Vol. 10 (2025): Proceedings of the 2nd International Conference on Business, Management and Sustainability (ICBMS 2025)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/icbms.2502.420427

Investigation and Analysis about Multimodal Fatigue Detection

Keran Huang*

XJTU-POLIMI Joint School of Design and Innovation, Xi'an Jiaotong University, Xi'an, China

**Corresponding author: Keran Huang*

Abstract

Fatigue driving accounts for a significant proportion of traffic accidents. Multimodal fatigue detection offers a new effective way to detect fatigue. This article selected six articles by some criteria and reviewed them carefully. The improvements of the method include using lightweight YOLOv8 networks to process complex images; others process multimodal EEG signals via non-smooth non-negative matrix factorization (nsNMF) and Gramian angular field imaging. Attention-based networks like MMA-Net and TMU-Net are designed to fuse EEG, EDA, PPG, and EOG signals. Additionally, LSTM-based models analyze PPG and facial features, while enhanced MTCNN and PFLD algorithms improve detection accuracy and reduce individual variability. Based on the methods, the article summarizes some challenges in the multimodal fatigue detection and proposes the future of the multimodal fatigue detection. In general, the article provides a comprehensive review to guide further research about multimodal fatigue detection.

Keywords

multimodal, fatigue driving, fatigue detection, EEG multimodal, multimodal improvement

1. Introduction

Nowadays, technology and economy are developing rapidly, private cars have gradually entered into daily life. Cars are always the first choice when people want to go out. From the National Bureau of Statistics of China's data, the traffic accidents are increased year by year, caused a large amount of unnecessary economic losses. Among all the factors that cause traffic accidents, fatigue is the most serious. According to the statistics of the National Highway Traffic Safety Administration of the United States, about 100,000 traffic accidents are caused by fatigue driving, the fatality rate is as high as 83% (Cheng et al., 2025). When the driver is in a fatigued condition, the ability to reaction and pay attention to the surrounding environment will decrease, which makes driving more dangerous. Fatigue cannot be avoided, but the fatigue detection in the car can alarm the drivers to stop driving. With the fatigue detection in the car, drivers will be more cautious in driving, and traffic accidents will be prevented. Hence the research in the detection of fatigue is significant and potential.

The traditional fatigue detection can be divided into three categories. The first category is based on physiological characteristics. Chang et al used Spatial Characteristics of EEG Signals to detect drivers' fatigue and got 70% average accuracy (Chang et al., 2024). The second category is based on the driving trajectory. Kai et al proposed a fatigue monitoring system that is based on lane line (Yan et al., 2022). The third category is based on facial recognition. Huang et al proposed a driver fatigue and dangerous driving detection system based on facial recognition (Huang et al., 2021). The three categories can detect driver's fatigue conditions.

However, they have some problems when using traditional methods. The traditional method uses the single model to detect fatigue, which can be affected by environment and individual differences easily that cause the detection accuracy decrease. For the last few years, multimodal detection gradually rose. Based on different kinds of data, multimodal systems can learn more features from the data, which can highly improve accuracy and robustness. Hence, applying the multimodal technique into the fatigue detection can play an important role in improving the fatigue detection.

There are many researchers currently engaged in multimodal research and have some results. Ge et al. used visible light and infrared images to achieve lightweight networks (Ge et al., 2025), Cheng et al. used multimodal Electroencephalogram (EEG) to detect fatigue (Cheng et al., 2025). By using multimodal, the accuracy of the fatigue detection system is improved. This article aims to focus on the industrial field condition, analyzes the six novel ways in multimodal fatigue detection and making conclusions, finally proposed the opinions of the challenge and the future of the multimodal fatigue detection.

1. Methods

1.1 Literation Selection

This review retrieved databases of CNKI, Web of Science databases to get the articles, and the time span of the publication time of the paper was from 06/2024 to 08/2025 to ensure the quality and timeliness. The language of the articles was limited into English and Chinese. There are two inclusion criteria. Firstly, multimodal fatigue detection that was proposed in the article should be used on cars. Secondly, the result of the research should have comparsion with original detection methods. There also have some exclusion criteria. Firstly, the article is dissertations. Secondly, the accuracy about the multifatigue detection is lower. Thirdly, the Impact of the factor of the source is lower. A total of 6 literatures were ultimately included.

