



## 基于自适应结构稀疏回归的异常脑电识别方法

吴涛, 刘夏, 孔祥增

引用本文:

吴涛, 刘夏, 孔祥增. 基于自适应结构稀疏回归的异常脑电识别方法[J]. *智能系统学报*, 2025, 20(6): 1432-1443.  
WU Tao, LIU Xia, KONG Xiangzeng. Abnormal electroencephalography recognition via adaptive structured sparse regression[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(6): 1432-1443.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202411006>

## 您可能感兴趣的其他文章

### 基于智能计算的脑机制研究

Brain mechanism research based on intelligent computing

智能系统学报. 2021, 16(5): 850-856 <https://dx.doi.org/10.11992/tis.202103029>

### 基于结构相似性与模板校正的织物瑕疵检测方法

Fabric defect detection based on structural similarity and template correction

智能系统学报. 2020, 15(3): 475-483 <https://dx.doi.org/10.11992/tis.201810011>

### 基于双向消息链路卷积网络的显著性物体检测

Salient object detection based on bidirectional message link convolution neural network

智能系统学报. 2019, 14(6): 1152-1162 <https://dx.doi.org/10.11992/tis.201812003>

### 面向自闭症辅助诊断的无监督模糊特征学习新方法

A novel unsupervised fuzzy feature learning method for computer-aided diagnosis of autism

智能系统学报. 2019, 14(5): 882-888 <https://dx.doi.org/10.11992/tis.201808005>

### 基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566-574 <https://dx.doi.org/10.11992/tis.201804056>

### 基于改进型BP神经网络的手部动作识别

Hand-motion recognition based on improved BP neural network

智能系统学报. 2018, 13(5): 848-854 <https://dx.doi.org/10.11992/tis.201703018>

DOI: 10.11992/tis.202411006

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20251009.1113.002>

# 基于自适应结构稀疏回归的异常脑电识别方法

吴涛<sup>1</sup>, 刘夏<sup>2</sup>, 孔祥增<sup>2</sup>

(1. 福建农林大学 计算机与信息学院, 福建 福州 350117; 2. 福建农林大学 未来技术学院, 福建 福州 350117)

**摘要:** 特征约简是提升脑电信号病理解码精度的一种重要手段。然而, 目前的异常脑电识别方法通常采用单尺度聚合实现特征降维, 并未考虑不同尺度聚合特征之间的互补性, 容易导致识别精度不高的问题; 此外, 现有方法在学习过程中常忽略了脑电数据特征内在的相关结构信息。为此, 提出一种基于自适应结构稀疏回归的异常脑电识别模型及其两阶段构造方法。采用自适应局部和全局特征聚合机制来有效融合从原始数据提取的小波统计特征, 为高精度脑电信号病理解码提供更具代表性的特征。定义了一种新的正则化稀疏模型, 它通过加权  $L_1$  范数约束剔除非重要特征, 同时利用加权成对结构正则化实现对高度相关特征的组选择。在实际异常脑电数据集上的实验结果表明, 新方法大幅度提高了分类精度和分类结果的稳定性。

**关键词:** 异常脑电识别; 聚合; 结构稀疏; 统计特征; 离散小波变换; 正则化; 组特征选择; 自适应聚合

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1673-4785(2025)06-1432-12

中文引用格式: 吴涛, 刘夏, 孔祥增. 基于自适应结构稀疏回归的异常脑电识别方法 [J]. 智能系统学报, 2025, 20(6): 1432-1443.

英文引用格式: WU Tao, LIU Xia, KONG Xiangzeng. Abnormal electroencephalography recognition via adaptive structured sparse regression[J]. CAAI transactions on intelligent systems, 2025, 20(6): 1432-1443.

## Abnormal electroencephalography recognition via adaptive structured sparse regression

WU Tao<sup>1</sup>, LIU Xia<sup>2</sup>, KONG Xiangzeng<sup>2</sup>

(1. College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350117, China; 2. College of Future Technology, Fujian Agriculture and Forestry University, Fuzhou 350117, China)

**Abstract:** Feature reduction is a critical technique for enhancing the decoding accuracy of electroencephalography (EEG) pathology. However, conventional abnormal EEG identification methods generally employ single-scale aggregation for dimension reduction, overlooking the complementary nature of multiscale aggregated features. This oversight frequently leads to suboptimal classification performance. Furthermore, current approaches often overlook the inherent correlation structure present among EEG features. To this end, an abnormal EEG recognition model based on adaptive structured sparse regression and its two-stage construction method is proposed. First, a novel adaptive local-global aggregation mechanism is employed to integrate wavelet statistical features extracted from the original data. This mechanism aims to provide representative features for high-precision EEG pathology decoding. Second, a novel sparse regularization model is defined, which can automatically eliminate unimportant characteristics by the  $L_1$ -norm constraint term. Concurrently, weighted piecewise structural regularization is employed to enable the group selection of highly correlated features. The experimental findings based on real-world abnormal EEG datasets demonstrate that the proposed method significantly enhances the accuracy and stability of the classification results.

**Keywords:** abnormal EEG recognition; aggregation; sparse structure; statistical features; discrete wavelet transform; regularization; group feature selection; adaptive aggregation

脑电 (electroencephalography, EEG) 是通过头

皮电极记录下来的脑细胞群的自发性、节律性电活动, 包含丰富的、可有效反映生理和病理变化的信息。由于 EEG 具有操作简单、时间分辨率高、成本低 (与正电子发射断层扫描、功能性磁共

收稿日期: 2024-11-08. 网络出版日期: 2025-10-09.

基金项目: 国家重点研发计划项目 (2022YFF1202400, 2023YFF1203900).

通信作者: 孔祥增. E-mail: xzkong@fafu.edu.cn.

