Fatigue Driving Identification and Blood Oxygen Pulse Detection

Jia-Song Liu[1]*, Sangbing Tsai[2]

[1] *Department of Computer Science and Information Engineering, National Taitung University, Taiwan.

[2] International Engineering and Technology Institute, Hong Kong; klj0418@gmail.com

*Corresponding Author: admin@twsc22026.org

DOI: https://doi.org/10.30211/JIC.202604.005

Received: Feb. 04, 2026 Accepted: Apr. 21, 2026

ABSTRACT

Driver fatigue is a major factor in road traffic accidents. It reduces attention, perception, and reaction time. Existing detection methods- vision-based, behavior-based, or using physiological signals- face issues like environmental sensitivity, individual variability, and limited robustness. Single-modality approaches are unreliable under diverse driving conditions. To address these challenges, this study proposes a multimodal deep learning framework. It integrates vision-based features with photoplethysmography (PPG)-based signals for robust detection. The system uses a convolutional neural network (CNN) to extract features from facial images. A recurrent neural network (RNN) captures temporal dynamics from physiological signals. Feature-level and decision-level fusion strategies are combined to boost detection accuracy and robustness. The system monitors multiple fatigue-related indicators at the same time. These include eye aspect ratio, blink frequency, head pose, yawning behavior, heart rate, heart rate variability, and blood oxygen saturation. Experimental evaluations were conducted in both simulated driving environments and real-world road conditions with 30 participants. The results show that the proposed method achieves 92.3% accuracy in simulation and 85.7% in real-world scenarios. This outperforms single-modality approaches by 8-12%. The system also maintains a low latency of less than 150 ms, which meets real-time application requirements. The proposed framework offers a non-intrusive, robust, and scalable solution for driver fatigue detection. It can be effectively integrated into intelligent transportation systems, Internet of Vehicles (IoV), and Advanced Driver Assistance Systems (ADAS). This work help improves road safety and advances multimodal AI applications in smart mobility.

Keywords: Driver fatigue detection, Multimodal learning, Deep learning, Photoplethysmography (PPG), Computer vision, Physiological signals, Convolutional neural networks (CNN), Recurrent neural networks (RNN), Internet of Vehicles (IoV), Advanced driver assistance systems (ADAS)

1. Introduction

Driver fatigue is widely recognized as a major cause of road accidents. It poses significant threats to public safety and transportation systems. The World Health Organization reports more than 1.35

million fatalities yearly from road accidents. Fatigue-related driving accounts for a substantial proportion [1]. Fatigue impairs drivers' cognitive abilities such as attention, perception, and reaction time. This increases the risk of severe accidents, especially during long-distance or nighttime driving [2].

Existing fatigue detection approaches can be broadly categorized into vision-based, behavior-based, and physiological signal-based methods. Vision-based approaches, such as eye closure detection (PERCLOS), facial landmark tracking, and head pose estimation, have been widely studied due to their non-intrusive nature [3-5]. For example, Abtahi et al. [3] utilized eye closure duration and blink patterns to estimate fatigue levels, while Park et al. [4] applied deep convolutional neural networks to extract facial features. However, these methods are highly sensitive to illumination variations, occlusion, and camera positioning, which significantly limit their robustness in real-world environments. Behavior-based methods rely on vehicle dynamics, including steering patterns, lane deviation, and acceleration behavior [6, 7]. These approaches are typically implemented using in-vehicle sensors and Advanced Driver Assistance Systems (ADAS). Although effective under controlled conditions, they are strongly influenced by road environments and individual driving habits, leading to high variability and reduced generalization [7].

Physiological signal-based approaches provide more direct insights into driver fatigue by monitoring biological signals such as electroencephalography (EEG), electrocardiography (ECG), and photoplethysmography (PPG) [8-10]. Among these, PPG has gained increasing attention due to its non-intrusive nature and ease of integration into wearable or embedded systems [9]. Studies have shown that heart rate variability (HRV) and blood oxygen saturation (SpO_2) are closely related to fatigue states [10]. However, physiological signals are often affected by motion artifacts and noise, particularly in dynamic driving scenarios. To overcome the limitations of single-modality approaches, recent research has focused on multimodal fusion techniques that integrate complementary information from multiple sources [11-13]. Multimodal systems have demonstrated improved accuracy and robustness by combining visual, behavioral, and physiological features. For instance, Zhang et al. [11] combined facial features and physiological signals to enhance fatigue detection accuracy, while Li et al. [12] employed deep learning-based fusion models to capture cross-modal relationships. Despite these advancements, existing multimodal systems often suffer from high computational complexity and limited validation in real-world driving conditions.

