EEG-based Driver Fatigue Detection with Eye-tracking Guided Weak Supervision

Fenglin Xu[1], Peng Yu[2,*], Hanying Guo[1]
[1]School of Automobile and Transportation, Xi Hua University, Chengdu 610039, China
[2]Sichuan Large Item Transportation Co., Ltd, Chengdu 610066, China
*Corresponding email: yupeng@scdj-trans.com

Abstract

Driver fatigue is a major cause of traffic accidents. Electroencephalography (EEG), which directly reflects neural activity, has been widely used for fatigue detection. However, existing methods still suffer from limited interpretability of features, high subjectivity in fatigue labeling, and insufficient capability in modeling temporal dependencies. To address these issues, this study proposes a fatigue detection method based on EEG band-ratio features and eye-movement-derived weak supervision labels. A deep learning model, namely the stacked long short-term memory attention network (stacked LSTM attention network, SLAN), is designed to capture temporal dependencies in EEG signals. After preprocessing, five EEG band-ratio features are extracted, and a dual-threshold pupil diameter strategy is used to construct three-level fatigue labels for weakly supervised learning. An attention mechanism is further introduced to enhance temporal feature modeling and improve interpretability by focusing on critical time segments. Experiments on a self-collected dataset of 20 subjects demonstrate that the proposed model achieves an accuracy of approximately 97.0% in the three-class fatigue classification task, outperforming EEG-based spatio-temporal convolutional neural network (ESTCNN), EEGNet, and Interpretable convolutional neural network (CNN) across all evaluation metrics. The proposed approach effectively improves temporal representation and interpretability of fatigue states and reveals dynamic temporal patterns of EEG activity under fatigue conditions, providing a foundation for continuous driver state monitoring and adaptive warning systems in intelligent cockpit applications.

Keywords

Intelligent transportation, Fatigue detection, Two-layer long short-term memory network, Multi-band features

Introduction

Driver fatigue is one of the major contributing factors to road traffic accidents [1]. Accurate and real-time monitoring of driver fatigue has become a key focus in the field of intelligent transportation systems [2]. At present, fatigue detection methods can be broadly classified into three categories: image-based behavioral monitoring, external behavior analysis based on vehicle operation parameters, and internal state recognition based on physiological signals [3,4].

Image-based methods rely on features such as eyelid movement, facial expressions, and head posture. They offer advantages including non-intrusive acquisition, low hardware cost, and strong generalizability, but are highly susceptible to environmental conditions such as illumination changes, occlusion, and eyeglass wear [5]. Vehicle-state-based methods, such as steering wheel angle and lane deviation, can reflect abnormal driving behaviors. However, they lack direct representation of neurophysiological fatigue changes [6]. In contrast, physiological signal-based approaches can directly capture internal states of the central and autonomic nervous systems, providing stronger temporal sensitivity and physiological validity for fatigue representation.

Common physiological signals include heart rate variability (HRV), electrodermal activity (EDA), electroencephalography (EEG), and electrooculography (EOG). Fowler et al. found that HRV is strongly negatively correlated with subjective mental workload, while EDA peaks can identify moments of uncertainty or time pressure, suggesting their usefulness for continuous workload monitoring [7]. Wang et al. using EEG signals and the percentage of eyelid closure over the pupil over

time (PERCLOS) eye-movement metric, verified that PERCLOS, as a weakly supervised label, is highly correlated with fatigue level [8]. Its variation exhibits strong stability and objectivity and can be employed to drive continuous prediction of fatigue severity. Among these signals, EEG is the most widely used for fatigue detection. Traditional EEG-based methods typically rely on handcrafted features such as band energy, power spectral density, and entropy, which suffer from high dimensionality, noise sensitivity, and strong inter-subject variability. Kalogeropoulos et al. systematically reviewed EEG analysis from neurophysiological foundations to classical methods and emerging AI approaches, highlighting that future EEG analysis will likely evolve toward neuro-symbolic architectures combining generative learning with interpretable signal theory [9].

With the advancement of deep learning, models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers have been widely applied to EEG-based fatigue detection [10]. In particular, LSTM networks are effective in modeling the progressive temporal evolution of fatigue states. Yang et al. proposed a fatigue detection approach combining kernel principal component analysis (KPCA) and LSTM, where KPCA is used for nonlinear feature extraction and dimensionality reduction, and LSTM captures temporal dependencies of fatigue progression, achieving 97.1% accuracy on a self-built dataset, demonstrating its effectiveness in modeling dynamic fatigue patterns [11].

In recent years, attention mechanisms have been introduced to further improve both performance and interpretability in fatigue recognition tasks. Some studies combine attention with LSTM or CNN architectures to enhance classification accuracy while providing partial interpretability. Zhang et al. proposed a Transformer-based multimodal attention network integrating EEG and eye-tracking signals [12]. The model employs causal convolution, convolutional sparse attention, and a Transformer encoder for spatiotemporal feature extraction, and uses cross-modal attention with uncertainty weighting for fusion, achieving high-accuracy cross-subject fatigue detection on the Shanghai Jiao Tong University emotion EEG dataset-vigilance estimation (SEED-VIG) dataset.

