Qu *et al. Journal of Big Data* (2024) 11:32
https://doi.org/10.1186/s40537-024-00890-0

Journal of Big Data

SURVEY Open Access

Comprehensive study of driver behavior monitoring systems using computer vision and machine learning techniques

Fangming Qu[1], Nolan Dang[1], Borko Furht[1] and Mehrdad Nojoumian[1*]

*Correspondence:
mnojoumian@fau.edu

[1] Department of Electrical Engineering and Computer Science, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, United States

Abstract

The flourishing realm of advanced driver-assistance systems (ADAS) as well as autonomous vehicles (AVs) presents exceptional opportunities to enhance safe driving. An essential aspect of this transformation involves monitoring driver behavior through observable physiological indicators, including the driver's facial expressions, hand placement on the wheels, and the driver's body postures. An artificial intelligence (AI) system under consideration alerts drivers about potentially unsafe behaviors using real-time voice notifications. This paper offers an all-embracing survey of neural network-based methodologies for studying these driver bio-metrics, presenting an exhaustive examination of their advantages and drawbacks. The evaluation includes two relevant datasets, separately categorizing ten different in-cabinet behaviors, providing a systematic classification for driver behaviors detection. The ultimate aim is to inform the development of driver behavior monitoring systems. This survey is a valuable guide for those dedicated to enhancing vehicle safety and preventing accidents caused by careless driving. The paper's structure encompasses sections on autonomous vehicles, neural networks, driver behavior analysis methods, dataset utilization, and final findings and future suggestions, ensuring accessibility for audiences with diverse levels of understanding regarding the subject matter.

Keywords: Driver behavior monitoring system, Driver behavior classification, Computer vision, Machine learning, Autonomous vehicles

Introduction

Recently, an increasing focus has been on creating self-driving or autonomous vehicles - vehicles that can operate without human intervention. This development has opened up new ways to increase safety in these vehicles. A key aspect is the capability to understand and keep an eye on what is happening inside the vehicle, particularly with the driver. A review conducted by researchers from Japan [1] shows that driver inattention was a leading cause of most traffic accidents. Researchers have extensively studied this issue, categorizing driver inattention into two primary types: visual distraction and fatigue. Detecting and mitigating driver inattention requires a multifaceted approach, incorporating subjective reports, driver biological indicators, physical measurements,

Springer Open

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

driving performance assessments, and hybrid measures that combine multiple indicators. Hybrid measures, in particular, offer more reliable and accurate solutions compared to relying solely on a single measure. However, commercial products for driver inattention monitoring exist, and their effectiveness in actual driving conditions may be limited. An ideal driver inattention monitoring system for safety enhancement integrates driver physical variables, driving performance metrics, and data from the In-Vehicle Information System (IVIS) while considering the driving environment. This research aims to develop AI-based monitoring software for autonomous vehicles, contributing to a safer and more secure transportation landscape.

In this survey paper, an AI-based driving assistant is proposed that can see and interpret the inside of the vehicle using a branch of artificial intelligence algorithms, which allows computers to learn and make decisions from data: an AI-based offline monitoring system to boost the safety of these autonomous vehicles. This system is designed to assist the driver and issue warning alerts if the driver seems to be not paying attention to the road without any data privacy concerns. By combining three distinct classification methods to detect fatigued drivers, a software system that works on an autonomous vehicle acts as an intelligent driving assistant. There are various machine learning methods, known as neural networks, for analyzing these behaviors. These include artificial neural networks (ANNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs).

Figure 1 below illustrates the roadmap of driver behavior monitoring systems.

This paper first consists of the basic definitions of complex terms in computer science. These include autonomous vehicles, vision systems, machine learning, driver behavior classification, deep learning, convolutional neural networks, recurrent neural networks, and artificial neural networks. Then, the paper elaborates upon the idea behind the proposed system, including the different step-by-step computational procedures, or algorithms, for each function and how they all come together. In the fourth section, the operation of this proposed system and how each dataset can contribute meaningfully to the training process of this algorithm are illustrated. Lastly, the survey shows how all these components come together into a completed product. This research aims to create an AI-based monitoring software to enhance the safety of autonomous vehicles, which will be a significant step in safer and more secure transportation.

Preliminary materials

This discussion begins by defining key terms crucial for understanding the topics explored in this article. These concepts form the bedrock of the ideas presented herein, and a detailed explanation of each follows in subsequent sections.

Autonomous vehicle

We should first emphasize that the driver behavior monitoring systems have been used in human-driving vehicles for a long time, for instance, by insurance companies through mobile apps or small hardware equipment. However, these systems are more critical for autonomous vehicles as well as vehicles that utilize advanced driver-assistance systems, i.e., using a level of autonomy, because they may face unpredictable situations requiring the driver's intervention.

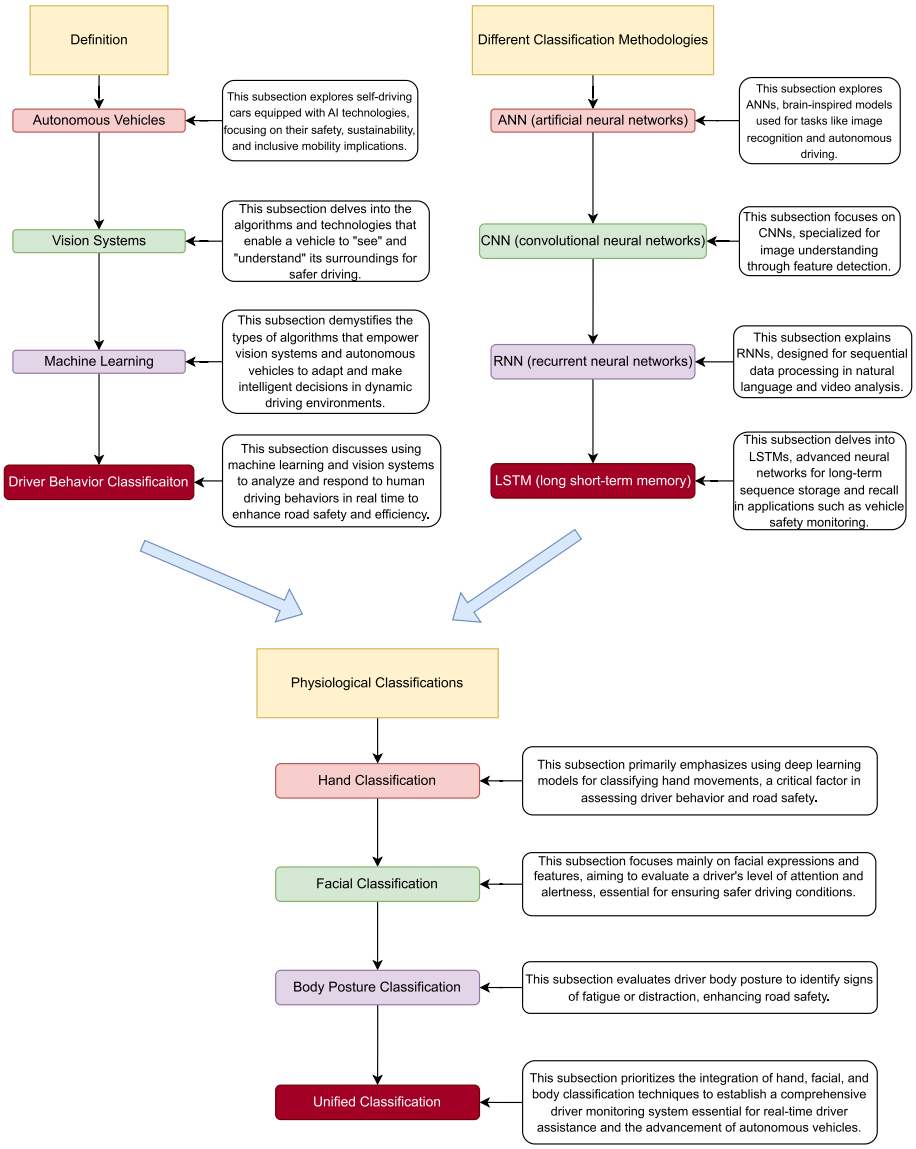


Fig. 1 Roadmap of driver behavior monitoring systems

An autonomous vehicle with any levels of autonomy revolves around a vehicle that can operate without (or with minimum) human intervention. These vehicles are designed to navigate and drive themselves, relying on advanced technologies and systems rather than requiring a human driver. These vehicles contain complex computer systems that often utilize artificial intelligence. AI acts similarly to the brain, processing information and making decisions. It gathers data from its surroundings via sensors, which function as the vehicle's eyes, then uses it to navigate safely, just as a human driver would. Vehicles that drive themselves have the potential to change our world. AI technologies could transform how people and goods travel, advance military and security operations, and provide a new level of freedom to those unable to drive. Furthermore, ADAS and AVs are potentially expected to make roads safer and

save fuel, providing better transportation options, especially for those who have difficulties in driving, and reshaping society's transportation approach entirely.

Autonomous vehicles offer potential benefits compared to human drivers, as reported in [2]. Firstly, while AVs can enhance safety by reducing accidents, it's essential to recognize that both AVs and human drivers share equal responsibility for preventing accidents. The report notes that many traffic accidents result from unsafe driving behaviors or drowsy driving. Secondly, they reduce traffic congestion by optimizing traffic flow and improving communication among road vehicles through in-between vehicles telecommunication, addressing the issue of inefficient traffic management. Additionally, using AI technologies, AVs enhance accessibility and mobility for individuals unable to drive, thus promoting inclusivity and providing cost savings for those who opt for AV transportation instead of owning a vehicle, addressing economic considerations.

There are also disadvantages to autonomous vehicles. For example, addressing the technological challenges of software and hardware systems is essential. This includes developing robust algorithms, ensuring effective communication between components, and enhancing overall system reliability and safety. Then, adopting AVs raises concerns about job displacement in the driving and transportation sectors, potentially resulting in unemployment and necessitating retraining or job transition programs. Also, ethical and legal considerations arise when adapting laws to accommodate autonomous technologies. Addressing liability in accidents, decision-making processes in critical situations, and establishing an ethical framework for AV behavior are crucial for public trust and safety. Additionally, cybersecurity threats must be considered, as hackers targeting AVs could gain control over operations and endanger passengers and other road users. Robust cybersecurity measures should be implemented to prevent such risks. Furthermore, privacy concerns of AV drivers require attention, with clear guidelines and safeguards to protect personal information collected by autonomous vehicles. Moreover, the reliance on infrastructure and connectivity challenges widespread adoption and effectiveness. Consistent and reliable support, including road markings, traffic signals, and communication networks, is vital for successfully integrating autonomous technologies on a large scale. Despite these challenges, the overall advantages of AVs outweigh the disadvantages. These are in addition to other human–machine interaction (HMI) challenges such as cross-cultural expectations from self-driving cars [3, 4], passengers' trust [5, 6], social acceptability [7, 8], and customized autonomous driving technologies [9, 10].

While complete vehicle automation is yet to be commonplace soon, the interpretation of driver behavior is crucial for partially and conditionally automated vehicles. These vehicles, which require either the driver's readiness to regain control at any moment or their intervention when the vehicle cannot perform certain critical operations, are predicted to be the dominant form in the market until 2030. Since these systems are automated, they still heavily rely on human supervision and intervention [11–13].

Machine learning

Machine learning is a sub-field of computer science that gives computers the capacity to learn from data and subsequently make informed decisions or predictions. This concept embodies the machine equivalent of a human brain, utilizing a variety of algorithms and

statistical models to learn and adapt over time. Three primary types of learning exist in this context: supervised, unsupervised, and reinforcement learning.

The algorithm assumes a student-like role in supervised learning, with data presented as question-and-answer pairs serving as the tutor. Through exposure to this data, the algorithm learns the pattern of problem-solving and, over time, acquires the ability to solve similar problems independently. Supervised learning is analogous to a learning paradigm in which a student learns by being presented with a problem and the corresponding solution.

Unsupervised learning, in contrast, presents the algorithm with a dataset without the provision of predefined solutions. The focus falls on the algorithm to identify patterns, correlations, and relationships within the data. This method is similar to a student given tasks without an explicit solution, necessitating independent pattern recognition and problem-solving.

Reinforcement learning takes a different approach, embodying a process of repetitive learning through trials and errors, reminiscent of learning to ride a bicycle. Each attempt and subsequent failure offers the algorithm new insights, adjusting its future decisions based on gathered experiences.

The utility of machine learning pervades a multitude of sectors in contemporary times. From facial recognition capabilities on smartphones to stock market trend predictions, its applications span sectors including healthcare, finance, retail, and transportation. Machine learning equips us with the tools to identify patterns within extensive data sets and informs data-driven decision-making processes.

In the following sections, this paper will comprehensively explore different learning types, their associated algorithms, and the vast array of applications where machine learning is harnessed. This survey aims to present an inclusive examination covering the extensive spectrum of machine learning. By examining the diverse algorithms, learning models, and machine learning applications, this paper aims to provide a holistic view of this rapidly evolving field.

Driver behavior classification

Recognizing and understanding the unique driving behaviors of individuals is crucial for enhancing driver awareness. This is where driver behavior classification draws its significance, acting as a sophisticated observer that continually monitors and analyzes the driver's actions, including hand movements, facial expressions, and body posture. By leveraging advanced technologies including vision systems, machine learning algorithms, and sensor data, driver behavior classification aims to provide real-time feedback and promote safer driving practices.

Driver behavior classification encompasses multiple aspects. Firstly, hand classification monitors the driver's hand movements. For instance, if the system detects the driver's hand off the wheel, including reaching for a coffee cup, it gently reminds them to keep their hands on the wheel. Secondly, facial classification focuses on analyzing facial expressions that may indicate fatigue or distraction. If the driver's eyes frequently close, indicating drowsiness, the system can trigger an alert or activate the autonomous system if necessary. Lastly, body posture classification examines the driver's posture and movements. If the driver starts slouching after extended periods of driving, it suggests

the need for a break or seat adjustments to enhance comfort. These driver behavior classifications serve as indicators of potential safety concerns. Considering the importance of transportation safety and the seamless integration of autonomous vehicles, it is essential to also address the legal perspectives and regulatory challenges associated with these advanced systems. As autonomous vehicles become more prevalent, their successful integration relies on the ability to identify and adapt to human driving behaviors. This adaptability ensures that autonomous vehicles can assimilate naturally into existing traffic flows, promoting safer and more efficient travel.

Artificial neural networks

Artificial neural networks are computational models directly inspired by the structural and functional characteristics of the human brain. They embody a network of interlinked artificial neurons that collaboratively function to learn from data and generate predictions. Similar to the data processing mechanism of the human brain, an ANN accepts inputs, performs calculations, and yields outputs. ANNs can be compared to a well-coordinated team of professionals, each having a defined role, passing information in a relay. Each member takes input, performs a specific calculation, and forwards the resulting data to the next member. This process iterates until the final member delivers the output. During its training phase, an ANN adjusts its calculations based on provided examples, essentially learning through fine-tuning. It aims to find the most effective methodology to make accurate predictions or decisions.

A noteworthy aspect of ANNs is their ability to comprehend complex patterns and relationships in data, even revealing associations that may not be immediately apparent. As a result, ANNs are invaluable in diverse applications including image recognition, natural language understanding, and even autonomous driving. Integrating different types of ANNs, including recurrent neural networks for sequential data processing and convolutional neural networks for image analysis, has made significant advancements in domains like natural language processing, image recognition, and autonomous vehicles.

In conclusion, ANNs are potent tools for learning from data and making predictions or decisions. They are universally employed across many fields to unravel complex problems and augment our understanding of and interaction with the world.

Convolutional neural network

A convolutional neural network stands as an essential component in the toolkit of deep learning techniques, especially in domains necessitating image understanding and interpretation. For humans, interpreting a picture is an intuitive process; however, for computers, the task is significantly more complex as it perceives the image as an array of tiny points or "pixels." CNNs facilitate computers in comprehending these pixels and their interrelationships.

