Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/31/20230123

Driver's hazardous state detection in human-computer interaction of automotive cockpits

Xin Zhang

College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong, China.

20201209425@stu.sxau.edu.cn

Abstract. Today, the smart car industry is growing rapidly, the functions of the intelligent cockpit based on human-computer interaction are more and more extensive, and the sales volume of intelligent vehicles continues to rise. The incidence of traffic crashes caused by the unsafe state of drivers remains high. The different behavioral states that drivers may emit during driving is a necessary consideration in the design of the intelligent cockpit. This paper takes the driver's state as the starting point to systematically consider the driver's state detection. Summarizing the driver's state detection from four parts: eye state, limb state, facial state, and language state. This paper introduces the development status of the current four types of detection systems, focusing on eye state recognition and limb state recognition. The key driver's characteristic signals are mainly collected by the camera. The driver's state is judged by deep learning, machine learning, and database. This paper is more systematic and comprehensive than the existing literature. Comprehensive consideration of the driver's state contributes to the driver and passengers.

Keywords: driver status detection, eye state recognition, limb state recognition, driver safety.

1. Introduction

The automobile is an indispensable means of transportation for modern people to travel. With the development of today's intelligent technology and the advanced design of automobile intelligent cockpit, the car cabin has become the third living environment for people in addition to living accommodations and work [1].

More and more people buy cars, making the number of cars on the road grow rapidly. The World Health Organization (WHO) report shows, about 1.3 million people die in traffic accidents worldwide each year, and between 200,000 and 500,000 non-fatal injuries occur every year and many people are disabled as a result [2]. In the current detection system applied to the automotive cockpit, The driver's dangerous state detection system is relatively fixed, and many detections are static detections, which cannot dynamically capture the dangerous state that the driver may emit. For automobile enterprises, in the design of automobile cockpit, designers cannot fully and systematically consider how to completely detect the dangerous state of drivers, resulting in insufficient auxiliary safety functions of automobiles. For the driver, the imperfection of the detection system makes the driver unable to be fully detected when there are various dangerous states, and driving safety cannot be improved. When driving a car, the driver needs to focus more attention on driving, and the judgment of the external environment cannot be accurately judged. The driving of the vehicle is determined by the driver's operation. Compared with other objective factors, the dangerous state of normal drivers during driving is more due to their own subjectivity [3]. It is necessary to systematically examine the hazardous states of drivers in the driving cabin and conduct danger state detection.

The driver's dangerous state detection system is mainly composed of two parts: dangerous action state detection and dangerous emotional state detection. The state of dangerous action is divided into the eye state and the limb state. Dangerous emotional states are divided into facial expressions, and language states. For eye detection, parameters can be obtained from the driver's eye movement signal to analyse whether the driver's cognitive load affects driving [4]. Eye aspect ratio real-time parameters [5], analyse whether the driver is in fatigue. Limb danger is mainly due to the process of drinking water, answering the phone, smoking and other actions, the hand from the steering wheel [6], which may cause driver distraction and affect driving safety. Emotion is one of the important factors affecting safe driving. Through the analysis of eye and mouth, the driver's emotions are obtained from the face, and the driver's emotions are divided into happy, angry, sad, and anger [7]. Language plays a crucial role in understanding the driver's emotions and is an important avenue for assessing their emotional state. Combining EEG and language signals [8], the real emotional state of the driver can be accurately identified. In the existing research, the work of dangerous state recognition is mainly based on CNN deep learning method [8], which judges the driver's state through static image analysis. At present, the driver's dangerous state detection system is not perfect, and the literature research is relatively single. Designers cannot comprehensively consider the design of a safety system. In this paper, the existing relevant literature is reviewed, and the dangerous state detection technology of automobile cockpit drivers is systematically reviewed. Intelligent vehicle cockpit designers can consider the driver's dangerous state detection system more comprehensively and improve the safety of intelligent vehicles.

