

基于便携式脑电和心电的双层级融合抑郁症智能检测方法

应屹扬¹, 裘生梁^{1,2}, 陈熙昂³, 江海腾⁴

(1. 浙江中医药大学第一临床医学院, 杭州 310053; 2. 浙江省中医院(浙江中医药大学附属第一医院), 杭州 310006; 3. 浙江大学生物医学工程与仪器科学学院, 杭州 310027; 4. 浙江大学教育部脑与脑机融合前沿科学中心, 杭州 310058)

摘要: 重度抑郁障碍(Major depressive disorder, MDD)的传统筛查与诊断主要依赖结构化访谈和量表评估, 易受主观性与场景可及性限制, 难以满足基层与社区场景下的早期识别需求。为此, 本文面向真实世界可扩展筛查场景, 构建并验证一种基于便携式单通道脑电(Electroencephalography, EEG)与可穿戴心电(Electrocardiography, ECG)的多模态抑郁识别框架P-MFCNet(Portable multimodal fusion with CLS network)。在数据层面, 设计了包含睁眼静息、负性情绪诱发、闭眼静息和半结构化访谈的多任务实验范式, 并在两家医疗机构完成同步多模态采集; 方法层面, 提出基于Transformer与分类标记(Classification token, CLS)的双层级融合策略, 分别从模态维度与任务态维度学习互补信息与条件差异, 再通过层次化CLS生成样本级融合表示, 并结合ResNet分类头实现MDD与HC判别。结果表明, P-MFCNet在单中心五折交叉验证中取得 $ACC=0.842\ 4$ 、 $AUC=0.796\ 7$, 在跨中心外部测试中取得 $ACC=0.810\ 0$ 、 $AUC=0.875\ 0$ 。整体结果显示, P-MFCNet在准确率、灵敏度及综合表现方面优于单模态模型, 并与对比方法相比表现出较好的性能平衡。进一步结合置零消融与注意力可视化开展事后可解释性分析, 结果显示模型整体上更关注ECG/HRV相关信息, 而EEG在负性诱发和静息条件下提供补充判别证据。上述结果表明, 便携式EEG-ECG多模态方案在抑郁快速筛查中具有良好的应用潜力, 并具备一定的跨中心泛化能力和可解释性。

关键词: 重度抑郁障碍; 便携式脑电; 可穿戴心电; 多模态融合

中图分类号: TP391

文献标志码: A

引用格式: 应屹扬, 裘生梁, 陈熙昂, 等. 基于便携式脑电和心电的双层级融合抑郁症智能检测方法[J]. 数据采集与处理, 2026, 41(4): 966-980. YING Yiyang, QIU Shengliang, CHEN Xi'ang, et al. A hierarchical fusion method for intelligent depression detection based on portable EEG and ECG[J]. Journal of Data Acquisition and Processing, 2026, 41(4): 966-980.

引言

重度抑郁障碍(Major depressive disorder, MDD)是一类高患病率、高致残性的精神疾病, 典型表现为持续性情绪低落、快感缺乏以及社会与职业功能显著受损^[1-2]。当前临床筛查与诊断主要依赖医生结构化访谈、标准化问卷及自评量表等方式, 这些方法在实际应用中虽具有可操作性与普及性, 但仍不可避免地受到主观判断差异、回忆偏差与信息不完整的影响, 尤其在初级保健与社区场景中更易出现漏诊与延迟识别^[3]。以中国为例, 有研究报告抑郁症终生患病率约为6.9%, 但接受专业干预的比例不足10%, 大量患者长期处于未确诊或未规范治疗状态^[4]。基于抑郁症造成的巨大疾病负担, 研发一种客

观、快速、可扩展且有助于早期识别与干预的筛查工具具有明确的公共卫生与临床价值。

近年来,多模态生理监测技术与人工智能方法的快速发展,为抑郁症的客观评估提供了新的技术路径^[5]。相较于单一模态,融合心理、行为与生理信号如脑电图(Electroencephalography, EEG)、心电图(Electrocardiography, ECG)、文本、语音与视频等的多模态方法在准确性与鲁棒性方面已展现出优势^[6-8],能够更全面地刻画抑郁相关的多维表型。其中,电生理信号具有直接、定量及可连续采集的特点,因而在客观生物标志物探索中尤具潜力^[9]。2022年,Zitouni等^[10]采集包括ECG、光电容积脉搏波描记法(Photoplethysmography, PPG)和呼吸信号(Respiratory signal, RSP)在内的多模态生理信号,采用离散小波变换提取非线性特征,基于支持向量机和径向基函数对健康与MDD患者进行分类,准确率较高。2023年,Qayyum等^[11]整合不同层次的语音和EEG特征,并应用视觉Transformer和各种预训练网络,显著提高了抑郁症患者的诊断效能。同年,Zheng等^[12]提出了一种新型的基于知识嵌入的时间卷积Transformer,用于连接音频和EEG,从而提取有效特征并提升抑郁症检测模型的性能。2025年,Zhang等^[13]采集MDD患者和健康对照组(Healthy control, HC)的EEG和ECG数据,使用收敛交叉映射方法证明了异常的心脑相互作用在抑郁症的病理生理学中起着重要作用。同年,徐庆等^[14]将语音-文本-面部表情为模型的输入特征,提出一种基于自注意力模型和残差记忆网络的多模态抑郁症检测模型TF-RMN,在DAIC-WOZ数据集上进行抑郁症检测二分类实验,准确率为0.85。

然而,现有证据仍存在关键局限:其一,多数研究依赖实验室环境下的多通道EEG系统,设备昂贵、佩戴复杂、采集成本高,难以支持大规模或真实世界筛查,比如Qayyum等^[11]和Zheng等^[12]采用的EEG数据多达128通道,难以在基层医疗与社区等真实世界场景中推广部署;其二,尽管便携式单通道EEG与可穿戴ECG设备为院外应用提供了更友好、可扩展的替代方案,早期结果也提示简化硬件条件下仍可能提取到有意义的抑郁相关特征,但便携式EEG在较低空间分辨率与噪声敏感性条件下能否提供临床可用的诊断价值,仍缺乏系统验证。2022年,王肖文等^[15]设计了一种抑郁症患者睡眠质量评估系统,包含便携式睡眠心电监测模块和深度学习评估模块,但缺乏与标准化临床诊断流程的一致性验证,例如基于结构化临床访谈的诊断结论及临床量表评估结果,其对抑郁筛查与诊断的有效性仍不明确。

基于上述背景与研究空白,本文面向真实世界可扩展筛查需求,构建并验证一种基于可穿戴便携设备的EEG-ECG多模态抑郁识别框架。方法上,本文基于抑郁相关脑电节律异常与HRV变化的理论和经验基础,采用可穿戴EEG和ECG设备,在贴近真实应用的多情境条件下获取生理数据;在信号处理与建模层面,构建标准化预处理流程,提取EEG与ECG特征,并通过跨模态融合策略建立联合模型,以实现MDD与HC的快速筛查,同时结合跨中心验证与事后可解释性分析评估模型的泛化能力与决策依据。

本文的实验流程和研究框架如图1所示,主要工作和贡献如下:

(1)构建面向真实世界筛查场景的便携式多模态采集范式与数据体系。基于便携式单通道EEG与可穿戴ECG设备,设计了包含睁眼静息、负性情绪诱发、闭眼静息和半结构化访谈的多任务实验范式,并在两家医疗机构完成数据采集与质量控制,形成用于单中心训练验证与跨中心外部测试的多模态数据基础。