In recent years, attention mechanisms have been introduced into fatigue recognition models, further improving model performance and interpretability. Bello et al. proposed a FAST-Net model that integrates frequency attention, spatial feature extraction (including weighted feature fusion), and temporal attention convolution [13]. Under five-fold cross-validation on the SEED-VIG dataset, the model achieved an average accuracy of 93.8%±0.5%, outperforming the baseline models.

Despite promising progress, several challenges remain in EEG-based fatigue detection. First, traditional EEG features are sensitive to noise and inter-subject variability, limiting their ability to represent fatigue-related neural activity. Second, driver fatigue exhibits strong temporal accumulation characteristics, yet existing studies are still insufficient in capturing long-term dependencies and identifying critical fatigue-related temporal segments. Third, although eye-movement indicators such as PERCLOS have been shown to reliably reflect fatigue levels, how to effectively leverage their weak supervisory signals to enhance EEG-based learning remains underexplored [14,15]. Finally, deep learning models generally suffer from limited interpretability, making it difficult to explain their decision mechanisms.

To address these issues, this study proposes a fatigue detection method that integrates EEG band-ratio features with eye-movement-based weak supervision labels. Specifically, EEG signals are decomposed into five canonical frequency bands (Delta, Theta, Alpha, Beta, and Gamma), and five physiologically meaningful band-ratio features (α/β, θ/β, θ/α, δ/θ, γ/β) are computed to emphasize relative changes among fatigue-sensitive bands, thereby improving interpretability. Subsequently, pupil diameter variation is used to construct three-level fatigue labels via a dual-threshold strategy, enabling automated weak supervision and reducing subjectivity in manual annotation. For model design, a Stacked LSTM Attention Network (SLAN) is constructed by stacking a two-layer LSTM and incorporating a customized attention mechanism to capture the most discriminative temporal segments in EEG sequences.

Pre-processing and experiments

Experimental equipment

This study collected driving experimental data from 20 (male and female, aged 20-30 years old) participants with a certain level of driving experience, including two

modalities: EEG and EOG.

To ensure data reliability, participants were instructed to avoid consuming stimulants such as strong tea, high-caffeine products, and alcohol within 24 hours prior to the experiment. The following equipment was used in this study.

(1) Tobii Pro Glasses 3

A wearable eye-tracking device developed by Tobii was used in this study. It adopts a glasses-based design and enables real-time acquisition of gaze position, fixation points, and pupil dynamics in natural environments. The device is based on infrared eye-tracking technology with a sampling rate of up to 100 Hz, providing high tracking accuracy and stability. It also supports first-person video recording and multi-sensor data synchronization. It is widely applicable to driving, cognitive workload, and fatigue detection scenarios, and can be integrated with EEG and other physiological signals (Figure 1).

Figure 1. Tobii Pro Glasses 3.

(2) The eego^{TM}mylab

The eego^{TM}mylab portable EEG system is a new-generation electroencephalography recording and analysis platform developed by ANT Neuro. The system consists of four core components: the Waveguard™ Original EEG cap, the eego amplifier, the eego recording software, and the ASA data analysis software.

(3) Experimental scenario

The experiment was conducted using Euro Truck Simulator 2 as the driving simulation platform, which provides a high-fidelity, controllable highway driving environment with realistic visual rendering, vehicle dynamics, and road infrastructure. A continuous, long-distance highway route was selected to design a round-trip driving task, ensuring sufficient driving duration to elicit natural driver fatigue progression while maintaining consistent road geometry and traffic conditions throughout the session (Figure 2).

To eliminate confounding variables and ensure experimental controllability and procedural stability, all stochastic environmental events were explicitly disabled during the simulation. These included dynamic weather changes (such as rain, fog, or snow), random traffic congestion, variable speed limits, and unexpected traffic incidents, so that the only major environmental variation during the experiment was the natural progression of the day-night cycle.

To systematically investigate the dynamic effects of circadian variation on driver fatigue, the simulation was configured to follow a complete 24-hour day-night cycle. This setting allowed the driving environment to transition continuously from bright daytime, through dusk with gradual dimming ambient light, to nighttime with low illumination and reduced visibility, and back to daytime conditions over the course of the experiment. As illustrated in Figure 2, the three typical scenarios captured during the experiment include: (a) daytime, characterized by high ambient brightness and clear visibility, (b) dusk, with intermediate light intensity and soft, diffused lighting conditions, (c) night, with significantly reduced ambient light and reliance on vehicle headlights for forward visibility. This controlled variation in lighting conditions enabled the study to examine how changes in visual environment, as driven by the circadian cycle, interact with and influence the development of driver fatigue.

(a) daytime

(b) dusk

(c) night

Figure 2. Typical scenario: (a) daytime; (b) dusk; (c) night.