CNN's operation begins with the input layer. The initial step is where the image data comes into the CNN model. In instances of a color image, the input layer receives information on the three primary colors: red, green, and blue. Following the input layer is the convolutional layer. During this phase, the layer traverses the image in small sections, referred to as filters or kernels, to construct a feature map. Feature maps enable CNN to recognize fundamental shapes or patterns, including lines or textures. Subsequently, the

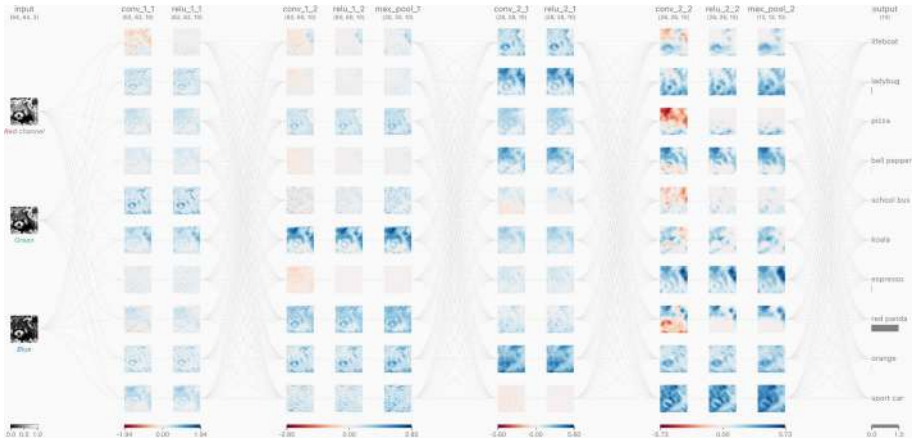


Fig. 2 Visualization of CNN sample based on the reference of [14]

activation function, typically the ReLU (Rectified Linear Unit), is applied. This function enhances the CNN's learning capabilities by nullifying any negative pixel values, thereby increasing the effectiveness of the CNN in identifying complex patterns. The subsequent pooling layer condenses the information derived from the convolutional layer by downsizing the feature map but preserving the essential features. The pooling process increases the efficiency of CNNs. Max pooling is a widespreadly deployed method, which retains only the highest value from each section of the feature map. After several iterations of these steps, the fully-connected layer is activated. This layer assimilates the information collected thus far to arrive at a final decision, categorizing the image appropriately. Finally, the Softmax function generates a probability distribution for each category, indicating the likelihood of the image belonging to each class. Essentially, a CNN transforms the raw pixels of an image into a categorized output, enabling the computer to interpret the image. In essence, CNNs can identify complex patterns and objects within an image, similar to human visual interpretations.

Figure 2 below provides a visual representation of a CNN sample, as elaborated in reference [14], offering a graphical insight into the model's structure and function.

Recurrent neural networks

Recurrent neural networks are powerful tools that allow computers to understand sequences of data, similar to a time radar. Sequences can range from sentences in text, frames in a video, or a series of numbers. RNNs enable computers to comprehend these sequences as a series of data points.

The initial step is the input layer. In the case of a sentence, the input layer receives vectors, which are representations of the input data. The next step is the recurrent layer, where RNNs pass information from one point to the next in the sequence. The sequential information exchange aids in understanding the order with the context of previous information. Sequentially, RNNs apply a comprehending function, including "ReLU", to the output of the previous layer, in order to understand the contextual meaning of the data by identifying intricate patterns within the sentence. Next, RNNs progress to the fully-connected layer, consolidating the knowledge into the decision-making process

which involves tasks including prediction and classification. Finally, RNNs employ the Softmax function, which assigns a probability score to each possible outcome.

The potential of RNNs extends even further as researchers continually explore ways to enhance their capabilities. One notable advancement is integrating spatiotemporal graphs into RNNs, enabling the capture of intricate, high-level structures involving both space and time. The results obtained from this novel technique have surpassed existing methods, whether in understanding human movements or interpreting interactions among objects. This significant progress represents a substantial leap forward, providing a potent tool to augment machine learning models and laying the foundation for more precise predictions and analyses in complex scenarios [15].

Long short-term memory (LSTM)

Long short-term memory (LSTM) units, a form of artificial intelligence, embody a neural network capable of storing, learning, and recalling patterns over prolonged periods. LSTMs use their data inputs to predict sequences, a mechanism that could bring significant advancements in speech recognition, natural language processing, and time series prediction. A groundbreaking use of LSTMs lies in vehicle safety, where real-time driver distraction detection becomes possible through analyzing long-term patterns in driving and head tracking data. With an impressive accuracy rate of up to 96.6 percent, LSTM-based approaches outdo traditional methods like support vector machines and show remarkable utility in handling time-series data. This notable application can lead to advanced vehicle safety systems development, improving road safety and further reflecting LSTM units' potential in enhancing AI applications [16].

Classifying drivers' behaviors in autonomous vehicles

Research conducted by scholars from leading universities broadly focuses on hand, facial, and body posture classifications to understand driver behaviors in autonomous vehicles. These classifications play a pivotal role in recognizing distracted or fatigued driving actions-like texting, yawning, or slouching-which could lead to accidents. Several machine learning techniques are repurposing the ways in which autonomous vehicle cabins can be monitored. Some involve video data to model human movements while others utilize a sensor network for improved hazard identification. Cost-effective, vision-based systems that detect driver inattention further endorse road safety. Collectively, these methods offer an encompassing understanding of driver behavior, contributing decisively to the evolution of safer autonomous driving systems [17–20].

Introduction to driver behavior classification

According to the World Health Organization (WHO) data for 2023, road traffic injuries emerge as a substantial global concern, accounting for approximately 1.3 million fatalities annually, with an additional 20 to 50 percent of cases resulting in non-fatal injuries that often lead to disabilities. It is noteworthy that over half of these road traffic deaths occur in low-income and middle-income countries, primarily affecting individuals with lower socioeconomic statuses who are more susceptible to being involved in such accidents [21].

In response to these alarming statistics, the National Highway Traffic Safety Administration (NHTSA) is actively addressing unsafe driving behaviors that are significant contributors to road accidents, injuries, and fatalities. These behaviors encompass drug-impaired driving, involving both alcohol and drugs; distracted driving, which includes activities like texting while driving; aggressive driving, characterized by behaviors such as tailgating and excessive speeding; drowsy driving; and the failure to use seat belts. These risky driving behaviors not only pose a grave threat to passenger safety but can also result in severe injuries and tragic fatalities [22].

Driver behavior classification is a pivotal aspect of studying human interaction with autonomous vehicles; it aims to distinguish the safe and risky driving behaviors. These behaviors are typically classified into two states in the datasets: safe and unsafe. "1" represents safe behaviors, and "0" denotes unsafe behaviors. This system utilizes binary classification. Safe behaviors include maintaining a steady speed, maintaining an appropriate gap from the vehicle in front, and regularly checking rear-view mirrors. In contrast, unsafe behaviors can involve texting, talking on the phone, operating the radio, drinking, reaching in back, doing hair and makeup, and talking to passengers. The causes of unsafe driving behavior are multifarious, ranging from driver fatigue and drowsiness to distraction and impairment due to substance use. Under specific circumstances, hostile driving or vehicular aggression could also lead to unsafe behaviors. Identifying the causes for unsafe driving is crucial to developing effective interventions and preventive measures. The classification of driver behaviors has the potential to indisputably enhance the protection of the driver. The classification can facilitate real-time monitoring and feedback, alerting drivers to potentially dangerous behaviors, and prompting corrective action. This classification additionally maintains the potential for long-term benefits, including providing driver education and training programs, as well as the ability to contribute to the design and development of enhanced, intuitive, and safer autonomous vehicle interfaces. Further, this classification also has the ability to feed into the development of advanced driver-assistance systems, thereby enabling these systems to better understand and predict human behavior; this may also serve to prevent accidents.

In this context, the review presented in [23] offers an inclusive framework that addresses the regulation of autonomous vehicles and associated challenges, both in the United States and Europe. Regulatory issues, particularly those pertaining to the legal interpretation of driver behavior, present complex problems. However, the potential for autonomous vehicles to impressively enhance road safety in developed countries justifies these efforts. In anticipation of this emerging technology, countries in the European Union are being primed to adjust their legal structures, and they are collaborating with lawmakers and technical experts to establish unambiguous guidelines and practical solutions for optimization of the AVs on the road.

Despite the challenges posed by regulatory issues, driver behavior technology is making significant strides in understanding and improving the interaction between drivers and autonomous vehicles. Research in driver behavior technology now focuses on distraction and fatigue issues. Driver state-of-mind analysis involves a blend of self-reports, biological metrics, driving performance indicators, and hybrid methods. The latter combines multiple data sources for a clearer picture and shows more accurate results by cutting down on false alerts and keeping high-performance ratings [24].

While deep learning excels in interpreting complex data for self-driving cars, it's crucial not to overlook the importance of traditional machine learning techniques like SVM, Decision Trees, and KNN. These methods play a vital role due to their computational efficiency, easy comprehensibility, and minimal data requirements. They form the backbone of strengthening the robustness and reliability of AI-powered monitoring systems in diverse driving situations [25–27].

Researchers have developed a new vision system using "vehicle dynamics data", which eliminates the need for cumbersome eye-tracking hardware. This approach with "support vector machines" (a machine learning model) has shown promising high classification rates, boosting the development of the next generation of autonomous driving aid systems. On another front, Kinect V2 sensors, commonly used in video gaming, have created a database of standard upper limb movements in healthy individuals. Initially proposed for rehabilitation, this method might help understand how drivers interact with vehicle controls. Understanding driving behavior has evolved with machine learning and deep learning models that draw on large-scale vehicle data. Deep learning methods, including neural networks, have shown high accuracy levels, signaling they may soon dominate driving behavior analysis as the technology progresses [28–30].

Hand classification

In 2018, a recent incident involving a Tesla Model S striking a fire truck in "Autopilot mode," which is a system in vehicles that automates certain aspects of control, including maintaining speed and staying within lanes, with human supervision, highlights the danger of keeping hands off the wheel, even in autonomous vehicles. According to the National Transportation Safety Board (NTSB), the driver had his hands off the wheel for most of the trip, receiving multiple alerts to place his hands back on the wheel. The vehicle accelerated towards the driver-set cruise control speed and collided with the parked fire truck while the Autopilot system failed to detect the driver's hands on the wheel. This incident, along with previous fatal crashes involving Tesla vehicles, emphasizes the potential importance in hand classification in autonomous vehicles for drivers' safety [31, 32].

Hand classification in driver monitoring systems involves monitoring and analyzing a driver's hand movements while operating a vehicle. It holds a pivotal position in evaluating the driver's behavior and ensuring their safety on the road. The classification system of hands provides valuable insights into the driver's actions and level of engagement by detecting whether the driver is holding the steering wheel properly, using turn signals appropriately, or reaching for objects within the vehicle. Deep learning models have emerged as promising approaches for recognizing specific hand actions and movements. In [42], with remarkable accuracy, a "pre-trained Keras Neural Network" was employed to classify hand presence: a pre-trained Keras Neural Network model was an already trained model on a large dataset; Keras, a Python Programming Language library, allows for the quickly building and testing these networks; Pre- trained, in this instance, means the model's initial weights come from an earlier training run, often from an embracing dataset like ImageNet. These pre-trained models already recognize common patterns, which can be adapted to new tasks, reducing training time and computational resources; Keras provides various pre-trained models, especially beneficial when the dataset is not

large enough to train a whole network from the beginning. This model is able to distinguish between one hand on the wheel, two hands on the wheel, or no hands on the wheel. By utilizing this deep learning model and a carefully selected hand- classifying dataset comprising data from 30 volunteers, the system achieved an impressive 100 percent accuracy. Despite reaching 100% accuracy with a limited dataset, the experiment showcases the promise of transfer learning for visual tasks. Its successful predictions for consecutive images suggest broader applicability beyond controlled conditions.

Another area of exploration involves using a multi-camera framework complemented by contact sensors for hand classification in driver monitoring systems. This approach enables a fusion of visual and sensory data for more precise and thorough hand detection and tracking by utilizing multiple cameras mounted at different locations inside the cabinet strategically placed within the vehicle cabin. It enhances the system's ability to accurately classify hand movements and gestures, thereby seriously improving traffic safety and reducing accidents caused by distracted driving. In [43], the recent development in hand classification using contact sensors has demonstrated high accuracy, with the Adaptive Least-Squares Support Vector Machine model achieving up to 92.9% accuracy in gesture recognition. Multi-camera framework along with contact sensor data aligns with the broader objective of vision systems and machine learning analysis within autonomous vehicle cabins, ensuring a complete understanding of the hand classification and facilitating a more accurate hand classification. Furthermore, another paper [44] demonstrates that the new method of multi-modal fatigue detection, RPPMT-CNN-BiLSTM, is able to combine improved feature extraction and 1D neural networks for enhanced accuracy, achieving 98.2% accuracy on the Multi-Modal Driver Alertness Dataset (MDAD), showing significant development in the accuracy of fusion approaches. In summary, hand classification in driver monitoring systems benefiting from both contact sensors and multi-model deep learning approaches is critical for assessing driver behavior and promoting safe driving practices. Deep learning models, including the pre-trained Keras Neural Network and RestNet CNN, demonstrate the potential for accurate hand presence classification when integrated with contact sensor technology [40].

Table 1 offers an organized summary of key research studies focusing on hand classification, presenting an overview of significant contributions in this area.

A powerful approach to enhance on-wheel hand action recognition and prioritize driver safety is utilizing feature trajectories, which is the technique to track the path or progression of specific features or characteristics over time. Wang et al. propose a method that analyzes video actions using dense trajectories, which is a method in computer vision and video analysis that densely samples key points in a video sequence to capture motion information and track object movements. This approach efficiently evaluates hand motions and quick movements in hazardous circumstances, determining potential risks. Additionally, the implementation of convolutional neural network models, including spatio-temporal multiplier networks (STMNs) introduced by Zolfaghari et al., emphasizes the importance of hand classification for driver safety. By combining temporal convolution with spatial convolution, STMNs offer an inclusive approach to analyzing spatiotemporal patterns in driver behavior, enabling the identification of unsafe driving actions hands-off-the-wheel [33, 35].

Table 1 Summary of research on hand classification

Refs.	Summary	Methodology	Relevance
[33]	Enhances action recognition using dense trajectories, improving understanding of physical actions in videos.	Dense trajectories	Enhances action recognition models for accurate action detection, relevant to driver behavior analysis in autonomous vehicles.
[34]	Combines temporal and spatial convolution in a new CNN model to learn spatiotemporal features from videos.	Spatiotemporal Multiplier Networks	Proposes Spatiotemporal Multiplier Networks (STMNs) for video data analysis within autonomous vehicle cabins, extracting important features for in-cabin analysis.
[35]	Uses EfficientNet, a highly efficient ConvNet family, to achieve state-of-the-art accuracy through systematic model scaling.	EfficientNet	EfficientNet's superior accuracy and efficiency are relevant for developing robust vision systems within autonomous vehicle cabins.
[36]	Utilizes MobileNets for efficient and lightweight deep neural networks design, suited for mobile and embedded vision applications.	MobileNets	MobileNets' efficient, lightweight architecture is suitable for real-time vision systems in autonomous vehicle cabins, minimizing computational resources.
[37]	Combines GoogLeNet and LSTM models to classify self-efficacy levels through human body gesture and movement recognition, achieving high accuracy.	CNN (GoogLeNet) and LSTM	Provides an effective approach to monitor and analyze driver behaviors, enhancing safety and efficiency within autonomous vehicles.
[38]	Uses a pre-trained Keras neural network to classify hand presence in a controlled hand-washing dataset, achieving perfect accuracy.	Neural Network using Keras	Demonstrates an effective approach for hand presence classification, potentially enhancing safety and efficiency by monitoring driver actions in autonomous vehicles.
[39]	Introduces a novel hard attention network for Driver Action Recognition (DAR), effectively recognizing driver behaviors in real-world conditions and reducing computational complexity.	Bidirectional LSTM (Bi-LSTM)	Investigates deep learning for driver behavior monitoring and action recognition, aligning with the goal of in-cabin analysis in autonomous vehicle cabins.
[40]	Uses a multi-camera framework for hand classification in driver monitoring systems, potentially enhancing traffic safety and reducing distracted driving.	RestNet CNN	Discusses a multi-camera framework for hand classification in driver monitoring systems, aligning with the topic of vision systems and machine learning analysis in autonomous vehicle cabins.
[41]	TPresents a CNN-based system for abnormal driving behavior recognition, emphasizing the importance of monitoring and preventing potential accidents caused by distractions.	CNN	Detects abnormal driving behaviors through physiological character classification using deep learning, contributing to understanding of vision systems for driver behavior analysis in autonomous vehicle cabins.

