

DOI:10.20033/j.1003-7241.(2026)03-0099-05

基于脑电信号的多维融合情绪识别模型

曹雅茜

(上海电子信息职业技术学院,上海 201411)

摘要:在进行情绪识别任务时,脑电信号来源于与情绪密切相关的中枢神经系统,与个体情绪具有较强的关联度,具有时-频-空多维度信息。但其信息维度复杂、易被噪声干扰,给情绪识别的准确性和可靠性带来挑战。为了充分挖掘情绪深层次的细节信息,提出一种融合胶囊网络、注意力机制与时序卷积网络的脑电情绪识别模型,首先提取脑电信号的微分熵特征,该特征可以反映不同频段脑电活动的能量分布,结合频域与空间域信息映射为多频带特征矩阵,其次通过胶囊网络获取导联的空间信息,并利用注意力机制给予不同脑区不同的权重,使得分类胶囊聚焦关键导联,然后使用时序卷积网络对时序动态特征进行学习,捕捉情绪随时间动态变化的特征,最终获取情绪分类结果。实验结果表明,在 DEAP 数据集上的效价-唤醒度二分类准确率为 94.97%、95.28%,四分类准确率为 94.62%,相较于单一维度,所提融合模型可以有效提升情绪识别的精度。

关键词:情绪识别;脑电信号;胶囊网络;时序卷积网络;特征提取;特征融合

中图分类号: TP391 文献标志码: A 文章编号: 1003-7241(2026)03-0099-05

Multidimensional fusion emotion recognition model based on EEG signals

CAO Yaxi

(Shanghai Technical Institute of Electronics & Information, Shanghai 201411, China)

Abstract: In the task of emotion recognition, EEG signals are derived from the central nervous system which is closely associated with emotions, exhibit a strong correlation with individual emotional states, and possess time-frequency-spatial multi-dimensional information. However, their complex information dimensions and susceptibility to noise interference pose challenges to the accuracy and reliability of emotion recognition. To fully explore the deep-seated details of emotions, it proposes an EEG emotion recognition model integrating capsule network, attention mechanism and temporal convolutional network. First, it extracts the differential entropy features of the EEG signals, which can reflect the energy distribution of EEG activities in different frequency bands, and frequency domain and spatial domain information are combined to map into multi-band feature matrices. Subsequently, spatial information is acquired through the CapsNet, and an attention mechanism is employed to assign different weights to various brain regions, enabling the classification capsule to focus on key leads. Then, temporal convolutional network (TCN) is utilized to learn the temporal dynamic features, capturing the dynamic changes of emotions over time. Finally, obtaining the emotion classification results. The experimental results show that the proposed fusion model achieves an accuracy of 94.97% and 95.28% for Valence-Arousal binary classification tasks, respectively, along with an accuracy of 94.62% for four-class classification tasks. Compared with single-dimensional models, the proposed fusion model can effectively improve the accuracy of emotion recognition.

Keywords: emotion recognition; EEG; CapsNet; temporal convolutional network; feature extraction; feature fusion

情绪识别是通过分析采集到的多种信号,提取并整合反映情绪的关键特征,最后通过情绪量化模型识别出个体的情绪状态的一种过程。在情绪识别技术发展的早期,研究者大多是采用面部表情、语音、动作姿态等情绪识别方式,这种方式易于获取和分辨,但由于个体情绪的主观性以及易于隐藏掩饰的特性,情绪识别的客观性难以保证,因此采用这类非生理信号的情绪识别方式存在一定的局限性^[1-2]。而脑电、皮电、心电等生理信号是由内分泌系统和中枢神经系统协同影响的,不易受个体意志支配,因此取得了更好的效果^[3]。情绪的产生主要与中枢神经系统神经元

的电位活动相关,大脑活动时神经元会产生相应的电势差,然后通过各个相应位置的导联电极所采集的电信号就可以构成脑电信号(electroencephalography, EEG),因此,相比其他生理信号,EEG最能客观反映神经元的活动^[4]。