Experimental setup

The simulated driving experiment was conducted in a low-light, low-noise, and well-ventilated environment. To induce fatigue, each participant was required to complete approximately 120 minutes of continuous simulated driving. The experiment was scheduled between 13:00 and 14:30, a post-lunch period during which monotonous driving tasks are more likely to induce fatigue.

The experimental procedure consisted of three stages: preliminary preparation, driving task execution, and data recording. During the preparation stage, participants were familiarized with the driving simulator, including steering wheel operation, brake sensitivity, and key functions on the steering wheel. They were also introduced to the experimental route and procedure to ensure smooth execution and avoid operational errors. The driving stage involved a 120-minute monotonous simulated driving task after participants had mastered the system operations. During the data recording stage, EEG and eye-tracking signals were continuously collected. For each participant, EEG feature sequences and corresponding fatigue label sequences were constructed for subsequent deep learning model training and validation.

Data pre-processing

Due to the non-stationary nature and susceptibility to artifacts of EEG signals, preprocessing was performed using EEGLAB in MATLAB to minimize or remove noise and artifacts. The preprocessing pipeline included band-pass filtering, down-sampling, re-referencing, and artifact removal. The processed data was then prepared for subsequent frequency-band feature analysis.

Dataset

This study employed a self-collected fatigue dataset consisting of 20 valid participants. All data were strictly time-aligned in the temporal domain.

For EEG signals, after artifact removal, re-referencing, and filtering, five band-ratio features were extracted, including α/β, θ/β, θ/α, δ/θ, and γ/β.

For eye-tracking data, the PERCLOS metric was used as a fatigue reference indicator. Based on predefined thresholds, the outputs were categorized into three fatigue states: alert, mild fatigue, and severe fatigue.

Method

This study proposes a temporal deep learning framework based on EEG band-ratio features and automatically generated eye-movement threshold labels. The method consists of three key stages (Figure 3).

(1) EEG feature extraction: The preprocessed EEG signals are decomposed into canonical frequency bands, and physiologically meaningful band-ratio features are extracted to enhance representational capacity and interpretability.

(2) PERCLOS weak labeling: The PERCLOS metric is derived from eye-tracking data, and a dual-threshold strategy is applied to construct three-class fatigue labels, enabling automatic weak supervision and reducing manual annotation effort.

(3) Temporal modeling: Feature sequences are generated using a sliding window strategy and fed into a stacked two-layer LSTM network. An attention mechanism is further introduced to emphasize key temporal segments, improving both classification performance and interpretability.

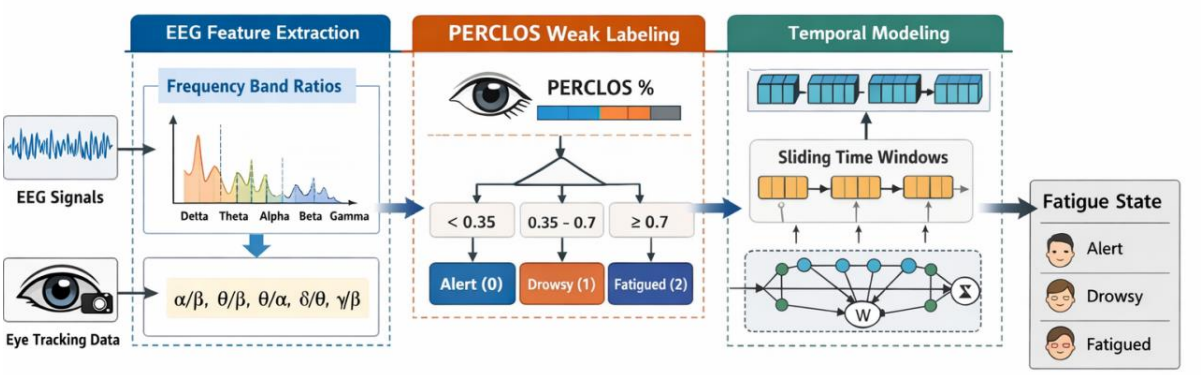

Figure 3. Model processing chart.

Feature extraction

After preprocessing, the cleaned EEG signals were band-pass filtered into five canonical frequency bands: Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-13 Hz), Beta (13-30 Hz), and Gamma (30-45 Hz). Each band was stored as an independent EEG dataset for subsequent feature extraction.

To enhance the physiological relevance of EEG features, five fatigue-sensitive band-ratio features were computed from these frequency bands, including α/β, θ/β, θ/α, δ/θ and γ/β.

The feature extraction procedure consisted of the following steps:

(1) Segmentation: Continuous EEG signals were divided into 900 equal-length segments, with one feature value computed per segment.

(2) Power spectral density estimation: For each segment, the power spectral density was calculated using the fast Fourier transform (FFT).