Moreover, advancements in efficient CNNs like EfficientNet and lightweight deep neural networks like MobileNets reinforce the significance of hand classification for in-cabin analysis and real-time monitoring, enhancing autonomous vehicles' safety and efficiency. The utilization of complex attention networks, as presented in Driver Action Recognition (DAR), further underscores the importance of hand classification for driver safety. By focusing exclusively on vital behavioral elements including the hand and head, this approach aligns perfectly with the analysis of data inside of the cabinet, ensuring an exhaustive understanding of driver behavior. The studies above indicate that more than

relying solely on hand classification for driver behavior detection is required as it fails to capture the full spectrum of driver behavior and intention. While hand movements provide valuable insights into driver actions, an all-inclusive understanding requires considering factors including head position, eye gaze, and body posture. Incorporating multiple factors leads to a better understanding of a driver's cognitive state, attention level, emotional response, and fatigue level [36, 45].

As technology advances, the focus on comprehending the holistic range of driver behavior has led to integrating these multiple factors into driver assessment models. Notably, the area of "Hand Classification" has seen significant breakthroughs, revolutionizing our grasp of driver interactions. Recent developments in the field of hand classification present intriguing innovations using machine learning and computer vision for understanding hand gestures and driver behavior. One study created an algorithm that tracks a driver's right hand and ear in real-time, processing video frame images to identify if the driver is distracted. With an impressive accuracy score of 74 percent, this algorithm can classify various actions, including everyday driving, touch screen interaction, and phone conversations. Another study formulated an algorithm that identifies hand gestures using three specific characteristics of a hand's shape, achieving 91 percent classification from a test set of 200 images [46–48].

Researchers in the realm of sign language innovated a system that interprets gestures by thinning a segmented image resulting in a communication breakthrough for sign language users. Further studies rolled out an Urdu Alphabet translation recognition system with an accuracy rate of 97.4 percent, proving extremely helpful for individuals with vocal and hearing disabilities. To improve human-computer interaction, a trailblazing system recognizes hand gestures as an alternative to classical mouse and keyboard inputs. This system uses the AdaBoost algorithm to identify the hand in a video feed and then applies multi-class support vector machines to understand the gesture [49–51].

In a similar vein, an approach for recognizing moving hand shapes was developed. Keeping the focus on real-time image processing, researchers first extracted the hand region and then identified the hand's shape. In an effort to boost secure access, a hand image-based identification system was developed, achieving confident recognition in groups of about 500 people. Finally, the creation of a detailed video-based dataset stands as a pioneering venture for hand detection in varied driving settings. This dataset, encompassing various backgrounds, lighting conditions, users, and viewpoints, serves as a potent tool for fine-tuning machine learning algorithm performance. Notably, it also features annotations that offer detailed hand related insights, marking significant advancements in the field of hand classification [52–54].

Facial classification

In the vision system, cameras and sensors function as the eyes of the computer, with sophisticated software algorithms acting like a human brain to interpret the data they capture. These systems analyze images, breaking them down to understand each element, enabling the recognition of faces, objects, and navigation paths. Such vision systems empower machines to "envision" and "understand" their surroundings, playing a pivotal role in diverse applications such as robotics, security systems, autonomous vehicles, and mobile phones.

Driver distraction, a major contributor to vehicular accidents, is well-documented and can be effectively monitored by these vision systems. However, it's important to note that while these systems have achieved impressive performance levels, there are ongoing concerns about their evaluation methodologies. Often, AI-based models in vision systems are assessed using datasets involving drivers who were part of the training set. This practice can lead to a potential 'memory' effect, where the models are fine-tuned to the characteristics of these specific drivers, raising questions about their ability to generalize to new, unseen drivers [23, 55–58].

To counter this, it is crucial to incorporate evaluation strategies that ensure these models can effectively adapt to diverse driving behaviors not represented in the training data. This involves using more heterogeneous datasets and applying rigorous cross-validation techniques. Such approaches are essential to evaluate the real-world applicability and robustness of these vision systems, ensuring they remain effective across a broader range of scenarios and driver behaviors. Recent reviews, like Ji's panoramic study, have highlighted various non-invasive approaches for detecting signs of fatigue using vision systems. These methods leverage video analysis of a driver's visual characteristics to identify fatigue levels. However, the success of these techniques hinges on the models' ability to generalize their learning to a wide array of drivers. The development of adaptable and broad-based AI models remains a key area of research in enhancing the effectiveness of vision systems in real-world applications [59].

Visual perception, fundamental to driving, relies heavily on visual sensors for data capture. However, this data contains abundant indirect information that machine vision and image understanding techniques handle. Smart vehicles, from advanced assistance systems to autonomous vehicles, leverage machine vision to distill and categorize video data, making it useful for driving. Techniques like convolutional neural networks play a crucial role in identifying specific objects in traffic and aiding in the mapping and positioning of self-driving cars. They also incorporate the discussion of real-time computing architectures backed by real-world experiments. However, the field of vision systems is witnessing an inventive shift where, instead of trying to identify objects universally, the system should adapt its method based on the size and context of the object under observation. It requires the system to be adaptable enough to modify its strategy according to the target, using different techniques for smaller or less transparent objects than for larger, more detailed ones. More than just theoretical, this concept has been tested and has outperformed other methods on popular benchmarks. The adaptable vision system reinforces that there is always room for innovation and performance enhancement, especially as vision systems tackle various real-world situations [60, 61].

The integration of vision systems with facial classification technologies represents a paramount advancement in addressing critical challenges within the transportation sector. However, as discussed, concerns arise regarding the generalization of AI models in these systems, particularly in the context of driver distraction. The development of driver fatigue monitoring systems, paired with facial detection, emerges as a pivotal solution for detecting drowsy and inattentive driving. By combining the adaptability of vision systems with facial indicators, these integrated technologies enhance real-time monitoring, ensuring road safety by adeptly assessing drivers' drowsiness levels and mitigating risks associated with such accidents.

As the global economy rapidly expands, the transportation sector is also swiftly advancing. Specifically, heavy trucks stand out for their impressive cargo capacity and have become crucial in logistics and road transportation. In China, with 2022 sales projected at 1.2 million units, the country's total heavy truck ownership will reach 11.7 million by 2025. However, this growth has resulted in a corresponding rise in traffic incidents, often due to drowsy driving. Fatigue is notably prevalent among these drivers who undertake extensive drives to make ends meet, leading to exhaustion, decreased attention, and potential accidents. Data from the US National Highway Traffic Safety Administration (NHTSA) indicates that 91,000 crashes involved drowsy driving, leading to approximately 50,000 injuries and nearly 800 deaths. The traffic safety, sleep science, and public health communities generally agree that these figures underestimate the true impact of drowsy driving. Additionally, a Chinese study highlighted the propensity for such incidents to occur at any time, especially during early morning or mid-afternoon hours. These statistics underscore the serious threat posed by drowsy driving, particularly with heavy trucks, marking it as a significant factor in critical traffic accidents. In recent years, there have been driver fatigue monitoring systems with facial detection technology and precise infrared sensors developed to effectively identify signs of driver drowsiness. The system monitors facial movements and detects subtle fatigue indicators, including increased blinking, drooping eyelids, or prolonged eye closures, often unnoticed by the drivers. The detection system remains unaffected by external variables, including the time of day, the presence of glasses, or reflective light. Enhanced by integrated pre-trained artificial intelligence with advanced facial recognition capabilities, the system operates even without a WiFi connection due to its inbuilt algorithms. The AI system is programmed to recognize and audibly alert drivers about signs of drowsiness, providing real-time and reliable monitoring. This deployment of facial detection technology significantly boosts road safety by adeptly assessing drivers' drowsiness levels and reducing the risks associated with drowsy driving [62–64].

Table 2 provides a comprehensive summary of various studies centered on facial classification research, offering a detailed overview of this specific area.

Facial Classification entails detecting and analyzing a driver's facial features and expressions, including yawning, blinking, or looking away from the road. Such analysis can offer insights into the driver's level of attention and alertness, which are vital for ensuring responsible driving. One facial classification approach employs the FaceNet system to efficiently carry out facial recognition, clustering, and verification. The Euclidean Embedding method simplifies complex data by representing it more linearly while keeping the critical relationships intact through a convolutional neural network to tackle facial recognition challenges. By examining feature vectors, the FaceNet system can provide solutions that enhance facial recognition accuracy under various conditions. Another multiple-resolution cascade network method combines different layers with varying levels of detail to efficiently process and extract features from complex data, with high differentiating capabilities. This CNN Cascade technique uses a sequence of convolutional neural networks to progressively filter and refine object detection. The result addresses challenges associated with pose, expression, and lighting. This cascade system uses a sequence of classifiers to progressively refine and improve the classification accuracy of an object or pattern recognition task.

Table 2 Summary of research on facial classification

Refs.	Summary	Methodology	Relevance
[65]	Utilizes feature vectors in FaceNet for efficient face recognition, clustering, and verification tasks.	CNN-based Euclidean embedding	Explores facial recognition, presenting solutions to pose and illumination issues, relevant for enhanced biometric systems in autonomous vehicles.
[66]	Presents a multi-resolution cascade network to handle pose, expression, and lighting issues in facial recognition.	CNN Cascade with discriminative capabilities	Introduces a discriminative cascade system for efficient facial distinctions analysis, pertinent to vehicle cabin surveillance.
[67]	Explores appearance-based gaze estimation in non-lab conditions using the MPIIGaze dataset with 213,659 images from 15 participants.	Appearance-based gaze estimation	Investigates gaze estimation under everyday conditions, contributing to improved facial feature recognition within autonomous vehicle cabins.
[68]	Detects driver emotions unobtrusively via smartphone-captured contextual features, outperforming facial recognition by 7 percent and ensuring privacy.	YOLOv5, Microsoft Face Recognition, DeepLabV3, OpenCV	Unobtrusive Sensor Feed Pipeline analyzes driver emotions less intrusively, relevant to vision systems in autonomous vehicle cabins.
[69]	Proposes a hazardous driving image classification system using a modified ShuffleNet model, balancing speed and accuracy for real-time monitoring.	ShuffleNet	Proposes a solution for dangerous driving behavior monitoring, enhancing safety within autonomous vehicle cabins.
[70]	Suggests a deep-learning-based system for drowsiness detection using a novel CNN model for eye state classification, enhancing traveler protection.	CNN - Deep Driver Drowsiness Detector (4D) Model	Uses a deep-learning-based system for drowsiness detection via a novel CNN model, important for driver state analysis in autonomous vehicle cabins.
[71]	Offers a non-invasive approach for driver vigilance classification via deep learning HyMobLSTM (MobileNetV3 and LSTM model) model and transfer learning, analyzing facial and eye components.	HyMobLSTM model ()	Contributes to driver behavior analysis inside autonomous vehicle cabins via a non-invasive vision system, improving safety and alertness monitoring.
[72]	Presents Hypo-Driver, a real-time driver inattention and fatigue detection system using multi-view cameras and biosensors, outperforming existing solutions.	Hypo-Driver system: fused through CNN, RNN-LSTM, and DRNN (Deep Residual Neural Network)	Hypo-Driver system employs multimodal features for driver hypovigilance detection, aligning with vision-based systems in autonomous vehicles for safety enhancement.
[73]	Monitors driver behavior using image processing and computer vision techniques to prevent accidents, promising high accuracy and real-world application potential.	OpenCV, Support Vector Machine (SVM)	Describes real-time driver monitoring using computer vision techniques, relevant for understanding how such techniques can enhance safety and behavior analysis in autonomous vehicle cabins.

In addition to these approaches, appearance-based gaze estimations further augment facial recognition in everyday situations. By concentrating on real-world scenarios, this method enhances the recognition of facial features and contributes to a more accurate evaluation of a driver's attention and alertness levels. Moreover, driver emotion detection systems have been explored by analyzing different types of information or characteristics surrounding a particular subject or situation, which are considered together to gain a more comprehensive understanding. One research experiment utilizes advanced machine learning models, including YOLO (You Only Look Once)v5, Microsoft Face Recognition API classifier, VGG13-based image classifier, DeepLabV3 semantic segmentation, and OpenCV. The innovative unobtrusive sensor feed pipeline (USFP) developed in this research provides a less intrusive method for analyzing driver emotions inside an autonomous vehicle's cabin, contributing significantly to developing vision systems for autonomous vehicles [74–76].

A hazardous driving classification system based on a modified ShuffleNet lightweight model has been proposed. This system effectively reduces model complexity and increases operational speed without compromising classification accuracy, making it a potential solution for real-time monitoring of dangerous driving. Similarly, a deep-learning-based drowsiness detection system has been proposed using a novel CNN model to classify eye states. The HyMobLSTM model presents a non-intrusive method putting emphasis on analyzing facial features and eye localization in order to yield a more comprehensive interpretation. This model determines a driver's alertness by categorizing it into five levels based on head orientation and the eye position relative to the eyelids. Transfer learning extracts additional features from the driver's eyes, serving as input vectors for the LSTM network [39, 40].

Another real-time driver inattention and fatigue detection system, Hypo-Driver, utilizes multi-view cameras and biosignal sensors to extract hybrid features. The Hypo-Driver system uses a combination of CNNs, RNNs, and deep residual neural networks (DRNN). This system achieves a high accuracy rate of 96.5 percent and outperforms other top-rated driver fatigue detection systems. This is achieved by extracting multimodal features and using deep learning models for driver's decreased alertness levels in individuals, often through the analysis of behavioral or physiological indicators. In addition to the above, another project uses OpenCV, an open-source computer vision library that provides tools and functions for image and video processing, as well as support vector machine (SVM), a machine learning algorithm used for the classification and regression tasks. The described systems above contribute to understanding how computer vision and machine learning techniques can enhance safety and behavior analysis in drivers, thereby improving the overall safety measures within autonomous vehicles [73, 77].

Building on the use of machine learning and computer vision for driver monitoring through hand classification, researchers have expanded into the realm of facial classification. This advancement is opening new avenues for improving driver safety through cutting-edge recognition and emotion perception technologies. One study explored how blocking facial features affects emotion perception, while another mapped facial features to emotional recognition. Significant strides were made in

real-time driver distraction detection by analyzing visual cues from the face and tracking eye and head positions [78–80].

Furthermore, an AdaBoost algorithm-based system calculates gaze direction to assess if drivers maintain eye contact with the road. Another study improved distraction detection accuracy to 81.1 percent by analyzing eye activities and driving performance data. In contrast, others employed facial cues to develop highly precise classifiers for visual and cognitive distractions [81–83].

Researchers also optimized convolutional neural networks and introduced an unified face detection system through the wearable face recognition system, a notable development for blind and visually impaired individuals. The rapidly advancing field of facial classification is creating breakthroughs in autonomous driving, human-computer interaction, and communication aids for the visually impaired [84].

In summary, these Facial Classification methodologies have been put into practice to advance facial classification tasks. These strategies provide a comprehensive approach to monitoring and analyzing driver behavior inside autonomous vehicle cabins, enhancing safety measures and contributing to the development of autonomous vehicles.

Body posture classification

Driving a vehicle demands prolonged periods of intense focus and repeated sitting posture and movements. These factors inevitably cause fatigue. When the driver experiences fatigue, their ability to maintain focus diminishes, and their reaction times may suffer, posing potential safety risks. Therefore, it is crucial to devise methods that can promptly and accurately assess the level of vehicle drivers' fatigue. This assessment should be conducted to ensure operational safety and efficiency without interfering with the driver's routine tasks. Body posture classification involves analyzing a driver's body posture and movements, including slouching, leaning, or sudden jerky movements. The result can provide insights into the driver's fatigue, distraction, or impairment level.

Table 3 provides a consolidated overview of various studies focused on the classification of body posture, summarizing key research in this area.

Firstly, in [92], researchers from China have found a deep learning technique: by extracting features related to upper body posture, including the head, neck, chest, shoulders, and arms, from images captured of train drivers to detect drivers' fatigue level. In [93], the researchers from Beijing Jiaoton University introduced a method for detecting the fatigue state of drivers by analyzing their upper body postures extracted by OpenPose framework and a Deep Belief Network - Back Propagation Neural Network ("DBN-BPNN") model. The model takes a "9-dimensional principal eigenvector" of the driver's upper body posture as input: the 9-dimensional principal eigenvector is like the main road on a map with nine different directions, which provides the most efficient route to capture the essential features in a dataset or system.