1.1 Based on YOLOv8 and Lightweight Dual-Mode

Ge et al improved the original YOLOv8 algorithm to achieve the goal of lightweight. C2f-SCConv was proposed based on a novel perspective that analyzes the feature extraction mechanism of convolutional neural networks (CNNs). SCConv module is composed of Spatial Reconstruction Unit (SRU) and Channel Reconstruction Unit (CRU) (Ge et al., 2025), leading to the proposal of a method that more effectively exploits spatial and channel redundancy. To merge more feature information from different scales, bidirectional feature pyramids (BIFPN) is used into the YOLOv8 networks' head layer. Ge et al invented the lightweight neural networks (Ghost-Net) by stacking Ghost module. This novel network plays an important role in deploying the system to embedded hardware devices to achieve fatigue detection. Through the ablation experiment, the new algorithm has a stable loss function, the accuracy of the new algorithm rises to 98.1% and the system's capacity is decreased to 3.8MB. According to Table 1 (Ge et al., 2025), the new algorithm exhibits lower capacity and higher accuracy. This new algorithm can meet the requirements of real-time fatigue detection and provide a new model that combines visible light and infrared images to overcome the complex light environment.

Table 1: Comparative Trial Result

Algorithm	Accuracy	Weight
Faster-rcnn(origin)	95.8%	108MB
SSD(origin)	94.6%	92.1MB
YOLOv5 Dual-modal SSD	96.6% 96.32%	14MB 45.6MB
Mediapipe modal	94.62%	42MB
Bi-LSTM	96.7%	48.7MB
YOLOv8	98.1%	3.8MB

1.2 Based on EEG Channel-Weighted Multimodal

Cheng et al summarized the current mainstream method I n the fatigue detection based on EEG and found that few studies have considered feature-level multimodal fusion. To extract the frequency domain features of each channels EEG, differential entropy is used into the algorithm. Cheng et al proposed a new model that introduced the non-smooth non-negative matrix factorization (nsNMF) algorithm into the shallow layers of the networks that can acquire the contribution degree of each electrode channel (Cheng et al., 2025). Based on the theory of the EEG imaging, the Gramian angular field imaging method was introduced into the middle layer of the networks. The Gramian angular field imaging method can convert one-dimensional time series into two-dimensional images, which can retain the temporal characteristics of the signals and have more features to be extracted. The Parallel Convolutional Neural Network (PCNN) was used to process the EEG two-dimensional images, and the Gated Recurrent Unit (GRU) was used to process the channels' contribution. The PCNN and the GRU are adopted in a parallel approach as a feature-level multimodal data fusion module. The multi-head self-attention (MSA) mechanism was used in the deep layer of the network to deeply explore the dependency relationships of features, obtain multiple sets of attention results and perform concatenation and linear projection. Table 2 (Cheng et al., 2025) shows that the accuracy of the model can up to 93.37% on the SEED-VIG dataset, which is higher than the advance model in recent years. This method will provide certain theoretical reference value for the design of portable brain-computer interface intelligent driving safety prevention and control systems.

Table 2: Ablation experimental results using mixed data from all subjects on SEED-VIG

Algorithm	Accuracy	Precision	Recall
Without DE features	0.6651	0.8105	0.7022
Without channel contribution	0.8619	0.8964	0.8766
Only GASF	0.9174	0.9511	0.9176
Only GADF	0.8972	0.9087	0.9230
Without GRU	0.9064	0.9229	0.9247
Without MSA	0.9040	0.9259	0.9156
Original method	0.9337	0.9559	0.9345

1.3 Based on Multi-Modality Attention

Guo et al proposed an Attention network model that is based on three signals to detect fatigue. The frontal EEG, the electrodermal activity (EDA) and the photoplethysmography (PPG) are chosen to detect fatigue. The three signals have more advantages in data availability and driving adaptability, which is better than other signals. Three signals are mixed to form a multi-modality signal, which has comprehensive information. To excavate more features from the multi-modality signal and improve the accuracy of fatigue detection, a Multi-Modality Attention Network (MMA-Net) was designed. The MMA-Net is mainly composed of two modules including signal adaptive coding module (SAC-M) and attention-based feature dissimilation module (AFD-M) (Guo et al., 2025). SAC-M is composed of three distinct branches to process three signals. 1D-CNN was embedded in the SAC-M to process data. AFD-M uses attention mechanism to allow models to focus on the feature that is the most important. AFD-M will calculate the attention scores and distill the signal modality with high scores. The simple design of AFD-M performs well in multimodal fatigue detection, demonstrating the effectiveness of the attention mechanism in emphasizing key signal modes. Based on the MMA-Net, Guo et al conducted an ablation verified the effectiveness of the multi-modality signal and found that the 3s window length as the input of the MMA-Net show the best accuracy in the comparison experiment (Guo et al., 2025). Compared with other advanced methods, MMA-Net shows high accuracy and effectiveness.