Given a discrete EEG segment with sampling frequency f_s, and length N, the power spectral density (PSD) describes the distribution of signal energy across different frequency components. Its general form is given by:

$$PSD\ （k）= \frac{1}{Nf_s}|X[k]|^2 \quad (1)$$

where $|X[k]|^2$ denotes the signal energy at frequency bin k; and Nf_s is the normalization factor used to obtain the PSD in units of V^2/Hz.

Following PSD estimation, the average spectral power of each frequency band was extracted. Subsequently, fatigue-related band-ratio features were calculated to characterize the relative variations among different frequency bands, for example:

$$\text{Ratio}\theta/\beta = \frac{P_\theta}{P_\beta} \quad (2)$$

These band-ratio features effectively characterize the relative variations among different EEG rhythms and reflect the evolution of neural activity from alertness to fatigue.

Finally, multi-channel feature fusion was performed. Since each EEG segment contained 32 channels, the five band-ratio features extracted from all channels were concatenated to form a 900×160 feature matrix (32×5=160 features per segment). These features capture neural activity changes across different fatigue levels and serve as the primary input to the proposed SLAN.

Construction of eye-movement fatigue labels

Eye-movement fatigue labels were constructed using the percentage of eye closure (PERCLOS) as a vigilance marker to assess the state of fatigue for each participant. The PERCLOS value is calculated as follows:

$$PERCLOS = \frac{blink+clos}{fixation+saccade+blink+clos} \quad (3)$$

where: $blink$ = cumulative duration of short blinks (normal blinks); $clos$ = cumulative duration of eye closure $fixation$ = cumulative duration of gaze on a fixed point; $saccade$ = cumulative duration of rapid eye movements.

The PERCLOS value ranges from 0 to 1.00; a higher value indicates a greater fatigue level. Based on PERCLOS, a dual-threshold strategy is used to generate three-level fatigue labels, which are then aligned with EEG features for model training:

PERCLOS <0.35: alert (label=0);

0.35≤ PERCLOS <0.70: mild fatigue (label=1);

PERCLOS ≥0.70: moderate-severe fatigue (label=2);

Pupil data were time-aligned with EEG signals and cropped to the same length for automatic label generation.

Model design

To capture the dynamic characteristics of fatigue over time, this study constructs a temporal modeling framework based on sliding windows and SLAN and further introduces an attention mechanism to enhance discriminability and interpretability.

(1) Construction of temporal sliding windows

The feature matrices of all participants are stacked row-wise to obtain $\tilde{X} \in R^{M_{tot}*D}$. The mean and standard deviation of the j column is then computed as:

$$\mu_j = \frac{1}{M_{tot}} \sum_{i=1}^{M_{tot}} \tilde{X}_{i,j} \quad (4)$$

$$\sigma_j = \sqrt{\frac{1}{M_{tot}} \sum_{i=1}^{M_{tot}} (\tilde{X}_{i,j} - \mu_j)^2} \quad (5)$$

The Z-Score normalization transformation is:：

$$\tilde{X}_{i,j} = \frac{\tilde{X}_{i,j} - \mu_{i,j}}{\sigma_j} \quad (6)$$

For each normalized sequence $X^{(s)} \in R^{M_s*D}$, a series of segments (sequence samples) is generated using a window length T and step size S The sequence sample constructed from the i window is:

$$X_{(i)}^{(s)} = \left[X_{(t_i)}^{(s)}, X_{(t_i+1)}^{(s)}, X_{(t_i+2)}^{(s)} \ldots X_{(t_i+T-1)}^{(s)} \right] \in R^{T*D} \quad (7)$$

Where $X_{(t)}^{(s)}$ denotes the feature vector at row t.

Make the label of each window determined by the PERCLOS value at the end of the window, and define the mapping function $g(x)$ as:

$$y = g(x) = \begin{cases} 0, & p < \eta_1 \\ 1, & \eta_1 \le p < \eta_2 \\ 2, & p > \eta_2 \end{cases} \quad (8)$$

Where η_1 =0.35, η_2 =0.70 are the thresholds for fatigue label construction. The label vector $y \in \{0,1,2\}^N$ is then obtained.

(2) Stacked LSTM Temporal Modeling

Based on the temporal samples constructed above, this study employs a two-layer Long Short-Term Memory (LSTM) network to model fatigue states.

By incorporating gating mechanisms (forget gate, input gate, output gate), LSTM effectively alleviates the vanishing gradient problem in traditional RNNs and captures long-term dependencies. Its core computation is as follows:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (9)$$

Where f_t is the forget gate; i_t is the input gate; $\tilde{c}_t$ is the candidate state; c_t is the memory cell update.

The hidden state and memory cell are updated recursively at each time step. The final output hidden state is:

$$h_t = o_t \cdot tanh(c_t) \quad (10)$$

Where h_t is the hidden state and o_t is the output gate.

The above mechanism enables LSTM to preserve both long-term trends and short-term dynamics, making it suitable for modeling fatigue accumulation patterns in EEG and EOG.