©《智能系统学报》编辑部版权所有

振成像等其他神经影像技术相比)等优势,使其不仅是当前协助医生诊断阿尔茨海默病、精神分裂症、癫痫等各种神经系统疾病的一个重要手段<sup>[1]</sup>,还是研究人类脑功能的关键方法之一<sup>[2]</sup>。临床上,脑电专家常常通过遵循一种类似于决策树的复杂分析规则来判别 EEG 记录为正常或异常<sup>[3]</sup>。然而,EEG 信号的强时变性、低信噪比等特点使这一过程费时费力,并且易受评分者间低一致性的影响<sup>[4]</sup>。一种可靠且无需人工干预的异常脑电识别方法显得尤为重要。它可以作为基础工具,能够以更客观、准确和高效的方式分析 EEG 数据,尤其在寻求低成本远程诊断时具有显著优势。

脑电数据不仅具有传统生物数据的高复杂性,还表现出个体差异性大的特点。受此影响,当用传统算法分析此类数据时,效率低下且较为复杂<sup>[5]</sup>,尤其经典的统计模型可处理的特征(也称为变量)通常是有限的。另外,面对复杂多变的脑电数据,直接利用目前较为成熟的特征选择技术也难以有效降维<sup>[6]</sup>。如何从原始数据中自动提取鲁棒低维数据表征(也称数据特征),成了实现精准异常脑电识别的首要任务。

目前,脑电数据表征学习方法主要包括统计参数(均值、标准差、偏度和峰度)<sup>[5]</sup>,振幅相关参数(非线性能量、最大值和最小值)和熵相关度量<sup>[7]</sup>。这些方法可归纳为时域、频域和时频域特征学习三大类。由于 EEG 信号具有非线性、非平稳、混沌等特点,导致以小波变换为代表的时频分析方法近年来成为提取 EEG 特征的一个重要手段<sup>[5,8-10]</sup>。这是因为小波变换在高分辨率表示和细节定位方面具有出色性能。例如,Sharma 等<sup>[10]</sup>利用小波变换分解 EEG 信号为不同频率子带,并从特定的频率子带中提取相关特征,作为传统分类器的输入,以实现异常脑电识别。然而,通过小波变换等技术得到的小波系数常存在大量冗余<sup>[11]</sup>,致使这些方法往往难以获得令人满意的识别效果。最近,Gemein 等<sup>[5]</sup>和 Albaqami 等<sup>[8]</sup>利用单尺度聚合技术来降低提取特征的维度,以提升异常脑电分类精度。类似地,Zhong 等<sup>[12]</sup>在聚合 EEG 特征的基础上,使用克鲁斯卡尔-沃利斯检验进一步选取有用特征,提高了异常脑电识别效果。然而,一方面,文献<sup>[5,8,12]</sup>所提出方法均采用单一视图进行特征聚合,忽略了局部与全局聚合特征的互补性;另一方面,EEG 特征间通常包含许多重要的相关结构<sup>[13]</sup>,如分组结构<sup>[6,14]</sup>等,这些结构对于提高模型分类性能是有益的,但现有方法在特征选择过程中并未充分考虑这一关键信

息。此外,受颅内容积传导效应的影响,目前的异常脑电识别方法提取的 EEG 特征中往往存在大量噪声信息以及特征间高度相关,尤其是从空间相邻或对称电极抽取的 EEG 特征<sup>[15]</sup>。实际应用中,这些冗余信息的存在不仅无助于模型对类的鉴别,还将大大降低模式识别系统的有效性。

近年来,正则化稀疏逻辑回归模型由于其良好的稀疏性和模式识别效率在生物信息领域受到广泛关注<sup>[16-19]</sup>。与传统的分类算法不同,这类方法可以通过对模型的未知参数施以某种先验约束来使非重要特征对应的参数在训练过程中压缩为零,从而实现稀疏特征选择的同时,有效提升分类性能。这一特性弥补了传统表征学习的局限性。例如,DÍAZ 和 Olaya<sup>[19]</sup>通过利用  $L_2$  正则化逻辑回归模型实现了在减少脑电实验次数的情况下,仍取得了较高的分类精度;Garces 等<sup>[20]</sup>则使用弹性网惩罚逻辑回归模型(elastic net logistic regression, ENLR)获得了相比传统支持向量机(support vector machine, SVM)更优的异常脑电分类效果。然而,这些方法容易出现无效估计或特征选择不一致等现象,这是因为它们对所有特征对应的回归系数施加相同程度的惩罚,忽略了特征的差异性<sup>[21]</sup>。针对这一问题,学者们提出根据特征与响应向量之间的相关程度来确定各回归系数的惩罚强度,以减少估计偏差。例如,Algamil 和 Lee<sup>[16]</sup>采用基于相关性惩罚的逻辑回归的初始估计来构造特征权重,结果表明其可以显著提升模型分类准确性的同时却选择更少的特征;Patil 等<sup>[22]</sup>提出采用诸如 Fisher 分数、卡方检验和信息增益等手段来缓解传统 Lasso 逻辑回归模型的有偏估计问题,增强了模型的预测性能。其他常用于特征权重构造的统计分析方法还包括皮尔逊相关系数<sup>[17]</sup>、Wilcoxon 秩和检验<sup>[23]</sup>以及 T 检验<sup>[24]</sup>等。然而,这些稀疏逻辑回归模型在学习过程中缺乏考虑特征间的相关结构,并且所使用的权重构造策略并不能很好地抑制真实数据中的异常值和噪声影响<sup>[25]</sup>,从而导致它们在异常脑电检测任务中往往表现欠佳。最近,Wang 等<sup>[18]</sup>利用对称不确定性来构造每个特征对应的回归系数惩罚权重,并提出了结构稀疏逻辑回归模型(adaptive lasso and correlation based penalty, ALCP)。与多种经典稀疏学习模型相比,ALCP 在模拟数据和基准数据上均取得了更高的分类精度,并表现出更优的特征选择能力。然而,ALCP 在特征组选择过程中侧重于考虑特征间的相关性,而忽略了它们之间的交互信息,这是不完备的。

本文提出一种基于结构稀疏回归的异常脑电识别方法,用于对蕴含多种脑疾病信息的高维复杂脑电数据进行分类。首先,利用离散小波变换将 EEG 信号分解为多个小波系数,并从中提取出一组紧凑的统计特征集的同时,采用自适应特征聚合技术来增强特征的表达能力;随后,基于最大信息系数和多元交互信息 (three-way interaction information),提出成对特征相关权重和特征权重的两种权重构造策略,并以此为基础定义了新的正则化回归模型——自适应结构稀疏的逻辑回归模型 (adaptive structured sparse logistic regression, ASLR)。通过引入加权  $L_1$  范数约束和加权成对结构正则化项,实现了成组的重要 EEG 特征选择。在两个真实异常脑电数据集上的实验结果表明,ASLR 方法能够从复杂数据中挖掘出具有良好判别力的显著特征,有效提高了分类准确度和分类结果的稳定性。