1.4 Based on LSTM Model

Lu et al proposed a model that is based on Long Short-Term Memory (LSTM) and use PPG signals, facial features and head postures to detect fatigue detection. The author adopted suitable methods to process each signal. For PPG signals, Lu et al considered the small amplitude of the PPG signals and the interference from the noise, the author used Butterworth digital filter in MATLAB. For the facial features, Lu et al used facial features extraction that based on the Dlib library, the use of the facial features extraction mainly for the location of drivers' eyes and mouth. With the fatigue calculation formula, the system can calculate driver's fatigue condition. For the head postures, the author used perspective-n-point (PnP) method to calculate head postures. After the signal processing, the author used LSTM network to detect fatigue and tried different optimizers to compare the final accuracy. According to Table 3 (Yu et al., 2024), the Adam optimizer shows the best performance in the experiment. It is worth mentioning that the experiment was conducted in an actual driving environment, which can improve the data quality and find some problems easily.

Table 3: Result about different optimizers

Algorithm	Accuracy	Precision	Recall	
Adam	0.9778	0.9781	0.9777	
Momentum	0.8336	0.8349	0.8336	
Rmsporp SGD	0.7379 0.7087	0.7376 0.7070	0.7376 0.7079	

1.5 Based on the Improved Algorithm

Li proposed a high precise fatigue detection system based on the improved algorithm that overcomes the problem of system accuracy degradation in complex environments. The author started from the Temporal Convolutional Neural Network (TCNN) to achieve facial detection. Fig.1 shows the details about facial detection. Grouped convolution and depth-separation convolution are introduced to reduce complexity and improve system performance in complex environments. The facial detection data will be processed by Practical Facial Landmark Detector (PFLD) to locate the facial key point. The structure of the PFLD and the loss function are improved by the author to improve accuracy in the complex environments. The author uses the lightweight Mobilenet-V2 as the structure and an auxiliary network is introduced to predict the head posture. The sample weights and the geometric information are added into the loss function to assign higher weights to rare samples. Based on the accurate facial key point, the system calculates the fatigue characteristics of eyes, head and mouth and alarm drivers to keep away from the danger. Figure 2 shows the details about the features extraction about eyes and mouth. The system detects fatigue by multi-features and uses deep learning models to improve the generalization ability to the new data. The system's fatigue detection average accuracy can up to 97% (Li, 2025).

Figure1 *Facial detection (Picture credit: Original)*

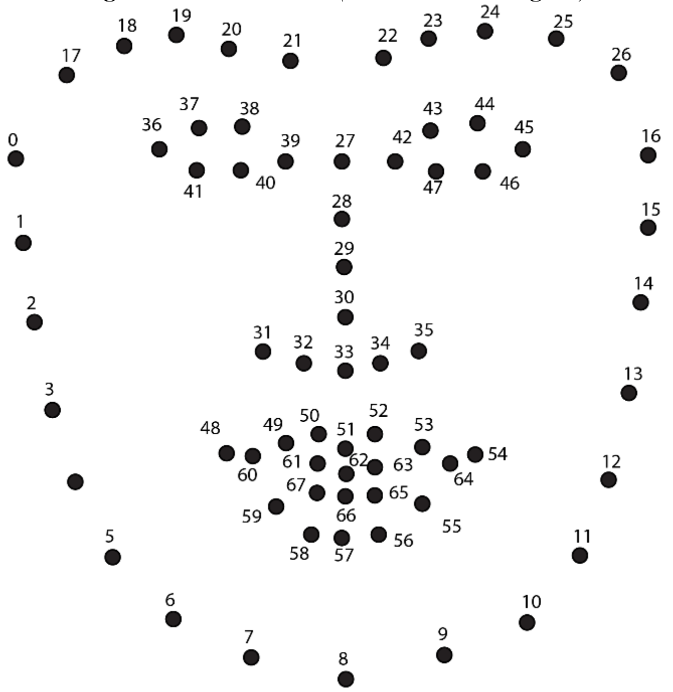
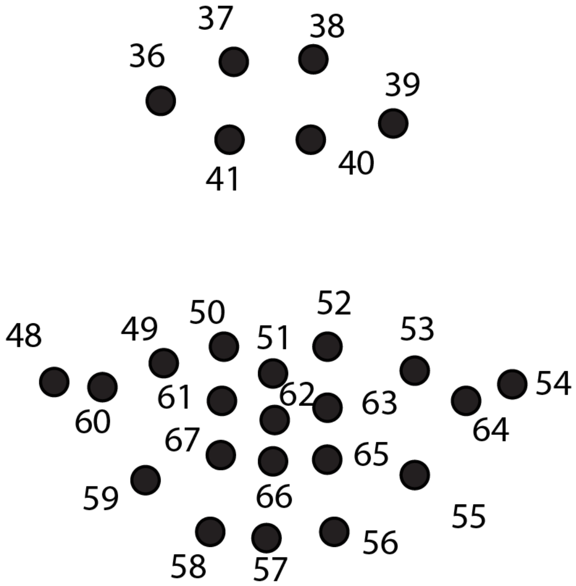


