Haddad C, Alam MR, et al. Unfolding the dynamics of driving behaviour: a machine learning analysis from Germany and Belgium. Eur Transp Res Rev. 2024;16(1):40.
36. Hosseini MK, Talebpour A, Devunuri S, Hamdar SH. An unsupervised learning framework for detecting adaptive cruise control operated vehicles in a vehicle trajectory data. Expert Syst Appl. 2022;208: 118060.
37. Ali Y, Hussain F, Bliemer MC, Zheng Z, Haque MM. Predicting and explaining lane-changing behaviour using machine learning: a comparative study. Transp Res C Emerg Technol. 2022;145: 103931.
38. Yu W, Huang Q. A deep encoder-decoder network for anomaly detection in driving trajectory behavior under spatio-temporal context. Int J Appl Earth Obs Geoinf. 2022;115: 103115.
39. Zhao X, Li Z, Zhao C, Wang C. Real-time multistep time-series prediction of Driver's head pose during IVIS secondary tasks for Human-Machine codriving and distraction warning systems. IEEE Sens J. 2022;22(24):24364–79.
40. Reyes Jr. FC, Cayabyab GT, Mendoza PJC. Analyzing truck driver's behavior on the road using YOLO v4 tiny algorithm. In: Proceedings of the 2022 6th International Conference on Computer Science and Artificial Intelligence, December 2022. p. 177–183.
41. Ghandour R, Potams AJ, Boulkaibet I, Neji B, Al Barakeh Z. Driver behavior classification system analysis using machine learning methods. Appl Sci. 2021;11(22):10562.
42. Xie M, Xiao Y, Wang H, Wang Y. Recognition and evaluation of driving behavior based on MEMS sensors and machine learning. Adv Transp Stud. 2021;55:185.
43. Sohail AM, Teh YW, Khan N, Khan Z, Koubaa A. A comprehensive review: analysis of machine learning, deep learning, and large language model techniques for revolutionizing driver identification. 2024. Authorea Preprints.
44. Li C et al. Learning to rewrite prompts for personalized text generation. In: Proceedings of the ACM on Web Conference 2024.
45. Pu H, Yang X, Li J, Guo R. AutoRepo: a general framework for multimodal LLM-based automated construction reporting. Expert Syst Appl. 2024;255: 124601.
46. Andreoni M, Lunardi WT, Lawton G, Thakkar S. Enhancing autonomous system security and resilience with generative AI: a comprehensive survey. IEEE Access. 2024;12:109470–93.
47. Behera S, Misra R. Generative adversarial networks based on remaining useful life estimation for IIoT. Comput Electr Eng. 2021;92: 107195.
48. Cui C, Ma Y, Cao X, Ye W, Zhou Y, Liang K, Chen J, Lu J, Yang Z, Liao K-D, Gao T, Li E, Tang K, Cao Z, Zhou T, Liu A, Yan X, Mei S, Cao J, Wang Z, Zheng C. A survey on multimodal large language models for autonomous driving. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024; p. 958–979.
49. Deeluea S, Jeenanunta C, Tunpun A. Fault prediction model for motor and generative adversarial networks for acceleration signal generation. In: 2022 17th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). November 2022; IEEE. p. 1–5
50. Ćelić J, Bronzin T, Horvat M, Jović A, Stipić A, Prole B, Maričević M, Pavlović I, Pap K, Mikota M, Jelača N. Generative AI in E-maintenance: myth or reality? Artif Intelligence. 2024;7:8–10. https://doi.org/10.1109/MIPRO60963.2024.10569282.
51. Bhattacharyya R, Wulfe B, Phillips DJ, Kuefler A, Morton J, Senanayake R, Kochenderfer MJ. Modeling human driving behavior through generative adversarial imitation learning. IEEE Trans Intell Transp Syst. 2022;24(3):2874–87.

52. Kumar V, Jain S, Soni N, Saran A. Drive GPT–an AI based generative driver model (No. 2024-26-0025). SAE Technical Paper; 2024.
53. Ma H, Li J, Zhan W, Tomizuka M. Wasserstein generative learning with kinematic constraints for probabilistic interactive driving behavior prediction. In: 2019 IEEE Intelligent Vehicles Symposium (IV) (p). June 2019; IEEE. p. 2477–2483.
54. Angelini SA. (2024). Generative AI for the Simulation of Autonomously Driven Vehicles (Doctoral dissertation, Politecnico di Torino).
55. Bao N, Carballo A, Kazuya T. Prediction of personalized driving behaviors via driver-adaptive deep generative models. In 2021 IEEE Intelligent Vehicles Symposium (IV). July 2021; IEEE. p. 616–621
56. Parameshachari BD, Srinivas C, Leena Shruti HM, Manasa DS, Kumar K, Rashmi P. GPS data for behaviour style prediction in driving using optimized AI model. In: 2023 International Conference on Data Science and Network Security (ICDSNS). July 2023; IEEE. p. 1–8.
57. Sun L, Zhan W, Hu Y, Tomizuka M. Interpretable modelling of driving behaviors in interactive driving scenarios based on cumulative prospect theory. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC). October 2019; IEEE. p. 4329–4335.