### 1 异常脑电识别方法

本文提出的异常脑电识别方法的总体工作流程如图 1 所示。新方法主要包括 3 个阶段:第 1 阶段是预处理,目的是保证处理数据的一致性;第 2 阶段为多尺度特征融合,主要是从每个离散

小波变换系数中提取若干个统计特征,然后通过自适应特征聚合机制增强 EEG 特征的表达能力;最后,使用结构稀疏逻辑回归模型对其进行分类,并筛选出具有强区分性的特征。在对新方法正式展开叙述之前,首先约定全文使用的记号。给定一组含有  $N$  个原始脑电样本的训练数据集  $T_r = \{(z_i, y_i)\}_{i=1}^N$ ,其中,  $z_i \in \mathbf{R}^{C \times N}$  为原始数据空间中的第  $i$  个样本,  $y_i \in \{0, 1, \dots, K-1\}$  是样本  $z_i$  对应的类标签或目标值 ( $i = 1, 2, \dots, N$ )。这里,  $C$  表示脑电信号采集的电极数目,  $K$  表示训练数据集  $T_r$  中包含的类数目。本文重点研究高维复杂脑电信号的二分类问题,故  $K$  取值为 2。设  $\mathbf{y} = (y_1, y_2, y_N)^T \in \mathbf{R}^N$  是由训练数据集中所有样本类标签值构成的响应向量。对于任意一个样本  $z_i$ , 将经模型学习而来的聚合特征组成一个向量  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T \in \mathbf{R}^p$ , 其中  $p$  代表数据特征的维数。进而,可以构建出一个与训练数据集  $T_r$  相对应的设计矩阵  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N]^T \in \mathbf{R}^{N \times p}$ , 同时将设计矩阵  $\mathbf{X}$  的第  $j$  列用  $\mathbf{x}_{(j)} = (x_{1j}, x_{2j}, \dots, x_{Nj})^T$  来表示 ( $j = 1, 2, \dots, p$ )。不失一般性,假设向量  $\mathbf{y}$  已作中心化处理,且设计矩阵  $\mathbf{X}$  中每个特征已被标准化,即  $\sum_{i=1}^N y_i = 0$ ,  $\sum_{i=1}^N x_{ij} = 0$  和  $\sum_{i=1}^N x_{ij}^2 = 1$  ( $j = 1, 2, \dots, p$ )。

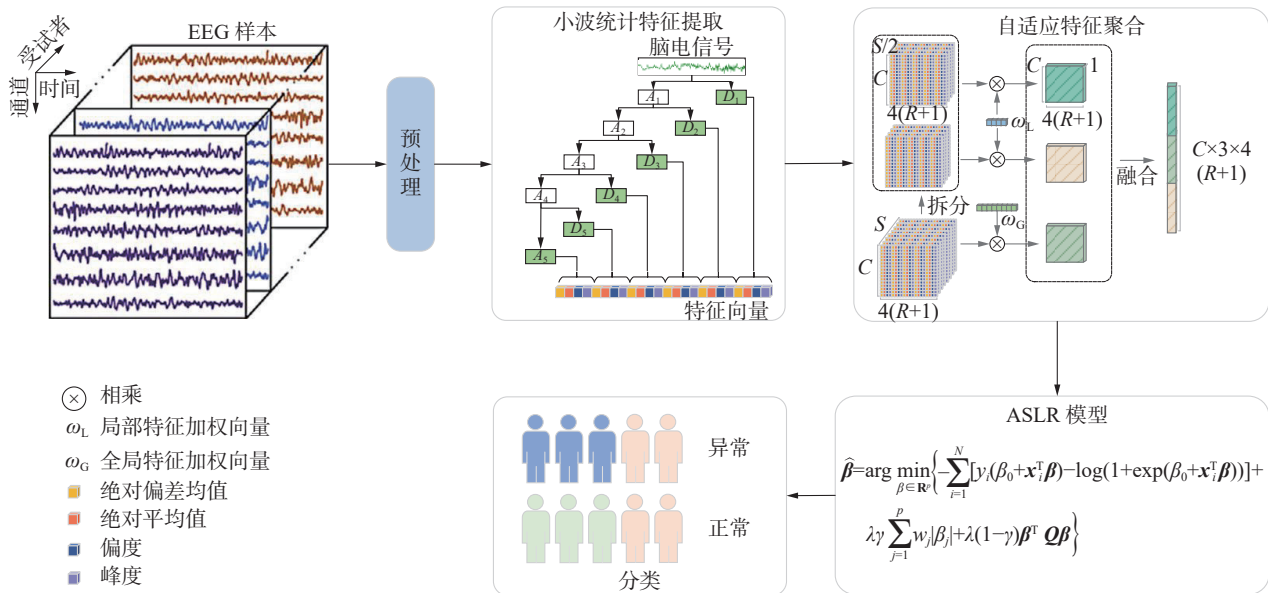


图 1 异常脑电识别方法流程  
Fig. 1 Flow of abnormal EEG recognition method

#### 1.1 多尺度特征融合

表征学习是解决复杂脑电数据问题的一个重要手段,可以为后续任务提供具有判别力的特征。为了更好地提取数据特征,预处理过程常将原始数据分割为若干个较短的脑电片段,并假设

每个片段蕴含有益于脑电模式识别的相似特征<sup>[26]</sup>;随后,采用时频分析技术分解脑电片段为不同频率子带,并从其中提取相关特征。这是因为 EEG 信号具有非平稳性,可能存在时域或频域特征量不足<sup>[2]</sup>,而时频分析方法能够有效应对这一

难题。目前,小波变换和短时傅里叶变换是最常用的时频分析方法。针对高时间分辨率的 EEG 信号,小波变换由于可以更好地平衡频率和时间分辨率,致使其相比短时傅里叶变换更为适用。因此,本文采用离散小波变换对分割后的 EEG 片段进行分解,并从选取的小波系数中提取一组有用的统计特征。虽然分解后的脑电数据可直接拼接成单个向量作为分类算法的输入,但由于其易受噪声影响,导致后续分类效果不佳。图 1 中给出了采用 5 阶离散小波变换从单通道 EEG 片段中提取统计特征的基本过程,其中符号“ $A$ ”和“ $D$ ”分别代表近似系数和细节系数。为了表征小波系数所蕴含的信息,本文选用了绝对偏差均值  $\gamma$  (mean absolute difference)、绝对平均值  $\xi$  (mean absolute value)、偏度  $\alpha$  和峰度  $\delta$  这 4 个经典的统计参数。给定一个含有  $C$  个通道的样本  $\mathbf{z}_i$ , 经特征提取后可以构建出一个与之相对应的特征矩阵  $\mathbf{M}_i \in \mathbf{R}^{C \times S \times 4(R+1)}$ , 其中,  $S$  代表每个样本被切割的片段总数,  $R$  代表离散小波变换的分解阶数,  $4(R+1)$  代表每个单通道 EEG 片段提取的小波统计特征数。由于头皮 EEG 信号具有任意的正或负电压值,故采用 Z-score 归一化方法对提取特征进行标准化处理,以解决特征缩放问题和消除偏移效应。