Figure2: Feature extraction about mouth and eyes (Picture credit: Original)



1.2 Based on the Transformer

Zhang et al proposed a novel Multimodal Attention Network (TMU-Net) based on the Transformer to achieve fatigue detection. The novel network processes the electroencephalogram (EEG) and electrooculogram (EOG) signals separately by the Unimodal Feature Extraction Module. A Transformer encoder, convolutional sparse attention mechanism and causal convolution are integrated into the Unimodal Feature Extraction Module to achieve the extraction of the time-frequency features. The outputs of the unimodal features extractors will be received by the multimodal fusion module. The multimodal fusion module will integrate different modalities dynamically through cross-modal attention mechanisms and dynamic weighted gating mechanisms based on uncertainty. The author chose SEED-VIG database to do the experiment. The result shows that the TMU-Net achieved high accuracy of the fatigue detection based on SEEF-VIG database, with an average RMSE of 0.1506 ± 0.0680 and MAE of 0.1238 ± 0.0640, outperforming baseline methods such as DResNet (RMSE: 0.1569 ± 0.0735) by reducing RMSE by 4.01% and standard deviation by 7.48% (Zhang et al., 2025). The novel system has potential for road safety applications.

2. Discussion

Six methods achieve high accuracy in fatigue detection. However, there still are some challenges in multimodal fatigue detection.

2.1 Challenges

2.1.1 Deployment Problem

Multimodal data is composed of different kinds of data, processing multimodal data will consume a large amount of capacity. Most car computers in the market cannot support the detection system. This disadvantage will influence the deployment in actual scenarios. The placement of the equipment, like camera will raise the price of the car. Hence, deploying multimodal fatigue detection system into commercialization still needs technological improvements.

2.1.2 Equipment Problem

The extraction of multimodal data faces some challenges. Drivers should wear relevant equipment to collect data. Wearing the equipment before starting the car will waste a lot of time and the experience will decline due to the wearing of the equipment. The bad experience may cause sale decline which is unfavorable for the system commercialization. The convenience of the equipment should be improved.

2.1.3 Modal Problem

Most multimodal fatigue detection systems mainly use PERCLO standard and facial features extractions to detect fatigue. These methods should use cameras to collect data, which may cause privacy leakage problems. Although there are some articles that improve algorithms, the interferences from complex driving environments and individual differences still exist. Hence choosing and improving modes is still a challenge to overcome in the future.

2.1.4 Evaluation Problem

There still do not have unified standards to evaluate the quality of the multimodal fatigue detection model. The absence of evaluation criteria is not conducive to the promotion and popularization of multimodal fatigue detection. And the robustness of the model still needs to be overcome to make the model adaptable to different individuals.

2.2 Future Prospectives

2.2.1 Future about Lightweight and Algorithm

As technology improves, suitable car computers will appear in the market and are able to support the system. The lightweight of the system should be considered. Ge et al proposed lightweight network based on YOLOv8 algorithm, which can reduce the capacity of the system. Yin proposed a new algorithm named YOLOv10-GMF that can finish the detection in 19ms (Yin et al., 2025). YOLO algorithm developed fast in recent years; it is worth developing YOLO algorithm to achieve the fast detection and low cost of computer capacity.

2.2.2 Future about Convenient Ways

The improvement of the physiological signal’s extraction equipment has a potential future. Compared with the modals based on cameras, physiological signals are not affected by the environment and are able to detect fatigue in advance. However, the deployment of the signal’s extraction equipment is the biggest pain point. To solve this problem, the improvement of equipment or changing location of data collection can be one of the solutions. Some authors like Wang et al designed a semi-dry electrode that can automatically replenish conductive fluid which can improve the convenience and comfort of the equipment (Wang et al., 2025). The new signal’s extraction equipment is developing rapidly, such as Ear EEG signals and electromyographic signals. This new equipment can be worn comfortably and quickly. In the future, developing Ear EEG signals or the electromyographic signals as new models can push the multimodal fatigue detection system commercialized.