围绕脑电信号,研究者们目前主要从时域和频域两个维度进行研究,提取其标准差、均值、能量、功率谱密度、微分熵等特征,然后结合深度学习算法进行情绪分类^[5-6]。但是由于大脑的复杂性,不同情绪状态下大脑特定区域的活动模式和空间分布可能存在差异,因此不同区域导联对情绪的表达能力可能是不一样的。现有研究较少关注空

收稿日期:2024-07-04;录用日期:2024-07-24

基金项目:上海电子信息职业技术学院校级科研项目(B24313)

作者简介:曹雅茜(1993—),女,硕士,助教,研究方向:模式识别、脑机接口。

引用本文:曹雅茜. 基于脑电信号的多维融合情绪识别模型[J]. 自动化技术与应用, 2026, 45(3): 99-103. (CAO Yaxi. Multidimensional fusion emotion recognition model based on EEG signals[J]. Techniques of Automation and Applications, 2026, 45(3): 99-103.)

间维度,如果能将时域、频域、空间域3个维度的信息进行有效融合,可以深度挖掘脑电信号中不明显的情绪特征,辅助了解不同脑区在情绪产生和调节过程中的活动模式,帮助研究者更好地理解情绪的神经机制,从而进一步提升情绪识别准确性。

考虑到以上问题,本文提出一种基于胶囊网络(capsule networks, CapsNet)和时序卷积网络(temporal convolutional network, TCN)的情绪识别方法,充分挖掘脑电信号时频域及空间域特征,并在模型中引入空间注意力机制应用于优势导联的选取,识别对情绪更敏感的优势导联,从而在保证准确率的前提下提高了计算效率。

1 情绪识别整体框架

本文选用来自 Queen Mary University of London 研究团队建立的多模态情绪数据集 DEAP 数据集^[7],该数据集在情绪识别研究中被广泛应用。情绪识别模型的整体框架如图1所示,主要包括脑电数据预处理、频带分解、特征提取、多频带特征拼接及基于注意力机制的 ACapsNet-TCN 融合网络5个环节。

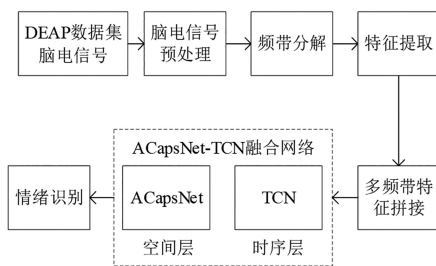


图1 基于 ACapsNet-TCN 的情绪识别模型框架

Fig. 1 Framework of the emotion recognition model based on ACapsNet-TCN

首先会将预处理后的 EEG 信号通过离散短时傅立叶变化划分为 θ 波(4~8 Hz)、 α 波(8~13 Hz)、 β 波(13~30 Hz)和 γ 波(30~42 Hz)4种对情绪影响较大的关键频段。为了融合导联空间信息,本文构建一个融合了导联空间位置的稀疏矩阵,计算各个导联的微分熵(differential entropy, DE)并注入在该矩阵中,然后将四个频段生成的稀疏矩阵拼接成一个多频带特征矩阵。此时该矩阵不仅融合了脑电信号的频域信息,也融合了分布在不同脑区的导联位置信息。

接着采用一种与 CNN 结构相似的卷积神经网络 CapsNet^[8]对融合后的多频带特征矩阵进行分类,并在其中融入空间注意力机制。相比于 CNN, CapsNet 使用胶囊单元代替神经元,将卷积作用所构成的单一特征扩展至胶囊单元所包含的一组特征信息,引入动态路由机制取代传统 CNN 中的池化层,胶囊中的多维向量能够充分保留空间信息的完整性,从而挖掘到潜在的情感特征。因此选择 CapsNet 学习融合了频域和空间信息的脑电特征矩阵,实现了对脑电信息的多维度解析。由于情绪状态的产生和变化会受到外界刺激的影响,被试者在观看音乐视频时所