The structure of this paper is as follows: Section 1 introduces the development current situation and background of driver status detection and reviews the existing literature. Section 2 systematically reviews and analyse the existing literature from different angles. Section 3 analyses the literature data and summarizes the literature research results. Section 4 puts forward the basic model of the detection system. Section 5 summarizes the thesis.

2. Driver hazardous state detection module system

The automobile industry has developed rapidly in recent years, the number of cars on the road has increased rapidly, and traffic accidents have occurred more frequently than before. Under normal circumstances (regardless of drunk driving, disability, serious illness, etc.), drivers may have a series of dangerous states that affect driving safety during driving. Improving the detection of dangerous states of automobile drivers is an important task that cannot be ignored now and in the future. At present, most of the intelligent vehicle cockpit design based on human-computer interaction is about entertainment systems and control systems. The application of driver safety detection systems is not popular enough. Even if there is, it is mainly based on simple fixed action detection systems. The development of intelligent vehicles requires not only the wide application of intelligent vehicle cockpits but also the comprehensive design and innovation of dangerous state detection systems.

The detection in the cockpit of the human-computer interaction vehicle is determined, and the driver's dangerous state is classified. The relevant classification is searched one by one in PubMed, IEEE, ACM, MDPI, Google Scholar WOS, and WOS. Use the following keywords to search in the above database: human-computer interaction vehicle, human-computer interaction vehicle cockpit, driver's dangerous behaviour, driver's dangerous state, driver's eye movement recognition, driver's dangerous action, driver's emotion recognition, driver's face recognition, driver's speech recognition, and other keyword combination search.

Combined with the main factors affecting driving safety, this paper divides the driver dangerous state detection system into two parts: 1. Driver dangerous action state. Detection technology includes eye movement state detection and limb movement state detection. 2. Driver emotional state. Detection techniques include facial state detection and language state detection. Combined with the review research, the driver dangerous state detection module diagram shown in Figure 1 is listed.

DOI: 10.54254/2755-2721/31/20230123

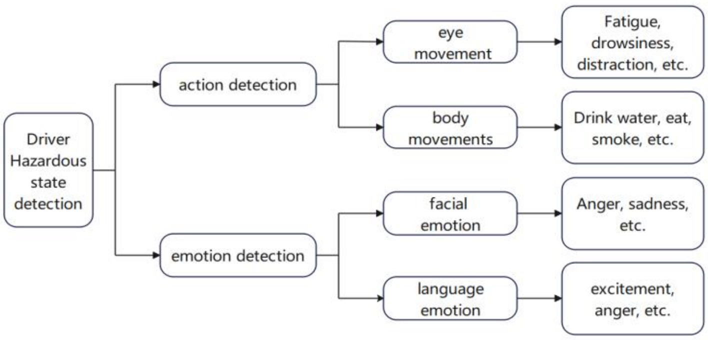

Figure 1. Driver dangerous state detection module

3. Research on dangerous action states

The driver's action state directly affects driving safety and passenger safety. The subtle driving action is mainly an eye movement. The eye movement index can show whether the driver's state is fatigue and drowsiness, and whether it affects driving safety. The obvious dangerous action is mainly that the driver leaves the steering wheel to drink water, eat, smoke and so on during the driving process, which affects driving safety.

3.1. Eye movement research

Rahman et al. studied enhancing the accuracy of classifying driver's cognitive load through eye movement parameters [4]. In the study, the face in the video is extracted, and then the image is processed to determine the eye position through Eye-Pupil Detection. Researchers divide the original eye movement signal into two types: fixation and saccade. When the eye continues to gaze at a certain position, it is called fixation. When two eyes move rapidly from one fixation stage to another, it is called saccade. The eye position extraction method is shown in Figure 2. Using image processing, the video file is first converted into a grayscale image, and finally, the circular eye target is detected for calculation. According to the adjacent position between the two eyes and the time of motion calculation speed, the eye movement position characteristics of saccade and gaze are calculated and judged. By comparing the three deep learning algorithms of CNN, LSTM , and AE, the final result shows CNN has the highest accuracy. Accuracy classification of cognitive load by SVM, LR, LAD, K-NN, and DT. Using the SVM classifier, the results with an average maximum accuracy of up to 92 % are shown by the final test of camera features.