经小波统计特征提取后,若直接展平特征矩阵  $\mathbf{M}_i$ , 则每个样本将产生一个长度为  $C \times S \times 4(R+1)$  的一维特征向量。在这种情况下,假定数据集中每个脑电样本的通道数  $C$  和分割片段数  $S$  分别为 21 和 100,  $R$  设置为 5, 则展平后获取的表征向量长度为  $21 \times 100 \times 24 = 50400$ , 意味着获得的数据表征维度仍极高。同时,高维数据特征中通常包含许多冗余和噪声信息。这些对传统分类模型提出了严重挑战,如性能退化、计算复杂等。鉴于单尺度聚合难以有效刻画小波统计特征中蕴含的丰富信息,提出了一种基于全局和局部聚合策略的自适应特征融合方法,如图 1 所示。具体地,新方法一方面沿所构建的特征矩阵  $\mathbf{M}_i$  的时间维度方向将其划分为前后两部分,并分别使用有序加权平均作为聚合函数对这两部分中各 EEG 片段所属的每个小波统计特征进行聚合,用于表示每个通道蕴含的局部信息;另一方面,仍沿时间维度方向对每个通道下各特征向量进行聚合,用于获取每个通道的全局信息。接下来,以  $\mathbf{M}_i$  中任意通道下所有 EEG 片段的第  $r$  个小波系数的峰度  $\delta$  特征为例,给出使用自适应聚合方法计算得到其对应的全局特征  $gf_r$  的具体过程:

$$gf_r = \omega_1 \hat{\delta}_{1r} + \omega_2 \hat{\delta}_{2r} + \cdots + \omega_s \hat{\delta}_{sr} + \cdots + \omega_S \hat{\delta}_{Sr}$$

式中:  $\hat{\delta}_{sr}$  是数据特征向量  $\{\delta_{1r}, \delta_{2r}, \cdots, \delta_{sr}\}$  降序排列后的第  $s$  个元素 ( $s=1, 2, \cdots, S$ ),  $\omega_s \in [0, 1]$  表示第  $s$  个特征对应的权重且满足  $\sum_{s=1}^S \omega_s = 1$ 。根据文献 [27], 权重  $\omega_s$  的计算方式为

$$\omega_s = Q\left(\frac{s}{S}\right) - Q\left(\frac{s-1}{S}\right)$$

式中  $Q(\cdot)$  是模糊语义量化算子。最后,将不同尺度聚合的视图特征相融合,并以此构建一个维数为  $C \times 3 \times 4(R+1)$  的一维特征向量,作为后续分类模型的输入。显然,自适应特征聚合策略不仅使约简后的特征很好地反映 EEG 通道中具有互补性的局部与全局信息,而且可以使特征表示更为紧凑,有利于降低分类模型的计算量及缩短训练时间。

## 1.2 模型估计

在利用多尺度特征融合方法从原始数据中提取代表性特征后,接下来的关键任务是基于这些特征构建分类模型,以实现异常脑电的精准识别。逻辑回归是统计学习中的一种经典的二类分类模型,它的样本对应类别的条件概率表示为

$$P_r(y_i = 1 | \mathbf{x}_i) = \frac{\exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta})}$$

式中:  $\beta_0$  为截距项,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T \in \mathbf{R}^p$  是一个  $p$  维的待估计系数向量,常可以应用极大似然估计法来获得其估计值。这里,逻辑回归模型的对数似然函数具体定义为

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^N [y_i (\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) - \log(1 + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}))] \quad (1)$$

尽管逻辑回归在数据挖掘、机器学习和生物信息学等领域得到了广泛的应用并取得了良好的分类效果,但当数据特征的维度高于训练样本量时,该模型容易出现过拟合 [17]。另外,受颅内容积传导效应影响,脑电特征中往往存在冗余或非重要的信息。这些不仅对样本类标号的识别没有帮助,还将影响模型的性能,起到负面作用。当前,正则化稀疏逻辑回归是解决上述难题的一个重要手段,它通过在式 (1) 的基础上引入正则化项,促使模型可以在学习过程中从数据特征空间剔除冗余信息和降低噪声干扰。通常,这类方法也被称为带惩罚函数的逻辑回归,其估计量一般定义为

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbf{R}^p}{\operatorname{argmin}} \{-\ell(\boldsymbol{\beta}) + P_\lambda(\boldsymbol{\beta})\}$$

式中:  $P_\lambda(\boldsymbol{\beta})$  为正则化项或惩罚项,  $\lambda > 0$  为调整损失函数和模型复杂度的正则化参数。显然,惩罚

函数  $P_\lambda(\boldsymbol{\beta})$  的定义决定了模型的特征选择能力。鉴于聚合特征中包含噪声和冗余信息以及特征间往往存在许多重要的相关结构,为此,在最大信息系数和多元交互信息基础上定义了成对特征相关权重和特征权重的度量策略,构造新的惩罚函数:

$$P_\lambda(\boldsymbol{\beta}) = \lambda(1-\gamma) \sum_{l < j} \alpha_{lj} \|\mathbf{x}_{(l)}\beta_l - \mathbf{x}_{(j)}\beta_j\|_2^2 + \lambda\gamma \sum_{j=1}^p w_j |\beta_j| \quad (2)$$