激发的情绪是随着视频片段的播放时间而逐渐发生变化,因此本文还采用 TCN^[9]来捕捉时序信息,TCN 相较于 RNN、LSTM 等时序处理算法而言,具有并行处理、长距离依赖捕捉等方面的优势,而且由于 TCN 的卷积操作可以在整个序列上并行执行且参数较少,因此可以显著加快模型的训练速度。

2 数据及预处理

DEAP 数据集共有 32 名被试者,需在预设条件下观看 40 段不同类型的视频,采集总时长为 63 s(前 3 s 为基线信号,后 60 s 为视频片段诱发信号)。参与者根据自我评估量表对自己的情绪状态进行评分,覆盖唤醒度、效价、支配性和喜好度等多个维度,评分为 1 至 9 分,其采集过程如图 2 所示。

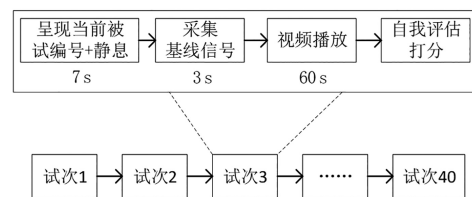


图2 DEAP 数据集信号采集流程

Fig. 2 Signal acquisition process of the DEAP dataset

2.1 标签确定

本文采用效价-唤醒度(valence-arousal, VA)构成的二维模型对情绪进行量化^[10],如图3所示,其中 Valence 值的大小代表积极或消极程度, Arousal 值的大小代表情绪的强弱。以 5 的分值作为阈值,四分类是将情绪状态根据评分分为低效价低唤醒(low valence low arousal, LVLA)、低效价高唤醒(low valence high arousal, LVHA)、高效价低唤醒(high valence low arousal, HVLA)、高效价高唤醒(high valence high arousal, HVHA)4种情绪标签。二分类则是分别对效价和唤醒度进行评估,分为低效价(low valence, LV)、高效价(high valence, HV)和低唤醒度(low arousal, LA)、高唤醒度(high arousal, HA)。

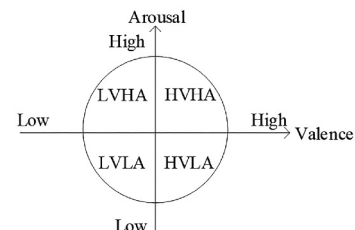


图3 效价-唤醒度二维情绪模型

Fig. 3 Two-dimensional emotion model of valence-arousal

2.2 预处理

DEAP 数据集的脑电信号是由 32 个电极帽采集得到,通过降采样将原始 512 Hz 采样频率降至 128 Hz,因此一个被试者一次视频可采集到 $128 \times 63 = 8\ 064$ 个数据点,每个被试者脑电数据为 $40 \times 32 \times 8\ 064$ 。63 s 包括 3 s 的被试静息状态数据和 60 s 的视频采集数据,为了降低静息状态时个体差异影响,本文去除了 3 s 的基线信号,并利用 max-min

标准化对特征进行归一化,那么视频采集阶段一个被试者40段视频的样本总数为 $40 \times 60 = 2400$,总数据为 $2400 \times 32 \times 128$ (样本数 \times 导联数 \times 采样频率)。由于脑电信号强度较小,在采集过程容易受到伪迹信号干扰,因此还采用盲源分离法去除眼电伪影^[11-13]。

模型选择微分熵作为 EEG 的情感特征,计算公式为

$$DE = - \int_{-\infty}^{\infty} p(x) \ln p(x) dx \quad (1)$$

式中, x 表示单通道脑电信号向量, $p(x)$ 表示概率密度函

数,计算方式为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[- \frac{(x - \mu)^2}{2\sigma^2} \right] \quad (2)$$

为了融入空间信息,本文按照导联电极分布,将计算所得的微分熵值映射到一个 9×9 的二维矩阵,4个频段的特征矩阵可以拼接成一个 18×18 的多频带矩阵,如图4所示,该矩阵包含了 EEG 的时域、频域及空间域信息。

