Driving Fatigue Detection Based on Hybrid Electroencephalography and Eye Tracking

Zequan Lian, Tao Xu , Zhen Yuan , Junhua Li , Nitish Thakor, and Hongtao Wang

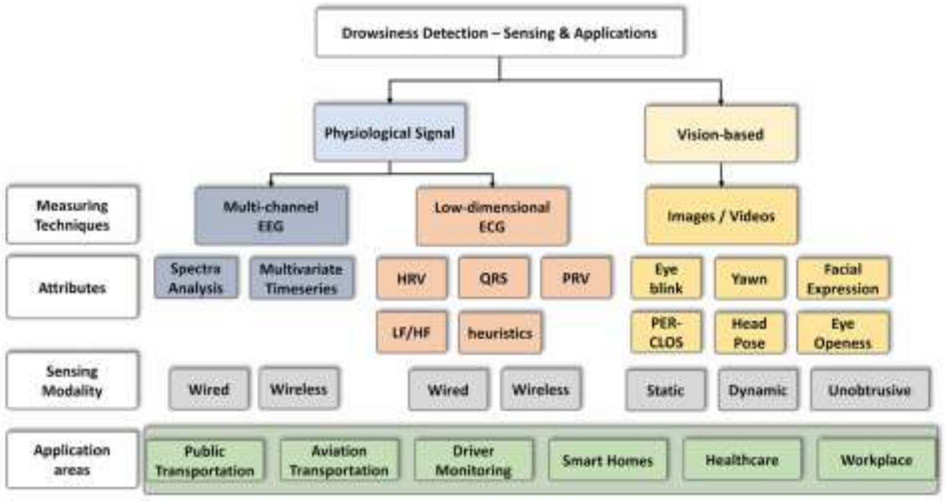
IJBHI 2024

Introduction

- Public Transportation
- Road Safety
- Workplace Safety
- ……

- 基于生理信号的检测：EEG（脑电图）、ECG（心电图）
- 基于视觉特征的检测：眼部特征、面部特征、嘴部特征

- 单模态方法存在明显局限（如只使用EEG信号）
- 同一个被试不同会话（cross-session）或不同被试（cross-subject）时，模型性能大幅下降。
- 引入多模态，能从多个视角捕获疲劳的不同方面（生理+行为），互补信息。
- 现有CCA-based多模态方法（如DCCA/DGCCA）仍有缺陷：强迫模态间最大相关性，容易导致特征坍缩，限制了特征的可区分性。



Method

(a) Experimental scenario (b) Close-up details

Fig. 1. The system of multimodal driving fatigue detection. (a) experimental scenario. (b) a subject wearing an EEG cap and a pair of Tobbi Glasses was driving a simulated vehicle in the simulation platform.

14名健康参与者（9男5女，平均年龄21.6 ± 1.4岁）
32通道 EEG帽，Tobii Glasses 2 眼动追踪眼镜，Thrustmaster T300 GT 驾驶模拟器
每位参与者进行2个会话（session），间隔一周内，每个会话持续90分钟连续驾驶。

Physical Data（物理数据，如瞳孔直径）。
Tracking Data（追踪数据，如眼跳、注视）经过特殊 I-VT（Velocity-Threshold Identification）注视滤波器处理，提取关键指标。