58. Parthasarathy VB, Zafar A, Khan A, Shahid A. The ultimate guide to fine-tuning llms from basics to breakthroughs: an exhaustive review of technologies, research, best practices, applied research challenges and opportunities. 2024; arXiv preprint arXiv:2408.13296.
59. Cui C, Ma Y, Cao X, Ye W, Wang Z. Receive, reason, and react: drive as you say, with large language models in autonomous vehicles. IEEE Intelligent Transp Syst Magazine. 2024;16(4):81–94.
60. Azadani MN, Boukerche A. Driving behavior analysis guidelines for intelligent transportation systems. IEEE Trans Intell Transp Syst. 2021;23(7):6027–45.
61. Mehdipour N, Althoff M, Tebbens RD, Belta C. Formal methods to comply with rules of the road in autonomous driving: State of the art and grand challenges. Automatica. 2023;152: 110692.
62. Wawage P, Deshpande Y. Smartphone sensor dataset for driver behavior analysis. Data Brief. 2022;41: 107992.
63. Chen JC, Lee CY, Huang PY, Lin CR. Driver behavior analysis via two-stream deep convolutional neural network. Appl Sci. 2020;10(6):1908.
64. Wu Y. Large language model and text generation. In: Natural language processing in biomedicine: a practical guide. Cham: Springer International Publishing; 2024. p. 265–97.
65. Liu Z, Yang Q, Wang A, Gu X. Vehicle driving safety of underground interchanges using a driving simulator and data mining analysis. Infrastructures. 2024;9(2):28.
66. Calvi A, Cafiso SD, D'Agostino C, Kieć M, Petrucci G. A driving simulator study to evaluate the effects of different types of median separation on driving behavior on 2+ 1 roads. Accid Anal Prev. 2023;180: 106922.
67. Chhabra R, Singh S, Khullar V. Privacy enabled driver behavior analysis in heterogeneous IoV using federated learning. Eng Appl Artif Intell. 2023;120: 105881.
68. Sethuraman R, Sellappan S, Shunmugiah J, Subbiah N, Govindarajan V, Neelagandan S. An optimized AdaBoost multi-class support vector machine for driver behavior monitoring in the advanced driver assistance systems. Expert Syst Appl. 2023;212: 118618.
69. Chen Y, Wang K, Lu JJ. Feature selection for driving style and skill clustering using naturalistic driving data and driving behavior questionnaire. Accid Anal Prev. 2023;185: 107022.
70. Lv B, Yue R, Zhang Y. The influence of different factors on right-turn distracted driving behavior at intersections using naturalistic driving study data. IEEE Access. 2019;7:137241–50.
71. Bandyopadhyay S, Datta A, Sachan S, Pal A. Unsupervised driving behavior analysis using representation learning and exploiting group-based training. 2022; arXiv preprint arXiv:2205.07870.
72. Hao H, Li YE, Medina A, Gibbons RB, Wang L. Understanding crashes involving roadway objects with SHRP 2 naturalistic driving study data. J Safety Res. 2020;73:199–209.
73. El-Maissi AM, Argyroudis SA, Nazri FM. Seismic vulnerability assessment methodologies for roadway assets and networks: a state-of-the-art review. Sustainability. 2020;13(1):61.
74. Geyer J, Kassahun Y, Mahmudi M, Ricou X, Durgesh R, Chung AS, Hauswald L, Pham VH, Mühlegg M, Dorn S, Fernandez T, Jänicke M, Mirashi S, Savani C, Sturm M, Vorobiov O, Oelker M, Garreis S, Schuberth P. A2d2: Audi autonomous driving dataset. 2020; arXiv preprint: arXiv:2004.06320.
75. Singh H, Kathuria A. Analyzing driver behaviour under naturalistic driving conditions: a review. Accid Anal Prev. 2021;150: 105908.
76. Kovaceva J, Isaksson-Hellman I, Murgovski N. Identification of aggressive driving from naturalistic data in car-following situations. J Safety Res. 2020;73:225–34.
77. Zhao Y, Li J, Huang R, Li B, Luo A, Li Y, Cheng H. Attention-based deep driving model for autonomous vehicles with surround-view cameras. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). October 2022, IEEE. p. 286–292
78. https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/FARS/.
79. Philips CR, Garrett RC, Tatro AJ, Fisher TJ. An analysis of crash-safety ratings and the true assessment of injuries by vehicle. Comput Statistics. 2021;36:1639–60.
80. https://dsri.uiowa.edu/nads-1.
81. Ziakopoulos A, Tselentis D, Kontaxi A, Yannis G. A critical overview of driver recording tools. J Safety Res. 2020;72:203–12.
82. Roussou S, Garefalakis T, Michelaraki E, Brijs T, Yannis G. Machine learning insights on driving behaviour dynamics among Germany, Belgium, and UK drivers. Sustainability. 2024;16(2):518.
83. Zahid M, Chen Y, Jamal A, Al-Ofi KA, Al-Ahmadi HM. Adopting machine learning and spatial analysis techniques for driver risk assessment: Insights from a case study. Int J Environ Res Public Health. 2020;17(14):5193.