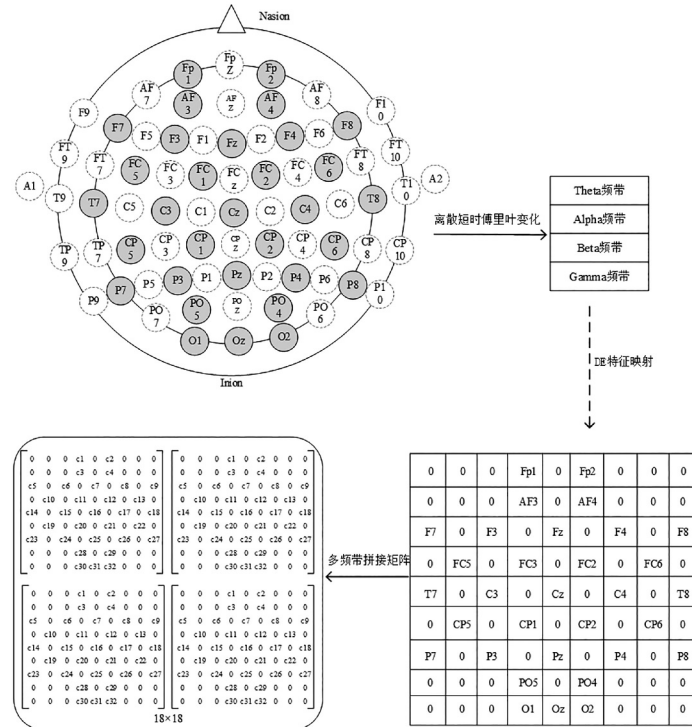


图4 脑电信号不同频段映射的特征矩阵

Fig. 4 Feature matrix mapped from different eeg frequency bands

3 ACapsNet-TCN 情绪识别模型

3.1 基于注意力机制的胶囊网络层

基于空间注意力机制的胶囊网络主要包括卷积层、空间注意力层、主胶囊层以及分类胶囊层,模型结构如图5所示。

卷积层能够提取数据的基本特征,对输入数据进行局部特征检测,Conv1 卷积核尺寸设定为 $3 \times 3 \times 32$,步长为1,对输入的多频带矩阵提取初级特征。

空间注意力层用于根据注意力权值动态估计不同脑区域的重要性,权值越大说明该区域的导联对情绪更加敏感。池化操作采用 Mean-pooling 对特征点邻域计算平均值,从而降低特征和参数的数量,Mean-pooling 可以降低噪声的影响,平滑时间维度上的特征,泛化能力更强。接着压缩池化后的混合特征,从而确定不同区域的重要性。空间注意力计算方法为

$$R_{ij} = \tanh(\mathbf{P}_{ij} \mathbf{w}_{ij}^T + b_{ij}) \quad (3)$$

$$I = \mathbf{P} \odot f(\tanh \mathbf{R}_{\text{mean}}) \quad (4)$$

式中, \mathbf{P} 为经过 Conv1 卷积操作后输出的特征矩阵, \mathbf{P}_{ij} 是 \mathbf{P} 位于 (i, j) 的通道向量, \mathbf{w}_{ij} 是用于融合通道信息的权重转换向量, b_{ij} 为偏置。经过平均池化压缩后得到 R_{mean} , f 为激活函数, \odot 表示哈达玛积,得到 I 之后传给主胶囊层。

主胶囊层除初级特征之外还存储了空间特征, Conv2 卷积核尺寸为 $3 \times 3 \times 32$,步长为2,使用256个 Conv2 进行卷积操作后输出尺寸为 $7 \times 7 \times 256$ 。为了后续分类胶囊元的接收,将其转换为 $7 \times 7 \times 32 \times 8$ 的向量形式,得到共1568个深度为8的胶囊,每个胶囊都保留了位置信息。

分类胶囊层主要是进行四分类任务,将主胶囊层的输入转换为 $1568 \times 4 \times 8 \times 1$ 的 Tensor, 将其与权重矩阵相乘得到预测向量,并对其进行扩维,扩维后乘以耦合系数并相加得到 16×4 的输出 s_j 。本文使用 Squash 函数对输出向量 s_j 进行压缩,使得分类胶囊长度介于 $0 \sim 1.0$ 之间,并保留了输入的方向信息,公式为