(2)提出便携式EEG-ECG双层级融合模型P-MFCNet。针对便携式单通道EEG空间分辨率有限、噪声敏感等特点,提出基于Transformer与分类标记(Classification token, CLS)的层次化融合框架,分别在模态维度和任务态维度进行自适应信息聚合,并结合ResNet分类头实现MDD与HC的判别学习。

(3)开展单中心验证、跨中心测试与系统可解释性分析。在单中心数据集上采用五折交叉验证评估模型性能,并在跨中心数据集上进行测试,以验证模型的泛化能力;同时结合置零消融和注意力可视

化,从特征贡献、模态偏好和任务偏好等多个层面分析模型决策依据,为临床解释与后续模型优化提供参考。

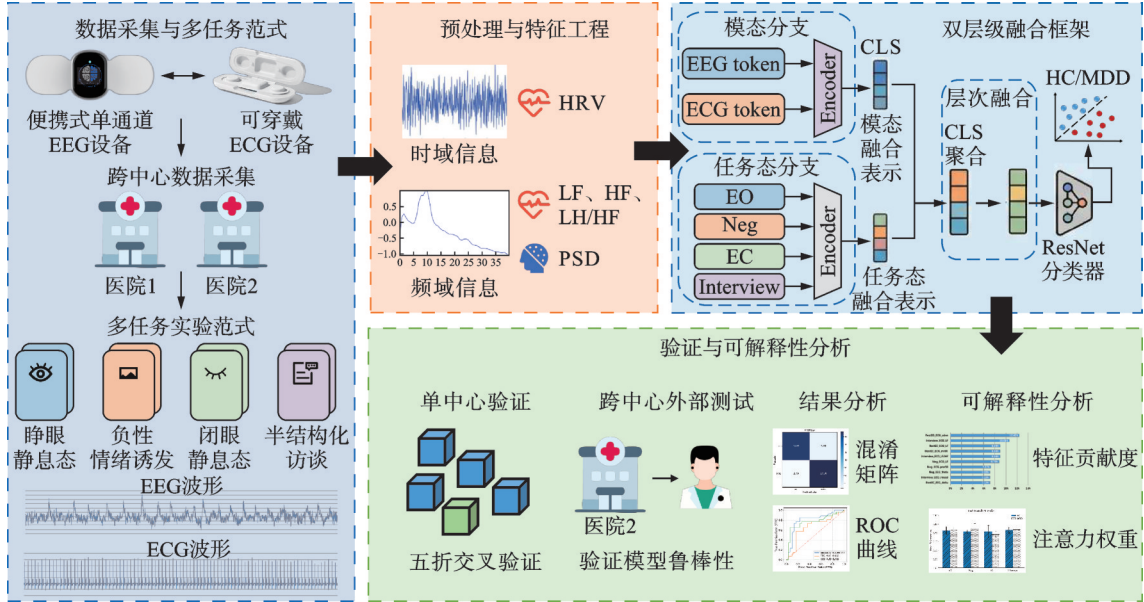


图1 实验流程和研究框架图

Fig.1 Experimental procedure and research framework diagram

1 数据采集与处理

1.1 实验范式设计

本研究采用病例-对照 (Case-control) 研究设计。受试者的初步诊断与筛查由精神科医生依据《精神障碍诊断与统计手册 (DSM-5)》研究版结构化临床访谈^[16]完成,并结合汉密尔顿抑郁量表 24 项 (HAMD-24) 进行症状严重程度评估。在此基础上,符合纳入标准的受试者进入同步多模态数据采集流程,使用便携式 EEG 与 ECG 设备进行连续记录。EEG 设备为 iFocus (Niantong Intelligence), 采用单通道前额贴片式电极; ECG 设备为 VV-330 (VivaLNK), 采用便携式电极贴片。数据采集在一系列实验条件下进行,涵盖静息态 (睁眼和闭眼)、情绪图片观看、结构化临床访谈,以及自评量表与他评量表的填写。所有实验流程均在同一时间段内一次性完成,以降低跨时段状态波动对生理信号的影响并提升可比性。

在正式任务开始前,受试者首先进行 ECG 记录,并设置 5 min 的稳定期以减少体位变化与适应阶段对心率变异性 (Heart rate variability, HRV) 指标的干扰;随后启动 EEG 记录。实验刺激呈现与多模态数据同步由 PsychoPy (v2023.2.3) 与 LabRecorder 共同完成,运行平台为华为 Matebook 14 笔记本电脑 (分辨率 2 160×1 440,刷新率 60 Hz)。为保证不同任务段落之间的精确对齐,在每个任务开始与结束时均记录绝对时间戳,用于后续数据分段与跨模态同步校准。

实验协议共包含 4 个条件,按顺序依次执行: (1) 睁眼静息 (Eyes open rest, EO)。受试者注视屏幕中央固定点 5 min; (2) 负性情绪图片观看 (Negative emotion picture viewing, Neg)。从中国面孔情绪图片系统 CFAPS^[17] 中选取 150 张图片,持续约 5 min; (3) 闭眼静息 (Eyes closed rest, EC)。受试者闭眼放松 5 min; (4) 临床访谈 (Clinical interview, Interview)。围绕抑郁相关主题设置 12 个问题,访谈提纲

参考 Zou 等^[6]并进行适配。完成上述 4 个实验条件后,受试者继续完成自评与他评量表测评,以获得症状维度与功能维度的临床表型信息,便于与生理特征进行联合分析。EEG 记录在最后一个实验条件结束后停止,而 ECG 记录则继续延长 5 min,以获取更稳定、长度更充足的 HRV 数据段,从而提高心电特征估计的可靠性。

1.2 数据集

本实验数据分别于两家医院先后采集,共纳入受试者 106 名,年龄分布为 16—60 岁。其中,杭州市第七人民医院(以下简称“杭七”)初始采集 HC 组 32 例、MDD 组 30 例;浙江省中医院(以下简称“省中”)初始采集 HC 组 21 例、MDD 组 23 例。受电极脱落、受试者中途退出以及 MDD 患者在实验范式执行过程中症状发作等因素影响,共有 10 例未纳入最终分析。最终纳入杭七 HC 组 30 例、MDD 组 26 例,省中 HC 组 20 例、MDD 组 20 例。

MDD 组患者在入组前 2 周内由精神科医生采用结构化临床访谈(SCID)确诊,诊断严重程度为 4 级;同时要求入组前 1 周内 HAMD-24 评分>18 分。所有 MDD 患者均接受综合临床评估,量表包括:匹兹堡睡眠质量指数(18 项版本,PSQI-18)用于评估睡眠质量;患者健康问卷 9 项(PHQ-9)用于抑郁症状筛查;广泛性焦虑量表 7 项(GAD-7)用于焦虑症状评估;个人与社会表现量表(4 项版本,PSP-4)用于社会功能评估;以及症状自评量表 SCL-90 的相关分量表,包括恐惧焦虑(SCL-90-PA)与精神质分量表(SCL-90-PS)。HC 组在年龄与性别方面与 MDD 组匹配,并经 SCID 确认无精神障碍病史。排除标准包括:双相情感障碍、精神分裂症及其他主要精神疾病;以及可能影响 EEG 和 ECG 采集或信号特征的治疗或用药史(如近期电休克治疗、ECT 等)。

所有受试者在入组前均签署书面知情同意书,本研究方案经杭州市第七人民医院伦理委员会审查批准(研 2023-03-02)。受试者的人口统计学和临床特征总结见表 1。

表 1 最终入组受试者人口统计学和临床特征总结
Table 1 Summary of the demographic and clinical characteristics of the final enrolled subjects