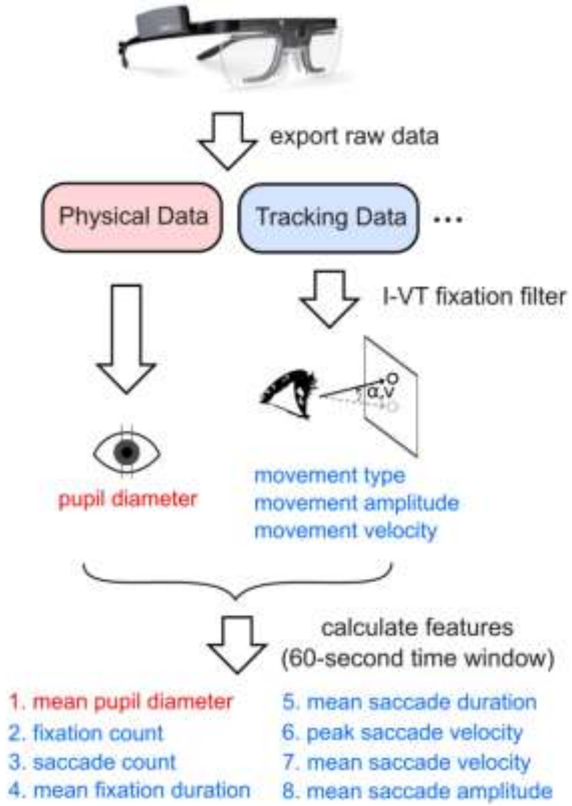
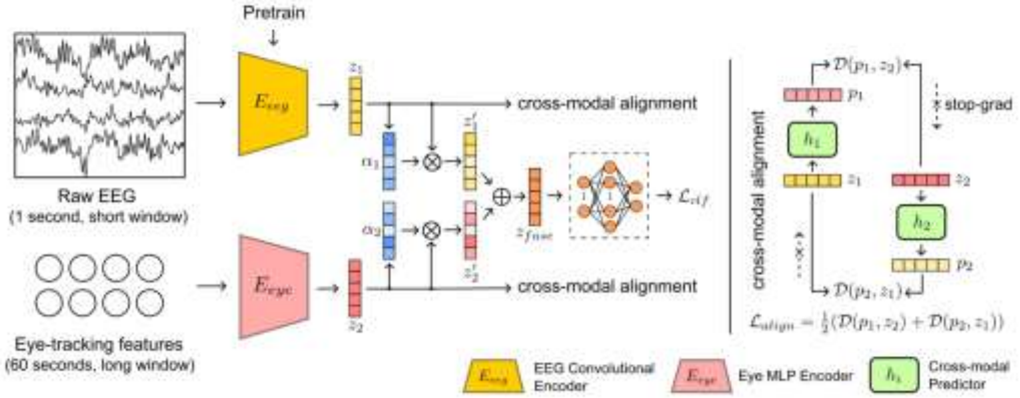
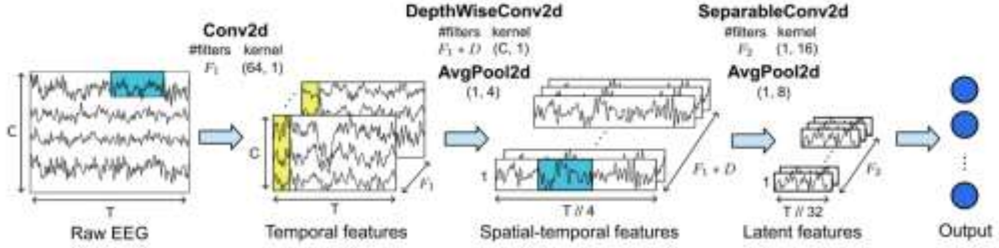


Fig. 2. The overall framework of eye-tracking data processing. The data was first recorded on an SDCard of Tobbi Glasses and then exported by Tobbi Pro Lab. Among the exported data, the tracking data was further processed to obtain several critical indices by a specially designed I-VT fixation filter. Finally, a 60-second sliding window with a 1-second time step was applied to generate 8 characteristics.