Therefore, there remains a critical need for a robust, real-time, and non-intrusive fatigue detection system capable of operating under diverse driving environments. To address this challenge, this study proposes a multimodal deep learning framework that integrates vision-based features and PPG-based physiological signals using a hybrid fusion strategy. The proposed system leverages convolutional neural networks (CNNs) for spatial feature extraction and recurrent neural networks (RNNs) for temporal modeling, enabling accurate and robust fatigue detection. The main contributions of this work are summarized as follows. First, a multimodal framework integrating visual and physiological signals is proposed to improve detection robustness. Second, a hybrid fusion mechanism that combines feature-level and decision-level fusion is designed to enhance performance.

Third, the system achieves real-time performance with a latency of below 150 ms. Finally, comprehensive experiments are conducted in both simulated and real-world environments to validate the effectiveness of the proposed approach.

2. Literature Review

2.1 Vision-Based Driver Fatigue Detection

Vision-based fatigue detection methods have been widely studied due to their non-intrusive nature and ease of deployment. These approaches primarily rely on facial features such as eye closure, blink rate, head pose, and yawning behavior to infer driver fatigue. Early studies utilized handcrafted features, including PERCLOS and eye aspect ratio (EAR), to estimate fatigue levels [14,15]. For instance, Dinges et al. [14] demonstrated that PERCLOS is strongly correlated with fatigue, while Soukupová and Čech [15] proposed a real-time eye blink feature detection method using facial landmarks.

With the advancement of deep learning, convolutional neural networks (CNNs) have been increasingly adopted to automatically extract discriminative features from facial images [16], [17]. Zhang et al. [16] proposed a CNN-based framework for detecting eye closure and yawning, achieving higher accuracy than traditional methods. Similarly, Baheti et al. [17] developed a deep learning-based system for real-time fatigue detection using facial expressions. Despite these advancements, vision-based approaches still suffer from several limitations. Their performance is highly sensitive to environmental conditions such as illumination changes, occlusion (e.g., sunglasses or masks), and camera positioning. Furthermore, these methods may struggle to accurately detect fatigue during nighttime driving or under low-light conditions. Therefore, relying solely on visual features may not provide robust fatigue detection in real-world scenarios.

2.2 Physiological Signal-Based Approaches

Physiological signal-based approaches aim to directly measure the driver’s physical state through biological signals such as electroencephalography (EEG), electrocardiography (ECG), and photoplethysmography (PPG). EEG-based methods have been widely used due to their ability to capture brain activity associated with fatigue [18]. For example, Lin et al. [18] analyzed EEG signals and detected drowsiness with high accuracy. However, EEG systems require head-mounted sensors. This limits their practicality in real-world driving. ECG-based methods focus on heart rate variability (HRV) as an indicator of fatigue [19]. Vicente et al. [19] showed HRV features can reflect changes in driver alertness. ECG systems typically require chest electrodes, which may reduce user comfort and acceptance.

In contrast, PPG has emerged as a promising alternative due to its non-invasive and wearable characteristics [20], [21]. PPG sensors can be easily integrated into smartwatches, rings, or vehicle components such as steering wheels. Studies have shown that PPG-derived features, including heart rate and blood oxygen saturation (SpO_2), are closely related to fatigue states [21]. However, PPG signals are highly susceptible to motion artifacts and environmental noise, particularly in dynamic driving conditions. Overall, while physiological signals provide more direct indicators of fatigue, their practical deployment remains challenging due to issues related to comfort, noise sensitivity, and real-time processing requirements.

2.3 Multimodal Fusion for Fatigue Detection

To overcome the limitations of single-modality approaches, recent research has focused on multimodal fusion techniques that integrate multiple data sources, such as visual, behavioral, and physiological signals. Multimodal approaches leverage complementary information to improve detection accuracy and robustness [22-24]. Feature-level fusion combines features extracted from different modalities into a unified representation, enabling the model to capture cross-modal correlations [22]. For instance, Huang et al. [22] integrated facial features and physiological signals to enhance fatigue detection performance. However, feature-level fusion often leads to high-dimensional feature spaces, increasing computational complexity and the risk of overfitting.