特征指标		HC (<i>n</i> = 50)	MDD (<i>n</i> = 46)	<i>ρ</i>
人口学	年龄	25.6±5.8	27.1±9.8	n.s.
	性别(男/女)	26/24	20/26	n.s.
临床量表	PSQI-18	3.49±2.56	12.4±3.84	<0.001
	PHQ-9	3.15±1.84	17.4±4.82	<0.001
	GAD-7	2.44±2.43	12.4±3.58	<0.001
	PSP-4	96.8±2.64	64.5±14.4	<0.001
	SCL-90-PA	7.47±3.16	14.2±5.93	<0.001
	SCL-90-PS	9.56±2.58	25.3±7.94	<0.001
	HAMD-24	N/A	27.13±3.24	N/A

1.3 数据预处理

对于 EEG 信号,先结合人工复核与自动筛查识别明显平线、饱和、严重基线漂移及工频干扰片段;在分段后进一步采用电压阈值与滑动窗峰峰值联合判定伪迹,即当单段绝对振幅超过±200 μV,或在 200 ms 滑动窗内峰值超过 100 μV 时,将该片段判定为伪迹片段并予以剔除,同时要求单名受试者伪迹片段比例控制在 10% 以内,超过该范围者不纳入后续分析。EEG 预处理在 EEGLAB(v2024.2)环境下完成。首先对原始 EEG 施加 1~30 Hz 带通滤波,以去除低频漂移和低频噪声;随后依据事件标记将信

号分割为4个任务片段,为后续条件级特征提取提供输入。

对于ECG信号,首先基于R峰检测结果计算相邻心搏间期(Inter-beat intervals, IBI),并按照常用HRV处理规范剔除异常IBI,即将超出局部均值 ± 3 个标准差的间期视为离群点并排除,以减少误检R峰、运动伪迹及信号丢失对HRV指标的影响。之后利用任务起始时间戳在ECG时间序列中进行最近邻匹配,实现与EEG的同步分段。通过该策略,两模态的对齐误差控制在10 ms以内,能够满足条件级特征分析与跨模态融合建模的精度需求。

上述预处理步骤对所有受试者与两组数据均采用完全一致的参数与流程,以保证组间可比性并降低处理偏差。最终获得了在4种实验条件下严格对齐、按任务分段的EEG与ECG数据,为后续特征提取、融合建模与统计分析提供了高质量输入。

2 P-MFCNet算法

2.1 特征工程

预处理后的EEG和ECG数据根据任务条件(EO、Neg、EC、Interview)被划分为4个片段,并分别进行特征提取。考虑到本研究面向便携式设备下的快速筛查场景,且当前样本规模相对有限,本文在特征工程阶段未采用大规模自动特征搜索或高维特征堆叠,而是依据抑郁相关自主神经异常与脑电节律异常的既有理论和经验基础,优先选取低维、稳健且具有明确生理意义的经典HRV与EEG频带特征,并由后续融合网络自适应学习不同特征在不同任务条件下的贡献。

ECG特征主要基于HRV,首先进行R峰检测:对原始ECG去均值后,为突出QRS复合波并抑制低频漂移与高频噪声,采用二阶Butterworth带通滤波(5~15 Hz),并对滤波后信号进行平方能量增强。R峰定位使用基于阈值与最小间隔约束的峰值搜索方法,其中阈值取能量序列的90%分位数,最小峰间距设为0.25 s,以避免重复检出。基于检测到的R峰样本点序列 p_i 及采样率为 f_s ,构建RR/NN序列。RR间期(s)、NN间期(ms)、相邻NN差分(ms)的定义分别为

$$RR_i = \frac{p_{i+1} - p_i}{f_s} \quad i = 1, 2, \dots, N-1 \quad (1)$$

$$NN_i = 1\,000 \cdot RR_i \quad (2)$$

$$\Delta NN_i = NN_{i+1} - NN_i \quad i = 1, 2, \dots, N-2 \quad (3)$$

在此基础上,从每个条件片段计算8个ECG特征(5个时域以及3个频域):平均心率(Mean heart rate, Mean HR),反映整体心率水平,是交感和副交感共同作用的结果,在抑郁或高压力状态下静息心率有升高的趋势;NN间期标准差(Standard deviation of NN intervals, SDNN),表征整体HRV总量,较低SDNN通常被解读为整体自主神经灵活性下降或应激负荷增加的信号;相差超过50 ms的连续NN间期百分比(pNN50),相邻心搏间期差异超过50 ms的比例,对快速心率波动敏感,常被作为副交感活动的一个近似指标。pNN50降低常被用作迷走调节减弱的证据之一;连续差值均方根(Root mean square of successive difference, RMSSD)和连续差值标准差(Standard deviation of successive difference, SDSD),二者主要反映短时HRV,若SDSD与RMSSD一致降低,通常支持短时迷走调节减弱的解释;低频(Low frequency, LF)0.04~0.15 Hz、高频(High frequency, HF)0.15~0.4 Hz、低/高频比值(LF/HF)三个频域特征,HF常与呼吸相关迷走调制联系更紧密,而LF往往反映交感-副交感混合调制,它们的比值是自主神经平衡的参考量之一。

之所以选择以上指标,是因为这类特征在短时HRV分析中应用成熟,与抑郁相关的HRV下降、自主神经失衡等现象具有较明确的生理学对应关系。相较于大量高维非线性特征或复杂组合特征,上述

经典HRV特征对便携式ECG短时记录更稳健,维度更紧凑,也更便于后续临床解释与低资源场景部署。

EEG为单通道额区信号,其特征主要基于功率谱密度(Power spectral density, PSD)。在每个任务条件片段内,先进行去均值处理以去除直流偏置,然后采用Welch方法估计功率谱密度。Welch参数设置与实现保持一致:窗长2 s,窗重叠50%,并采用常数项去趋势。在此基础上,对预设频段的功率谱密度进行积分以获得频带功率(Band power, BP)特征,即

$$BP_b = \int_{f_{\min}}^{f_{\max}} S(f) df \quad (4)$$

本文采用5个经典频带划分: δ (1~4 Hz)、 θ (4~8 Hz)、 α (8~13 Hz)、 β (13~30 Hz)、 γ (30~45 Hz)。一方面,这5个频带与抑郁研究中的节律异常、唤醒水平、注意控制及情绪加工差异具有较清晰的对应关系;另一方面,频带功率特征维度低、计算开销小、对便携式设备噪声条件相对稳健,更契合真实世界部署需求。

2.2 多模态特征融合

本文基于Transformer^[18]和CLS^[19],提出了一种双层级特征融合框架P-MFCNet(Portable multimodal fusion with CLS network)。该框架面向便携式单通道EEG与可穿戴ECG的多任务采集场景,分别从模态维度(EEG/ECG)与任务态维度(EO/Neg/EC/Interview)学习互补信息与条件差异,并通过层次化CLS汇聚获得最终的样本级融合表示,用于后续分类器判别MDD与HC。算法框架如图2所示。