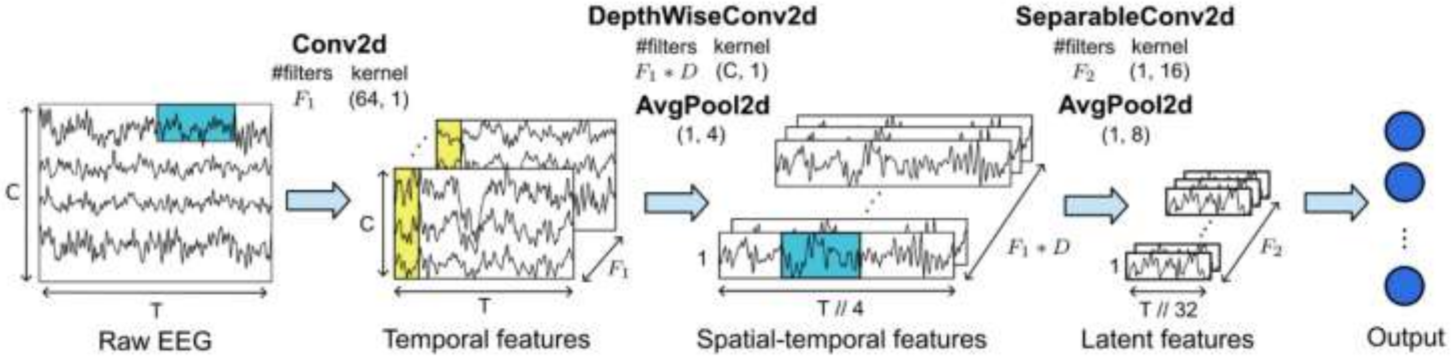
Method



(a) Overall Framework

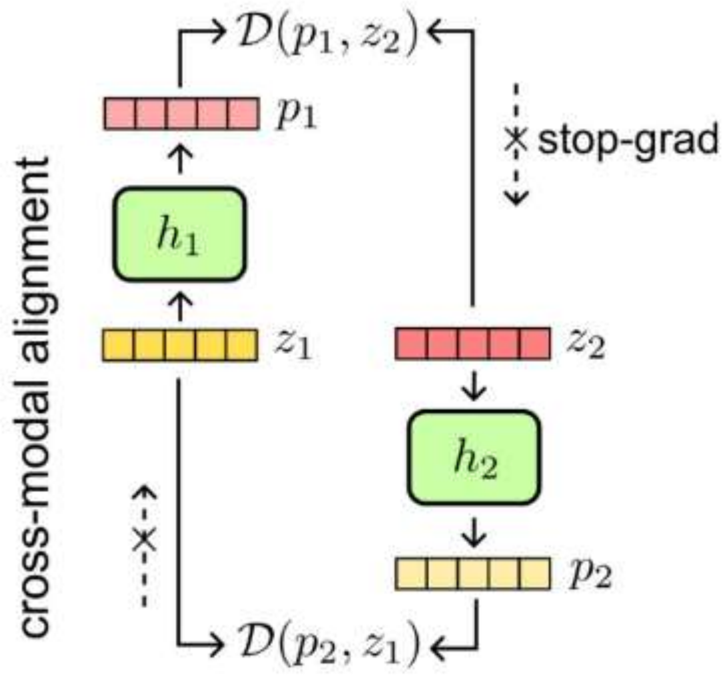


(b) EEG Convolutional Encoder

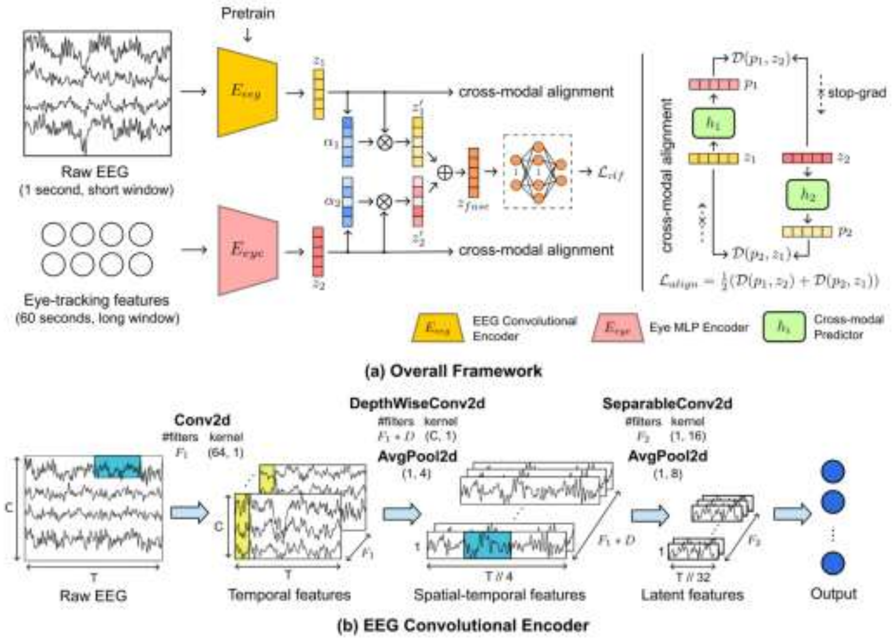


(b) EEG Convolutional Encoder

Method



$$\mathcal{L}_{align} = \frac{1}{2}(\mathcal{D}(p_1, z_2) + \mathcal{D}(p_2, z_1))$$



(a) Overall Framework

(b) EEG Convolutional Encoder

$$p_1 = h_1(z_1)$$

$$p_2 = h_2(z_2),$$

$$\mathcal{D}(p_1, z_2) = -\frac{p_1}{\|p_1\|_2} \cdot \frac{z_2}{\|z_2\|_2},$$

计算预测结果和目标特征的负余弦相似度

$$\mathcal{L}_{align} = \frac{1}{2}(\mathcal{D}(p_1, z_2) + \mathcal{D}(p_2, z_1)),$$

$$\mathcal{D}(p_1, \mathrm{stopgrad}(z_2)),$$

$$\mathcal{L}_{align} = \frac{1}{2}(\mathcal{D}(p_1, \mathrm{stopgrad}(z_2)) + \mathcal{D}(p_2, \mathrm{stopgrad}(z_1))), \tag{5}$$

Experiments

- 三个频段
  - Theta 波 (Theta band，通常 4–8 Hz)
  - Lower Alpha 波 (Lower Alpha band，通常 8–10 Hz)
  - Beta 波 (Beta band，通常 13–30 Hz)
- 五个脑区
  - Prefrontal Cortex (前额叶皮层)
  - Frontal (额叶)
  - Central (中央区)
  - Parietal (顶叶)
  - Occipital (枕叶)
- Prefrontal Cortex在三个频段上都有显着增加
- Parietal和Occipital的Lower Alpha波段表现出显着增加
- Frontal内的β带明显减少。