The proposed SLAN extends along a deeper dimension, where the final output hidden sequence of the previous layer serves as the input to the next layer.

Let the output of the first LSTM layer be: $h_t^{(1)}$

Then the second layer receives this sequence as its input: $x_t^{(2)} = h_t^{(1)}$

The update process of the second layer is: $h_t^{(2)} = LSTM^{(2)}\left(x_t^{(2)}, h_{t-1}^{(2)}, c_{t-1}^{(2)}\right)$

The hidden state at the last time step of the second layer is used for classification: $z = W_c h_T^{(2)} + b_c$

Finally, the classification result is: $\hat{y} = Soft\,max(z)$

Structurally, the first LSTM layer extracts basic temporal features and filters noise, while the second layer further learns high-level semantic information, thereby enhancing the model's ability to represent complex fatigue patterns. Finally, the model outputs the hidden state at the last time step for subsequent classification

tasks.

(3) Attention mechanism

To further improve the model's ability to capture critical temporal information, an attention mechanism is introduced at the output layer of the LSTM. Specifically, the hidden states at each time step are weighted, and attention weights are learned to measure the importance of different time segments for classification. The computation includes attention scoring, normalization, and weighted summation. A learnable attention mechanism is applied to each hidden state h_t of the last LSTM layer along the time dimension:

$$u_t = tanh(Wh_t + b) \in R^H \quad (11)$$

Score:

$$e_t = u^T u_t \in R \quad (12)$$

Normalize to obtain attention weights:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)}, \quad \sum_{t=1}^{T} \alpha_t = 1 \quad (13)$$

Finally, weighted aggregation is performed as:

$$h_{att} = \sum_{t=1}^{T} \alpha_t h_t \in R^H \quad (14)$$

With the attention mechanism, the model adaptively focuses on time regions critical for fatigue discrimination, enhancing classification performance.

Training employs sparse categorical cross-entropy loss and the Adam optimizer (learning rate 1×10^{-3}), with a batch size of 64 and 30 epochs. 20.0% of the training set is used for validation. The test set is partitioned via Stratified Shuffle Split (test_ratio=0.2, The dataset was divided into training and testing subsets with a ratio of 8:2 using Stratified Shuffle Split) to preserve class distribution. Model evaluation includes accuracy, precision, recall, F1-score, and miss rate.

Results and analysis

Feature-level analysis

To evaluate the effectiveness of multi-band ratio feature fusion, comparative experiments were conducted using five individual band-ratio features (α/β, δ/θ, γ/β, θ/α, and θ/β) and the proposed fusion of all five band-ratio features. Each feature set was separately fed into the model under the same network architecture and training settings. Model performance was evaluated using accuracy, precision, recall, miss rate, and F1-score. The experimental results are presented in Table 1.

Table 1. Performance comparison of different feature sets.

features	Accuracy	Precision	Recall	Miss rate	F1-score
Multi-band features	0.9745	0.9001	0.8888	0.1112	0.9001
α/β	0.9560	0.8315	0.8263	0.1737	0.8315
δ/θ	0.9669	0.8668	0.8451	0.1549	0.8668
γ/β	0.9627	0.8278	0.8129	0.1871	0.8278
θ/α	0.9697	0.8647	0.8554	0.1446	0.8647
θ/β	0.9641	0.8498	0.8486	0.1514	0.8498

Experimental results demonstrate that the multi-band ratio feature fusion model achieved the best performance across all evaluation metrics. The fused-feature model attained an accuracy of 0.9745 and an F1-score of 0.9001, outperforming all single band-ratio feature models. Compared with the best-performing individual feature, δ/θ (F1-score =0.8668), the proposed fusion strategy improved the F1-score by approximately 3.3%. In addition, recall increased to 0.8888, while the corresponding miss rate decreased to 0.1112.

From a physiological perspective, different EEG frequency bands reflect distinct aspects of neural activity. Theta activity is commonly associated with drowsiness and reduced attention, alpha activity reflects relaxed states, whereas beta activity is closely related to cognitive processing. Consequently, a single band-ratio feature can only capture a limited aspect of fatigue-related neural dynamics. In contrast, multi-band feature fusion integrates information from multiple neural processes, providing a more comprehensive representation of fatigue states.

Furthermore, the recall and miss rate results indicate that the fused-feature model exhibits more stable fatigue detection performance with fewer missed detections. This is particularly important in fatigue-monitoring applications, where missed detections may lead to potential safety risks. Therefore, the proposed multi-band ratio feature fusion strategy not only improves overall

recognition performance but also enhances the reliability of the model in practical applications.

Model-level analysis

To evaluate the effectiveness of the proposed SLAN for EEG-based fatigue recognition, three representative deep learning models, namely ESTCNN, EEGNet, and Interpretable CNN, were selected for comparative analysis. All models were trained and tested using the same dataset, feature inputs, and train-test partition settings. Performance was evaluated using accuracy, precision, recall, miss rate, and F1-score (Table 2 and Figure 4).