式中:  $\gamma \in [0, 1]$  为调整参数,而  $w_j$  和  $\alpha_{lj}$  的计算方式分别为

$$w_j = \ln \left( \frac{1}{\varrho_j^{\eta_1} + \epsilon} \right) \quad (3)$$

$$\alpha_{lj} = e^{\eta_2 [\hbar_{lj}]_+^{-1}} \quad (4)$$

式中:  $\varrho_j$  表示设计矩阵  $\mathbf{X}$  中特征  $\mathbf{x}_{(j)}$  与响应向量  $\mathbf{y}$  之间的最大信息系数;  $\epsilon$  是一个非常小的常数,用于避免  $\varrho_j$  为零而造成  $w_j$  的无意义,本文将其设为  $10^{-6}$ ;  $\eta_1 > 0$  和  $\eta_2 \geq 0$  为平滑系数;  $[\hbar_{lj}]_+$  为线性整流函数,其中  $\hbar_{lj}$  是特征  $\mathbf{x}_{(l)}$ 、 $\mathbf{x}_{(j)}$  与响应向量  $\mathbf{y}$  之间的多元交互信息。需要说明的是,本文采用最大信息系数来衡量特征权重,主要是因为其在应用范围和准确性方面均优于皮尔逊相关系数、互信息等其他传统方法。另外,本文称式 (2) 中等式右侧的第一项为加权成对结构正则化项。由此易知,成对特征相关权重  $\alpha_{lj}$  兼顾了响应向量  $\mathbf{y}$  所提供的信息,完全不同于 SPR(structured penalized regularization)<sup>[17]</sup> 和 ALCP<sup>[18]</sup> 等新近提出的方法,它们采用皮尔逊相关系数来衡量特征间的相似性。换言之,ALCP 等方法无法确保特征  $\mathbf{x}_{(l)}$ 、 $\mathbf{x}_{(j)}$  与响应向量  $\mathbf{y}$  均相关。当  $[\hbar_{lj}]_+$  值越大,表明特征之间依赖关系越紧密,意味着通过综合考虑两个特征可以获得更多的信息量<sup>[28]</sup>。此外,在响应向量  $\mathbf{y}$  已中心化和设计矩阵  $\mathbf{X}$  已标准化的情况下,加权成对结构正则化项可变换为

$$\sum_{l < j} \alpha_{lj} \left[ (1-\rho_{lj}) (\beta_l^2 + \beta_j^2) + \rho_{lj} (\beta_l - \beta_j)^2 \right] \quad (5)$$

式中:  $\rho_{lj}$  是设计矩阵  $\mathbf{X}$  中特征  $\mathbf{x}_{(l)}$  和  $\mathbf{x}_{(j)}$  两者之间的相关系数,即  $\rho_{lj} = \mathbf{x}_{(l)}^T \mathbf{x}_{(j)}$ 。若特征  $\mathbf{x}_{(l)}$  和  $\mathbf{x}_{(j)}$  之间强正相关 ( $\rho_{lj} \rightarrow 1$ ),则  $(\beta_l - \beta_j)^2$  成为主导,使得模型系数  $\beta_l$  和  $\beta_j$  相互收缩,即两个系数的估计是相似的 ( $\hat{\beta}_l \approx \hat{\beta}_j$ );若特征  $\mathbf{x}_{(l)}$  和  $\mathbf{x}_{(j)}$  之间是强负相关 ( $\rho_{lj} \rightarrow -1$ ),此时式 (5) 可变换为

$$\sum_{l < j} \alpha_{lj} \left[ (1-|\rho_{lj}|) (\beta_l^2 + \beta_j^2) + |\rho_{lj}| (\beta_l + \beta_j)^2 \right]$$

同样,在此情形下,  $(\beta_l + \beta_j)^2$  将成为主导,意味着模型激励系数  $\beta_l$  逼近  $-\beta_j$ 。换言之,高度相关的

特征所对应的模型系数估计值在绝对值上彼此相互接近 ( $|\hat{\beta}_l| \approx |\hat{\beta}_j|$ ),而符号由正相关或负相关所决定,即模型可以实现特征自动分组效应。此外,假设对于任意的  $l \neq j$ ,则式 (5) 可以写成简单的二次型:

$$\sum_{l < j} \alpha_{lj} \|\mathbf{x}_{(l)}\beta_l - \mathbf{x}_{(j)}\beta_j\|_2^2 = \boldsymbol{\beta}^T \mathbf{Q} \boldsymbol{\beta}$$

式中:  $\mathbf{Q} = (q_{lj})_{1 \leq l, j \leq p}$  是一个由成对特征权重构建的实对称半正定矩阵,其每个元素  $q_{lj}$  通过计算得到:

$$q_{lj} = \begin{cases} \sum_{s \neq l} e^{\eta_2 (\hbar_{ls} - 1)}, & l = j \\ -e^{\eta_2 (\hbar_{ls} - 1)} \rho_{lj}, & \text{其他} \end{cases}$$

因此,本文提出的结构稀疏逻辑回归模型可表示为

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ -\ell(\boldsymbol{\beta}) + \lambda\gamma \sum_{j=1}^p w_j |\beta_j| + \lambda(1-\gamma) \boldsymbol{\beta}^T \mathbf{Q} \boldsymbol{\beta} \right\}$$

式中: 加权  $L_1$  范数惩罚项  $\sum_{j=1}^p w_j |\beta_j|$  增加了 ASLR 模型估计的稀疏性。若一个特征  $\mathbf{x}_{(j)}$  与响应向量强相关,则  $w_j$  取值较小,意味着模型在训练过程中对相应的回归系数施加的惩罚强度较弱;相反地,一个特征与响应向量弱相关,甚至无相关,则对应的回归系数将被施加强惩罚,因此加权  $L_1$  范数约束确保了模型可以根据特征与响应向量间的相关性差异而自适应地选择重要特征。其次,加权成对结构正则化项将鼓励彼此高度相关的特征在模型训练时,它们所关联的系数亦相近 (即  $|\hat{\beta}_l| \approx |\hat{\beta}_j|$ ),也就是说,成对依赖性很高的特征将在训练过程被同时选择或剔除。当对任意的  $l, j$ , 均有  $q_{lj} = 0$  时,ASLR 退化为自适应 Lasso 惩罚逻辑回归模型;当  $\rho_{lj} \rightarrow 0$  时,ASLR 退化为自适应弹性网惩罚逻辑回归模型。

### 1.3 模型求解

对于 ASLR 模型中含有待估参数  $\boldsymbol{\beta}$  的问题,本文采用坐标下降法和牛顿法来求解。下面,给出求解 ASLR 模型参数的基本过程。