证明EEG模态能有效捕捉疲劳相关的脑生理变化

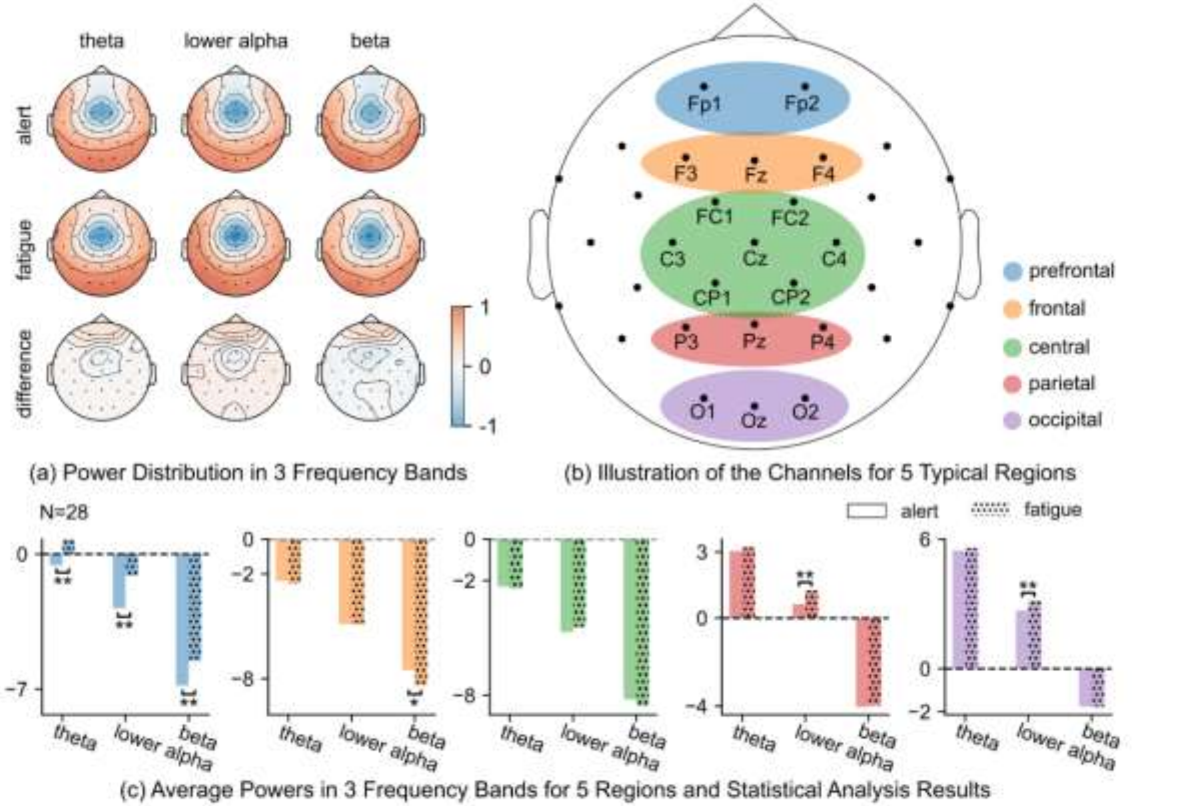
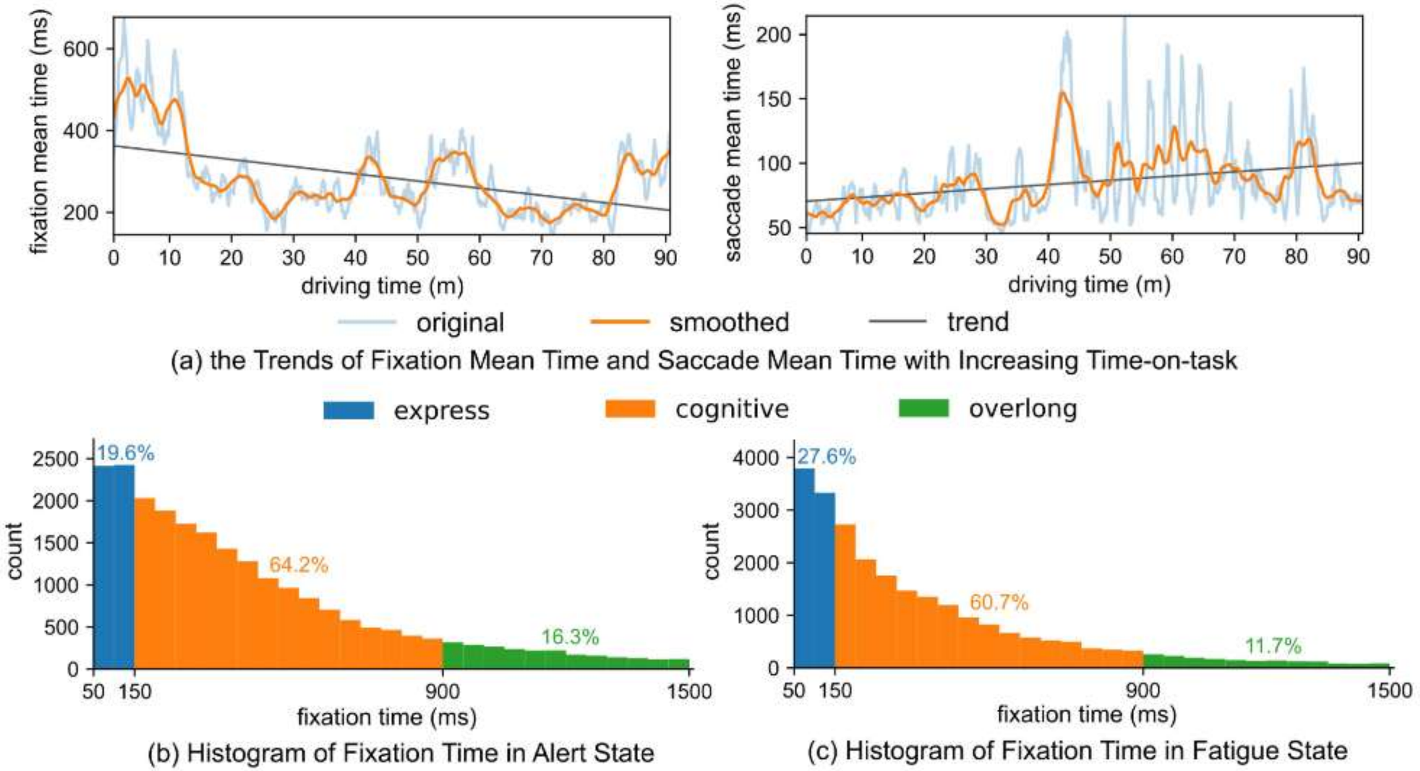