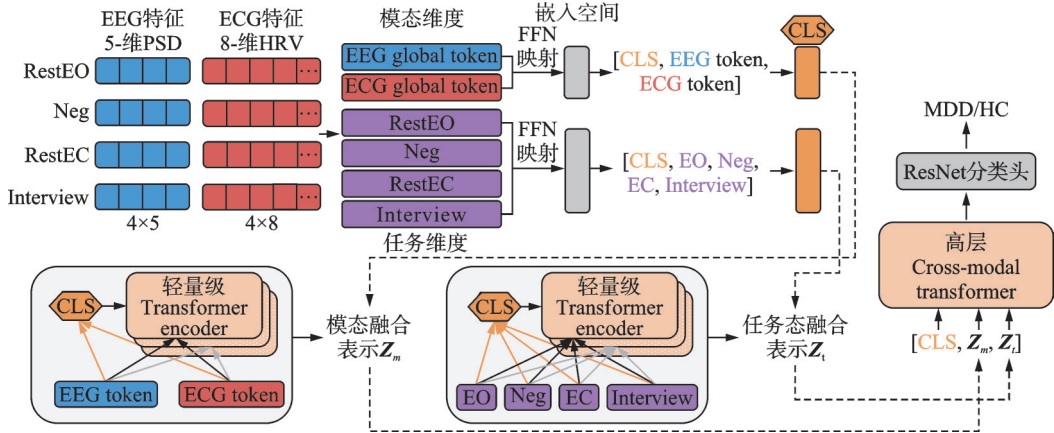


图2 P-MFCNet框架图

Fig.2 Framework diagram of P-MFCNet

具体来说,记4个任务维度 $t \in \{0, 1, 2, 3\}$ (分别对应EO、Neg、EC、Interview),对每个受试者,在每个任务条件下提取EEG频域特征(5维)与ECG特征(8维),从而得到两模态的条件级特征矩阵

$$E = \begin{bmatrix} e_0^T \\ e_1^T \\ e_2^T \\ e_3^T \end{bmatrix} \in \mathbb{R}^{4 \times 5}, C = \begin{bmatrix} c_0^T \\ c_1^T \\ c_2^T \\ c_3^T \end{bmatrix} \in \mathbb{R}^{4 \times 8} \quad (5)$$

式中 e_i 和 c_i 分别表示第 i 个任务条件下EEG与ECG的特征向量。为在任务条件粒度上形成跨模态表征,将同一任务条件下的EEG与ECG特征进行拼接,得到条件级跨模态特征

$$x_i = [e_i; c_i] \in \mathbb{R}^{13} \quad i = 0, 1, 2, 3 \quad (6)$$

综上所述,P-MFCNet的输入既包含“按模态组织”的特征集合(E, C),也包含“按任务组织”的条件

级跨模态特征集合 $\{x_i\}$ 。基于这一结构, P-MFCNet 在低层级分别建模两类互补信息: 一方面学习 EEG 与 ECG 的模态互补性, 即模态维度融合; 另一方面学习不同任务条件对判别的差异化贡献, 即任务态维度融合。最终在高层级进一步对两条分支的输出进行层次化融合, 得到样本级全局表征。

2.2.1 模态维度融合

模态维度融合的目标是在模态层面学习 EEG 与 ECG 的互补贡献, 并生成模态融合表示。考虑到两模态均包含 4 个任务条件, 为形成模态级的全局 token, 先将每个模态在任务维度上按行展开 (Vectorization) 为一维向量, 即

$$\bar{e} = \text{vec}(E) \in \mathbf{R}^{20}, \bar{c} = \text{vec}(C) \in \mathbf{R}^{32} \quad (7)$$

为消除维度不一致并增强表达能力, 本文采用两层多层感知机 (Multilayer perceptron, MLP) 的非线性映射 $\phi(\cdot)$ 将两者投影到同一嵌入维度 d , 以获得可比较、可交互的 token 表示, 即

$$\phi(u) = \sigma(uW_1 + b_1)W_2 + b_2, \mathbf{R}^{\phi(\bar{e})} = \mathbf{R}^{\phi(\bar{c})} = \mathbf{R}^d \quad (8)$$

在此基础上, 构造由 3 个 token 组成的序列 $[\text{CLS}^{\text{mod}}, \phi(\bar{e}), \phi(\bar{c})]$, 并输入轻量级 Transformer encoder 执行多头自注意力与前馈变换。由于该序列仅包含 EEG token 与 ECG token 两个信息源, CLS token 在自注意力机制下可自适应地聚合两模态证据, 从而显式建模 EEG-ECG 的交互关系。最终以编码器最后一层的 CLS 隐向量作为模态融合表示 h_{mod} , 用于表征跨任务汇总后的模态互补信息。

2.2.2 任务态维度融合

任务态维度融合的目标是在任务条件层面学习不同条件对判别任务的差异化贡献。与模态分支不同, 该分支以条件级跨模态向量 $x_i \in \mathbf{R}^{13}$, 作为基本单元。为增强表达并统一维度, 将每个条件 token 经两层 MLP 做非线性投影

$$\phi(x_i) = \sigma(x_iW_1 + b_1)W_2 + b_2 \quad \phi(x_i) \in \mathbf{R}^d \quad (9)$$

构造任务维度 5-token 序列 $[\text{CLS}^{\text{task}}, \phi(x_0), \phi(x_1), \phi(x_2), \phi(x_3)]$ 并输入轻量级 Transformer encoder。该编码器通过自注意力机制学习 4 种条件之间的相关性与互补性, 使 CLS token 能够对条件 token 进行自适应加权汇聚, 从而得到任务态融合表示 h_{task} , 该表示在语义上对应于跨条件的判别性摘要, 可用于分析 4 个任务条件对模型决策的相对重要性。

2.2.3 层次融合

在得到两条低层分支输出后, P-MFCNet 在高层级进一步融合 h_{mod} 和 h_{task} , 以同时保留模态互补信息和任务差异信息。具体地, 将 h_{mod} 和 h_{task} 视为两个高层 token 输入高层级编码器, 以最后一层 CLS 向量作为样本级融合表示 h_p , 该表示随后输入分类器进行 MDD/HC 判别。通过这一层的融合, 模型避免了简单拼接造成的冗余与尺度不一致问题, 并可在注意力空间中显式学习“模态-任务”两种维度的贡献分配, 从而提升融合表示的鲁棒性与可解释性。

2.3 ResNet 分类器

在 P-MFCNet 框架中, 最终通过层次化 CLS 融合得到的模态融合表示 h_p 被用于后续的分类任务。为进一步提升模型的性能, 并利用深度神经网络的表达能力, 本文采用残差网络 (Residual network, ResNet)^[20] 作为分类器。ResNet 是一种深度卷积神经网络架构, 其核心特点是引入了残差连接 (Skip connections), 以缓解深度神经网络训练中的梯度消失问题, 并提升网络的训练效率和泛化能力。

在本文框架中, ResNet 分类器主要负责将融合后的特征向量映射到 MDD 与 HC 的类别标签上。使用交叉熵损失函数 (Cross-entropy loss) 作为模型的优化目标, 即

$$L = - \sum_{i=1}^2 y_i \log \hat{y}_i \quad (10)$$

式中: y_i 为真实标签, $\hat{y}_i$ 为模型预测的概率, i 为类别索引。交叉熵损失函数能够度量预测概率与真实标签之间的差距,并通过梯度下降优化网络参数。

3 结果与分析

3.1 训练流程

本实验基于Pytorch框架进行模型训练,并在NVIDIA 4060T GPU上执行。采用Adam优化器,设置学习率为 $5e-4$,权重衰减为 $1e-4$,以帮助网络在训练过程中逐渐收敛。为系统评估模型的内部泛化性能及跨中心泛化能力,本文首先在杭七数据集上采用五折交叉验证(5-fold cross-validation)进行模型训练与验证。具体而言,将数据集划分为5个互斥子集,每次选取其中1个子集作为验证集,其余4个子集作为训练集,共进行5轮训练。每一折均训练70个epoch,并采用准确率(Accuracy, ACC)、受试者工作特征曲线下面积(Area under the curve, AUC)、灵敏度(Sensitivity, SEN)和特异度(Specificity, SPE)作为性能评价指标。其中,灵敏度用于衡量模型对MDD患者的识别能力,特异度用于衡量模型对健康对照个体的区分能力,从而更全面地反映模型在临床筛查场景下的应用价值。训练过程中依据验证集综合表现保存各折最优模型权重。最后,采用省中数据集作为外部测试集,对模型的跨中心泛化能力进行进一步验证。