首先, Friedman 等<sup>[29]</sup> 证明用于求解对数似然函数 (见式 (1)) 的拟牛顿法实际上等同于一个迭代加权最小二乘法。具体地,若  $\tilde{\boldsymbol{\beta}}$  为普通逻辑回归模型当前的数值解,则优化目标函数  $\ell(\boldsymbol{\beta})$  可以近似 (关于当前估计的泰勒展开) 为

$$\ell(\boldsymbol{\beta}) \approx \ell(\tilde{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T \nabla \ell(\tilde{\boldsymbol{\beta}}) + \frac{1}{2} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T \mathbf{H} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}),$$

式中:  $\nabla \ell(\tilde{\boldsymbol{\beta}})$  和  $\mathbf{H} = \mathbf{X}^T \boldsymbol{\Lambda} \mathbf{X} / N$  分别为  $\ell(\boldsymbol{\beta})$  在当前估计时的梯度和海森矩阵。这里,  $\boldsymbol{\Lambda} = \operatorname{diag}(\tau_1, \tau_2, \dots, \tau_N)$  是一个对角矩阵,其中第  $i$  个对角元素  $\tau_i$  为

$$\tau_i = \tilde{\pi}(\mathbf{x}_i)(1 - \tilde{\pi}(\mathbf{x}_i)),$$

式中:  $\tilde{\pi}(\mathbf{x}_i)$  表示在当前模型参数下计算  $\pi(\mathbf{x}_i)$  的值, 其中  $\pi(\mathbf{x}_i) = P_r(y_i = 1|\mathbf{x}_i)$ 。由此, 式 (1) 可近似为

$$\ell(\boldsymbol{\beta}) = -\frac{1}{2} \sum_{i=1}^N \tau_i (z_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta})^2$$

式中:  $z_i = \tilde{\beta}_0 + \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} + \frac{y_i - \tilde{\pi}(\mathbf{x}_i)}{\tilde{\pi}(\mathbf{x}_i)(1 - \tilde{\pi}(\mathbf{x}_i))}$  是根据当前参数值而估计的伪响应。进一步, 经整理可得 ASLR 模型的优化目标函数为

$$\hat{\boldsymbol{\beta}}(\lambda, \gamma) = \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ \frac{1}{2} \sum_{i=1}^N \tau_i (z_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \lambda \gamma \sum_{j=1}^p w_j |\beta_j| + \lambda(1 - \gamma) \boldsymbol{\beta}^T \mathbf{Q} \boldsymbol{\beta} \right\}$$

接下来, 坐标下降法可用于解决上述重加权最小二乘问题。假设给定当前解  $\tilde{\beta}_0$  和  $\tilde{\beta}_l (l \neq j)$ , 则求解非零参数  $\beta_j$  的优化目标函数为

$$\begin{aligned} J(\tilde{\boldsymbol{\beta}}) = & \frac{1}{2} \sum_{i=1}^N \tau_i \left( z_i - \tilde{\beta}_0 - x_{ij} \beta_j - \sum_{l \neq j} x_{il} \tilde{\beta}_l \right)^2 + \\ & \lambda(1 - \gamma) \sum_{l \neq j} \alpha_{lj} (\tilde{\beta}_l^2 + \beta_j^2 - 2\rho_{lj} \tilde{\beta}_l \beta_j) + \\ & \lambda(1 - \gamma) \sum_{s \neq j} \sum_{l \neq j} \alpha_{sl} (\tilde{\beta}_l^2 + \tilde{\beta}_s^2 - 2\rho_{ls} \tilde{\beta}_l \tilde{\beta}_s) + \\ & \lambda \gamma w_j |\beta_j| + \lambda \gamma \sum_{l \neq j} w_l |\tilde{\beta}_l| \end{aligned}$$

对  $\beta_j$  求导, 可得

$$\begin{aligned} \frac{\partial J(\tilde{\boldsymbol{\beta}})}{\partial \beta_j} = & -\sum_{i=1}^N \tau_i x_{ij} \left( z_i - \tilde{\beta}_0 - x_{ij} \beta_j - \sum_{l \neq j} x_{il} \tilde{\beta}_l \right) + \\ & 2\lambda(1 - \gamma) \sum_{l \neq j} \alpha_{lj} (\beta_j - \rho_{lj} \tilde{\beta}_l) + \lambda \gamma w_j \operatorname{sign}(\beta_j) \end{aligned}$$

令  $\frac{\partial J(\tilde{\boldsymbol{\beta}})}{\partial \beta_j} = 0$ , 推导得出  $\beta_j$  的更新方式为

$$\tilde{\beta}_j \leftarrow \frac{S(\Phi, \lambda \gamma w_j)}{\sum_{i=1}^N \tau_i x_{ij}^2 + 2\lambda(1 - \gamma) \sum_{l \neq j} \alpha_{lj}} \quad (6)$$

式中:  $\Phi = \sum_{i=1}^N \tau_i x_{ij} \left( z_i - \tilde{\beta}_0 - \sum_{l \neq j} x_{il} \tilde{\beta}_l \right) + 2\lambda(1 - \gamma) \sum_{l \neq j} \alpha_{lj} \rho_{lj} \tilde{\beta}_l$ , 而  $S(\varphi, \varsigma)$  为软阈值算子<sup>[29]</sup>。根据式 (6) 交替更新  $\boldsymbol{\beta}$  中的每个回归系数, 直到满足给定的收敛条件。算法 1 给出了求解 ASLR 模型的主要步骤, 其时间复杂度为  $O(Np)$ 。坐标下降法每次迭代只优化一维变量, 其余变量看作常量, 且优化系数可以在变量循环中更新, 故其是一种相对高效的求解方法, 适用于不同类型的问题。

### 算法 1 ASLR 模型求解算法

输入  $X, Y, \lambda, \gamma, \eta, \theta$ ;

控制算法终止的阈值  $\phi = 10^{-6}$ ;

最大允许迭代次数  $\xi_{\max}$ 。

输出 稀疏向量  $\hat{\boldsymbol{\beta}}$ 。

方法

初始化: 设  $\xi$  为迭代次数,  $\xi = 0$ ;

根据式 (3) 计算各个特征的权重  $w_j$ ;

根据式 (4) 计算各个成对特征的权重  $\alpha_{lj}$ ;

**Repeat**

for  $j = 1$  to  $p$  do

根据式 (6) 更新模型系数值  $\tilde{\beta}_j^{(\xi)}$ ;

$\hat{\boldsymbol{\beta}}^{(\xi)} \leftarrow (\tilde{\beta}_1^{(\xi)}, \tilde{\beta}_2^{(\xi)}, \dots, \tilde{\beta}_p^{(\xi)})$ ;