Fig. 4. The EEG topographies of power distribution in three frequency bands and the subsequent statistical analysis across five brain regions are presented. These topographies are organized in a 3x3 grid, as depicted in (a), with each column representing a distinct frequency band and each row signifying different cognitive states: the alert state, fatigue state, and the subtraction of the fatigue state from the alert state. Notably, the color representations in (b) and (c) are consistent. In (b), the channels chosen for analysis in the five brain regions are shown. Meanwhile, (c) illustrates the average power across three frequency bands for the five regions in two mental states, accompanied by the corresponding significance by Wilcoxon test. The Wilcoxon test was performed on 28 driving sessions ("*" indicates $p < 0.05$, and "**" indicates $p < 0.01$).

Experiments

- 随着任务时间的增加，平均扫视时间呈现出增加的趋势，平均注视时间呈下降趋势
- 注视时间分为三类：express（<150 ms）、cognitive（150 ms-900 ms）和overlong（> 900 ms）
- cognitive 64.2% to 60.7%,
- overlong 16.3% to 11.7%
- express 19.6% to 27.6%.
- 疲劳时，整体认知资源耗尽，导致主动认知型注视比例下降。眼睛虽然还在动，但更多是“走马观花”而非“深度解读”。
- 眼动特征能有效区分疲劳（眼跳增多、注视缩短，表示注意力分散）。



(a) the Trends of Fixation Mean Time and Saccade Mean Time with Increasing Time-on-task

(b) Histogram of Fixation Time in Alert State

(c) Histogram of Fixation Time in Fatigue State

Experiments

Intra-session（会话内评估）
含义：训练和测试数据来自**同一个session**（同一被试、同一时间段的连续驾驶数据）。

Cross-session（跨会话评估）
含义：训练和测试数据来自**同一个被试的不同session**（同一人，但不同时间，通常间隔几天到一周）。

Cross-subject（跨被试评估）
含义：训练和测试数据来自**不同被试**（不同人）。

	EEG	Eye Tracking	Multimodal -DCCA	Multimodal -DGCCA	Multimodal -Ours
Intra-session	98.84	91.61	99.60	99.72	**99.93**
Cross-session	80.15	82.75	88.25	87.83	**88.67**
Cross-subject	72.32	73.98	77.64	77.75	**78.19**