Next, the model applies a forward Restricted Boltzmann Machine (RBM) learning algorithm to reconstruct the eigenvector and extract high-level distribution features. The DBN-BPNN model includes four levels for classifying fatigue states. Results from the experiment demonstrate an average detection accuracy of 92.7 percent using the DBN-BPNN model, indicating the method's high accuracy in detecting fatigue among drivers. MoveNet, furthermore, is a deep neural network designed to predict subject-specific

Table 3 Summary of research on body posture classification

Refs.	Summary	Methodology	Relevance
[85]	Proposes an ensemble model using deep convnets for Human Body Posture Recognition (HBPR), relevant for posture-based analysis in autonomous vehicle cabins.	Bivarial Deep Convnet Model	Deep convnets in the ensemble model for HBPR, useful for posture analysis in autonomous vehicles.
[86]	Classifies seven common driving activities using pre-trained CNN models, contributing to driver behavior analysis within autonomous vehicles.	AlexNet, GoogLeNet, ResNet50	Uses AlexNet, GoogLeNet, ResNet50 to classify driving activities, aiding in driver behavior analysis in autonomous vehicles.
[87]	Proposes a drone surveillance system for human behavior analysis, including posture analysis via OpenPose, relevant for outdoor surveillance and interaction with autonomous vehicles.	Pose Estimation, OpenPose, Deep- Sort, YOLO	Applies OpenPose and other algorithms for human behavior analysis, useful in autonomous vehicle interactions.
[88]	Utilizes PoseNet for real-time 6-DOF camera relocalization from single RGB images, relevant for vehicle cabin monitoring and driver pose estimation.	23-layer deep CNN (convolutional neural network) trained in an end-to-end manner to regress the 6-DOF camera pose.	Utilizes PoseNet for 6-DOF camera relocalization, aiding in driver pose estimation in vehicle cabins.
[45]	Uses MoveNet to predict subject-specific joint angle profiles for different walking conditions, applicable for pedestrian behavior analysis around autonomous vehicles.	MoveNet	MoveNet predicts joint angles for pedestrian analysis, useful for passenger behavior study in autonomous vehicles.
[89]	Presents D3-Guard, a real-time drowsy driving detection system using built-in smartphone audio devices and LSTM networks, applicable for drowsiness monitoring in autonomous vehicle cabins.	LSTM networks	D3-Guard detects driver drowsiness using smartphone audio and LSTM, relevant for monitoring in autonomous vehicles.
[90]	Introduces BiRSwinT network for fine-grained driver behavior recognition, offering enhanced driver action learning and accuracy in driver behavior analysis.	Bilinear full-scale residual Swin-Transformer network (BiRSwinT)	BiRSwinT network enhances fine-grained driver behavior recognition, contributing to in-cabin analysis.
[91]	Proposes a driver distraction detection system using a blend of deep learning and machine learning models, relevant for enhancing roadway safety through distraction monitoring.	DenseNet and Genetic Algorithms (GA)	TCombines DenseNet and GA for driver distraction detection, focusing on posture-based insights for safety in autonomous vehicles.
[76]	Leverages MobileNetV2 for efficient classification of driver distraction behaviors, demonstrating potential for reducing accidents caused by distracted driving in autonomous vehicles.	MobileNetV2	MobileNetV2 classifies driver distractions by detecting subtle posture changes, aiding vision systems in autonomous vehicles.

joint angle profiles for various walking speeds and slopes, minimizing input data requirements. MoveNet's ability to predict highly user-specific profiles from minimal input data shows the potential for using similar approaches in vision systems analyzing the interior of autonomous vehicle cabins. By understanding and adapting to individual passengers' needs and preferences, MoveNet can contribute to a more personalized and comfortable ride in autonomous vehicles.

In fact, Beijing Institute of Technology researchers introduced D3-Guard, a system that detects driver drowsiness in real-time using the audio capabilities of a smartphone. It identifies unique sound patterns from behaviors like yawning and steering and uses long short-term memory (LSTM) networks for efficient detection. With an average accuracy of over 93 percent in real-world testing, D3-Guard suggests that sound-based detection can complement or even replace vision-based systems in self-driving cars. The scaling method of the system preserves the original aspect ratio of images or videos during resizing, ensuring high-quality output. Furthermore, another study proposes the Residual Swin-Transformer (BiRSwinT), a network that recognizes ten fine-grained driver behaviors. BiRSwinT employs a dual-stream structure to process and analyze various data types, performing exceptionally well on the AUC V1 and V2 datasets. This dual-stream design allows for the simultaneous processing of global and local cues of driver actions, enhancing the detection of subtle behaviors and improving the overall safety of autonomous vehicles [69, 89, 91].

Another highly accurate system for detecting driver distraction consists of a blend of deep learning and machine learning models, fine-tuned by a genetic algorithm. This system adapts to new datasets in real-time, aiming to enhance traffic precautions through the Hybrid Genetic Deep Network. This model uses principles from genetic algorithms and deep neural networks. It utilizes evolutionary techniques to optimize the architecture and processes of deep learning networks. This approach aims to enhance performance or efficiency when studying driver behavior within self-driving vehicle cabins [76].

Furthermore, the research from National Tsing Hua University (NTHU) [39] underscores the proficiency of the MobileNetV2 model in categorizing driver activities, achieving an impressive blend of speed and precision while preserving low computational demands – an essential characteristic for mobile system implementation. The research leveraged two distinct datasets for their experiment: a 10-class dataset from State Farm and a 2-class dataset. The clearly defined features in the State Farm dataset allowed the model to successfully differentiate between two classes, resulting in superior predictive accuracy. However, the NTHU drowsiness dataset, in its realistic depiction of driver behavior, offered a more authentic training environment, fostering progress toward real-world applications. In the context of mounting traffic fatalities worldwide, specifically in areas like Malaysia where distraction-induced accidents are prevalent, the application of deep learning techniques, specifically convolutional neural networks presents a promising avenue for efficient identification and classification of distracted driving behavior. Therefore, it contributes to the broader objective of promoting safety in autonomous vehicles [71].

While machine learning techniques, particularly convolutional neural networks, promise swift identification and classification of distracted driving behavior, innovations

extend beyond this realm to improve driver safety. These advanced systems now consider other vital parameters, including head and body movements, to gauge a driver's alertness, paving the way for comprehensive driver behavior assessment. In the pursuit of creating safer roads, advanced systems are analyzing drivers' alertness, including their head and body movements, to prevent accidents due to fatigue or distraction.

Procedures including the integration of the Microsoft Kinect range camera's capabilities of capturing and analyzing 3D shapes of drivers, and fitting a human skeleton model to this data have been beneficial in evaluating nuanced driving behaviors across varying demographics. When combined with machine learning techniques like K-means clustering, SVMs, and HMMs, the result is a highly accurate recognition of driving-related actions and postures [94, 95].

Algorithms like Part Affinity Fields (PAFs) have proven efficient in detecting 2-dimensional poses of multiple individuals in images, setting a benchmark for pose detection. The use of tools like head trackers and vision-based foot behavior analysis, along with video sequence trajectories, is enhancing the accuracy in action recognition and prediction of drivers' foot behavior. This data contributes to body posture classification as a significant component of autonomous vehicle safety. By detecting diversions or measuring the driver's head orientation, these advancements promise a safer future for autonomous driving [96, 97].

In summary, these body posture classification methodologies contribute to the advancement of safer driving. These strategies collectively provide a comprehensive approach for monitoring and analyzing driver behavior inside autonomous vehicle cabins, contributing to the development thereof and safer roads worldwide.

Integration of physiological indicators classification

Incorporating physiological indicators, including hand gestures, facial expressions, and body postures, is essential for developing an all-in-one driver monitoring system. This section outlines strategies for the physiological classification approaches to establish a cohesive driver monitoring system. To further elaborate on the integration of the classifications, it is important to note that driver behavior classification serves as a key component in predicting and preventing risky driving scenarios. Not only can driver behavior classification provide real-time assistance to human drivers, but it is also instrumental in shaping the development of autonomous vehicles. It is worthwhile to recognize that human interaction with self-driving cars is an emerging research area with profound implications for road safety, traffic efficiency, and overall driving experience. The classification schema that divides driving behaviors into safe and unsafe categories offers a practical and simplified representation of the complexities involved in everyday driving. This classification system is the basis for analyzing and predicting driver behavior. By assigning a binary value of "1" for safe driving behaviors, and "0" for unsafe driving behaviors, researchers can create a streamlined and consistent method of collecting and analyzing data. This data, in turn, provides valuable insights that can help develop various interventions to enhance road safety.

In [98], researchers shed light on the significance of different body parts, in the perception of emotions. While not directly addressing the process of integrating hand, facial, and body posture classifications, the research provides valuable insights for

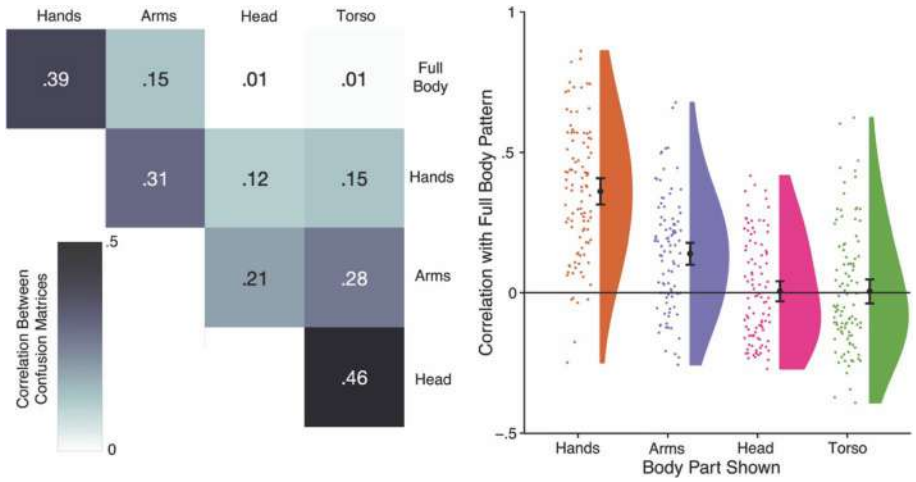


Fig. 3 Representational similarity analysis of confusions between isolated body parts and full body from [98]

developing an emotion detection system. The study highlights the importance of different body parts in accurately perceiving emotions, suggesting that emotion classification is an essential component in a multi-modal system for a comprehensive analysis. By incorporating these insights, a unified framework can be developed that accounts for the importance of hands, employs a multi-modal approach, harnesses shared mechanisms, and addresses challenges including the body inversion effect, leading to the creation of a robust and accurate system for emotion recognition.

Figure 3 demonstrates the representational similarity analysis of confusions between isolated body parts and full body from [98], providing a visual understanding of the complexity in emotion recognition related to different body parts.

Another study [99], conducted by Chinese researchers, emphasizes the integration of deep learning-based segmentation to isolate the driver's body parts, including the head and hands, which play critical roles in identifying distraction. Two segmentation architectures, Human Body Parts Segmentation (HBPS) and Cross-Domain Complementary Learning (CDCL), were investigated. Despite similar performance on the Pascal VOC dataset, the CDCL model performed significantly better under low light conditions in the study's specific dataset, efficiently segmenting critical body parts even in challenging lighting scenarios. This model facilitated the elimination of irrelevant image regions and concentrated on hands and head-related regions essential for safe driving. The system achieved an impressive average accuracy of over 96 percent on the authors' dataset and 95 percent on the public AUC dataset, indicating its substantial potential in developing comprehensive driver assistance systems by integrating physiological indicators for driver behavior classification.

Initiated in the early 2000s, the AWAKE project marked the European Union's pioneering effort in integrating driver state and performance metrics for effective fatigue detection, using measures like eyelid movement and steering grip changes. Building on AWAKE's foundational work, modern projects like Hi-Drive and Programmable Systems for Intelligence in Automobiles (PRYSTINE) have emerged, showing significant advancement in the field. Hi-Drive, focusing on integrating automated vehicles in mixed traffic,

emphasizes understanding user interactions with AVs, employing a multi-disciplinary approach with diverse studies to analyze user behavior, expectations, and limitations. PRYSTINE, notable in the Electronic Components and Systems for European Leadership (ECSEL) initiative, advances autonomous driving technology, aiming to enhance safe, clean, and efficient mobility through fail-operational behavior in AVs, achieved by integrating advanced Radar and Li-ght Detection And Ranging (LiDAR) sensor fusion. Both projects signify the evolution of vehicular technology and user interaction, enhancing road safety and the reliability of systems detecting driver fatigue and ensuring operational safety in automated driving [100–102].

The proposed method in [103] integrates detection and tracking algorithms to monitor distracted driving behavior based on facial and hand movements. The facial detection and tracking involve using the Viola-Jones algorithm to detect the driver's face and an algorithm described in reference [104] to detect key facial features like eyes, lips, and forehead. The center point of the forehead and lips are tracked using the KLT tracker algorithm. Hand detection focuses on a localized search region, typically the lower-right or lower-left quarter of the frame, and employs a hand detection algorithm from reference [105]. The center point of the hand is tracked using the KLT tracker as well. The tracking algorithms continuously estimate the displacement between consecutive center point and calculate the tracking error based on feature differences. If the tracking error exceeds certain thresholds, reinitialization is performed by redetecting the respective body part. The method emphasizes the significance of simultaneous tracking of these body parts in capturing distracted driving behaviors. By analyzing the trajectories and patterns of facial and hand movements, specific distracted driving behaviors including talking on the phone, eating, or texting can be recognized. This proposed method leverages the simplicity and effectiveness of the algorithms, taking into account the constrained setting of driving and marginal deformations of body parts. The integration of two physiological indicators together, hand gestures and facial expressions, enhances the understanding of distracted driving behaviors and contributes to the development of comprehensive driver monitoring systems.

This study [107] investigates how a driver's body and head characteristics can influence the categorization of driving tasks, beginning with evaluating depth information from facial landmarks and joints. The precision of task classification demonstrated substantial differences when relying exclusively on either head or body signals. The model, trained only with two-dimensional information like head rotation and joint coordinates, showed accuracy levels comparable to those trained with complete features. However, the classification accuracy decreased when only using head pose information. While the distracted driving behaviors were successfully detected, it was challenging to differentiate safe driving behaviors with similar head positions. In other words, using only body features (the coordinates of the hand, wrist, elbow, and shoulder joints) resulted in weaker detection of mirror-checking behaviors but a higher degree of accuracy for detecting distraction behaviors. So, the head and body characteristics are vital for comprehensively classifying driving tasks. Though there was a slight dip in the overall detection accuracy, the selection of 18 features, which includes yaw, pitch, roll, nose, hand, and shoulder coordinates, provided a reasonable balance between accuracy and computational speed. The final result demonstrates the potential of such a system in effectively

Table 4 Comparison of combined classifying models

Methods	Year	Dataset	Feature	Algorithm	Accuracy %
[98]	2023	Bochum Emotional Stimulus Set	Full body, particularly the hand	Isolated Body Part Emotion Recognition algorithm	64.7
[104]	2022	ORL Face Database from ATT and FEI dataset	Face and Hand	SVM	98.03
[99]	2021	Driver Monitoring Dataset and AUC	Full body	Two different pre-trained CNNs, VGG-19 and Inception-v3	96
[106]	2019	QVGA ToF image sequences	Body Key Points	3D CNN-LSTM	85

combining physiological indicators together for the classification of driver behaviors using the unification of various body part classifications.

In the study referenced as [106], the researchers built a model of the driver's posture classification consisting of nine key points – left/right shoulders, left/right elbows, left/right hands, left/right hips, and the right knee. They selected a combination of various body parts for their detectability in denoting driving behaviors. The team trained fully convolutional neural networks using their dataset to calculate the pivotal points for each frame of the body independently. Then, they transposed the data into three-dimensional camera coordinates using depth imagery, resulting in a real-time 3D rendering of the driver's physical stance. The focus is shifted from individual actions to the real-time tracking of a combination of various body parts, thereby enhancing the depth and precision of physiological indicator-based classification of driver behaviors. To provide a clearer comparison of these diverse methodologies, Table 4 presents a comparative analysis of combined classifying models, outlining the differences in datasets, features, algorithms, and accuracy across various studies.

While research like [106] illustrates how real-time 3D imaging and tracking of various body parts can enhance driver behavior analysis, the potential benefits of autonomous vehicles extend beyond improved safety measures. These breakthroughs not only revolutionize transportation policies and systems but also necessitate a thorough understanding of the legal regulations governing autonomous vehicles. This understanding is crucial to adapt the existing road traffic laws and navigate the regional differences in these regulations. Autonomous vehicle technologies potentially lower transportation costs and increase accessibility, particularly for those with mobility limitations. With a focus on the communication between autonomous vehicles and infrastructure, opportunities arise to develop efficient routing systems. Such technologies can revolutionize transportation policies. Meanwhile, in the U.K., connected and autonomous vehicles are triggering a transformative change in the economy, promising benefits like improved safety, reduced congestion, and increased productivity. Vital innovation and research capabilities in the U.K. automotive sector help leverage these benefits efficiently. From a legal perspective, a professional understanding of autonomous vehicles' legal regulations is crucial because it aids the discussions related to modifying existing road traffic laws and the navigation of the variations in the regulations across different regions, including the U.S. and Europe. In vision-based human action recognition or labeling image

sequences, varied advancements focus on image representation and the subsequent classification process. Despite current challenges and limitations, these advancements uncover potential areas for further exploration and improvement. Finally, the accuracy in recognizing subtle driver behaviors can improve significantly by using a network like BiRSwinT. This network combines global shape appearances and local discriminative cues of driver actions in its structure, effectively identifying multi-scale, local lines and can help drive future research in recognizing driver behaviors [108, 109].