3.2 单中心验证

首先,对构建的多模态P-MFCNet框架进行系统评估,结果如图3所示。在五折交叉验证设置下,汇总了每一折验证集的性能:平均ACC为 0.8424 ± 0.1058 ,AUC为 0.7967 ± 0.1376 ,SEN为 0.8467 ± 0.0869 ,SPE为 0.8333 ± 0.2041 。可见,该模型在不同折次下均保持了较为稳定的分类性能,说明所构建框架具有一定的鲁棒性和泛化能力。与此同时,模型在灵敏度和特异度两个指标上均取得了较高水平,表明其不仅能够有效识别MDD患者,还能够较准确地区分健康对照,整体误判较少。上述结果进一步说明,所提多模态融合策略能够有效挖掘EEG与ECG信号之间的互补信息,提升对抑郁相关病理特征的识别能力,从而在真实病例数据上表现出较好的诊断应用价值。

其次,为验证各模态对模型性能的贡献,本文开展了单模态消融实验(其余训练设置完全一致)。以EEG单模态为例:在完成特征工程后,将4个任务条件下的EEG特征按任务维度组织为token序列,输入编码器完成任务态汇聚,最终由ResNet分类头输出类别预测;ECG单模态采用相同流程,仅将输入替换为ECG特征。结果如图4所示:EEG单模态分类性能为 $ACC=0.7167$, $AUC=0.6600$, $SEN=0.6427$, $SPE=0.7624$;ECG单模态为 $ACC=0.7530$, $AUC=0.7644$, $SEN=0.6933$, $SPE=0.7814$ 。与单模态相比,多模态P-MFCNet的性能获得显著提升:相较EEG单模态,ACC提升0.1257(相对提升17.54%)、AUC提升0.1367(相对提升20.71%);相较ECG单模态,ACC提升0.0894(相对提升11.87%)、AUC提升0.0323(相对提升4.23%)。在灵敏度和特异度指标上,多模态维持了较高水平,且二者数值更为接近,说明其在MDD样本识别能力与HC样本区分能力之间取得了较好的平衡,体现出更稳定、更全面的判别效果。相比之下,单模态模型虽然在部分指标上仍具有一定识别能力,但SEN与SPE之间存在一定程度的不均衡,反映出单一生理信号在表征抑郁相关异常时仍具有局限性,难以同时兼顾对患者样本的敏感检出和对健康对照的准确排除。

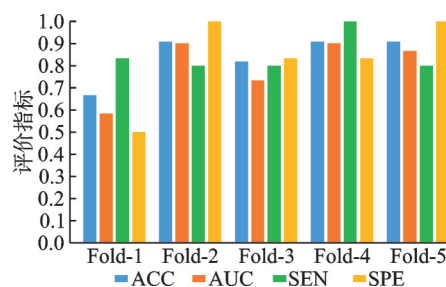


图3 P-MFCNet算法五折交叉验证结果
Fig.3 Result of the 5-fold cross-validation of P-MFCNet algorithm

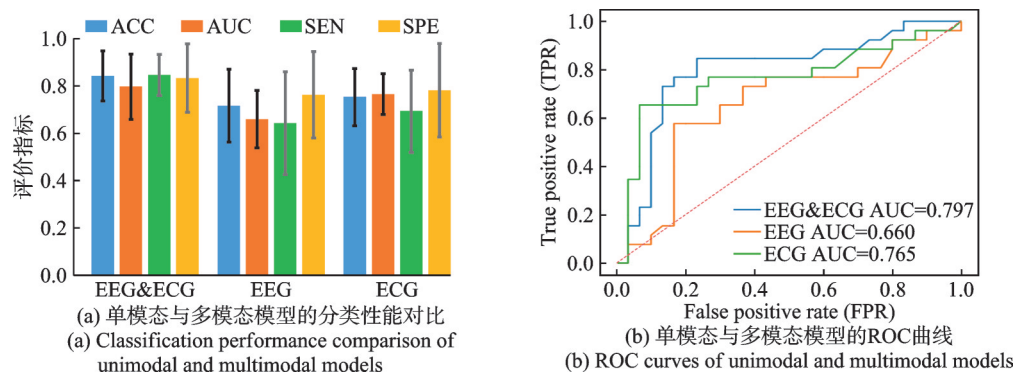


图4 模态消融实验
Fig.4 Modal ablation experiment

最后,选取 EEG 和 ECG 领域常用的卷积神经网络(CNN)与 MLP 作为基线模型,以及近两年发表的抑郁症检测模型 MDD-AttNet^[21]和 TabPFN^[22]作为对比模型。为保证对比公平,基线方法同样采用一致的特征工程流程,并将两模态特征向量直接拼接后输入分类器进行训练与测试。对比结果汇总见表 2。与传统基线模型相比,P-MFCNet 的优势更为明显。相较 CNN,模型 ACC 由 0.700 0 提升至 0.842 4, AUC 由 0.623 3 提升至 0.796 7,说明简单卷积结构难以充分建模多模态特征间的高阶关联;相较 MLP, P-MFCNet 同样在 ACC、AUC 和 SEN 等指标上均有明显提升,表明仅依赖浅层全连接映射难以有效挖掘 EEG 与 ECG 之间的互补信息。与近年来提出的 MDD-AttNet 和 TabPFN 相比,P-MFCNet 虽然在个别单项指标上并非全部最优,但在准确率这一综合指标上取得了最佳结果,并在灵敏度与特异度之间保持了相对均衡的表现,说明其不仅具备较好的整体分类能力,而且在误检与漏检之间取得了更合理的折中。上述结果表明,P-MFCNet 的性能提升并非单纯来源于多模态特征拼接,而是得益于基于 CLS token 的层次化融合机制对模态互补信息与任务态差异的显式建模,从而增强了对抑郁相关判别特征的表征能力。

表 2 单中心对比试验结果
Table 2 Results of single-center comparative trials

模型	ACC	AUC	SEN	SPE
CNN	0.700 0 ± 0.134 8	0.623 3 ± 0.191 3	0.400 0 ± 0.316 2	0.966 7 ± 0.074 5
MLP	0.736 4 ± 0.152 1	0.650 0 ± 0.216 7	0.673 3 ± 0.316 6	0.800 0 ± 0.074 5
MDD-AttNet	0.814 8 ± 0.140 7	0.760 8 ± 0.154 9	0.672 9 ± 0.241 8	0.886 4 ± 0.172 8
TabPFN	0.820 9 ± 0.094 9	0.807 6 ± 0.084 6	0.884 7 ± 0.123 7	0.764 8 ± 0.157 1
P-MFCNet	0.842 4 ± 0.105 8	0.796 7 ± 0.137 6	0.856 7 ± 0.086 9	0.843 3 ± 0.204 1

需要指出的是,尽管 P-MFCNet 在五折交叉验证中取得了较好的整体性能,仍存在少量误分类样本。结合本研究的采集方式与任务设计,我们推测,这些误判可能与以下因素有关:(1) 部分受试者在静息或访谈过程中存在注意波动、疲劳变化或情绪反应不稳定,导致 EEG 节律和 HRV 指标出现偏移;(2) 便携式单通道 EEG 与可穿戴 ECG 在空间分辨率、抗噪能力和佩戴一致性方面仍存在一定局限,可能削弱边界样本的关键判别信息;(3) 抑郁症具有较强异质性,部分轻症、恢复期或共病特征复杂的样本,可能与健康对照呈现相近的生理模式。