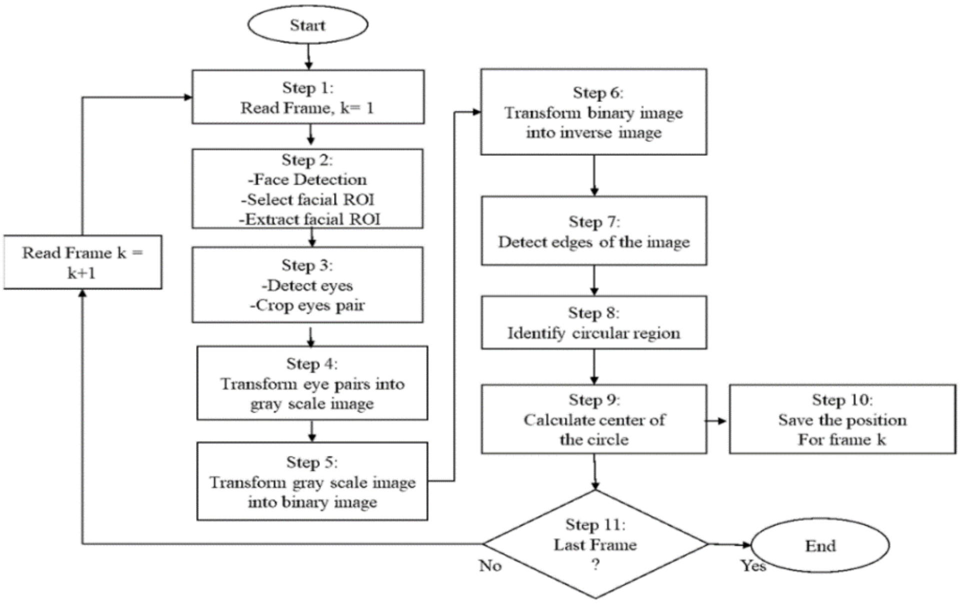

Figure 2. Eye position extraction method [4].

Wang et al. proposed a method for the driver to track the human eye fixation point using the nonlinear fixation point refinement of the dual camera system. [9]. Figure 3 is the proposed method system diagram. The two cameras are used to collect eye information and driving scene information respectively. This paper introduces driver state tracking, pre-training model, and nonlinear gaze refinement, aimed at improving driver state classification. The system of tracking module uses a Kalman filter to track and analyse the measured values to minimize random interference in the measured values. The corresponding driver feature gaze mapping is a pre-trained gaze model is created during offline training and gives the initial gaze point. The nonlinear gaze refinement model corrects the initial gaze point to the real eye gaze position, which effectively improves the results of the pre-training model and gaze prediction. Figure 3 showed Each module of the non-linear gaze point refinement system.