$\xi \leftarrow \xi + 1$ ;

**Until**  $\|\hat{\boldsymbol{\beta}}^{(\xi)} - \hat{\boldsymbol{\beta}}^{(\xi-1)}\|_{\infty} \leq \phi$  or  $\xi = \xi_{\max}$ 。

## 2 实验与分析

本节主要验证本文所提新方法在多个实际异常脑电数据集上的有效性, 并与若干相关工作比较。所有实验在配置为 3.7 GHz CPU 和 32 GB RAM 的计算机上运行。

### 2.1 实验数据

实验采用了 2 个主流的脑电数据集对所提方法进行验证, 以检验新方法的性能。2 个数据集分别来源于美国天普大学医院收集的异常脑电数据集 (Temple University Hospital EEG Abnormal Corpus, TUHEAC)<sup>[3]</sup> 和巴基斯坦-阿联酋军事医院收集的 NUST-MH-TUKL (NMT) 数据集<sup>[30]</sup>, 它们在采集设备、数据规模、数据记录协议、人口统计学特征和患者构成方面均存在明显差异, 如表 1 所示。

表 1 实际应用数据集的有关信息

Table 1 Some information about the real-world datasets			
数据集	性别	正常 EEG 数量	异常 EEG 数量
TUHEAC	男性	853	742
	女性	668	730
NMT	男性	489	122
	女性	998	196

1) TUHEAC 数据集: 已公开发表的重要脑电数据集, 被广泛用于异常脑电的分类研究中<sup>[31]</sup>。该数据集按照国际标准 10-20 系统电极放置法, 以 250 Hz 或更高的采样率从 2329 位男性和女性患者头部采集而来, 年龄范围在 7 天到 96 岁之间 (平均年龄为  $48.69 \pm 17.89$  岁)。患者的脑部病理包

括但不限于癫痫、中风和抑郁症。每位患者的 EEG 记录持续 15 min 或更长,且已被神经学专家标注为正常或病理记录。鉴于 TUHEAC 数据集中的脑电记录通道数介于 21~31,为了确保数据的一致性,实验选择了脑电记录中均存在的 21 个脑电通道,并将所有数据均重采样至 250 Hz。

2) NMT 数据集:一个专注于南亚人口的公开脑电异常数据库,由 Khan 等<sup>[30]</sup>贡献。该数据集的记录来自 2417 名独立患者(男性和女性),平均年龄为 24.64 岁。原始头皮 EEG 信号通过国际标准 10-20 系统的 21 个通道采集,采样率为 200 Hz。整个数据集中的每个脑电记录均已被神经学专家标注为正常或病理性。此外,由于记录时长较短的脑电数据无法提供充足的信息,类似于其他研究<sup>[5,8]</sup>,实验中仅采用时长超过 10 min 的记录进行实验验证。因此,实验使用的 NMT 数据集包含 1805 条 EEG 记录,其中 82.38% 为正常记录,剩余约 17.62% 为病理记录,详细信息见表 1。

## 2.2 实验设置

为检验所提方法的有效性,实验使用了 2 个真实异常脑电数据集。首先,采用 5 s 非重叠滑动窗口将每个 EEG 通道数据分割为 100 个长度均等的片段,丢弃剩余数据;其次,采用以 Sym6 为母小波的离散小波变换,分解层级设置为 5,获得每个通道信号片段的细节系数和近似系数,并从  $D_1$ 、 $D_2$ 、 $D_3$ 、 $D_4$ 、 $D_5$  和  $A_5$  系数中计算出相应的统计参数,构建特征矩阵;最后,采用特征聚合方法增强特征表达能力,并使用 ASLR 进行分类。其中,采用有序加权平均作为聚合函数,并使用与文献<sup>[27]</sup>一样的参数设置。而对于 ASLR 的参数,本文通过参数敏感性实验确定了它们的最佳值。具体地,调整参数  $\gamma$  和  $\lambda$  分别设置为 0.5 和 0.05,平滑系数  $\eta_1$  和  $\eta_2$  分别设置为 8 和 5。实验采用 Python3 语言来实现。此外,由于对比方法大多是开源的,故实验中它们均采用推荐参数,以确保实验结果的可比较性。

本文采用 5 折交叉验证来验证不同方法的性能优劣,并使用 F1-Score 和 G-mean 作为评价指标,这是因为它们在分析不平衡数据集时(即正常脑电类别占比较大,见表 1),可以更准确地反映模型的识别效果。具体地,F1-Score 是精确度和召回率的加权平均值,当精确度和召回度都很高时,F1-Score 值较大,因而可以在两者之间保持良好的平衡;G-mean 是一种通过计算两种脑电类

别的召回率的几何平均值来衡量模型的性能。

## 2.3 性能对比实验

为了验证所提方法的先进性,实验选择了若干相关工作进行比较。主要包括小波包分解(wavelet packet decomposition, WPD)+集成学习分类器 CatBoost<sup>[8]</sup>、小波分解(wavelet decomposition, WD)+SVM<sup>[10]</sup>、混合特征+SVM<sup>[5]</sup>、ChronoNet<sup>[31]</sup>和混合网络<sup>[30]</sup>。其中,前 3 种采用的基于用户定义脑电特征的异常脑电模式识别方法具有代表性,尤其是 WD+SVM 使用的统计特征计算过程与所提方法较为接近;后两种是基于深度学习的脑电异常检测方法,它们的表征学习方式具有代表性,特别是混合网络是新近提出的。

表 2 给出了 6 种方法在两个实际应用数据集上的分类结果,以“均值 $\pm$ 1 个方差”的格式呈现。由表 2 可见,本文所提方法在真实异常脑电数据集上获得了较好的分类效果。表 2 以 TUHEAC 数据集为例,进一步分析表 2 后,可以获得以下观察结论:1)ASLR 方法的 F1-Score 值比基于用户定义特征的基线方法(混合特征+SVM)高约 8.3%,比代表性的深度学习方法 ChronoNet 显著提高了 23.1%;2)与新近提出的混合深度学习方法相比,本文方法也实现了更好的分类性能。另外,对于 NMT 数据集,从表 2 可以看出,ASLR 在分类效果方面也明显优于其他 5 种异常脑电识别方法。特别地,ASLR 与最近发表的方法相比,G-mean 值提升了 6.48%,F1-Score 值大约提升了 2%。这一结果充分表明了新方法在解决异常脑电识别中所面临问题的有效性和适用性。通过实验分析,ASLR 取得如此较好结果的原因主要归结于其能够有效挖掘出具有明显鉴别能力的显著特征。具体而言,对于 NMT 数据集,ASLR 可以从原始数据中提取的 1512 个小波统计特征中最终平均选择 586 个特征作为最优特征子集,而对于 TUHEAC 数据集,则从提取的小波统计特征集中最终平均选择 475 个特征作为最优子集,表明所提出的结构稀疏逻辑回归模型可以极大地剔除提取特征中的冗余信息,从而提高模式识别的准确性。此外,通过比较表 2 中不同数据集的实验结果,可以发现 ASLR 在 TUHEAC 数据集上表现出的分类效果略优于 NMT 上的分类。有趣的是,这一结果与文献<sup>[30]</sup>中的实验结果相一致。其主要原因是不同 EEG 数据集中的受试者之间差异较大,以及 NMT 数据集相对 TUHEAC 规模较小。