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \times \frac{s_j}{\|s_j\|} \quad (5)$$

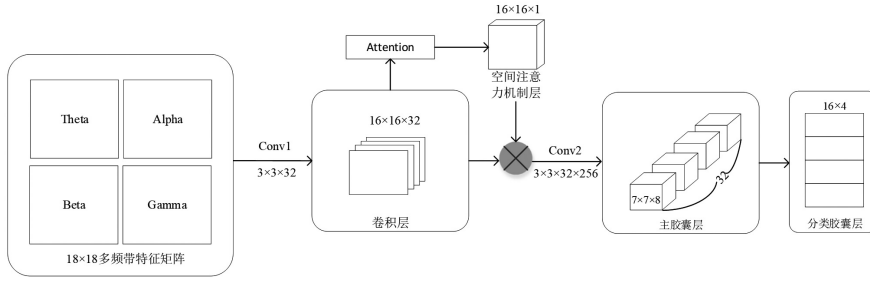


图5 基于空间注意力的胶囊网络层

Fig.5 Capsule network layer based on spatial attention

胶囊损失评估采用交叉熵损失函数,其计算方法为

$$\text{loss} = \sum_i [y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})] + \|R_F\|_2 \quad (6)$$

式中, y 为实际值, \hat{y} 为胶囊的预测值, $\|R_F\|_2$ 为正则化项,避免过拟合陷入局部最优。

3.2 时序卷积层

TCN在处理序列数据时,利用因果卷积(casual convolution)改进了CNN没有考虑时间顺序和因果关系的不足,确保当前的输出仅依赖当前时刻以及之前的数据。通过膨胀卷积(dilated convolution)灵活调整感受野,利用扩张率扩大卷积核感知范围,从而捕捉长距离依赖关系。表达式为

$$y_t = \sum_{i=0}^{k-1} w_i x_{t-d \times i} \quad (7)$$

式中, k 为卷积核大小, d 为膨胀因子, $x_{t-d \times i}$ 表示当前时间之前第 $d \times i$ 的输入。两者结合能够高效地处理被试者在情绪实验中从静息状态到情绪唤起再到静息状态过程中的变化。

模型首先需要将一段时间间隔的特征矩阵通过胶囊网络层提取空间维度的信息,然后将输出向量通过 Flatten

操作转化成一维脑电序列。TCN网络中使用步长为1,卷积核大小为4,并使用了残差模块缓解梯度消失问题,通过TCN网络后获取脑电信号的时序特征序列,使用ReLU作为激活函数,最后输入到全连接层进行分类,使用Softmax获取该时段情绪识别的结果。

4 实验设计与结果分析

实验搭建在pytorch3.7的框架下,使用DEAP数据集16个被试四分类的平均识别率,以及低效价低唤醒、低效价高唤醒、高效价低唤醒、高效价高唤醒这四种情绪的识别率作为评价指标。将训练集和测试集按照样本数据4:1的比例划分,在训练集中随机抽取其中10%作为验证集。模型最大迭代次数设置为100,为避免重复训练,损失函数连续10次没有下降时停止训练。

对于EEG这种时序信号来说,不同的时间分段影响着样本所包含的信息,会影响情绪分类精度。实验中时间序列分别选择间隔为1s,5s,10s,20s,图6为对比不同时间间隔模型四分类的测试结果。