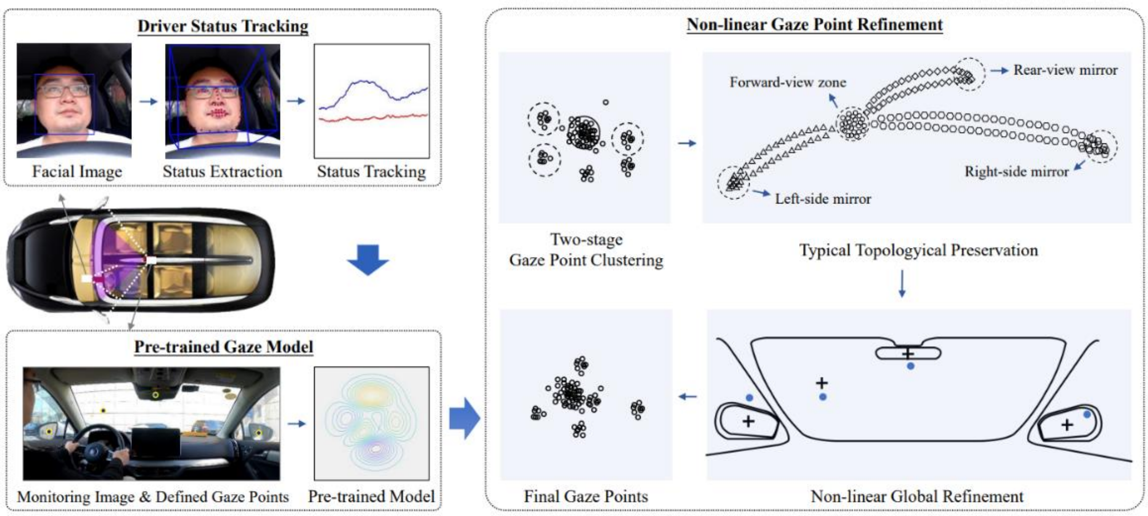

Figure 3. Overview system [9].

The blink phenomenon can reflect whether the driver is in a drowsy state. Dewi et al. proposed a method for real-time recognition of blinks in video recorded by car cameras [5]. In the study, the Eye aspect ratio (EAR) is utilized for assessing the extent of eye closure. The EAR liminal value can reflect whether the state of the driver's eyes is open or closed. In the experiment, the position of facial landmarks is estimated by EAR, a single scalar is extracted, and the proximity of eyes is recognized in each frame. Finally, the blink recognition in a short time is achieved by combining the modified EAR threshold with the EAR value pattern. The final experiment shows that the accuracy and performance of AUC (model evaluation index) are affected by the EAR threshold, 0.18 is the best EAR threshold. Figure 4 showed the facial features in the state of the opining eye and closing eye.

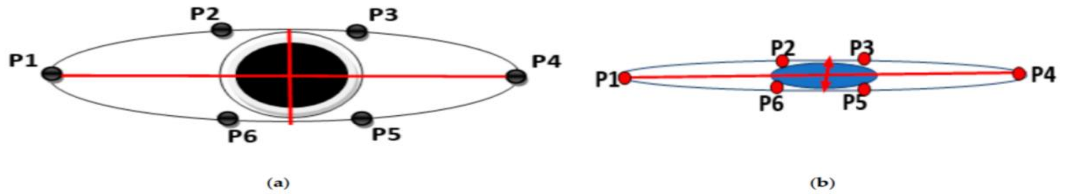

Figure 4. P1-P6 are six coordinates around the eye: (a) Open eyes. (b) Close eyes [5].

3.2. Research on limb status

The dangerous limb state of the driver during driving is mostly related to distraction, such as drinking water, eating, smoking, and so on. In Yan et al.' s study [6], classification is conducted based on three aspects: hand position, head position, and object position relative to the hand and head. The authors identified range of areas for four main locations. This paper proposes a cascade CNN model (CCNN), which is optimized under conditions of compatibility and mutual exclusivity recognition of different limb positions. Figure 5 shows, If the gesture corresponds to a two-hand driving pose from the hand pose set, the pose contradicts the five poses of the behavioral pose set. Before the result is output, the input pose data will only pass through four CNN word models. In this way, the consumption of time and resources is reduced, and the computational efficiency of the model will be improved. Figure 5 shows the CNN models mentioned.

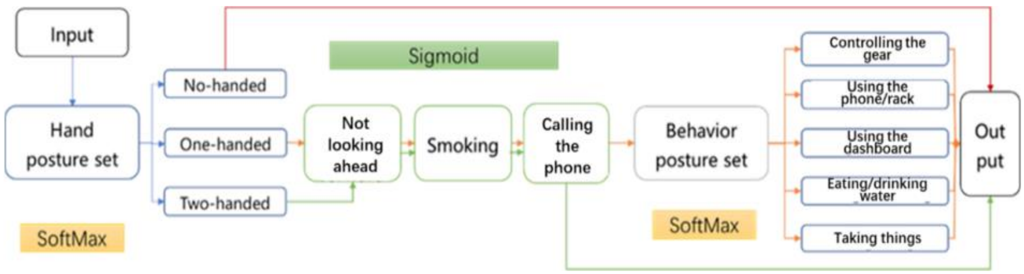

Figure 5. Cascaded CNN models [6].