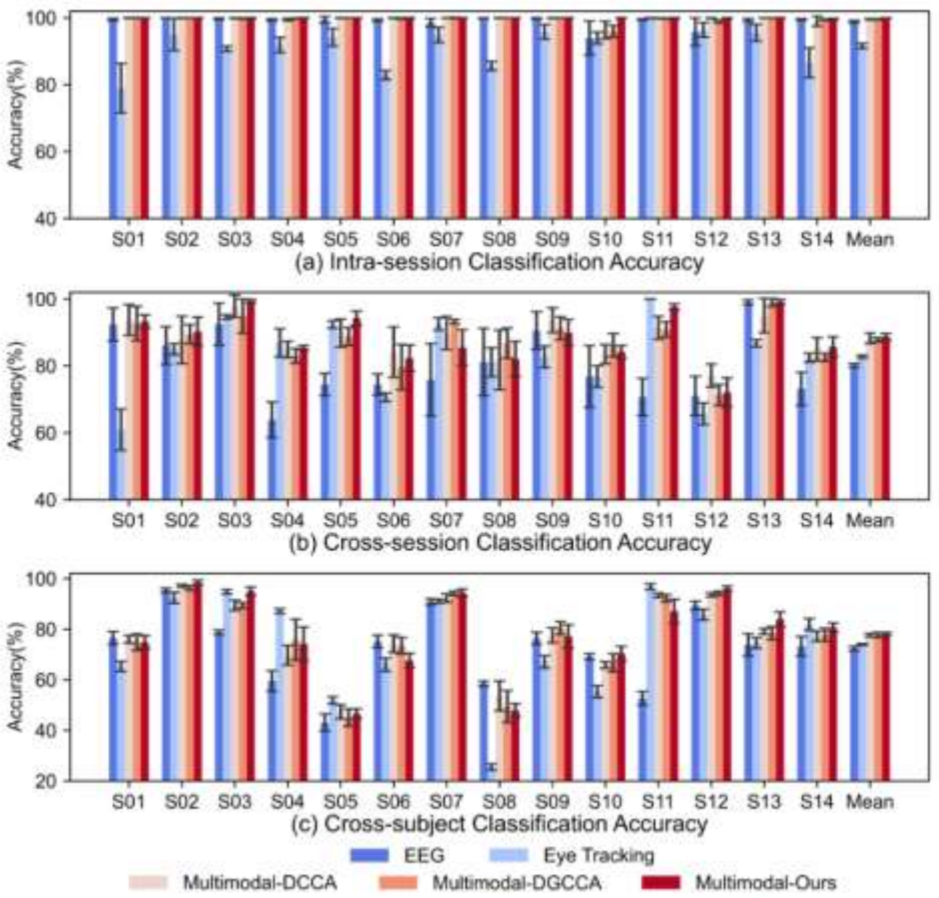


Fig. 6. Classification results of methods in intra-session, cross-session, and cross-subject evaluation tasks. (a) depicts the intra-session results. The average intra-session accuracy of EEG, eye tracking, DCCA multimodal, DGCCA multimodal and our multimodal approaches is 98.84 ± 0.39%, 91.61 ± 0.76%, 99.60 ± 0.16%, 99.72%± 0.15%, and 99.93 ± 0.05%, respectively. (b) depicts the cross-session results. The average cross-session accuracy of the five methods is 80.15 ± 0.58%, 82.75 ± 0.44%, 88.25 ± 1.58%, 87.83% ± 0.63%, and 88.67 ± 1.00%, respectively. (c) depicts the cross-session results. The average cross-subject accuracy of the five methods is 72.32% ± 0.86%, 73.98% ± 0.32%, 77.64% ± 0.69%, 77.75% ± 1.00%, and 78.19% ± 0.56%, respectively. Our multimodal method performs consistently well among the five approaches in three evaluation tasks.

TABLE IV
PERFORMANCE (IN PERCENT) ON SEED-VIG DATASET

Method	Accuracy
Multimodal-DCCA	86.99
Multimodal-DGCCA	87.83
Multimodal-Ours	**96.62**

The bold values indicates the highest accuracy among the methods.

Experiments

- 单模态（EEG或眼动）：警觉/疲劳两类点重叠严重（边界模糊）。
- 多模态：两类点分离明显，同一类聚得更紧,边界清晰。

- DCCA 强迫最大化模态间相关性,容易导致特征“坍缩”（太相似，丢失区分信息），类间边界模糊,图中表现为簇拉长、重叠。
- 本文方法更灵活：让模态互相“能预测对方”，但不强制完全一样，保留了更多互补和区分信息 → 图中簇更紧、分离更干净。