Developing a proficient and effective driver monitoring system requires the concurrent examination and integration of numerous physiological indicators. The singular analysis of distinct body parts, including facial expressions, hand movements, and overall postures, can indeed yield meaningful insights into a driver's actions. However, this individualized focus might need to look into the larger, more complex picture of driving behavior due to the multifaceted nature of human actions and responses. By integrating the findings from the analysis of various body parts, researchers can achieve a more complete understanding of a driver's behavior. This comprehensive view enables them to design more precise, well-rounded interventions and assistance systems. The full-body analysis holds the potential to uncover nuanced and complex driving behaviors, enhancing the capability to predict and prevent risky scenarios that might otherwise remain undetected. In summary, the core of driving behavior is not encapsulated solely in the isolated movements of the hands or the face. Instead, it is embodied in the complex interactions among all body parts. As the future is approaching, marked by autonomous vehicles and sophisticated driver-assist systems, a comprehensive, full-body analysis becomes increasingly significant in promoting safer and more efficient roads. In the light of this, focusing on the interplay of full-body indicators becomes a crucial step toward a future characterized by increasing safety and efficiency in driving.

Deep learning framework for distracted driver classification

Present research efforts are specifically centered on detecting distracted driver postures, as opposed to fatigue or drowsiness, predominantly employing the analysis of image-based spatial characteristics using convolutional neural networks. However, convolutional neural networks alone must improve their ability to systematically analyze location-based objects, which has spurred the demand for more advanced techniques. The challenge centers on effectively incorporating the full-body indicators of a distracted driver [110].

In [112], the researchers from the University of Nottingham focused solely on distracted driver postures, distinct from fatigue or drowsiness indicators, presenting a revolutionary methodology using CNNs and stacked bidirectional long short-term memory (BiLSTM) Networks. The presented technique efficiently utilizes both location-based and time-based attributes inherent in the images. The process begins with extracting location-based posture characteristics using pre-configured CNNs, succeeded by implementing the BiLSTM architecture to discern time-based features from the stacked feature maps obtained from the CNNs. An example of Long-term Recurrent Convolutional Networks for Visual Recognition, which supports such a methodology, is illustrated in Fig. 4. The potency of the proposed methodology has been put through rigorous tests using the Distracted Driver Dataset from the American University in Cairo (AUC).

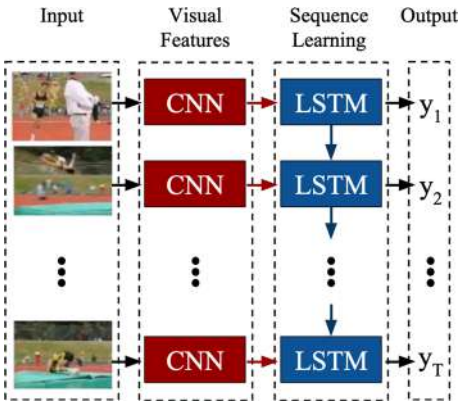


Fig. 4 Long-term recurrent convolutional networks for visual recognition [111]

Compared with existing advanced CNN models, combining CNN and BiLSTM method is superior, yielding an impressive average classification accuracy of 92.7 percent [113].

To verify the effectiveness of CNN with LSTM stack, another study from the University of Nottingham evaluated ten different deep learning methods for their effectiveness in classifying driver distraction postures. The methods incorporated in this study are AlexNet, Inception-V3, ResNet-50, VGG-19 networks, and DenseNet-201. Additionally, the researchers used Inception-V3 in conjunction with different versions of recurrent neural networks, namely RNNs, GRU, LSTM, BiGRU, and BiLSTM. Among all these models, Inception-V3 coupled with BiLSTM, denoted as CNN-BiLSTM emerged as the top performer. It achieved an average loss of 0.292, an average accuracy of 91.7 percent, and an average F1 score of 93.1 percent, highlighting its superior ability in image classification tasks. The success of CNN-BiLSTM comes from its capability to extract patterns from which elements or patterns change over time or follow a specific sequence, which typical CNN architectures might miss. It was more effective than unidirectional RNN architectures due to its ability to capture additional sequential pattern features by processing data in both forward and backward directions. Moreover, it was the best model for identifying challenging postures, including "drivers reaching behind." Despite being on par with CNN-BiGRU (InceptionV3 coupled with bidirectional gated recurrent unit) in terms of performance, CNN-BiLSTM had a longer average training time due to the LSTM's three-gate mechanism, making GRU models a potential choice for real-time distraction detection systems seeking computational efficiency. In situations where the emphasis lies on fast computation and training times, particularly in real-time scenarios, CNN-BiGRU may be favored. However, if the ultimate goal is to achieve the highest possible accuracy, with less regard for computation time, CNN-BiLSTM would be the more appropriate choice. Overall, the CNN-BiLSTM model exhibited remarkable performance in detecting distracted driver behaviors. Future research intends to apply these methods to video streams to capture the temporal dynamics of driving and extend them to anomaly detection techniques to recognize new types of distracted behaviors [114, 115].

Another recent study presented at the 14th International Conference brought to light the utilization of CNNs and LSTM algorithms in an intelligent real-time video surveillance system. This advanced model integrates the strengths of CNN for

extracting spatial information and LSTM for quick and accurate sequential tracking of detected objects. Implementing this combined CNN-LSTM methodology minimizes model complexity, enhances accuracy, and facilitates real-time operation. Feature extraction is the first step in this process using various CNN architectures, including VGG16 and MobileNetV2. Given the constraints of our dataset size, the researchers opted for transfer learning and trained our model on the MobileNet V2 CNN architecture. This choice was motivated by MobileNet V2's excellent fit for real-time applications and its ability to handle any input image more significant than 32 x 32 pixels. The researchers used an input image shape of 128 x 128 pixels in this study. Video segments were standardized for feature extraction, with 20 frames uniformly selected and resized from each video to match the input requirements of our chosen architecture. Following feature extraction, the next stage is classification, which uses LSTM. The LSTM's inherent capabilities for recognizing patterns in time-sequence data make it ideal for activity detection. LSTM, explicitly designed to handle long-term dependencies, distinguishes itself from typical feedforward neural networks, including DenseNet, through its feedback connections [116–118].

These two key components, CNN for feature extraction and LSTM for classification, operate together to deliver a real-time surveillance solution. The CNN, specifically MobileNet V2 in this study, is responsible for extracting spatial features from input images. The partial features are how pixels are arranged and related to each other in a two-dimensional layout. The extraction process ensures that the essential visual information is drawn from each video frame. Then, the extracted features are sequenced and passed to the LSTM network for classification. Equipped with the ability to manage sequential data and long-term dependencies, the LSTM can process image features over time. Using its gates to control how information flows, the LSTM can learn patterns across the sequence of frames. Thus, combining feature extraction via CNN and sequence pattern learning through LSTM enables the system to recognize and classify activities effectively in real-time video streams. The CNN-LSTM approach has shown promising results in real-world applications, detecting suspicious activities at 10–13 frames-per-second in real-time under various conditions. This study combined CNN and LSTM on a Raspberry Pi, demonstrating the possibility of a self-contained system using these two technologies. The study also includes a human action recognition (HAR) methodology that combines CNN and LSTM for optimal speed and precision, demonstrating the significant potential for real-time applications. The HAR approach achieved remarkable accuracy, reaching up to 98 percent on the Peliculas dataset and 91 percent on complex real-life datasets with variable backgrounds, thus showcasing improvements over earlier techniques. This study's findings are particularly relevant for the research into vision systems within autonomous vehicle cabins using machine learning, highlighting the practicality and real-time capability of a combined CNN-LSTM model in recognizing and classifying driver behaviors [119, 120].

Building on CNN-LSTM algorithms in real-time video surveillance, similar strategies surface in studies analyzing driver behavior in autonomous vehicles. From capturing smooth spatial patterns and fine-grained motion details to addressing scene and representation bias, these methods continue to enrich autonomous vehicle safety

systems, demonstrating the innovative use of machine learning models in detecting driver behavior [121].

Various studies have explored methods to analyze driver behavior in autonomous vehicles, particularly focusing on identifying driver distractions. Some used a dual processing system called ConvNets, which works with both still images and motion in videos to perform high-level action recognition, even with limited datasets. Others have evaluated CNNs for video classification, leveraging spatial-temporal information to amplify training and performance. Furthermore, the concept of SlowFast networks uses a dual-pathway system to capture both detailed images and videos, providing a unique perspective for a better understanding of behavior within a vehicle's cabin. Simultaneously, representation bias, where datasets are not fully representative, can be mitigated through the RESOUND method, which quantifies and minimizes bias [122–124].

Weight distribution optimization, known as REPAIR, has also been employed to penalize examples that are easy for specific classifiers, increasing the accuracy and transferability in classifying driver behaviors by not overly relying on specific patterns. Scene bias in video representation learning is tackled by introducing adversarial loss for scene types and human mask confusion loss for masked videos. This approach guides the model to concentrate more on ongoing activities, optimizing its ability to differentiate distinct actions. Unitedly, these diverse methods enhance autonomous vehicle safety systems and create more accurate and reliable datasets, improving the precision and effectiveness of machine learning models for detecting distracted driving behavior [125, 126].

Continuing the discussion on autonomic vehicles' safety systems, research into ConvNets and CNNs uncovers fascinating pathways. By weaving in more spatial-temporal data, mitigating bias, and incorporating changes like human pose data and complex systems like SlowFast networks, these studies significantly advance AI-based vision systems to improve autonomous vehicle safety. By integrating spatial and temporal data and combating bias, this research is further refining AI-based vision systems and safety measures in autonomous vehicles. Next, Research into deep ConvNets and CNNs provides critical insights for utilizing machine learning for cabin analysis in autonomous vehicles. Notably, a two-stream ConvNet architecture that integrates spatial and temporal networks successfully analyzes images and videos. The employment of multi-frame dense optical flow as training data and multi-task learning optimizes datasets and enhances performance, while evaluations of CNNs highlight the crucial role of spatio-temporal data and multi-resolution architecture in bolstering training efficiency [127, 128].

Enhancements to driver behavior analysis are also evident in initiatives integrating human pose data into models and implementing complex systems like SlowFast networks to capture spatial patterns and motion details, progressing driver safety in autonomous vehicles. Furthermore, strategies including RESOUND and REPAIR are adopted to combat representation bias and model over-reliance on specific patterns, substantially improving the accuracy of driver behavior detection. Simultaneously, the examination of scene bias places focus on activities, thereby enhancing nuanced behavior differentiation. These collective insights significantly contribute to advancements in AI-based vision systems, improving safety, and driving performance within autonomous vehicles [34, 129, 130].

In conclusion, integrating advanced machine learning models, including CNN-LSTM algorithms, is making noteworthy strides toward improving real-time video surveillance and safety systems in autonomous vehicles. The combination of spatial and temporal data, integration of human pose data, and implementation of complex network systems including SlowFast networks are emphatically contributing to detecting and classifying activities and driver behavior. At the same time, measures to counter representation and scene bias are enhancing the performance and precision of these models. This combination of distinct strategies and methods enables unprecedented advancements in autonomous vehicle safety systems and surveillance technology. Continued research in these areas is paramount to the further enhancement of AI-based vision systems and safety measures, driving us toward an increasingly autonomous future.

Proposed AI application and datasets for driver behavior classification in autonomous vehicles

To address reliability and privacy concerns in driver behavior tracking systems, an offline AI application that operates in autonomous vehicles is proposed. This system is based on the open-source Pose-Monitor AI Application, and it does not necessitate an internet connection, thus providing uninterrupted monitoring and privacy safeguarding for the driver. An "open-source Pose-Monitor AI Application" denotes a class of software utilities that leverage artificial intelligence methodologies to track and examine human postures and movements. Its underlying source code is accessible for inspection, modification, and distribution by any interested party. Such an application can be a digital animation to monitor drivers in automotive environments for safety considerations: the term "open-source" implies that the software's database is transparent online, promoting collaborative practice and allowing enhancement by a worldwide network of developers. The term "pose" relates to the spatial orientation and position of an object, with a focus on human body parts in this context. "Pose-Monitor" is an aspect of computer vision that aims to ascertain an object's spatial orientation and position, particularly a person, in a given visual medium; "AI Application" points towards a software solution that integrates artificial intelligence functionalities, such as computer vision, to execute tasks typically associated with human intelligence, including pattern recognition, experiential learning, and predictive decision-making. Thus, an open-source Pose-Monitor AI Application employs artificial intelligence technologies to analyze and monitor human poses. At the same time, its source code remains freely accessible for assessment and customization. Lastly, the offline software application protects the users' online privacy and blocks potential cyber-attacks.

Expanding on the concept of AI-driven monitoring and behavior classification within autonomous vehicles, several essential studies have taken this research further. These investigations underscore the unrealized potential of an offline AI application grounded in the open-source Pose-Monitor AI application. Such integration can further enhance safety measures within autonomous machines and address pressing concerns about privacy, while also making notable strides in improving road safety. Attention now turns to the specifics of these critical studies, exploring a broad spectrum of applications from detecting unconventional driving behaviors to anticipating pedestrian movements. These studies all contribute substantially to the development of artificial intelligence

applications for driver behavior classification in autonomous vehicles. One study proposes an AI system using real-time data to detect unusual driving behaviors, including drunk or reckless driving. Another focuses on identifying driver distractions utilizing vehicle movement data and machine learning models, helping tackle the growing problem of in-vehicle distractions [28, 131, 132].

A third study introduces a system that monitors both driver drowsiness and distraction simultaneously, markedly enhancing detection capabilities and accuracy. The researchers developed a new eye-detection algorithm and enhanced its accuracy with a support vector machine. Another project brings forward a novel 3D hand pose estimation method using 3D CNNs, which could provide a new way to interpret driver hand movements. Lastly, a study introduces a new predictive pedestrian path model, helping autonomous vehicles better anticipate pedestrian movements, and therefore improving overall road safety [20, 133, 134].

Pose monitor AI program

The Pose-Monitor AI Application is an open-source project that evaluates the user's body posture and provides real-time feedback to improve posture. The system utilizes image processing techniques to distinguish between proper and improper postures, generating a score based on the evaluation. If the rating falls beneath a pre-established limit after 30fps, the system warns the user; if the score remains below the threshold after another 30fps, it alerts the user to adjust their posture, using either a familiar voice or a more severe tone if necessary. Incorporating the Pose-Monitor AI Application into the Proposed AI system allows for adequate assessment of a driver's posture and behaviors in an autonomous vehicle. This integration enables the AI system to leverage existing image processing techniques and real-time feedback mechanisms to understand the driver's overall condition better. Consequently, the Proposed AI system can detect potential indications of fatigue, distraction, or impairment, ultimately improving safety in autonomous vehicles. Additionally, the open-source nature of the Pose-Monitor AI Application ensures that the AI system remains customizable and adaptable, permitting developers to continuously refine the algorithms and techniques used to analyze a driver's body posture. This adaptability is crucial for tailoring the AI system to address the specific needs of various autonomous vehicle manufacturers and user groups.

With the potential to be added based on the hand classification and facial classification algorithms [103], the Pose-Monitor AI Application emerges as a practical solution, leveraging real-time feedback of the driver's driving state. The technology's continuous monitoring at a rate of 30 frames per second ensures rapid detection of any shifts in the driver's physiological states or overall behavior. These changes could indicate the onset of fatigue or distractions, triggering the system to alert the driver or activate autonomous controls for enhanced safety.

The functionality of the Pose-Monitor AI Application extends beyond simple posture monitoring. It uncovers insightful behavioral indicators tied to the driver's physical disposition. For instance, drivers' subtle body adjustments can signal anxiety or unease with the autonomous vehicle's decisions. This understanding of drivers' behavior empowers the system to respond to drivers' needs proactively. When integrated with other AI modules, like emotion detection [135], the Pose-Monitor

AI-based Application contributes to a sweeping human behavior classification system. This integrated system can offer a more accurate and in-depth understanding of the driver's state, considerably improving the autonomous vehicle's interaction with its human occupants.