3.3. Facial emotion detection

Ryu et al. studied a method of using facial images and physiological skin signals to judge driver's emotions [7]. The recognizer is based on driver's facial expression recognition (FER) and Sensor Fusion Emotion Recognition (SFER). In the study, A new method using multiple convolutional neural networks is proposed. The facial image obtained by the camera is used as the input of the model, and the deep neural network (DNN)-based FER model outputs the driver's face's expression state. The recognized expression state is combined with the driver's physiological signal processed by the skin electrical sensor, and the SFER model determines the driver's true emotion. A DRER model with excellent accuracy and recognition of many emotional states is presented using the best accuracy of the FER model and the SFER model. The model only uses face images and EDA signals to identify the driver's induced emotions in the driving state, and the accuracy rate reaches 88.6 %. Figure 6 shows the proposed steps.

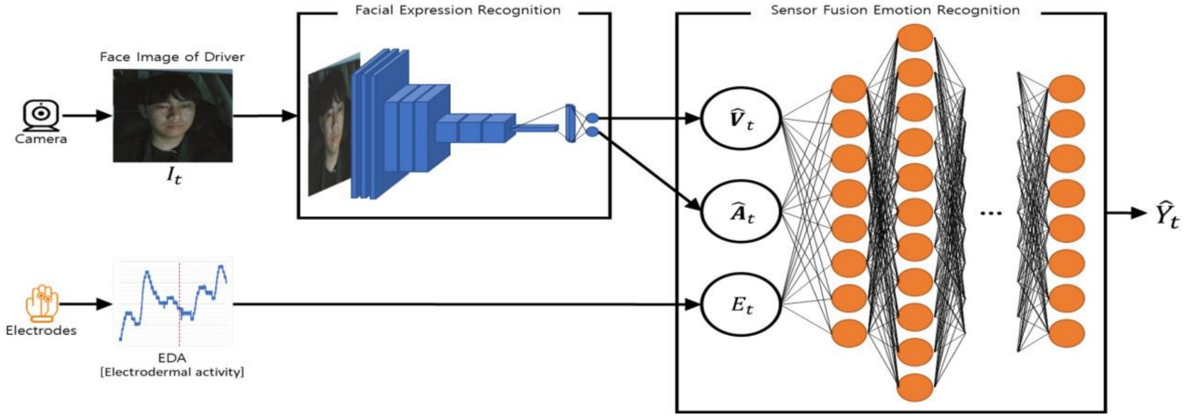

Figure 6. The work includes two main steps: FER and SFER [7].

3.4. *Language emotion detection*

John et al. proposed the EmNet model [8]. The model combines 1: it preserves the temporal information for effective emotion recognition by using common features that are useful for emotion recognition without eliminating their temporal information; 2: Deep neural networks capture temporal patterns of features and correlate them with emotional states, extracting higher-level representations. The model includes multiple processing stages: feature extraction, feature normalization, local convolutional CNN layer, global convolutional CNN layer, LSTM layer, and feedforward layer. Figure 7 describes the EmNet architecture designed for emotion recognition.