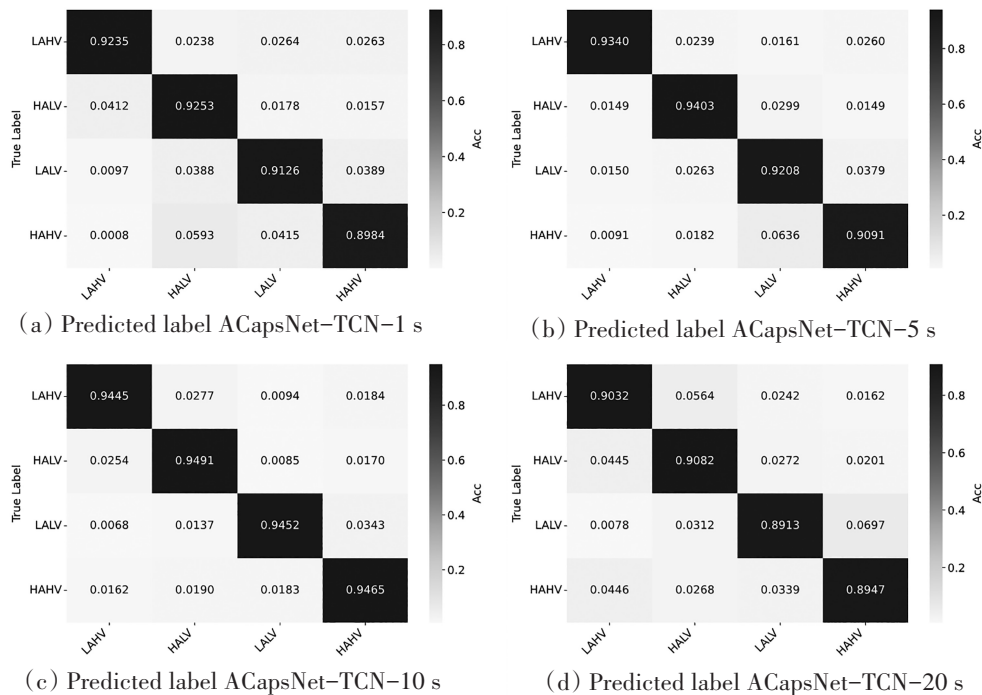


图6 不同时间间隔情绪四分类结果

Fig.6 Four-class emotion classification results at different time intervals

可以看出当间隔长度为 10 s 时,四分类识别效果最佳,因此在后续实验中,脑电信号的时间间隔均选择 10 s。

4.1 消融实验

为验证融合模型的必要性和有效性,分别将 CapsNet 模型、ACapsNet 模型与本文所提 ACapsNet-TCN 模型的性能进行对比,如表 1 所示。

表 1 消融实验

Tab. 1 Ablation experiments					%
模型	V-A	LVLA	LVHA	HVLA	HVHA
CapsNet	90.47	89.63	90.10	90.91	91.23
ACapsNet	92.73	90.58	91.61	93.94	93.82
ACapsNet-TCN	94.62	93.81	94.66	94.63	94.72

ACapsNet 模型是在原始的 CapsNet 模型中增加空间注意力机制所得,平均识别率上升了 2.50%,由此可见增加空间注意力机制后,通过给予不同脑区特征不同的权重可以有效提升情绪识别率。ACapsNet-TCN 融合模型通过增加时序卷积网络捕捉时序特征,在平均准确率以及不同情绪的识别率上均有提升。

4.2 同类研究对比

本文将所提 ACapsNet-TCN 模型分别与其他使用 DEAP 数据集的情绪识别算法进行对比,包含传统情绪识别算法 SVM 以及文献[14]、文献[15]、文献[16] 3 种多维融合模型。其中,SVM 算法设置惩罚因子为 2,核函数选择高斯核函数;文献[14]是将注意力机制应用于 LSTM 上,提取不同导联的空间特征,捕捉脑电信号的空间与时序信息进行情绪识别,其四分类准确率为 88.03%,在效价和唤醒度上的二分类准确率分别为 90.92%、94.21%;文献[15]在利用电极拓扑结构融入频域特征构建拓扑图,然后利用 GLFANet 学习 EEG 信号的全局和局部特征,四分类准确率为 92.92%,在效价和唤醒度上的二分类准确率分别为 94.53%、94.91%;文献[16]则是将残差注意力卷积神经网络与双向长短期记忆网络相结合,从时空维度获取脑电信息,其四分类准确率为 93.71%。5 种模型的四分类与二分类准确率对比结果如表 2 所示,本文所提模型显著提高了情绪识别的准确性。