The performance of the Pose-Monitor AI Applications in driver behavior analysis sets the stage for exploring cutting-edge machine learning and AI systems aiming to foster road safety. Distinct research approaches form an intriguing landscape of advanced tools; these range from real-time warning systems and neural network-based activity recognition to novel hand pose estimation using 3D Convolutional Neural Networks. Delving into the individual contributions of these unique studies reveals their remarkable potential in expanding the scope of AI and machine learning in understanding driver behavior and thus enhancing autonomous vehicle safety. The proposed AI Monitor Program makes extensive use of machine learning and AI systems to enhance road safety. An activity recognition system based on deep CNNs successfully identifies seven common driver activities. Four of these activities are classified as normal driving tasks and the rest as distractions, achieving an impressive 91.4 percent accuracy rate. Another approach uses CNN to develop a real-time warning system for driver distraction detection. A unique approach inhabits the use of cellular neural networks and electrostatic sensors on the steering wheel for real-time stress level monitoring, improving detection accuracy by up to 92 percent [136–138].

Understanding driving behaviors like car-following, lane-changing, and risky driving can be improved using sensor data, onboard vehicle computer data, and feature extraction methods. Deep-learning models have shown exceptional accuracy in identifying these behaviors, hence, implying their potential as a primary tool for understanding driver behavior. There also exists a new approach for real-time hand pose estimation using 3D CNNs, which enhances real-time monitoring of human activity. An open-source tool, VGG Image Annotator (VIA), which operates in any web browser, provides an efficient way to manage labeled data required for AI systems. Finally, a unique application tracks gaze direction to guide an automated surveillance system and represents a novel approach in AI surveillance. Each research study offers unique insights and significant contributions to autonomous vehicle safety, extending the potential use of AI and machine learning in understanding driver behavior [133, 139, 140].

In conclusion, the open-source nature of the Pose-Monitor AI Application opens up avenues for customization. Developers can adapt and enhance the system to cater to specific requirements, creating a versatile AI that fits various vehicle models, driving conditions, and user preferences. This ability to customize is pivotal in a rapidly evolving landscape of autonomous vehicle technology. Furthermore, the Pose-Monitor AI Application supports data-driven adjustments. Over time, accumulated posture data can guide refinements in AI algorithms, fostering a more nuanced understanding of human behavior patterns. This continuous learning process boosts the system's overall performance and safety measures. Incorporating the Pose-Monitor AI Application into a vision system for autonomous vehicles, as part of a broader machine learning-based approach, can yield safer and more interactive autonomous driving experiences.



Fig. 5 State farm dataset's example



Fig. 6 Driver Distraction dataset's example

Data collection for the proposed AI system

Our proposed AI system employs two datasets: the dataset from the State Farm Distracted Driver Detection contest hosted on Kaggle and the Distracted Driver Dataset, developed by researchers at the American University in Cairo, the Technical University of Munich, and Valeo Egypt. Both datasets consist of images depicting drivers engaging in various distracting activities, which can be used to train and validate our AI model to recognize driver behaviors. The dataset consists of images depicting drivers engaging in various distracting activities, which can be used to train and validate our AI model to recognize driver behaviors [113, 141].

Examples of the State Farm dataset can be seen in Fig. 5, while examples from the Distracted Driver Dataset are illustrated in Fig. 6.

The State Farm dataset encompasses:

- 22,424 annotated training images
- 79,726 unannotated testing images

The driver distraction dataset composes:

- 12,555 annotated training images
- 1923 unannotated testing images

Both datasets contain 10 behaviors organized into 2 broad categories:

1. Safe driving
   - Attentive driving (as shown below 'C0')
2. Unsafe driving
   - Phones using
     - Right-handed texting (as shown below 'C1')
     - Right-handed phone conversation (as shown below 'C2')
     - Left-handed texting (as shown below 'C3')
     - Left-handed phone conversation (as shown below 'C4')
   - Entertainment system
     - Using the radio (as shown below 'C5')

- Personal Matter
  - Consuming beverages (as shown below 'C6')
  - Reaching for objects in the back (as shown below 'C7')
  - Adjusting hair or applying makeup (as shown below 'C8')
  - Conversing with passengers (as shown below 'C9')

Both datasets contain 10 behaviors organized into 2 broad categories, as illustrated in Fig. 7, which shows the various classes included in the training dataset.

The diverse classes of driver behaviors contribute remarkably to the analysis and identification of distracted driving, each functioning as unique data classes that provide a multidimensional perspective on driver distraction: attentive driving serves as a comparative benchmark, characterizing the optimum driver behavior and facilitating the identification of anomalous, distracting actions; texting and phone conversation, both right and left-handed, typify manual distractions, where the hand's displacement affects vehicular control Usage data like hand position and device interaction frequency can infer such behavior; radio usage, another form of manual and cognitive distraction, can be inferred from irregular vehicle behavior concurrent with sound system activation data; consuming beverages, a manual and visual distraction, may be identified via motion data showing repeated hand-to-mouth movements; reaching for objects in the back seat signifies a severe distraction, identifiable via significant body and head movement data away from the driving orientation; adjusting hair or applying makeup exemplifies manual, visual, and cognitive distractions; the proposed detection can be achieved through vision system.

Each data class contributes to creating a total distraction model, enabling realistic driver behavior analysis. Each image is linked to one of the ten classes mentioned above. The images are in JPEG format and have a resolution of 1080x1920 or 640x480 pixels. The combined dataset becomes a powerful resource for extracting full-body features. With its extensive range of behaviors, this datasets offer a multi-faceted view of driver posture, allowing our AI system to monitor and analyze beyond mere hand or head movements. Actions including reaching for objects in the back seat or abrupt postural changes, which necessitate whole-body analysis, are accounted for in this dataset. Consequently, the dataset enables a better understanding of the driver's state. This dataset's richness in image count and behavioral variety facilitates training advanced machine learning models. Utilizing these two datasets, such models can



Fig. 7 Driver distraction dataset's training example

be trained to discern and categorize various driver behaviors, enhancing autonomous driving safety.

Despite being predominantly image-based, the datasets' high frame rate provides a basis for time-series analysis, particularly relevant to the foundational Pose AI monitor system. Analyzing a series of images can reveal patterns over time, including a gradual attention drift or fatigue onset. This aspect is crucial since it enables the capture of behavioral alterations that might be overlooked in individual frames. Given its broad spectrum of driver behaviors, the Pose-Monitor-AI-based Application can leverage these datasets for training and validation. By integrating real-world image data, the application's capabilities in real-time assessment of driver posture and behavior can be enhanced, thus augmenting the overall safety quotient of autonomous driving. While these two datasets include invaluable asset for model development and initial testing, the robustness and reliability of the developed AI system necessitate validation in real-world scenarios. Field testing under varied conditions and scenarios and a range of driver behaviors complement the initial validation performed using these two datasets. Such a thorough approach to validation ensures that the AI system accurately classifies driver behavior, thus reinforcing better autonomous driving.

Challenges and future directions

Artificial Intelligence plays a crucial role in analyzing and monitoring driver behavior in autonomous driving. These advanced systems must be agile, robust, and capable of evolving based on new datasets without compromising detection speed or accuracy.

One technological approach that holds promise in addressing these requirements is the use of Deep Neural Networks (DNNs). DNNs eliminate the need for handcrafted features, leading to improved recognition accuracy. However, they also come with their own set of limitations. They require a substantial amount of data for training, and insufficient data can result in over fitting. There are also computational uncertainties, such as the feasibility of training and updating DNNs effectively in a smartphone environment. Furthermore, these networks have to function in real-time, adding another layer of complexity to the system [142].

Ensuring coordination between multiple elements, such as facial analysis and the monitoring of hand movements, is also essential for providing a comprehensive assessment. Despite significant progress, there are still key issues to be addressed. For example, the acquisition of behavioral data presents certain barriers and unknown complexities when dealing with the security and privacy of the driver. Protecting sensitive information while maintaining data quality over time can pose its own set of challenges. Furthermore, most existing literature and technology have primarily focused on passenger vehicles, neglecting other forms of transportation such as heavy trucks and cargo transport vehicles. Operational constraints, such as inadequate performance under varying light conditions, difficulties in identifying individuals wearing eye-wear, and potential biases in facial recognition, continue to pose challenges [143].

After addressing the challenges of light conditions, biases in facial recognition, and ensuring robustness and transparency, future research should also consider the detection of moods and emotions. This aspect can be crucial in further personalizing and securing the in-cabin experience. By understanding the driver's or passengers' emotional

state, AI systems can offer more responsive and adaptive support, enhancing safety and comfort. This can lead to more sophisticated and user-centered autonomous mobility solutions, expanding the capabilities of AI beyond mere operational efficiency to encompass a more comprehensive understanding of human factors in the context of autonomous transportation [144].

As AI-driven systems penetrate the intimate environment of a vehicle, earning user trust becomes non-negotiable. These systems must be reliable and robust, especially against unexpected sensor failures or road conditions, and they must also be designed for transparency. Users and regulators must be equipped to understand and trust the AI's decision-making processes. In-built verification mechanisms are critical to ensure safety and effectiveness [145].

The adoption of autonomous mobility solutions is gaining remarkable traction globally, especially in dominant markets like China, the United States, and Europe. Future research should endeavor to advance guidance technologies and fine-tune AI algorithms, opening unprecedented opportunities in the logistics, retail, security, maintenance, and agriculture sectors. By addressing these multifaceted challenges, the research community can pave the way for more efficient, transparent, and universally applicable in-cabin AI systems. Whether it is making AI more explainable, the challenges are steep but surmountable, promising a future where AI not only coexists but thrives in synergistic harmony with human needs and safety concerns [146].

Significance of our paper

The field of driver behavior monitoring, essential for the advancement of autonomous vehicles and intelligent transportation systems, is rapidly evolving with diverse methodologies and focuses. Our paper contributes significantly to this domain by offering an in-depth analysis of neural network-based methodologies, specifically ANNs, CNNs, RNNs, and LSTMs, for monitoring driver behavior. It is distinctively focused on computer vision and machine learning technologies, emphasizing practical applications in autonomous driving and advanced driver assistance systems. We hope this paper can serve as a valuable resource for future researchers interested in learning more about the influence of neural networks in driver behavior monitoring systems and autonomous vehicles.

In contrast, other survey papers present a wider range of methodologies and applications not inherently focused on neural network methodologies and computer vision applications. Papers including "A Survey on Driver Behavior Detection Techniques for Intelligent Transportation Systems" and "Driver Behavior Analysis for Safe Driving: A Survey" offer broader surveys of technologies and techniques, covering ADAS, mobile phone sensors, and various detection techniques, including the use of smartphones and wearable devices. These papers provide comprehensive overviews of driver safety technologies and the role of various methodologies in intelligent transportation systems, but with less depth in machine learning specifics compared to our paper. Others, including "Toward Vehicle Occupant-Invariant Models for Activity Characterization" and "A Comprehensive Review of Driver Behavior Analysis Utilizing Smartphones," explore niche aspects including the development of occupant-invariant models and the use of smartphone technologies in driver behavior analysis, respectively. These studies address

Table 5 Comparative analysis of driver behavior monitoring research

Paper	Year	Focus	Methodology	Scope	Significance	Key Contributions	Differences
Ours	2023	Monitoring driver behavior using computer vision and machine learning techniques.	Neural network-based methodologies (ANNs, CNNs, RNNs, LSTMs).	Specifically focused on computer vision and machine learning technologies for driver behavior monitoring.	Relevant for developing specific driver behavior monitoring systems in autonomous vehicles and driver assistance systems.	In-depth insights into specific technologies with practical applications in autonomous driving.	Baseline for comparison
[147]	2023	Systematic literature review of driver behavior classification methods.	Various AI algorithms used for driver behavior classification, including machine learning and deep learning techniques.	Broader in scope, covering a variety of driver behaviors and classification methods.	Comprehensive view of the field of driver behavior classification.	Systematic overview of the field, identifying key contributions and potential research directions.	Systematic literature review, wider AI algorithm scope.
[148]	2022	Developing vehicle occupant-invariant models for activity characterization, focusing on violence and non-violence detection.	Adversarial learning, bilevel learning, and integration of pose information.	Creating actor-invariant models, emphasizing the integration of body posture data.	Insights into the challenges of actor bias in activity characterization.	Novel approaches to enhance the accuracy and transferability of activity recognition systems in vehicles.	Focuses on actor-invariant models, different methodological approach.
[149]	2021	Utilization of in-vehicle cameras for analyzing driver behavior, focusing on distracted and drowsy driving.	Traditional handcrafted feature-based methods and advanced deep learning techniques like CNNs.	Evolution of driver behavior analysis techniques, from traditional methods to advanced machine learning algorithms.	Comprehensive overview of driver behavior analysis using in-vehicle cameras for road safety.	Highlights gaps in NDS video data analytics and potential areas for future computer vision research.	Focuses on in-vehicle cameras, different data analytics approach.
[150]	2019	Reviews driver behavior analysis methods using smartphones for drowsiness and abnormal driving detection.	Smartphone-based sensing schemes and detection algorithms.	Explores smartphone technologies and potential integration with other systems.	Broad applicability and advantages of smartphone solutions in driver behavior analysis for enhancing road safety.	Comprehensive overview of smartphone-based methods, addressing challenges and future directions.	Smartphone-based methods, different technological scope
[151]	2018	Range of driver behavior detection techniques and their application in road safety.	Sensor-based, vision-based, and hybrid systems for driver behavior detection.	Broader range of technologies and approaches for detecting driver behavior.	Comprehensive overview of driver behavior detection technologies for road safety improvements.	Broad overview but may lack depth specific to computer vision and machine learning techniques.	Broader range of detection methods, less ML-specific.
[152]	2017	Comprehensive survey of various techniques for detecting driver behavior in intelligent transportation systems.	Covers a range of technologies including ADAS, mobile phone sensors, and more.	Wider focus on intelligent transportation systems.	Provides a broader understanding of technologies in intelligent transportation systems.	Overview of various technologies, less depth in machine learning specifics.	Broader focus on technologies, less depth in machine learning

Table 5 (continued)

Paper	Year	Focus	Methodology	Scope	Significance	Key Contributions	Differences
[153]	2016	Advancements in driver behavior modeling for ADAS and autonomous vehicles.	Big data, mobile cloud computing, and context-aware services for driver behavior modeling.	Broader scope, including mobile cloud computing and various approaches in modeling driver behavior.	Comprehensive overview of current developments and future directions in driver behavior modeling. Broader perspective on the evolving field of driver behavior modeling, highlighting current challenges and future research directions. Broader scope with big data and cloud computing.	Broader perspective on the evolving field of driver behavior modeling, highlighting current challenges and future research directions.	Broader scope with big data and cloud computing
[154]	2015	Addresses driver drowsiness and distraction, covering a wide range of detection techniques and the use of mobile technologies.	Visual and non-visual feature-based detection techniques, and the use of smartphones and wearable devices.	Broad scope including mobile technologies and vehicular ad hoc networks for driver safety.	Comprehensive overview of driver safety technologies.	Insights into various detection methods and the potential integration of driver behavior analysis into car-to-car communication.	Covers mobile technologies, broader safety focus.
[155]	2013	Reviews various methods for monitoring driver and driving behavior, with a focus on drowsiness and distraction detection.	Visual and non-visual features of driver behavior, driving performance behaviors, and vehicle-based features.	Wider range of methods, including physiological signals and vehicle-based features.	Comprehensive overview of methods for enhancing driver safety systems. Different approaches for driver behavior monitoring and their significance in road safety. Varied monitoring methods, including physiological signals	Different approaches for driver behavior monitoring and their significance in road safety.	Varied monitoring methods, including physiological signals

specific challenges, including actor bias in activity characterization and highlighting the non-intrusive nature of smartphones in driver behavior analysis.

Overall, our paper stands out for its focused approach to advanced machine learning techniques and their practical application in autonomous vehicle systems, offering detailed insights into recent techniques as well as known approaches. For a comprehensive and structured comparison of our work with other significant researches in this domain, refer to Table 5, which presents a comparative analysis of our paper with other current driver behavior monitoring research.