3.3 跨中心测试

在完成杭七数据集五折交叉验证后,本文进一步采用省中数据集进行跨中心外部测试。具体而言,将五折训练过程中依据验证集表现保存的最优模型分别用于省中数据集预测,并使用各折对应的标准化参数处理外部样本,最终对5次测试结果求均值与标准差,结果汇总见表3。与此同时,本文进一步给出了外部测试集上的混淆矩阵,如图5所示。由于省中数据集中HC与MDD样本数量基本均衡,混淆矩阵能够更直观地揭示模型是否存在对某一类别的偏置。

表3 跨中心对比试验结果
Table 3 Results of cross-center comparative trials

模型	ACC	AUC	SEN	SPE
MDD-AttNet	0.756 0 ± 0.065 2	0.835 0 ± 0.060 2	0.734 8 ± 0.146 8	0.786 7 ± 0.124 8
TabPFN	0.772 4 ± 0.044 7	0.845 0 ± 0.037 0	0.805 0 ± 0.106 7	0.780 4 ± 0.094 7
P-MFCNet	0.810 0 ± 0.089 4	0.875 0 ± 0.036 9	0.865 0 ± 0.096 2	0.830 0 ± 0.120 0

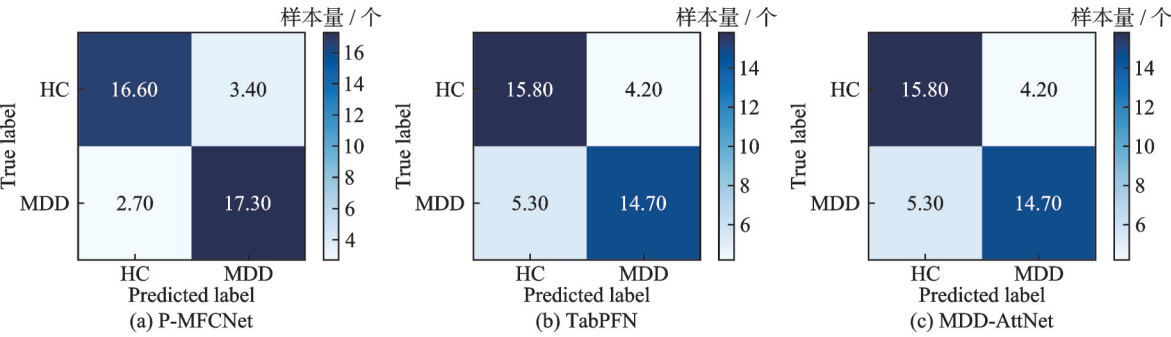


图5 混淆矩阵结果

Fig.5 Confusion matrix results

综合表3和图5的结果可知,P-MFCNet在跨中心测试中仍取得了最佳综合表现,这表明所提模型在外部独立中心数据上仍能保持较好的判别能力,其学习到的并非仅是单中心样本的特异性模式,而是具有一定跨中心稳定性的抑郁相关表征。尤其在AUC和SEN两项指标上的领先,说明P-MFCNet在中心分布发生变化时,仍能够较好地地区分MDD与HC,并保持较强的患者检出能力。

3.4 可解释性分析

为增强模型决策过程的可解释性,本文采用两类互补方法:(1) 基于置零消融(Zero-out ablation)的特征贡献度分析,用于评估不同任务-特征分量对模型性能的影响;(2) 基于Transformer注意力权重的可视化分析,用于考察模型在模态维度、任务态维度及高层融合层面的信息聚合偏好。二者分别从输入贡献和内部表示两个角度提供解释证据,从而增强对P-MFCNet决策机制的理解。本文所采用的可解释性方法均属于事后可解释(Post-hoc interpretability)范式,即在模型训练完成后,通过分析模型内部的注意力分配模式以及输入分量被移除后对分类性能的影响,来解释模型可能依赖的判别依据。

对于置零消融分析,在保持网络结构与输入维度不变的前提下,逐一将某一任务-特征态分量置零(其余分量保持不变),在验证集上重新计算分类准确率。将置零前后的准确率差值定义为该分量的重要性评分,即

$$\Delta ACC_k = ACC_{full} - ACC_{(-k)} \tag{11}$$

ΔACC 越大,表示该分量被移除后性能下降越明显,对模型越关键。取 Top-10 结果如图 6 所示。整体上来看,Top-10 的特征贡献度集中在 7%~13% 之间,说明模型并非依赖单一指标,而是多个任务态下的关键生理特征共同支撑。

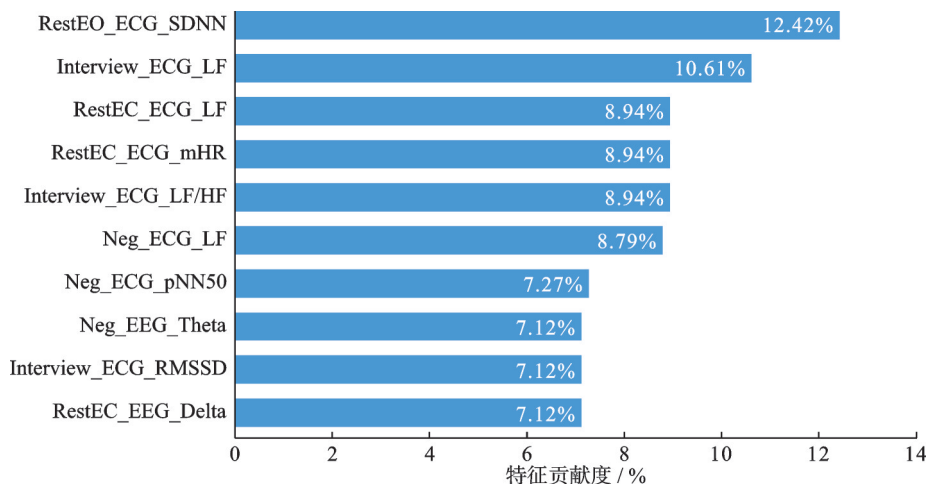


图 6 特征贡献度

Fig.6 Feature contribution degree

对于注意力权重的可视化分析,分别统计模态分支中 CLS token 对 EEG token 和 ECG token 的注意力权重,任务分支中 CLS token 对 EO、Neg、EC 和 Interview 4 个任务 token 的注意力权重,以及高层融合中 CLS token 对模态融合分支和任务融合分支的注意力权重,并在五折交叉验证的验证集上分别对 HC 和 MDD 样本求平均,结果如图 7 所示。误差线表示五折交叉验证结果的标准差。总体来看,不同折次下的注意力分配模式较为一致。

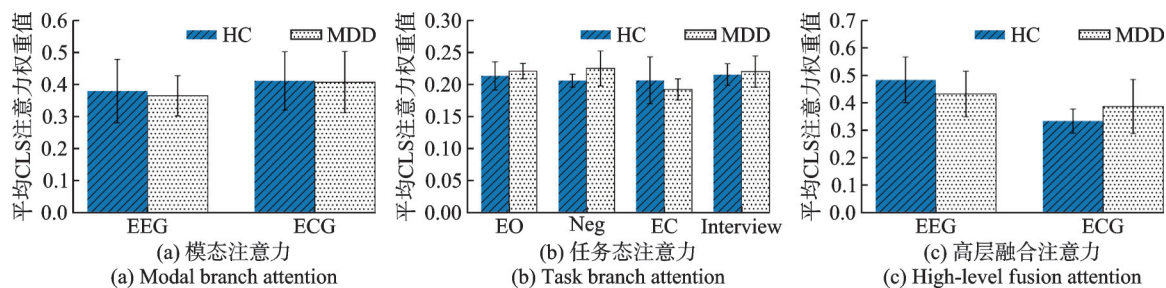


图 7 注意力权重可视化

Fig.7 Visualization of attention weights

3.4.1 模态维度

从模态维度来看,注意力可视化与特征贡献度分析呈现出较一致的趋势。图 7 显示,无论在 HC 还是 MDD 中,模态分支内 CLS token 对 ECG token 的关注均略高于 EEG token,说明模型在进行跨任务信息汇聚时,整体上更倾向优先整合 ECG/HRV 相关线索。这一结果与图 6 中 Top-10 关键分量多数来自 ECG 模态的现象相一致,也与前述单模态实验中 ECG 性能优于 EEG 的结果相互印证,提示自主神经调节异常可能是当前便携式筛查框架下更稳定的判别依据。同时,图 7 也表明 EEG 分支并未被模型忽略。尽管其注意力权重整体低于 ECG,但在两组样本中仍保持较稳定水平,说明 P-MFCNet 并未退