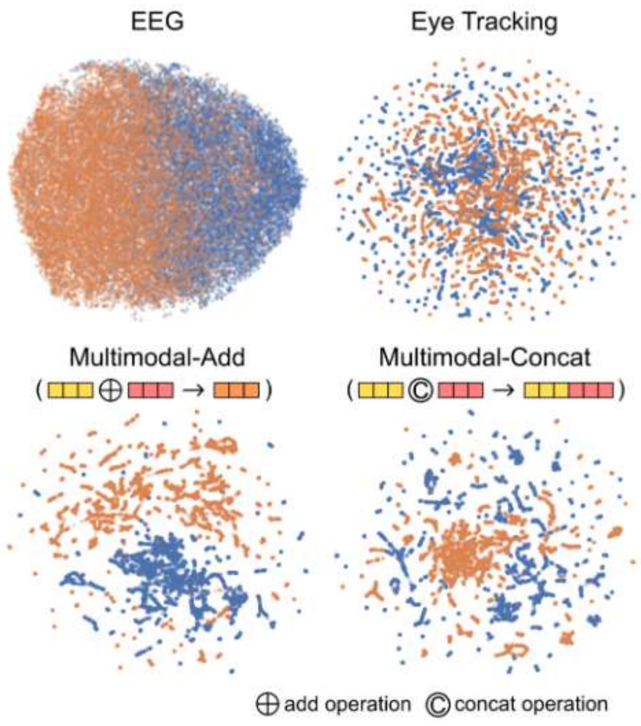


Fig. 8. Feature visualization of unimodal and multimodal methods for all subjects. The visualization represents the features learned by the two unimodal methods, namely “EEG” and “Eye Tracking,” as well as the features learned by the proposed multimodal methods using additive fusion and concatenative fusion as fusion methods denoted as “Multimodal-Add” and “Multimodal-Concat,” respectively. Notably, the visualization intentionally omits the modules of 1D attention and cross-modal alignment in “Multimodal-Add” and “Multimodal-Concat” to provide a clear and intuitive assessment of the efficacy of adopting multimodal approaches. To avoid the overfitting of models, the data of Subject 2 was used as validation and model selection.

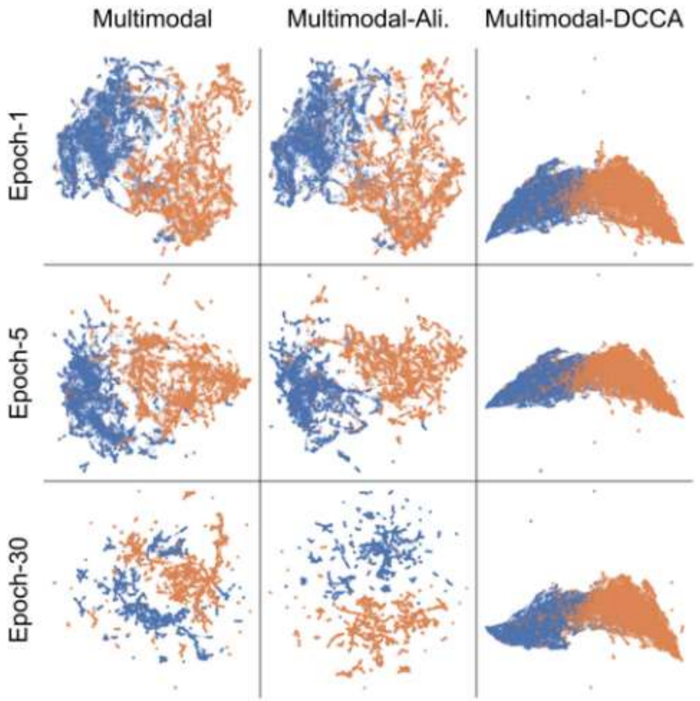


Fig. 10. Impact on features (UMAP) through cross-modal predictive alignment and Deep Canonical Correlation Analysis (DCCA). The visualization encompasses data collected from all subjects. The term “Multimodal” pertains to the proposed multimodal model without cross-modal alignment or DCCA, while “Multimodal-Ali.” and “Multimodal-DCCA” refer to the models that incorporate cross-modal alignment and DCCA, respectively. The top, middle, and bottom rows of the visualization correspond to the features extracted from epoch 1, epoch 5, and epoch 30, respectively. Notably, the module of 1D attention was deliberately excluded to facilitate a clear and direct comparison among the methods.

Thanks