Concluding remarks

This survey paper initiates a thorough exploration of safety concerns related to autonomous vehicles and highlights the critical necessity of an AI-based system for driver behavior classification. It provides a detailed explanation of each key term to ensure complete understanding. This exploration induces the development of a unified AI system. As a result of studying three primary classification methodologies, this new system can exceed current classification techniques in depth and breadth. Two primary components, the core program, and the primary database, make up the base of the proposed system. The study delves deep into the strategies for classifying and evaluating driver behavior in autonomous vehicles and provides a detailed record of these techniques. Lastly, it anticipates a growing focus on safety within autonomous vehicles and suggests that this research could steer the development of more advanced, nuanced systems for driver behavior detection. This foresight lays the groundwork for ongoing and future research in this rapidly evolving field.

Abbreviations

ADAS	Advanced driver-assistance systems
ANN	Artificial neural networks
AI	Artificial intelligence
AUC	American University in Cairo
AV	Autonomous vehicle
AWAKE	System for effective assessment of driver vigilance and warning according to traffic risk estimation
Bi-LSTM	Bidirectional LSTM
BiR- SwinT	Bilinear full-scale residual swin-transformer network
CDCL	Cross-domain complementary learning
CNN	Convolutional neural networks
DAR	Driver action recognition
DBN-BPNN	Deep belief network - back propagation neural network
DRNN	Deep residual neural networks
ECSEL	Electronic components and systems for European leadership
GA	Genetic algorithms
HBPR	Human body posture recognition
HBPS	Human body parts segmentation
HyMobLSTM	MobileNetV3 and LSTM model
IVIS	In-Vehicle information system
LiDAR	Light detection and ranging
LSTM	Long-short term memory
NHTSA	National highway traffic safety administration
NTHU	National Tsing Hua University
OpenCV	Open source computer vision library
PAFs	Part affinity fields
PRYSTINE	Programmable systems for intelligence in automobiles
RBM	Restricted Boltzmann machine
ReLU	Rectified linear unit
RNN	Recurrent neural networks
STMNs	Spatiotemporal multiplier networks
SVM	Support vector machine
USFP	Unobtrusive sensor feed pipeline

VGG	Visual geometry group
VIA	VGG image annotator
WHO	World Health Organization
YOLO	You only look once

Acknowledgements
We would like to thank the anonymous reviewers for their constructive feedback and inspiring comments. The reviewers invaluable comments eminently improved this survey paper.

Author contributions
All authors contributed equally.

Funding
Not applicable.

Availability data and materials
The datasets analyzed during the current study are derived from two publicly available repositories: The State Farm Distracted Driver Detection dataset is publicly available and can be accessed through Kaggle at the following URL: https://www.kaggle.com/competitions/state-farm-distracted-driver-detection. The AUC Distracted Driver dataset is publicly available and can be accessed at the following URL: https://abouelnaga.io/projects/auc-distracted-driver-dataset/. Both datasets are open-access and available for research purposes. Therefore, they can be accessed and analyzed by any researcher who wishes to use them in their study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 26 October 2023 Accepted: 31 January 2024
Published online: 22 February 2024

References

1. Dong Y, Hu Z, Uchimura K, Murayama N. Driver inattention monitoring system for intelligent vehicles: a review. IEEE Trans Intell Transport Syst. 2011;12(2):596–614. https://doi.org/10.1109/TITS.2010.2092770.
2. Bagloee S, Tavana M, Asadi M, et al. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies,. J Modern Transport. 2016;24:284–303. https://doi.org/10.1007/s40534-016-0117-3.
3. S. Tolbert, M. Nojoumian, Cross-cultural expectations from self-driving cars, Preprint (Version 1) available at Research Square. 2023.
4. Craig J, Nojoumian M. Should self-driving cars mimic human driving behaviors?, International on HCI in mobility, transport and automotive systems (MobiTAS), LNCS 12791. Berlin: Springer; 2021.
5. S. Shahrdar, C. Park, M. Nojoumian, Human trust measurement using an immersive virtual reality autonomous vehicle simulator, in: 2nd AAAI/ACM Conference on AI, Ethics, and Society (AIES), 2019. pp. 515–520.
6. Shahrdar S, Menezes L, Nojoumian M. A survey on trust in autonomous systems, computing conference (CC). Berlin: Springer; 2018.
7. C. Park, M. Nojoumian. 2022. Social acceptability of autonomous vehicles: Unveiling correlation of passenger trust and emotional response International Conference on HCI in Mobility, Transport and Automotive Systems (MobiTAS). Springer. Berlin
8. C. Park, S. Shahrdar, M. Nojoumian, EEG-based classification of emotional state using an autonomous vehicle simulator, in: 10th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM). 2018. pp. 297–300.
9. M. Nojoumian, Adaptive driving mode in semi or fully autonomous vehicles, US Patent 11,221,623. 2022.
10. M. Nojoumian, Adaptive mood control in semi or fully autonomous vehicles, US Patent 10,981,563. 2021.
11. Kirkpatrick K. Still waiting for self-driving cars. Commun ACM. 2022;65(4):12–4.
12. J. Leech, G. Whelan, M. Bhaiji, M. Hawes, K. Scharring, Connected and autonomous vehicles-the uk economic opportunity, KPMG. https://www.smmt.co.uk/wp-content/uploads/sites/2/CRT036586F-Connected-and-Autonomous-Vehicles-%E280.2015.
13. S. International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE Int. 2018;4970(724):1–5.
14. J. W. et al., Cnn explainer, Web, Accessed: 09 Sep 2023. https://poloclub.github.io/cnn-explainer/.
15. A. Jain, A. R. Zamir, S. Savarese, A. Saxena, Structural-rnn: Deep learning on spatio-temporal graphs, in: Proceedings of the ieee conference on computer vision and pattern recognition, 2016, pp. 5308–5317.
16. Wollmer M, Blaschke C, Schindl T, Schuller B, Farber B, Mayer S, Trefflich B. Online driver distraction detection using long short-term memory. IEEE Trans Intell Transport Syst. 2011;12(2):574–82.

17. Nascimento JC, Figueiredo MA, Marques JS. Trajectory classification using switched dynamical hidden markov models. IEEE Trans Image Process. 2009;19(5):1338–48.
18. Calderara S, Prati A, Cucchiara R. Markerless body part tracking for action recognition. Int J Multimedia Intell Secur. 2010;1(1):76–89.
19. Ohn-Bar E, Tawari A, Martin S, Trivedi MM. On surveillance for safety critical events: in-vehicle video networks for predictive driver assistance systems. Computer Vision Image Understand. 2015;134:130–40.
20. Jo J, Lee SJ, Jung HG, Park KR, Kim J. Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. Optical Eng. 2011;50(12):127202–127202.
21. W. H. Organization, Road safety, Web, Accessed: 02 Sep 2023.2023. https://www.who.int/health-topics/road-safety.
22. N. H. T. S. Administration, Risky driving, Web, Accessed: 02 Sep 2023. 2023. https://www.nhtsa.gov/risky-driving.
23. Fletcher L, Apostoloff N, Petersson L, Zelinsky A. Vision in and out of vehicles. IEEE Intell Syst. 2003;18(3):12–7. https://doi.org/10.1109/MIS.2003.1200722.
24. Dong Y, Hu Z, Uchimura K, Murayama N. Driver inattention monitoring system for intelligent vehicles: a review. IEEE Trans Intell Transport Syst. 2010;12(2):596–614.
25. B. K. Savaş, Y. Becerikli, Real time driver fatigue detection based on svm algorithm, in: 2018 6th International Conference on Control Engineering & Information Technology (CEIT), IEEE, 2018, pp. 1–4.
26. C. Agarwal, A. Sharma, Image understanding using decision tree based machine learning, in: ICIMU 2011: Proceedings of the 5th international Conference on Information Technology & Multimedia, IEEE, 2011, pp. 1–8.
27. Li Z, Zhang Q, Zhao X. Performance analysis of k-nearest neighbor, support vector machine, and artificial neural network classifiers for driver drowsiness detection with different road geometries. Int J Distributed Sensor Networks. 2017;13(9):1550147717733391.
28. Tango F, Botta M. Real-time detection system of driver distraction using machine learning. IEEE Trans Intell Trans Syst. 2013;14(2):894–905.
29. Darby J, Sánchez MB, Butler PB, Loram ID. An evaluation of 3d head pose estimation using the microsoft kinect v2. Gait & Posture. 2016;48:83–8.
30. D. Zhao, Y. Zhong, Z. Fu, J. Hou, M. Zhao, et al., A review for the driving behavior recognition methods based on vehicle multisensor information. J Adv Transport 2022.
31. K. Allen, Tesla model s in autopilot mode in utah crash; driver had hands off wheel, Web, Accessed: 02 Sep 2018. (2018). https://abcnews.go.com/US/tesla-model-autopilot-mode-utah-crash-driver/story?id=55168222.
32. N. T. S. Board, Collision between a sport utility vehicle operating with partial driving automation and a crash attenuator, Investigation report, National Transportation Safety Board, Accessed: 02 Sep 2023. https://www.ntsb.gov/investigations/Pages/HWY18FH011.aspx.
33. H. Wang, A. Kläser, C. Schmid, C.-L. Liu, Action recognition by dense trajectories, in: CVPR 2011, 2011, pp. 3169–3176. https://doi.org/10.1109/CVPR.2011.5995407.
34. C. Feichtenhofer, A. Pinz, R. P. Wildes, Spatiotemporal multiplier networks for video action recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4768–4777.
35. M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR, 2019, pp. 6105–6114.
36. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint. 2017. arXiv:1704.04861.
37. A. Mujahid, M. Aslam, M. U. G. Khan, A. M. Martinez-Enriquez, N. U. Haq. Multi-class confidence detection using deep learning approach. Appl Sci. https://www.mdpi.com/2076-3417/13/9/55671
38. R. Bakshi, Hand hygiene video classification based on deep learning. Name of the Journal . arXiv:2108.08127.
39. Jegham I, Alouani I, Khalifa AB, Mahjoub MA. Deep learning-based hard spatial attention for driver in-vehicle action monitoring. Expert Syst Appl. 2023;219: 119629.
40. R. Greer, L. Rakla, A. Gopalan, M. Trivedi, (safe) smart hands: Hand activity analysis and distraction alerts using a multi-camera framework. arXiv preprint arXiv:2301.05838. 2023.
41. Abosaq HA, Ramzan M, Althobiani F, Abid A, Aamir KM, Abdushkour H, Irfan M, Gommosani ME, Ghonaim SM, Shamji VR, Rahman S. Unusual driver behavior detection in videos using deep learning models. Sensors. 2013. https://doi.org/10.3390/s23010311.
42. R. Bakshi, Hand pose classification based on neural networks, arXiv preprint arXiv:2108.04529. 2021.
43. Colli Alfaro JG, Trejos AL. User-independent hand gesture recognition classification models using sensor fusion. Sensors. 2021;22(4):1321.
44. Kong L, Xie K, Niu K, He J, Zhang W. Remote photoplethysmography and motion tracking convolutional neural network with bidirectional long short-term memory: Non-invasive fatigue detection method based on multimodal fusion. Sensors. 2024;24(2):455.
45. Bajpai R, Joshi D. Movenet: a deep neural network for joint profile prediction across variable walking speeds and slopes. IEEE Trans Instrument Measurement. 2021;70:1–11.
46. Li L, Zhong B, Hutmacher C Jr, Liang Y, Horrey WJ, Xu X. Detection of driver manual distraction via image-based hand and ear recognition. Accident Anal Prevention. 2020;137: 105432.
47. A. Jinda-Apiraksa, W. Pongstiensak, T. Kondo, A simple shape-based approach to hand gesture recognition, in: ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, IEEE, 2010, pp. 851–855.
48. A. Jinda-Apiraksa, W. Pongstiensak, T. Kondo, Shape-based finger pattern recognition using compactness and radial distance, in: The 3rd International Conference on Embedded Systems and Intelligent Technology (ICESIT 2010), Chiang Mai, Thailand, 2010, pp. –.
49. R. Rokade, D. Doye, M. Kokare, Hand gesture recognition by thinning method, in: 2009 International Conference on Digital Image Processing. IEEE. 2009, pp. 284–287.
50. Tauseef H, Fahiem MA, Farhan S, Recognition and translation of hand gestures to urdu alphabets using a geometrical classification, in,. Second International Conference in Visualisation. IEEE. 2009;2009:213–7.

51. Y. Liu, P. Zhang, Vision-based human-computer system using hand gestures, in: 2009 International Conference on Computational Intelligence and Security, Vol. 2, IEEE, 2009, pp. 529–532.
52. N. Yasukochi, A. Mitome, R. Ishii, A recognition method of restricted hand shapes in still image and moving image as a man-machine interface, in: 2008 Conference on Human System Interactions, IEEE, 2008, pp. 306–310.
53. Yoruk E, Konukoglu E, Sankur B, Darbon J. Shape-based hand recognition. IEEE Trans Image Process. 2006;15(7):1803–15.
54. Das N, Ohn-Bar E, Trivedi MM, On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics, in,. IEEE 18th international conference on intelligent transportation systems. IEEE. 2015;2015:2953–8.
55. Connor J, et al. The role of driver sleepiness in car crashes: a systematic review of epidemiological studies. Accident Anal Prevent. 2001;33(1):31–41. https://doi.org/10.1016/s0001-4575(00)00013-0.
56. K. Hayashi, K. Ishihara, H. Hashimoto, K. Oguri, Individualized drowsiness detection during driving by pulse wave analysis with neural network, in: Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005., 2005, pp. 901–906. https://doi.org/10.1109/ITSC.2005.1520170.
57. T. Ito, S. Mita, K. Kozuka, T. Nakano, S. Yamamoto, Driver blink measurement by the motion picture processing and its application to drowsiness detection, in: Proceedings. The IEEE 5th International Conference on Intelligent Transportation Systems, 2002, pp. 168–173. https://doi.org/10.1109/ITSC.2002.1041208.
58. Smith P, Shah M, da Vitoria Lobo N. Determining driver visual attention with one camera. IEEE Trans Intell Trans Syst. 2003;4(4):205–18. https://doi.org/10.1109/TITS.2003.821342.
59. Ji Q. Non-invasive techniques for monitoring human fatigue. Reno: University of Nevada; 2003.
60. Ranft B, Stiller C. The role of machine vision for intelligent vehicles. IEEE Trans Intell vehicles. 2016;1(1):8–19.
61. D. Park, D. Ramanan, C. Fowlkes, Multiresolution models for object detection, in: Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11, Springer, 2010, pp. 241–254.
62. R. I. China, Global and china heavy truck industry report, 2021-2027, report ID: 6228542, Number of Pages: 130, Format: PDF (January 2022). https://www.reportlinker.com/p06228542/Global-and-China-Heavy-Truck-Industry-Report.html.
63. N. H. T. S. A. (NHTSA), Risky driving: Drowsy driving, Accessed: 02 Sep 2023. https://www.nhtsa.gov/risky-driving/drowsy-driving.
64. Zhu M, Liang F, Yao D, Chen J, Li H, Han L, Liu Y, Zhang Z, Heavy truck driver's drowsiness detection method using wearable eeg based on convolution neural network, in,. IEEE Intelligent Vehicles Symposium (IV). IEEE. 2020;2020:195–201.
65. Schroff F, Kalenichenko D, Philbin J. Facenet: a unified embedding for face recognition and clustering, in. IEEE Conf Computer Vision Pattern Recogn (CVPR). 2015;2015:815–23. https://doi.org/10.1109/CVPR.2015.7298682.
66. Li H, Lin Z, Shen X, Brandt J, Hua G. A convolutional neural network cascade for face detection, in. IEEE Conf Computer Vision Pattern Recogn (CVPR). 2015;2015:5325–34. https://doi.org/10.1109/CVPR.2015.7299170.
67. X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Appearance-based gaze estimation in the wild, in: Proc. IEEE Conference on computer vision and pattern recognition (CVPR), 2015, pp. 4511–4520. https://doi.org/10.1109/CVPR.2015.7299081.
68. Bethge D, Coelho LF, Kosch T, Murugaboopathy S, U. v. Zadow, A. Schmidt, T. Grosse-Puppendahl. Technical design space analysis for unobtrusive driver emotion assessment using multi-domain context. Proc ACM Int Mobile Wearable Ubiquitous Technol. 2023;6(4):1–30.
69. Song W, Zhang G, Long Y. Identification of dangerous driving state based on lightweight deep learning model. Comput Electrical Eng. 2023;105: 108509.
70. Jahan I, Uddin K, Murad SA, Miah M, Khan TZ, Masud M, Aljahdali S, Bairagi AK. 4d: a real-time driver drowsiness detector using deep learning. Electronics. 2023;12(1):235.
71. Akrout B, Fakhfakh S. How to prevent drivers before their sleepiness using deep learning-based approach. Electronics. 2023;12(4):965.
72. Abbas Q, Ibrahim ME, Khan S, Baig AR. Hypo-driver: a multiview driver fatigue and distraction level detection system. CMC-computers Mater Contin. 2022;71(1):1999–2017.
73. Patil D, Lokhande V, Patil P, Patil P, Gaikwad S. Real-time driver behaviour monitoring system invehicles using image processing. Int J Adv Eng Manag(IJAEM). 2022;4(5):1890–4. https://doi.org/10.35629/5252-040518901894.
74. H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, A convolutional neural network cascade for face detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5325–5334.
75. B. Esmaeili, A. AkhavanPour, A. Bosaghzadeh, An ensemble model for human posture recognition, in: 2020 International Conference on Machine Vision and Image Processing (MVIP), IEEE, 2020, pp. 1–7.
76. Fodli MHZM, Zaman FHK, Mun NK, Mazalan L, Driving behavior recognition using multiple deep learning models, in,. IEEE 18th international colloquium on signal processing & applications (CSPA). IEEE. 2022;2022:138–43.
77. Oliver N, Rosario B, Pentland A. A bayesian computer vision system for modeling human interactions. IEEE Trans Pattern Anal Machine Intell. 2000;22(8):831–43. https://doi.org/10.1109/34.868684.
78. Quettier T, Gambarota F, Tsuchiya N, Sessa P. Blocking facial mimicry during binocular rivalry modulates visual awareness of faces with a neutral expression. Sci Rep. 2021;11(1):9972.
79. Karthik L, Kumar G, Keswani T, Bhattacharyya A, Chandar SS, Bhaskara Rao K. Protease inhibitors from marine actinobacteria as a potential source for antimalarial compound. PLoS one. 2014;9(3):e90972.
80. L. Alam, M. M. Hoque, Real-time distraction detection based on driver's visual features, in: 2019 International Conference on Electrical. Computer and Communication Engineering (ECCE), IEEE, 2019, pp. 1–6.
81. P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, Vol. 1, Ieee, 2001, pp. I–I.
82. Liang Y, Reyes ML, Lee JD. Real-time detection of driver cognitive distraction using support vector machines. IEEE Trans Intell Trans Syst. 2007;8(2):340–50.