Table 2. Miss rate comparison of different models.

	ESTCNN	Interpretable CNN	EEGNet	SLAN (Proposed)
Miss rate	0.1728	0.1282	0.1895	0.1112

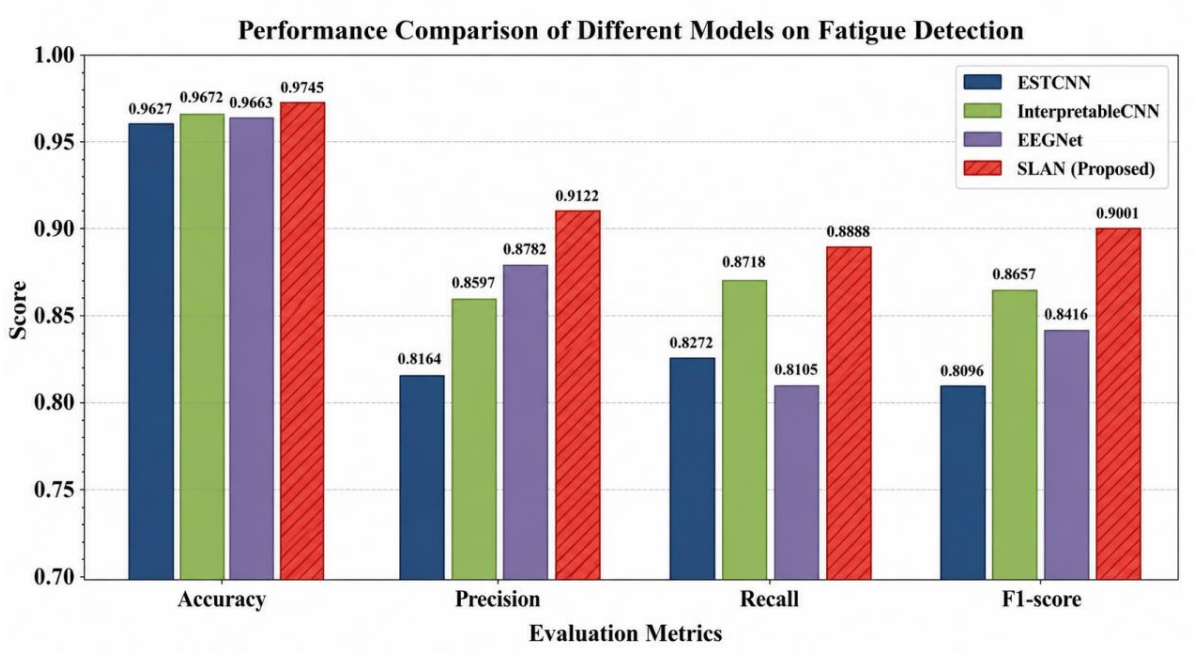

Figure 4. Performance comparison of different models.

As shown in Table 2 and Figure 4, the performance of different models varies in the fatigue recognition task. ESTCNN achieved an Accuracy of 0.9627. However, its relatively lower precision and F1-score indicate a higher tendency toward misclassification of fatigue samples. Interpretable CNN outperformed ESTCNN across all evaluation metrics, achieving an accuracy of 0.9672 and an F1-score of 0.8657, suggesting improved capability in extracting informative EEG features. Although EEGNet achieved a relatively high precision of 0.8782, its recall was only 0.8105, resulting in a miss rate of 0.1895, which indicates a higher proportion of missed fatigue detections. In contrast, the proposed SLAN achieved the best performance across all metrics. Specifically, it obtained an accuracy of 0.9745, a precision of 0.9122, a recall of 0.8888, and an F1-score of 0.9001, while reducing the miss rate to 0.1112. These results demonstrate that SLAN is more effective in identifying fatigue states and significantly reduces missed detections compared with the competing models.

The superior performance of SLAN can be attributed to its ability to model temporal dependencies in EEG signals. Conventional CNN-based models primarily focus on spatial feature extraction and are less effective at capturing sequential information. By contrast, the stacked LSTM architecture learns long-term temporal patterns associated with fatigue progression, while the attention mechanism further emphasizes informative temporal segments, thereby enhancing fatigue recognition performance.

Attention analysis

To further understand the decision-making process of the proposed model, the attention weights generated by the attention layer were visualized and analyzed.

The visualization results show that the attention weights gradually increase over time, with higher weights assigned to the latter portion of the temporal window. This indicates that the model places greater emphasis on EEG signals closer to the current time when determining fatigue states, as illustrated in Figure 5.

The observed pattern suggests that fatigue exhibits a cumulative temporal characteristic, whereby recent neural activity contributes more substantially to fatigue recognition.