表 2 同类研究情绪分类能力对比

Tab. 2 Comparison of emotion classification performance with similar studies

模型	V-A	Valence	Arousal
SVM	84.89	90.10	89.63
文献[12]	88.03	90.92	94.21
文献[13]	92.92	94.53	94.91
文献[14]	93.71	—	—
ACapsNet-TCN	94.62	94.97	95.28

5 结论

本文基于脑电生理信号,结合胶囊网络在捕捉空间结构信息的能力以及时序卷积网络在时序数据处理上的优势,提出 ACapsNet-TCN 模型用于情绪识别,有效挖掘了脑电信号在频域、时域及空间域上的情绪特征。考虑到不同脑区对情绪识别的贡献度不同,还在胶囊网络中融合了空间注意力机制用于优势导联的选择。在 DEAP 数据集上的实验结果表明该模型能够进一步提升情绪分类的精度。

参考文献

- [1] LI T Y, FU B L, WU Z X, et al. EEG-based emotion recognition using spatial-temporal-connective features via multi-scale CNN [J]. IEEE Access, 2023(11):41859-41867.
- [2] ZHAO S, JIA G, YANG J, et al. Emotion recognition from multiple modalities: fundamentals and methodologies [J]. IEEE Signal Processing Magazine, 2021, 38(6):59-73.
- [3] BHOSALE S, CHAKRABORTY R, KOPPARAPU S K. Calibration free meta learning-based approach for subject independent EEG emotion recognition for subject independent EEG emotion recognition [J]. Biomedical Signal Processing and Control, 2022(72):103289.
- [4] KEELAWAT P, THAMMASAN N, NUMAO M, et al. A comparative study of window size and channel arrangement on EEG-emotion recognition using deep CNN [J]. Sensors, 2021, 21(5):1678.
- [5] RANJAN R, SAHANA B C, BHANDARI A K. Ocular artifact elimination from electroencephalography signals: a systematic review [J]. Biocybernetics and Biomedical Engineering, 2021, 41(3):960-996.
- [6] 吕宝粮, 张亚倩, 郑伟龙. 情感脑机接口研究综述[J]. 智能科学与技术学报, 2021, 3(1):36-48.
- [7] KOELSTRA S, MUHL C, SOLEYMANI M, et al. A database for emotion analysis using physiological signals [J]. IEEE Transactions on Affective Computing, 2012, 3(1):81-89.
- [8] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules [J]. Advances in Neural Information Processing Systems, 2017, 30:3856-3866.
- [9] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. (2018-03-04) [2018-04-19]. <https://arxiv.org/pdf/1803.01271>.
- [10] RUSSELL J A. A circumplex model of affect [J]. Journal of Personality and Social Psychology, 1980, 39(6):1161-1178.
- [11] 李锦瑶, 杜肖兵, 朱志亮, 等. 脑电情绪识别的深度学习研究综述 [J]. 软件学报, 2023, 34(1):255-276.
- [12] 周晶, 周丽丽. 基于非侵入式多模态 BCI 的情绪识别与情感分析 [J]. 自动化技术与应用, 2024, 43(4):186-188.
- [13] 覃浩, 刘振华, 苏立伟. 基于多模态融合的电力智能客服系统情绪识别方法 [J]. 自动化技术与应用, 2024, 43(4):169-172.
- [14] DU X, MA C, ZHANG G, et al. An efficient LSTM network for emotion recognition from multichannel EEG signals [J]. IEEE Transactions on Affective Computing, 2020, 13(3):1528-1540.
- [15] LIU S, ZHAO Y, AN Y, et al. GLFANet: a global to local feature aggregation network for EEG emotion recognition [J]. Biomedical Signal Processing and Control, 2023(85):104799.
- [16] 王伟, 周建华, 刘紫恒, 等. 基于时空 Inception 残差注意力网络的脑电情绪识别 [J]. 重庆邮电大学学报(自然科学版), 2024, 36(1):68-75.