83. N. Li, C. Busso, Analysis of facial features of drivers under cognitive and visual distractions, in: 2013 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2013, pp. 1–6.
84. Neto LB, Grijalva F, Maike VRML, Martini LC, Florencio D, Baranauskas MCC, Rocha A, Goldenstein S. A kinect-based wearable face recognition system to aid visually impaired users. IEEE Trans Human-Machine Syst. 2016;47(1):52–64.
85. B. Esmaeili, A. Akhavanpour, A. Bosaghzadeh, An ensemble model for human posture recognition, 2020 International Conference on Machine Vision and Image Processing (MVIP). 2020. PP, 1–7.
86. Xing Y, Lv C, Wang H, Cao D, Velenis E, Wang F-Y. Driver activity recognition for intelligent vehicles: a deep learning approach. IEEE Trans Vehicul Technol. 2019;68(6):5379–90. https://doi.org/10.1109/TVT.2019.2908425.
87. Lee M-FR, Chen Y-C, Tsai C-Y. Deep learning-based human body posture recognition and tracking for unmanned aerial vehicles. Processes. 2022. https://doi.org/10.3390/pr10112295.
88. A. Kendall, M. Grimes, R. Cipolla, Posenet: A convolutional network for real-time 6-dof camera relocalization, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 2938–2946.
89. Y. Xie, F. Li, Y. Wu, S. Yang, Y. Wang, D3-guard: Acoustic-based drowsy driving detection using smartphones, in: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, 2019, pp. 1225–1233. https://doi.org/10.1109/INFOCOM.2019.8737470.
90. Yang W, Tan C, Chen Y, Xia H, Tang X, Cao Y, Zhou W, Lin L, Dai G. Birswint: Bilinear full-scale residual swin-transformer for fine-grained driver behavior recognition. J Franklin Instit. 2023;360(2):1166–83.
91. Aljohani AA. Real-time driver distraction recognition: a hybrid genetic deep network based approach. Alexandria Eng J. 2023;66:377–89.
92. Fan C, Huang S, Lin S, Xu D, Peng Y, Yi S. Types, risk factors, consequences, and detection methods of train driver fatigue and distraction. Comput Intell Neurosci. 2022. https://doi.org/10.1155/2022/8328077.
93. Z. Zheng, S. Dai, Y. Liang, X. Xie. 2019. Driver fatigue analysis based on upper body posture and dbn-bpnn model, in: 2019 IEEE 4th Advanced Information Technology. Electronic and Automation Control https://doi.org/10.1109/IAEAC47372.2019.8997925.
94. Kondyli A, Sisiopiku VP, Zhao L, Barmpoutis A. Computer assisted analysis of drivers' body activity using a range camera. IEEE Intell Transport Syst Magazine. 2015;7(3):18–28.
95. Gaglio S, Re GL, Morana M. Human activity recognition process using 3-d posture data. IEEE Trans Human-Machine Syst. 2014;45(5):586–97.
96. Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2d pose estimation using part affinity fields, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7291–7299.
97. M. Rezaei, R. Klette, Look at the driver, look at the road: No distraction! no accident!, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 129–136.
98. Blythe E, Garrido L, Longo MR. Emotion is perceived accurately from isolated body parts,. Especially Hands. 2023. https://doi.org/10.2139/ssrn.4008097.
99. Ezzouhri A, Charouh Z, Ghogho M, Guennoun Z. Robust deep learning-based driver distraction detection and classification. IEEE Access. 2021;9:168080–92. https://doi.org/10.1109/ACCESS.2021.3133797.
100. A. Heitmann, R. Guttkuhn, A. Aguirre, U. Trutschel, M. Moore-Ede, Technologies for the monitoring and prevention of driver fatigue, in: Driving Assessment Conference, 1, University of Iowa, 2001, pp. 81–86. doi:https://doi.org/10.17077/drivingassessment.1013.
101. R. Madigan, Y. M. Lee, N. Merat, C. Goodridge, E. Lehtonen, S. Wolter, M. Wilbrink, M. Oehl, M. Dozza, A. Edelmann, et al., Deliverable d4.4 - user evaluation methods, Project Report D4.4, Name of the Institution (Month (if available) 2023).
102. N. Druml, G. Macher, M. Stolz, E. Armengaud, D. Watzenig, C. Steger, T. Herndl, A. Eckel, A. Ryabokon, A. Hoess, et al., Prystine-programmable systems for intelligence in automobiles, in: 2018 21st Euromicro Conference on Digital System Design (DSD), IEEE, 2018, pp. 618–626.
103. Billah T, Rahman SM, Ahmad MO, Swamy M. Recognizing distractions for assistive driving by tracking body parts. IEEE Trans Circuits Syst Video Technol. 2018;29(4):1048–62.
104. Rahman SM, Howlader T, Hatzinakos D. On the selection of 2d krawtchouk moments for face recognition. Pattern Recogn. 2016;54:83–93.
105. M. Panwar, P. S. Mehra, Hand gesture recognition for human computer interaction, in: 2011 International Conference on Image Information Processing, IEEE, 2011, pp. 1–7.
106. Weyers P, Schiebener D, Kummert A. Action and object interaction recognition for driver activity classification, in. IEEE Intell Transport Syst Conf (ITSC). 2019;2019:4336–41. https://doi.org/10.1109/ITSC.2019.8917139.
107. Xing Y, Lv C, Zhang Z, Wang H, Na X, Cao D, Velenis E, Wang F-Y. Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition. IEEE Trans Computat Soc Syst. 2018;5(1):95–108. https://doi.org/10.1109/TCSS.2017.2766884.
108. Gjoreski M, Gams MŽ, Luštrek M, Genc P, Garbas J-U, Hassan T. Machine learning and end-to-end deep learning for monitoring driver distractions from physiological and visual signals. IEEE Access. 2020;8:70590–603.
109. Ohn-Bar E, Martin S, Tawari A, Trivedi MM, Head, eye, and hand patterns for driver activity recognition, in,. 22nd international conference on pattern recognition. IEEE. 2014;2014:660–5.
110. Baheti B, Talbar S, Gajre S. Towards computationally efficient and realtime distracted driver detection with mobilevgg network. IEEE Trans Intell Vehicles. 2020;5(4):565–74.
111. J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description. 2016. arXiv:1411.4389.
112. J. Mafeni Mase, P. Chapman, G. P. Figueredo, M. Torres Torres, A hybrid deep learning approach for driver distraction detection, in: 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 1–6. https://doi.org/10.1109/ICTC49870.2020.9289588.
113. H. M. Eraqi, Y. Abouelnaga, M. H. Saad, M. N. Moustafa, Driver distraction identification with an ensemble of convolutional neural networks. Journal of Advanced Transportation 2019.
114. Liu Q, Zhou F, Hang R, Yuan X. Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification. Remote Sensing. 2017;9(12):1330.

115. J. Mafeni Mase, P. Chapman, G. P. Figueredo, M. Torres Torres, Benchmarking deep learning models for driver distraction detection, in: Machine Learning, Optimization, and Data Science: 6th International Conference, LOD 2020, Siena, Italy, July 19–23, 2020, Revised Selected Papers, Part II 6, Springer, 2020, pp. 103–117.
116. W. Iqrar, M. Z. Abidien, W. Hameed, A. Shahzad, Cnn-lstm based smart real-time video surveillance system, in: 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2022, pp. 1–5. https://doi.org/10.1109/MACS56771.2022.10023018.
117. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. Mobilenetv 2: inverted residuals and linear bottlenecks, in. IEEE/CVF Conf Computer Vision Pattern Recogn. 2018;2018:4510–20. https://doi.org/10.1109/CVPR.2018.00474.
118. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
119. Kulshrestha A, Chang L, Stein A. Use of lstm for sinkhole-related anomaly detection and classification of insar deformation time series. IEEE J Selected Topics Appl Earth Observ Remote Sensing. 2022;15:4559–70. https://doi.org/10.1109/JSTARS.2022.3180994.
120. Abbasimehr H, Paki R. Improving time series forecasting using lstm and attention models. Journal of Ambient Intell Human Comput. 2022;13(1):673–91. https://doi.org/10.1007/s12652-020-02761-x.
121. A. Koesdwiady, S. M. Bedawi, C. Ou, F. Karray, End-to-end deep learning for driver distraction recognition, in: Image Analysis and Recognition: 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings 14, Springer, 2017, pp. 11–18.
122. Chen J-C, Lee C-Y, Huang P-Y, Lin C-R. Driver behavior analysis via two-stream deep convolutional neural network. Appl Sci. 2020;10(6):1908.
123. Arbabzadeh N, Jafari M. A data-driven approach for driving safety risk prediction using driver behavior and roadway information data. IEEE Trans Intell Transport Syst. 2017;19(2):446–60.
124. Y. Li, Y. Li, N. Vasconcelos, Resound: Towards action recognition without representation bias, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 513–528.
125. Y. Li, N. Vasconcelos, Repair: Removing representation bias by dataset resampling, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 9572–9581.
126. J. Choi, C. Gao, J. C. Messou, J.-B. Huang, Why can't i dance in the mall? learning to mitigate scene bias in action recognition. Advances in Neural Information Processing Systems 32. 2019.
127. K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos. Advances in neural information processing systems 27. 2014.
128. C. Feichtenhofer, H. Fan, J. Malik, K. He, Slowfast networks for video recognition, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6202–6211.
129. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-scale video classification with convolutional neural networks, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2014, pp. 1725–1732.
130. Feichtenhofer C, Pinz A, Wildes RP, Zisserman A. Deep insights into convolutional networks for video recognition. Int J Computer Vision. 2020;128:420–37.
131. Al-Sultan S, Al-Bayatti AH, Zedan H. Context-aware driver behavior detection system in intelligent transportation systems. IEEE Trans Vehic Technol. 2013;62(9):4264–75.
132. H. Wang, C. Schmid, Action recognition with improved trajectories, in: Proceedings of the IEEE international conference on computer vision, 2013, pp. 3551–3558.
133. L. Ge, H. Liang, J. Yuan, D. Thalmann, 3d convolutional neural networks for efficient and robust hand pose estimation from single depth images, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1991–2000.
134. J. F. P. Kooij, N. Schneider, F. Flohr, D. M. Gavrila, Context-based pedestrian path prediction, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13, Springer, 2014, pp. 618–633.
135. D. Bethge, C. Patsch, P. Hallgarten, T. Kosch, Interpretable time-dependent convolutional emotion recognition with contextual data streams, in: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–9.
136. Xing Y, Lv C, Wang H, Cao D, Velenis E, Wang F-Y. Driver activity recognition for intelligent vehicles: a deep learning approach. IEEE Trans Vehic Technol. 2019;68(6):5379–90.
137. Tran D, Manh Do H, Sheng W, Bai H, Chowdhary G. Real-time detection of distracted driving based on deep learning. IET Intell Trans Syst. 2019;12(10):1210–9.
138. Mühlbacher-Karrer S, Mosa AH, Faller L-M, Ali M, Hamid R, Zangl H, Kyamakya K. A driver state detection system-combining a capacitive hand detection sensor with physiological sensors. IEEE Trans Instrumentat Measurement. 2017;66(4):624–36.
139. A. Dutta, A. Zisserman, The via annotation software for images, audio and video, in: Proceedings of the 27th ACM international conference on multimedia, 2019, pp. 2276–2279.
140. Benfold B, Reid I. Guiding visual surveillance by tracking human attention. BMVC. 2009;2:7.
141. State farm distracted drivers dataset, Retrieved 15 Sep 2023 from https://urlis.net/odgzqf60.2023.
142. Chan TK, Chin CS, Chen H, Zhong X. A comprehensive review of driver behavior analysis utilizing smartphones. IEEE Trans Intell Transport Syst. 2020;21(10):4444–75. https://doi.org/10.1109/TITS.2019.2940481.
143. Michelaraki E, Katrakazas C, Kaiser S, Brijs T, Yannis G. Real-time monitoring of driver distraction: state-of-the-art and future insights. Accident Anal Prevention. 2023;192: 107241.
144. Ceccacci S, Maura M, Generosi A, Roberta P, Giuseppe C, Andrea C, Roberto M, et al. Designing in-car emotion-aware automation. Eur Trans Eur. 2021;84:1–15.
145. Holzinger A, Saranti A, Angerschmid A, Retzlaff CO, Gronauer A, Pejakovic V, Medel-Jimenez F, Krexner T, Gollob C, Stampfer K. Digital transformation in smart farm and forest operations needs human-centered ai: challenges and future directions. Sensors. 2022;22(8):3043.

146. Hemmati A, Rahmani AM. The internet of autonomous things applications: a taxonomy, technologies, and future directions. Int Things. 2022;20: 100635.
147. S. Bouhsissin, N. Sael, F. Benabbou, Driver behavior classification: a systematic literature review, IEEE Access. 2023.
148. Capozzi L, Barbosa V, Pinto C, Pinto JR, Pereira A, Carvalho PM, Cardoso JS. Toward vehicle occupant-invariant models for activity characterization. IEEE Access. 2022;10:104215–25.
149. Wang J, Chai W, Venkatachalapathy A, Tan KL, Haghighat A, Velipasalar S, Adu-Gyamfi Y, Sharma A. A survey on driver behavior analysis from in-vehicle cameras. IEEE Trans Intell Transport Syst. 2021;23(8):10186–209.
150. Chan TK, Chin CS, Chen H, Zhong X. A comprehensive review of driver behavior analysis utilizing smartphones. IEEE Trans Intell Transport Syst. 2019;21(10):4444–75.
151. Alluhaibi SK, Al-Din MSN, Moyaid A. Driver behavior detection techniques: a survey. Int J Appl Eng Res. 2018;13(11):8856–61.
152. R. Chhabra, S. Verma, C. R. Krishna, A survey on driver behavior detection techniques for intelligent transportation systems, in: 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, IEEE, 2017, pp. 36–41.
153. N. AbuAli, H. Abou-Zeid, Driver behavior modeling: Developments and future directions. International journal of vehicular technology 2016.
154. Kaplan S, Guvensan MA, Yavuz AG, Karalurt Y. Driver behavior analysis for safe driving: a survey. IEEE Trans Intell Transport Syst. 2015;16(6):3017–32.
155. H.-B. Kang, Various approaches for driver and driving behavior monitoring: A review, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 616–623.

Publisher's Note

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