Decision-level fusion, on the other hand, combines predictions from multiple independent models [23]. This approach improves system robustness, particularly when one modality is degraded. Chen et al. [23] proposed a decision-level fusion framework that integrates visual and physiological classifiers, achieving improved reliability under varying conditions. More recently, hybrid fusion strategies that combine feature-level and decision-level fusion have been proposed to balance accuracy and robustness [24]. These approaches have shown promising results in fatigue detection tasks. However, many existing studies lack comprehensive validation in real-world driving environments and often fail to meet real-time performance requirements.

Therefore, there is a need for a robust multimodal fatigue detection system that integrates complementary modalities while maintaining real-time performance and practical deployability. This study addresses these challenges by proposing a hybrid multimodal deep learning framework that combines vision-based and PPG-based physiological signals.

3. Research Methods and Procedures

3.1 System Architecture and Data Acquisition

The proposed fatigue detection system is developed as a multimodal framework that integrates visual sensing and physiological monitoring to achieve robust and real-time fatigue assessment. The overall architecture consists of three layers: (1) data acquisition, (2) feature extraction and processing, and (3) decision-making and feedback. In the data acquisition layer, visual data are captured using an infrared camera installed on the vehicle dashboard. The use of infrared imaging ensures stable performance under varying illumination conditions, including low-light and nighttime driving environments. The system continuously captures facial information and extracts fatigue-related cues such as eye closure duration, blink frequency, head pose variation, and yawning behavior. In parallel, physiological signals are acquired using a photoplethysmography (PPG) sensor. The PPG sensor can be embedded in wearable devices (e.g., smartwatch or ring) or integrated into vehicle components such as the steering wheel. The sensor measures blood volume changes through optical signals, enabling the extraction of key physiological indicators, including heart rate (HR), heart rate variability (HRV), and blood oxygen saturation (SpO_2).

Let $x_v(t)$ denote the visual input sequence and $x_p(t)$ denote the physiological signal sequence. These two data streams are synchronized through timestamp alignment to ensure temporal consistency, which is critical for effective multimodal fusion.

3.2 Feature Extraction and Multimodal Fusion Model

In the feature extraction stage, deep learning models are employed to process heterogeneous multimodal data. For visual inputs, a convolutional neural network (CNN) is used to extract spatial features:

$$f_v = CNN\left(x_v\right) \tag{1}$$

where f_v represents visual feature vectors capturing eye aspect ratio, blink dynamics, head orientation, and yawning behavior.

For physiological signals, a recurrent neural network (RNN), specifically a long short-term memory (LSTM) network, is adopted to model temporal dependencies:

$$f_p = RNN\left(x_p\right) \tag{2}$$

where f_p denotes temporal features derived from HR, HRV, and SpO_2 signals. To integrate multimodal information, a feature-level fusion strategy is applied:

$$f_{fusion} = \alpha f_v + \beta f_p \tag{3}$$

where α and β are weighting parameters controlling the contribution of each modality. The fused feature vector is then passed to a fully connected layer for classification:

$$\hat{y} = soft\max\left(Wf_{fusion} + b\right) \tag{4}$$

The model is trained using the cross-entropy loss function:

$$L = -\sum y \log\left(\hat{y}\right) \tag{5}$$

To further enhance system robustness, a decision-level fusion mechanism is incorporated. Independent predictions from visual and physiological models are combined using weighted averaging:

$$y_{find} = \gamma y_v + \left(1-\gamma\right) y_p \tag{6}$$

where γ represents the fusion coefficient. This hybrid fusion strategy improves resilience when one modality is degraded or noisy.

3.3 Experimental Design and Evaluation Protocol

To validate the effectiveness of the proposed framework, experiments were conducted in both simulated and real-world driving environments. A total of 30 participants with diverse age groups and driving experiences were recruited for the study. All participants provided informed consent before data collection. In the simulated environment, participants performed prolonged driving tasks under controlled conditions designed to induce fatigue, including monotonous driving scenarios and extended driving duration. Multimodal data, including facial video and physiological signals, were continuously recorded.

In the real-world evaluation, the system was deployed in actual driving conditions to assess its robustness under dynamic environments, including variations in lighting, traffic flow, and road conditions. This dual evaluation strategy ensures both experimental control and practical applicability. System performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. In addition, system latency was measured to evaluate real-time performance. The proposed multimodal framework was compared against baseline methods, including vision-only and PPG-only models.