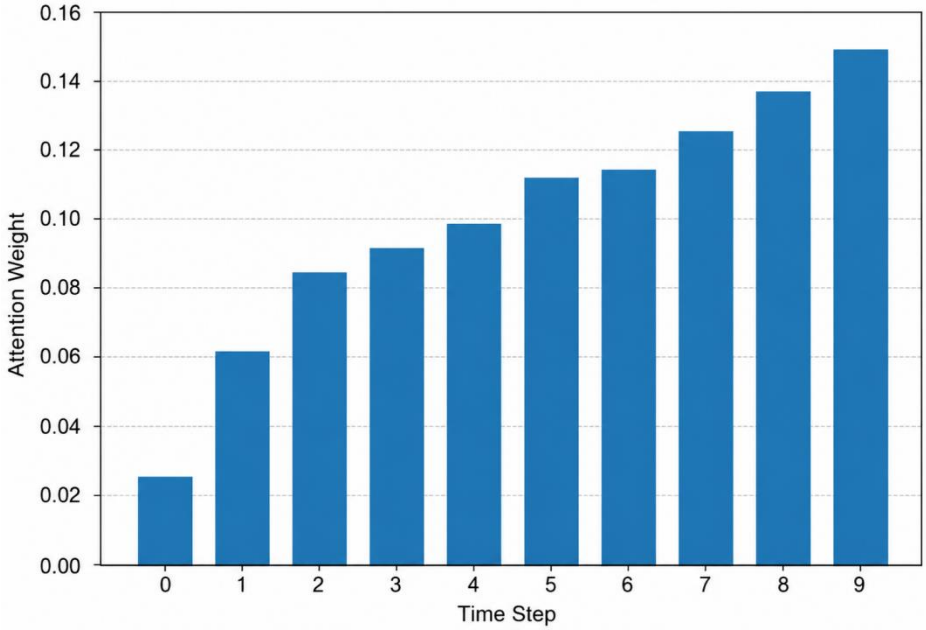

Figure 5. Temporal distribution of attention weights.

In addition, the attention weights are smoothly distributed across multiple time steps rather than being concentrated at a single point. This finding indicates that the model makes decisions by integrating information from a continuous temporal interval, rather than relying on isolated EEG observations. Such behavior is consistent with the gradual development of driver fatigue and further supports the effectiveness and interpretability of the proposed attention-based temporal modeling framework.

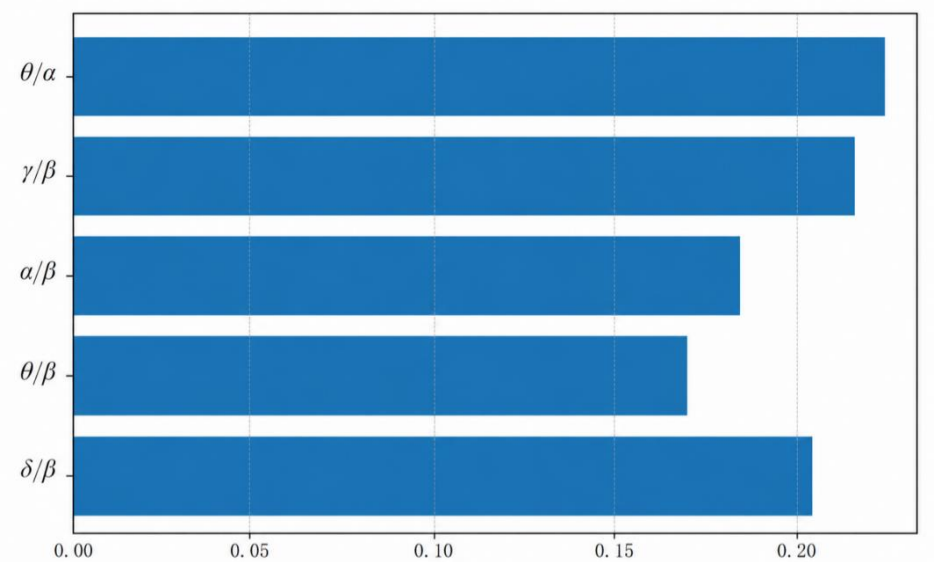

Figure 6. Distribution of feature contributions across band-ratio features.

The combined results of temporal attention distribution and feature importance analysis indicate that the attention mechanism can learn fatigue-related information from both temporal and feature dimensions. From the temporal perspective, the model assigns greater attention to EEG signals in the later portion of the sliding window, suggesting that fatigue exhibits a cumulative temporal characteristic. From the feature perspective, the model primarily relies on the θ/α, γ/β, and α/β band-ratio features, which are consistent with the known physiological changes in EEG rhythms associated with fatigue.

These findings demonstrate that the attention mechanism effectively enhances the model's focus on informative temporal segments and discriminative EEG features, thereby improving fatigue recognition performance. It

should be noted that attention weights represent the model's degree of focus rather than strict causal contributions. Therefore, the analysis provides an auxiliary interpretability perspective rather than a definitive explanation of the model's decisions.

Conclusion

This study proposes a fatigue detection framework that integrates EEG band-ratio features with eye-movement-based weak supervision labels and develops a SLAN for temporal fatigue classification.