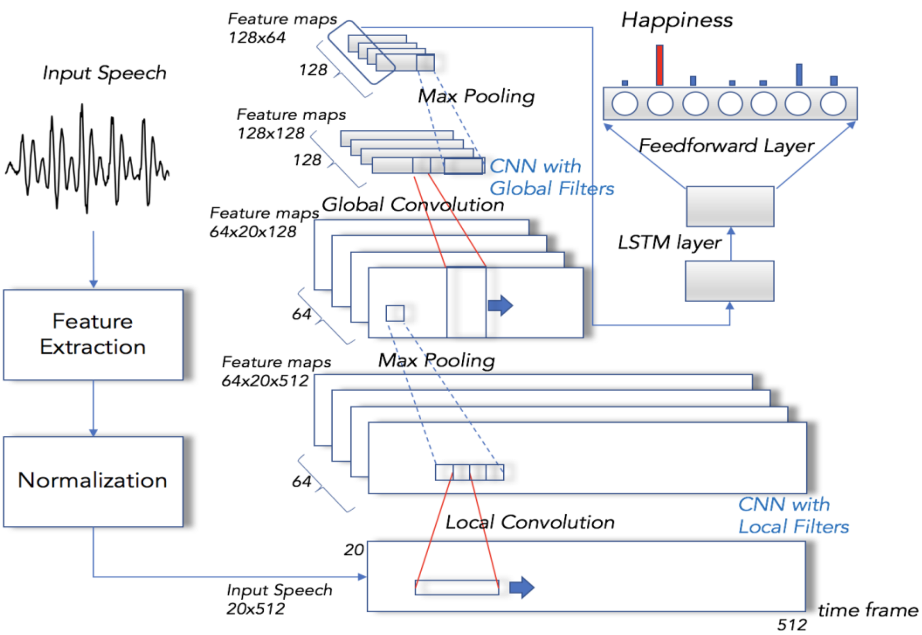

Figure 7. EmNet model structure [8].

Ancilin and A. Milton studied speech emotion recognition based on improved Mel frequency amplitude coefficient (MFAC) [10]. In this paper, the extraction of Mel frequency amplitude coefficients excludes the use of energy spectrum and discrete cosine transform. Three conventional spectral features of Mel frequency amplitude coefficient and Mel frequency cepstrum coefficient (MFCC), logarithmic frequency power coefficient, and linear prediction cepstrum coefficient are evaluated on multiple emotional speech databases of different languages. It is verified that the Mel frequency amplitude coefficient has a significant performance in classifying emotion. Compared with traditional features,

DOI: 10.54254/2755-2721/31/20230123

MFAC proves to be a superior spectral feature for emotion recognition in language. Figure 8 showed the Extraction process of MFCC and MFMC.

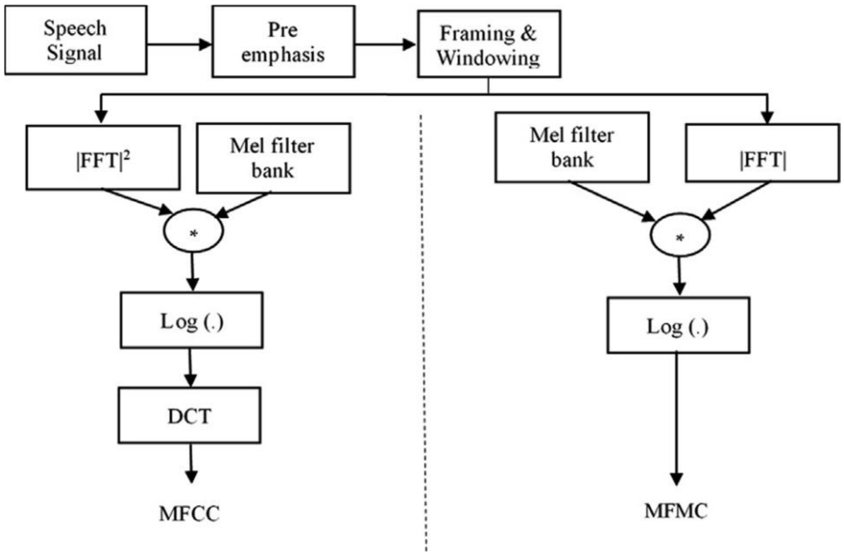

Figure 8. Extraction of MFCC and MFMC [10].

4. Discussion

Eye state: It is difficult to observe the features. In the detection, it is necessary to intercept the image with reference value in the camera video based on the algorithm, and then process the image to filter out the influencing factors, reduce the error to the minimum to extract the subtle features of the eye, and use the eye threshold and machine learning to judge the driver's eye state. Limb state: The detection divides the detection area. The proposed CCNN model can efficiently detect the driver's limb state and has good detection accuracy for the continuity. The driver's limb state has diversity and superposition. Facial state: The detection accuracy of the combination of the two is higher, but in practical applications, the EDA sensor needs to be in contact with the driver and needs to be improved. Language state: In the study, the improvement of the Mel frequency amplitude coefficient further improved the performance of speech emotion recognition. The EmNet model links the extracted emotional features to time.