化为单纯依赖心电信号的分类器。这一点与图6中Neg_EEG_Theta和RestEC_EEG_Delta仍进入高贡献分量的结果相一致。

从高层融合结果进一步看,HC和MDD两组样本对模态融合分支的注意力均高于任务融合分支,说明模型最终判别首先依赖于跨模态互补表示本身,而非仅依赖某一单独任务条件。与此同时,MDD组中任务分支与模态分支之间的差距相对缩小,说明抑郁样本的识别除依赖总体模态互补外,还更依赖不同任务状态下生理反应模式的变化。

3.4.2 任务态维度

从任务态维度来看,图7显示HC和MDD两组样本的注意力并未集中于单一任务,而是在4种任务条件之间呈现相对分散但具有差异的分布特征,说明模型对抑郁识别依赖的是多情境信息的联合积累。对于HC组,EO与Interview获得的注意力相对更高,而Neg与EC较为接近,提示健康对照样本的判别更多依赖较稳定的基线状态以及访谈情境下整体生理调节的一致性。相比之下,MDD组在Neg、EO和Interview上的注意力相对更高,而EC相对较低,表明模型在抑郁样本中更关注负性情绪诱发和访谈等更容易激发情绪加工、社会认知或应激调节差异的任务条件。

这一结果与图6中的特征贡献分析能够形成较好的对应关系。一方面,RestEO_ECG_SDNN作为全局贡献度最高的特征,说明睁眼静息状态下的基线自主神经差异具有较强判别力,这与图7中EO在两组样本中均保持较高注意力相一致;另一方面,Neg相关的脑电和心电特征、Interview相关的HRV频域及迷走调节指标在图6中均表现出较高贡献度,而图7中MDD组对Neg和Interview的关注也相对更强,说明模型在抑郁样本中确实更依赖这些能够放大任务反应性差异的情境。

3.5 实验结论

综合上述实验结果可以得出如下结论:

(1)便携式EEG-ECG多模态融合能够有效提升MDD/HC识别性能。相较于EEG或ECG单模态模型,P-MFCNet在单中心五折交叉验证中取得了更优且更稳定的分类结果,说明中枢神经活动信息与自主神经调节信息具有明显的互补性,多模态联合建模能够更全面地表征抑郁相关异常。与此同时,结合折间结果可见,模型性能仍存在一定波动,提示其在小样本和数据异质性条件下的稳定性仍有进一步提升空间。

(2)所设计的多任务实验范式有助于增强模型的判别能力。不同任务条件能够从基线状态和任务反应性两个层面提供互补信息,使模型不仅能够捕捉相对稳定的生理差异,还能够利用情绪加工与交互情境下的状态变化,从而提升筛查效果。

(3)基于CLS的层次化融合策略优于简单特征拼接。在相同特征工程前提下,P-MFCNet在准确率、灵敏度及整体平衡性等方面表现较优,表明其性能提升不仅来源于多模态输入本身,更得益于模型对模态互补信息和任务态差异的显式建模,能够学习到更具判别性的融合表示。

(4)模型具有一定的跨中心泛化能力,并具备初步可解释性。在外部中心测试中,P-MFCNet仍保持了较好的分类性能,说明其学习到的表征并非完全局限于单中心样本的特异性模式,而在一定程度上具有跨中心可迁移性。同时,结合注意力可视化与特征消融分析可见,模型整体上更关注ECG相关信息,EEG则在特定任务条件下提供补充证据,表明P-MFCNet能够从模态维度和任务维度自适应整合多源生理线索,并为模型决策提供一定的解释依据。

4 结束语

本文面向真实世界可扩展筛查需求,围绕便携式单通道EEG与可穿戴ECG构建了“数据-算法-验证-解释”一体化研究框架。首先,基于精神科临床评估、结构化访谈及标准化量表所构成的综合临床判

断,设计了包含睁眼静息、负性情绪诱发、闭眼静息及半结构化访谈的多任务实验范式,并采集了同步多模态电生理数据。其次,在统一的预处理与特征工程基础上,提出了基于 Transformer 与 CLS 的双层级融合模型 P-MFCNet,分别在模态维度与任务态维度进行自适应信息汇聚,并通过层次化 CLS 获得样本级融合表示,实现对 MDD 与 HC 的判别学习。实验结果表明,P-MFCNet 在单中心五折交叉验证中取得了较好的总体分类性能,并在外部中心测试中表现出一定的泛化潜力,验证了便携式 EEG-ECG 方案用于快速抑郁筛查的可行性。进一步地,本文结合注意力可视化与置零消融分析,为后续临床解释、特征优化及模型改进提供了参考依据。

尽管本文取得了较为积极的结果,仍存在若干限制。首先,本研究总体样本量仍然较小,且训练与验证主要基于单中心数据,折间性能存在一定波动,尚需在更大样本、多中心队列中进一步检验模型的稳健性与泛化性。其次,本文虽然采用手工特征与轻量级 Transformer 相结合的设计,以兼顾性能与解释性,但层次化 Transformer 与 ResNet 分类头仍增加了参数量、存储开销和推理计算量,这可能限制其在低算力、低功耗设备上的直接部署。未来可进一步探索特征压缩、模型剪枝、知识蒸馏和轻量化注意力模块,以在尽量保持性能的同时降低模型复杂度。最后,本文目前聚焦 MDD 与 HC 二分类,尚未覆盖与 MDD 鉴别诊断高度相关的其他精神障碍类别,如双相情感障碍和精神分裂症等。未来,我们计划在不同医院与社区场景开展更大样本、多中心验证,评估模型在设备差异、环境噪声与人群异质性下的稳健性,并探索域自适应迁移学习策略。

参考文献:

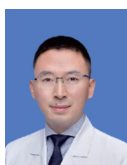
- [1] PRIZEMAN K, MCCABE C. Anhedonia and its sub-component processes predict clinically significant symptoms of Major Depressive Disorder (MDD) and loneliness in young people[J]. *British Journal of Clinical Psychology*, 2026, 65(1): 143-159.
- [2] WU C, MU Q, GAO W, et al. The characteristics of anhedonia in depression: A review from a clinically oriented perspective [J]. *Translational Psychiatry*, 2025, 15(1): 90.
- [3] RICHARDSON E, PATTERSON R, MELTZER-BRODY S, et al. Transformative therapies for depression: Postpartum depression, major depressive disorder, and treatment-resistant depression[J]. *Annual Review of Medicine*, 2025, 76: 81-93.
- [4] CHEN S, SUN H, SONG Y, et al. Transition and trend analysis of the burden of depression in China and different income countries: Based on GBD database and joinpoint regression model[J]. *Journal of Affective Disorders*, 2024, 362: 437-449.
- [5] 张学军, 刘济玮, 李夏芸. 基于注意力机制的双通道多特征卷积脑电信号分类[J/OL]. 数据采集与处理, 1-13[2026-01-14]. <https://link.cnki.net/urlid/32.1367.TN.20250925.0939.002>.
ZHANG Xuejun, LIU Jiwei, LI Xiayun. Dual-channel multi-feature convolutional EEG signal classification based on attention mechanism[J/OL]. *Journal of Data Acquisition and Processing*, 1-13[2026-01-14]. <https://link.cnki.net/urlid/32.1367.TN.20250925.0939.002>.
- [6] ZOU B, HAN J, WANG Y, et al. Semi-structural interview-based Chinese multimodal depression corpus towards automatic preliminary screening of depressive disorders[J]. *IEEE Transactions on Affective Computing*, 2022, 14(4): 2823-2838.
- [7] HABIB A, VANIYA SN, KHANDOKER A, et al. MDDBranchNet: A deep learning model for detection major depressive disorder using ECG signal[J]. *IEEE Journal of Biomedical and Health Informatics*, 2024, 28(7): 3798-3809.
- [8] 张学军, 王天晨, 王泽田. 基于多域信息融合的卷积 Transformer 脑电情感识别模型[J]. 数据采集与处理, 2024, 39(6): 1543-1552.
ZHANG Xuejun, WANG Tianchen, WANG Zetian. Convolutional Transformer EEG emotion recognition model based on multidomain information fusion[J]. *Journal of Data Acquisition and Processing*, 2024, 39(6): 1543-1552.
- [9] ELNAGGAR K, EL-GAYAR M M, ELMOGY M. Depression detection and diagnosis based on electroencephalogram (EEG) analysis: A systematic review[J]. *Diagnostics*, 2025, 15(2): 210.
- [10] ZITOUNI M S, LIH OH S, VICNESH J, et al. Automated recognition of major depressive disorder from cardiovascular and respiratory physiological signals[J]. *Frontiers in Psychiatry*, 2022, 13: 970993.

- [11] QAYYUM A, RAZZAK I, TANVEER M, et al. High-density electroencephalography and speech signal based deep framework for clinical depression diagnosis[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023, 20(4): 2587-2597.
- [12] ZHENG W, YAN L, WANG F Y. Two birds with one stone: Knowledge-embedded temporal convolutional transformer for depression detection and emotion recognition[J]. IEEE Transactions on Affective Computing, 2023, 14(4): 2595-2613.
- [13] ZHANG H, ZHOU H, LIAO Q, et al. Abnormal nonlinear brain-heart interactions in major depressive disorder: Evidence from resting-state EEG and ECG[J]. Clinical Neurophysiology, 2025: 2111371.
- [14] 徐庆,李鸿燕,梁卓,等.基于TF-RMN的多模态抑郁症检测模型[J].计算机仿真, 2025, 42(6): 564-569.
XU Qing, LI Hongyan, LIANG Zhuo, et al. A multimodal depression detection model based on TF-RMN[J]. Computer Simulation, 2025, 42(6): 564-569.
- [15] 王肖文,肖碧波,陈旭创.基于便携式睡眠心电采集及深度学习的抑郁症患者睡眠质量评估系统[J].医疗装备, 2022, 35(1): 28-29.
- [16] FIRST M B. Structured clinical interview for the DSM (SCID)[J]. The Encyclopedia of Clinical Psychology. DOI: 10.1002/9781118625392.wbecp351.
- [17] SLOBODSKOY-PLUSNIN J. Behavioral and brain oscillatory correlates of affective processing in subclinical depression[J]. Journal of Clinical and Experimental Neuropsychology, 2018, 40(5): 437-448.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 6000-6010.
- [19] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association For Computational Linguistics: Human Language Technologies, volume 1 (long and short papers). [S.l.]: [s.n.], 2019: 4171-4186.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 770-778.
- [21] CHEN Y, DU H, WANG H, et al. A lightweight variable-channel deep self-attention model for depression classification based on raw EEG signals[J]. Biomedical Signal Processing and Control, 2026, 113: 109227.
- [22] GUO X, ZHENG Z, DAI A, et al. Tabular prior-data fitted networks for objective depression screening with portable EEG and ECG[J]. Biomedical Signal Processing and Control, 2026, 117: 109598.

作者简介:



应伧扬(2004-),男,本科生,研究方向:中医学、精神疾病, E-mail: 2148966185@qq.com。



裘生梁(1978-),男,主任中医师、硕士生导师,研究方向:呼吸、消化、内分泌等疾病。



陈熙昂(2000-),男,博士研究生,研究方向:精神疾病脑电解码。



江海腾(1985-),通信作者,男,研究员,博士生导师,研究方向:医学大数据的人工智能、神经工程、计算神经科学、脑科学与认知神经科学, E-mail: h.jiang@zju.edu.cn。

(编辑:王静)

A Hierarchical Fusion Method for Intelligent Depression Detection Based on Portable EEG and ECG

YING Yiyang¹, QIU Shengliang^{1,2}, CHEN Xi'ang³, JIANG Haiteng^{4*}

(1. The First Clinical Medical College, Zhejiang Chinese Medical University, Hangzhou 310053, China; 2. Zhejiang Provincial Hospital of Traditional Chinese Medicine (The First Affiliated Hospital of Zhejiang Chinese Medical University), Hangzhou 310006, China; 3. College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou 310027, China; 4. MOE Frontier Science Center for Brain Science and Brain-Machine Integration, Zhejiang University, Hangzhou 310058, China)

Abstract: Major depressive disorder (MDD) is a prevalent psychiatric disorder. Conventional screening and diagnosis mainly rely on structured interviews and rating scales, whose subjectivity and limited accessibility hinder early identification in primary care and community settings. To address this problem, this study develops and validates P-MFCNet, a portable multimodal framework for depression recognition based on single-channel electroencephalography (EEG) and wearable electrocardiography (ECG). A multi-task protocol comprising eyes-open rest, negative emotion elicitation, eyes-closed rest, and a semi-structured clinical interview was designed, and synchronized multimodal data were collected at two medical centers. P-MFCNet employs a hierarchical fusion strategy based on Transformer encoders and classification tokens (CLS) to model complementary information across modalities and differences among task states. Modality-level and task-level representations are learned separately and integrated through hierarchical CLS aggregation to generate a subject-level representation for MDD-versus-HC classification using a ResNet head. In single-center five-fold cross-validation, P-MFCNet achieves an accuracy of 0.842 4 and an AUC of 0.796 7. In cross-center external testing, it achieves an accuracy of 0.810 0 and an AUC of 0.875 0. The model outperforms unimodal EEG and ECG models in accuracy, sensitivity, and overall performance, while maintaining a favorable balance between sensitivity and specificity compared with baseline methods. Post-hoc analyses using zero-out ablation and attention visualization show that ECG/HRV features provide discriminative evidence, whereas EEG contributes complementary information under negative emotion elicitation and resting conditions. These findings demonstrate the feasibility of portable EEG-ECG fusion for depression screening and indicate cross-center generalizability and interpretability.

Highlights:

1. A portable single-channel EEG and wearable ECG framework is developed for scalable and accessible MDD screening.
2. Hierarchical CLS-based fusion effectively integrates complementary modality information and task-specific physiological responses.
3. P-MFCNet achieves strong cross-center performance and provides interpretable evidence from ECG/HRV and EEG features.

Key words: major depressive disorder (MDD); portable EEG; wearable ECG; multimodal fusion

Foundation items: Brain Science and Brain-like Intelligence Technology—National Science and Technology Major Project (No. 2022ZD0212400); College Students' Innovative Entrepreneurial Training Plan Program (No.202510344031).

Received: 2026-02-06; **Revised:** 2026-07-06

***Corresponding author, E-mail:** h.jiang@zju.edu.cn.