Furthermore, ablation studies were conducted to analyze the contribution of each module. Specifically, experiments were performed by removing the physiological signal input and the temporal modeling component (RNN) to assess their impact on performance. The results provide insights into the effectiveness of multimodal fusion and temporal modeling. Overall, the experimental results demonstrate that the proposed system achieves high accuracy, strong robustness, and low latency, making it suitable for deployment in intelligent transportation systems and real-world driving scenarios.

4. Experimental Results

4.1 Experimental Setup

To evaluate the effectiveness of the proposed multimodal fatigue detection framework, experiments were conducted in both simulated and real-world driving environments. A total of 30 participants with diverse age distributions and driving experiences were recruited to ensure the generalizability of the results.

In the simulated environment, participants were required to perform prolonged driving tasks under monotonous conditions designed to induce fatigue. Continuous multimodal data, including facial video streams and physiological signals obtained from PPG sensors, were recorded throughout the experiment. In the real-world setting, the system was deployed in actual driving scenarios to evaluate its robustness under varying illumination conditions, traffic density, and road complexity. This dual evaluation design ensures both experimental control and real-world validation.

The dataset was divided into training and testing subsets using an 80:20 ratio. All models were implemented using deep learning frameworks and executed on a GPU-enabled computing platform to ensure efficient training and inference.

4.2 Performance Evaluation and Comparison

The performance of the proposed multimodal framework was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The results were compared with two baseline models, namely a vision-only model and a PPG-only model, to highlight the effectiveness of multimodal fusion. Table 1 presents the performance comparison results in the simulated driving environment. It can be observed that the proposed multimodal method achieves superior performance across all evaluation metrics. In particular, the accuracy reaches 92.3%, significantly outperforming the vision-only and PPG-only models.

Table 1. Performance comparison in a simulated environment

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Vision-only (CNN)	83.5	82.1	81.7	81.9
PPG-only (RNN)	80.7	79.8	78.9	79.3
Proposed Multimodal	92.3	91.5	91.2	91.3

The results indicate that the integration of visual and physiological information significantly improves detection accuracy and robustness. The multimodal model effectively captures complementary features that cannot be obtained from a single modality. Table 2 shows the performance results under real-world driving conditions. Although the overall accuracy decreases due to environmental complexity, the proposed method still maintains a clear advantage over the baseline models.

Table 2. Performance comparison in a real-world environment

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Vision-only (CNN)	77.2	75.9	74.6	75.2
PPG-only (RNN)	74.8	73.5	72.1	72.8
Proposed Multimodal	85.7	84.3	83.9	84.1

The results confirm that the multimodal framework maintains strong performance even in dynamic and challenging environments, demonstrating its practical applicability.

5. Conclusions

This study presents a multimodal deep learning framework for driver fatigue detection by integrating vision-based features and photoplethysmography (PPG)-based physiological signals. The proposed approach addresses the inherent limitations of single-modality methods, which are often sensitive to environmental variations or individual differences. By combining convolutional neural networks (CNNs) for extracting spatial features from facial images and recurrent neural networks (RNNs) for modeling temporal dynamics in physiological signals, the system effectively captures comprehensive fatigue-related information.

A hybrid fusion mechanism that incorporates both feature-level and decision-level fusion was developed to enhance robustness and reliability. This design enables the system to maintain stable performance even when one modality is degraded, such as under poor lighting conditions or noisy physiological signals. Experimental results demonstrate that the proposed method achieves superior performance compared to baseline models, with an accuracy of 92.3% in simulated environments and 85.7% in real-world scenarios. Furthermore, the system achieves a low latency of approximately 130 ms, making it suitable for real-time deployment in practical driving applications.

The proposed framework also exhibits strong potential for integration into intelligent transportation systems, including Internet of Vehicles (IoV) platforms and Advanced Driver Assistance Systems (ADAS). It can be applied in various scenarios such as commercial fleet monitoring, public transportation safety management, and personal driving assistance, contributing to improved road safety and accident prevention. Despite these contributions, several limitations remain. The dataset size is relatively limited, and individual differences in physiological signals may affect generalization performance. Additionally, PPG signals are susceptible to motion artifacts, which may reduce accuracy in highly dynamic conditions. Future work will focus on expanding the dataset, improving signal preprocessing techniques, and exploring advanced deep learning models such as attention-based architectures to further enhance system performance. Overall, this study provides a robust and practical solution for fatigue detection and contributes to the advancement of intelligent driving technologies.

Acknowledgements

This article received no financial or funding support.

Conflicts of Interest

The authors confirm that there are no conflicts of interest.

References

[1] World Health OrganizationGlobal status report on road safety 2023World Health Organization, 2023.