Experimental results demonstrate that: (1) among the extracted features, the θ/α, γ/β, and α/β band-ratio features contribute most significantly to fatigue recognition, which is consistent with established physiological mechanisms of fatigue-related EEG changes. (2) The proposed model achieves approximately 97.0% classification accuracy in the three-class fatigue detection task and consistently outperforms baseline and representative deep learning models across all evaluation metrics. (3) Attention weight analysis reveals that the model assigns higher importance to EEG signals in the latter part of the temporal window, further confirming the cumulative nature of fatigue development.

Despite these promising results, several limitations remain. The dataset is collected under controlled experimental conditions and contains a limited number of participants, which may affect the generalizability of the model. In addition, although the proposed temporal framework effectively captures fatigue-related sequential dependencies, future studies may incorporate Transformer-based architectures to further enhance long-range temporal modeling and improve interpretability. Developing more reliable and theoretically grounded explainable methods for EEG-based fatigue detection remains an important direction for future research.

Funding

This work was supported by Natural Science Foundation of Sichuan Province (Grant No. 2026NSFSC0319, entitled "Research on Driving-chain Interaction Mechanisms and Brain-inspired Decision-making Methods for Multi-vehicle Collaboration in Mixed Traffic Scenarios").

Acknowledgements

The authors would like to thank Prof. Hanying Guo and all the volunteers who participated in the driving fatigue experiments and data collection process.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Nguyen, K., Dunbar, C., Guyett, A., Bickley, K., Nguyen, D. P., Reynolds, A. C., Vakulin, A. (2025) Poorer objective but not subjective driving performance in drivers vulnerable to sleep loss effects during extended wake. *Journal of Sleep Research*, 34(4), e14455.

[2] Sikander, G., Anwar, S. (2018) Driver fatigue detection systems: a review. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2339-2352.

[3] Hassan, O. F., Ibrahim, A. F., Gomaa, A., Makhlouf, M. A., Hafiz, B. (2025) Real-time driver drowsiness detection using transformer architectures: a novel deep learning approach. *Scientific Reports*, 15(1), 17493.

[4] Mora, H. J., Echaveguren, T. B., Pino, E. J. (2026) Advanced forecasting of driver drowsiness events: non-intrusive data and multimodal BiLSTM-based modeling. *Biomedical Signal Processing and Control*,119, 109793.

[5] Skaiky, A. A., Ayad, H. (2025) Deep learning approaches for fatigue detection: a traditional review of models, datasets, and applications. *Al-Farahidi Expert Systems Journal*, 1(2), 10.

[6] AlArnaout, Z., Zaki, C., Kotb, Y., AlAkkoumi, M., Mostafa, N. (2025) Exploiting heart rate variability for driver drowsiness detection using wearable sensors and machine learning. *Scientific Reports*, 15(1), 24898.

[7] Fowler, A., Harvey, C., Wilson, M. L., Sharples, S. (2026) Using wearable measures to infer moments in workload from Electrodermal Activity and individual workload from Heart Rate Variability during a simulated railway signalling task. *Applied Ergonomics*, 133, 104710.

[8] Wang, Z., Du, X., Jiang, C., Sun, J. (2025) Research on the prediction of driver fatigue degree based on EEG signals. *Sensors*, 25(23), 7316.

[9] Kalogeropoulos, C., Theofilatos, K., Mavroudi, S. (2026) From neurons to networks: a holistic review

of electroencephalography (EEG) from neurophysiological foundations to AI techniques. *Signals*, 7(1), 17.

[10] Li, A., Wang, Z., Xu, T., Zhou, T., Zhao, X., Hu, H., Van Hulle, M. M. (2026) A cross-subject band-power complexity metric for detecting mental fatigue through EEG. *Brain Sciences*, 16(2), 199.

[11] Yang, B., Ding, Y., Cui, C., Guo, T. (2025) A hybrid kernel principal component analysis and long short-term memory approach for EEG signal-based fatigue evaluation. *Alexandria Engineering Journal*, 131, 209-217.

[12] Zhang, Y., Xu, X., Du, Y., Zhang, N. (2025) TMU-net: a transformer-based multimodal framework with uncertainty quantification for driver fatigue detection. *Sensors*, 25(17), 5364.

[13] Bello, I., Sehnan, M., Dang, W., Banzi, J. F., Aminu, S. A., Gao, Z. (2026) A frequency-attentive spatio-temporal convolutional neural network with weighted feature fusion for EEG-based driver fatigue detection. *Biomedical Signal Processing and Control*, 113, 108979.

[14] Kim, S., Wisanggeni, I., Ros, R., Hussein, R. (2020) Detecting fatigue driving through PERCLOS: a review. *International Journal of Image Processing (IJIP)*, 14(1), 1-7.

[15] Quddus, A., Zandi, A. S., Prest, L., Comeau, F. J. (2021) Using long short term memory and convolutional neural networks for driver drowsiness detection. *Accident Analysis & Prevention*, 156, 106107.