This study comprehensively starts from the driver's own state and systematically summarizes the state of the driver's driving process. Compared with the existing research, it is complete and convenient for the integrity reference design and improvement of the driver's dangerous state detection system. In research, the robustness of the system, the performance of different driver's dangerous states, and the application cost are not considered enough. In the future research, these aspects need to be further studied. At the same time, more data sets and multiple driving scenarios need to be established to establish a more comprehensive driver's dangerous state detection system.

5. Conclusion

This study comprehensively considers the different dangerous states of human-computer interaction drivers, classifies different states, and systematically reviews the driver's dangerous state detection. Through the eye pupil detection technology, the driver's fixation point, fixation time, and eye movement trajectory can be obtained. Through the vehicle's internal camera or sensor, the driver's posture, hand movements, body movements, etc. can be detected. Abnormal limb posture, or hand operation may indicate that the driver is in a dangerous driving state, which may affect driving performance and safety. Through facial expression analysis and emotion recognition algorithms, it can be judged whether the driver is in a state of anger, pleasure or depression that affects driving. Through speech signal processing and emotion recognition technology, the driver's voice emotion, speech speed, tone change, and

DOI: 10.54254/2755-2721/31/20230123

language content can be identified, and the tension, confusion, or anxiety that may affect the driver's emotional state can be detected.

In the future, the driver's dangerous state detection based on human-computer interaction will be the key direction of intelligent vehicle design. The research will focus on multimodal data fusion, deep learning, artificial intelligence, real-time warning, and the combination of advanced driving assistance systems and automatic driving technology. The state detection system based on human-computer interaction will continue to develop and improve, providing more support and guarantee for drivers.

References

[1] Cai M, Wang W. Summary of research on the interactive design of automobile intelligent cockpit Packaging. Engineering, 2023, 44(06), 430-40.
[2] World Health Organization. Road Traffic Injuries. Available online: http://www.who.net/news-room/fact-sheets/detail/road-traffic-injuries (accessed on 8 June 8, 2023).
[3] Liu S, Wang X, Ji H, Wang L, Hou Z. A novel driver abnormal behaviour recognition and analysis strategy and its application in a practical vehicle. Symmetry, 2022, 14(10), 1956.
[4] Rahman H, Ahmed M. U, Barua S, Funk P, Begum S. Vision-based driver's cognitive load classification considering eye movement using machine learning and deep learning. Sensors, 2021, 21(23), 8019.
[5] Dewi C, Chen R-C, Chang C-W, Wu S-H, Jiang X, Yu H. Eye aspect ratio for real-time Drowsiness detection to improve driver safety, Electronics, 2022, 11(19), 3183
[6] Yan X, He J, Wu G, Zhang C, Wang C. A proactive recognition system for detecting commercial vehicle driver's distracted behaviour. Sensors, 2022, 22(6), 2373.
[7] Agrawal U, Giripunje S, Bajaj, P. Emotion and gesture recognition with soft computing tool for driver's assistance system in human-centered transportation. IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 2013, pp. 4612-4616.
[8] Ali M, Mosa A. H, Machot F. A, Kyamakya K. Emotion recognition involving physiological and speech signals: A comprehensive review. Recent Advances in Nonlinear Dynamics and Synchronization: With Selected Applications in Electrical Engineering, Neurocomputing, and Transportation, 2018, 287-302.
[9] Wang Y, Ding X, Yuan G, and Fu X. Dual-cameras-based driver's eye gaze tracking system with non-linear gaze point refinement. Sensors, 2022, *22*(6), 2326.
[10] Ancilin J, & Milton A. Improved speech emotion recognition with Mel frequency magnitude coefficient. Applied Acoustics, 2021, 179, 108046.