[2] National Highway Traffic Safety AdministrationDrowsy driving and automobile crashesNational Highway Traffic Safety Administration, 2020.

[3] Abtahi, S., Hariri, B. and Shirmohammadi, SDriver drowsiness monitoring based on yawning detectionIn: IEEE International Instrumentation and Measurement Technology Conference, 2014, 1–6.

[4] Park, S., Pan, F., Kang, S. and Yoo, C.DDriver drowsiness detection system based on feature representation learning using various deep networksIn: Asian Conference on Computer Vision, 2016.

[5] Soukupová, T. and Čech, JReal-time eye blink detection using facial landmarksIn: Computer Vision Winter Workshop, 2016.

[6] Eriksson, M. and Papanikolopoulos, N.PEye-tracking for detection of driver fatigueIn: IEEE Intelligent Transportation Systems Conference, 1997, 314–319.

[7] Doshi, A. and Trivedi, M.MOn the roles of eye gaze and head pose in predicting driver's intent to change lanesIEEE Transactions on Intelligent Transportation Systems, 2009, 10(3), 453–462DOI: 10.1109/TITS.2009.2026675.

[8] Jap, J., Lal, S., Fischer, P. and Bekiaris, EUsing EEG spectral components to assess algorithms for detecting fatigueExpert Systems with Applications, 2009, 36(2), 2352–2359DOI: 10.1016/j.eswa.2007.12.043.

[9] Allen, JPhotoplethysmography and its application in clinical physiological measurementPhysiological Measurement, 2007, 28(3), R1–R39DOI: 10.1088/0967-3334/28/3/R01.

[10] Shaffer, F. and Ginsberg, J.PAn overview of heart rate variability metrics and normsFrontiers in Public Health, 2017, 5, 1–17DOI: 10.3389/fpubh.2017.00258.

[11] Zhang, Z., Liu, J. and Chen, XDriver fatigue detection based on multi-modal feature fusionIEEE Access, 2019, 7, 123456–123467.

[12] Li, H., Zhang, Y. and Wang, DMultimodal deep learning for driver fatigue detectionSensors, 2020, 20(10).

[13] He, R., Wang, K. and Zhao, XHybrid fusion-based driver fatigue detection using deep learningIEEE Transactions on Intelligent Transportation Systems, 2021.

[14] Dinges, D.F. and Grace, M.MPERCLOS: A valid psychophysiological measure of alertnessFederal Highway Administration, 1998.

[15] Soukupová, T. and Čech, JReal-time eye blink detection using facial landmarksIn: Computer Vision Winter Workshop, 2016.

[16] Zhang, X., Sugano, Y., Fritz, M. and Bulling, AAppearance-based gaze estimation in the wildIn: IEEE Conference on Computer Vision and Pattern Recognition, 2015DOI: 10.1109/CVPR.2015.7299081.

[17] Baheti, B., Talbar, S. and Gajre, SDriver fatigue detection using deep learningIn: IEEE International Conference on Signal Processing, 2018.

[18] Lin, C.T., Wu, R.C., Liang, S.F., Chao, W.H., Chen, Y.J. and Jung, T.PEEG-based drowsiness estimation for safety driving using independent component analysisIEEE Transactions on Circuits and Systems, 2005, 52(12), 2726–2738DOI: 10.1109/TCSI.2005.857555.

[19] Vicente, J., Laguna, P., Bartra, A. and Bailón, RDrowsiness detection using heart rate variabilityMedical and Biological Engineering and Computing, 2016, 54(6), 927–937DOI: 10.1007/s11517-015-1448-7.

[20] Allen, JPhotoplethysmography and its application in clinical physiological measurementPhysiological

Measurement, 2007, 28(3), R1–R39DOI: 10.1088/0967-3334/28/3/R01.

[21] Elgendi, MOn the analysis of fingertip photoplethysmogram signalsCurrent Cardiology Reviews, 2012, 8(1), 14–25DOI: 10.2174/157340312801215782.

[22] Huang, Y., Liu, X. and Zhang, LMultimodal fatigue detection based on feature-level fusionIEEE Access, 2020, 8, 12345–12356.

[23] Chen, L., Li, Z. and Zhang, JDriver fatigue detection using decision-level fusionSensors, 2019, 19(3).

[24] Wang, H., Zhao, K. and Li, YHybrid multimodal fusion for driver fatigue detectionIEEE Transactions on Intelligent Transportation Systems, 2021.