2.2.3 Future Prospects about Model Choosing

The standard of fatigue judgement is the most important in the system. Some standards like PERCLO, head position have been proven to be effective. These standards require cameras to collect data, which may increase the price of the data. As more information is collected from multimodal data, the privacy leakage problem will be more serious. Inventing new standard of fatigue or developing new models becomes a new way to solve the problem. Some researchers like Xiang et al proposed a new model that is based on the drivers’ wrist movement to detect fatigue (Xiang et al., 2025). This new model does not need camera and is suitable for the driving environment which can protect drivers’ privacy. In the future, as different models appear, it is worth combining each model to find efficient modal combinations. The data breaches will be prevented due to privacy protection and multimodal system improvement.

2.2.4 Future Prospects about Evaluation

Establishing an effective standard to evaluate multimodal fatigue detection is an urgent matter to do, which is helpful for some companies to occupy a dominant position in this field. The robustness of the models is also a significant factor to overcome. It is worth improving the methods of fatigue detection and finding better modal to improve robust. When the work of the evaluation is well finished and the robustness is improved, the multimodal fatigue detection can be promoted and reduce car accidents caused by fatigued driving.

3. Conclusion

This article provides a comprehensive review about multimodal fatigue detection. The improvement in algorithms like MMA-Net, YOLOv8 networks and MTCNN enhance the processing capacity and make the system lightweight. The use of new multimodal data like visible light and infrared light dual-models and the combination of physiological signals improves the accuracy of the fatigue detection result. Through careful analysis, some problems still exist in the fatigue system. Most of the car computers cannot support the system and the wearing of the equipment is time-consuming. The use of cameras and multimodal data will cause privacy leakage. These problems can be solved through technological improvements, the use of new models and the new fatigue judgement standards.

References

Chang, W., Nie, W., Lv, R., Zheng, L., Lu, J. and Yan, G., (2024). Fatigue driving state detection based on spatial characteristics of EEG signals. *Electronics,* vol. 13, p. 3742.

Cheng, W. X., Yan, G. H., Chang, W. W., Wu, B. J. and Huang, Y. N., (2025). Channel-weighted multimodal feature fusion for EEG-based fatigue driving detection *Journal of Zhejiang University(Engineering Science),* vol. 59**,** no. 9, pp. 1775-1783+1802.

Ge, W., Ma, M. and Qu, L., (2025). Research on lightweight bimodal fatigue driving detection algorithm. *Computer Simulation,* vol. 42**,** no. 6, pp. 551-556.

Guo, Y., Yang, K. and Wu, Y., (2025). A Multi-Modality Attention Network for Driver Fatigue Detection Based on Frontal EEG, EDA and PPG Signals. *IEEE Journal of Biomedical and Health Informatics,* vol. 29**,** no. 6, pp. 4009-4022.

Huang, Z. C., Zhao, H. M., Chen, Y. X. and Tang, Y. S., (2021). Driver fatigue and dangerous driving detection system based on facial recognition. *Mechanical & Electrical Engineering Technology,* vol. 50**,** no. 12, pp. 143-146.

Li, W., (2025). A high-precision fatigue driving monitoring system based on the improved MTCNN algorithm and PFLD algorithm. *Intelligent Transportation Systems,* vol. 9**,** no. 2, pp. 1-13.

Wang, F., Guo, Q. and Fu, R., (2025). A new semi-dry electrode for driving fatigue detection on urban roads. *Sensors and Actuators A: Physical,* vol. 393, p. 116794.

Xiang, H., Sun, L., Wang, Q., Liao, Z., Xie, J., Zhao, W., Qiu, J. and Zhao, H., (2025). Research on driver's wrist motion pattern and fatigue characterization methods. *International Journal of Occupational Safety and Ergonomics*, pp. 1-15.

Yan, K., Zhao, C., Shen, C., Wang, P. and Wang, G., (2022). Published. Driver Status Monitoring System with Feedback from Fatigue Detection and Lane Line Detection. *2022 IEEE 15th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, Penang, Malaysia. IEEE, pp. 167-173.

Yin, X. P., Zhao, X. Q., Wang, X. F., Ruan, Q. and Zhang, W., (2025). Improved YOLOv10 Algorithm for Driver Fatigue Detection *Journal of Electronic Measurement and Instrumentation*, pp. 1-13.

Yu, L., Yang, X., Wei, H., Liu, J. and Li, B., (2024). Driver fatigue detection using PPG signal, facial features, head postures with an LSTM model. *Heliyon,* vol. 10**,** no. 21, p. e39479.

Zhang, Y., Xu, X., Du, Y. and Zhang, N., (2025). TMU-Net: A Transformer-Based Multimodal Framework with Uncertainty Quantification for Driver Fatigue Detection. *Sensors,* vol. 25**,** no. 17, p. 5364.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.