表 2 实际应用数据集上的分类结果比较  
Table 2 Comparison of the classification results on the real-world datasets

方法	TUHEAC		NMT	
	G-mean	F1-Score	G-mean	F1-Score
混合特征+SVM	0.734 1±0.009 2	0.703 3±0.012 5	0.135 1±0.072 6	0.045 4±0.027 5
WPD+CatBoost	0.783 1±0.006 0	0.771 6±0.007 4	0.553 7±0.053 3	0.444 5±0.064 0
WD+SVM	0.777 2±0.016 4	0.749 7±0.023 3	0.699 4±0.058 1	0.323 8±0.046 2
ChronoNet	0.621 5±0.104 5	0.555 5±0.137 4	0.436 7±0.076 2	0.324 1±0.088 8
混合网络	0.745 9±0.034 4	0.724 8±0.049 2	0.676 5±0.043 1	0.555 9±0.036 7
ASLR	<b>0.794 0±0.017 7</b>	<b>0.786 5±0.023 2</b>	<b>0.741 3±0.027 7</b>	<b>0.574 9±0.037 8</b>

注: 加粗数字表示最优结果。

### 2.4 不同正则化稀疏逻辑回归方法的对比实验

为进一步说明 ASLR 的特征选择能力以及模型的性能, 本节实验选择了 Lasso 惩罚逻辑回归 (Lasso logistic regression, LLR)<sup>[20]</sup>、ENLR<sup>[32]</sup>、SPR<sup>[17]</sup> 和 ALCP<sup>[18]</sup> 这 4 种正则化稀疏逻辑回归模型作为对比对象。LLR 是传统逻辑回归的扩展, 引入了额外的  $L_1$  正则化项, 迫使其在训练过程中部分系数被收缩至零, 大大提升了模型的可解释性和预测性能。ENLR 是在 LLR 基础上, 额外对模型的回归系数施加  $L_2$  惩罚, 促使模型具有分组

效应性质。ALCP 和 SPR 是新近提出的正则化逻辑回归方法, 它们各自采用的结构稀疏惩罚项与 ASLR 方法比较接近。实验均使用经多尺度特征融合技术提取的 EEG 特征作为每种回归模型的输入, 以及它们采用文献推荐的默认参数。表 3 给出了各种方法在不同数据集上获得的分类结果以及对应选取特征的平均数量 (average number of selected features, AveSN)。其中, AveSN 表示模型在多次实验中从特征空间选取特征数量的平均数, 值越小则意味着模型具有更好的稀疏性。

表 3 不同稀疏回归模型在公共数据集上的性能比较  
Table 3 Performance comparison between ASLR and baselines on public datasets

模型	TUHEAC			NMT		
	AveSN	G-mean	F1-Score	AveSN	G-mean	F1-Score
LLR	801.20±7.30	0.751 7±0.016 0	0.744 6±0.019 6	589.20±33.62	0.637 1±0.041 7	0.442 3±0.052 1
ENLR	736.20±12.08	0.790 2±0.020 0	0.784 2±0.024 7	627.40±22.91	0.559 6±0.060 1	0.419 1±0.074 6
SPR	515.80±8.16	0.763 2±0.020 6	0.767 1±0.017 7	710.00±26.40	0.279 9±0.047 1	0.416 2±0.038 3
ALCP	1 121.00±15.97	0.782 7±0.018 8	0.764 8±0.023 6	663.00±46.57	0.718 2±0.028 9	0.527 0±0.042 1
ASLR	<b>475.80±11.90</b>	<b>0.794 0±0.017 7</b>	<b>0.786 5±0.023 2</b>	<b>586.00±14.73</b>	<b>0.741 3±0.027 7</b>	<b>0.574 9±0.037 8</b>

注: 加粗数字表示最优结果。

由表 3 可见, ASLR 在选择最少特征的情况下, 取得了显著优于其他 4 种方法的分类效果。以 NMT 数据集为例, 分析可得以下观察结果: 1) 与经典的 LLR 方法相比, ASLR 在实际异常脑电数据集上的 G-mean 和 F1-Score 分别提高了 10.4% 和 13.2%; 2) 与新近提出的稀疏回归模型 ALCP 相比, ASLR 同样展现出明显优势; 3) 相较于其他 4 种稀疏逻辑回归模型, ASLR 获得了更为稳定的分类结果。一方面是因为 ASLR 采用了基于信息度量的特征权重构造策略, 不依赖于特征的真实值, 促使模型可以自适应地选择重要特征; 另一方面是由于 ASLR 采用了自适应  $L_1$  惩罚项, 可以改善甚至避免现有稀疏逻辑回归模型中的有偏估计。上述结果表明, 针对复杂的异常脑

电模式识别问题, ASLR 相较于现有主流稀疏逻辑回归方法更为有效。

### 2.5 消融实验

#### 2.5.1 不同特征融合方式对实验结果的影响

为了验证所提出的特征融合方法的有效性, 本文通过移除特征聚合过程中的部分组件进行了消融实验。具体地, 本文主要设计并比较了 4 种不同的情况: 未使用特征聚合、仅使用局部或全局聚合特征, 以及同时使用局部和全局聚合特征。表 4 给出了在两个实际异常脑电数据集上每种情况对应的分类效果。

由表 4 可见, 直接使用统计特征的分类方法性能整体相对较差, 而基于聚合特征的方法均获得了较好的识别效果。这意味着所提出的聚合机

制在实现维度约简的同时,能够显著提升特征的质量。此外,基于局部和全局聚合的方法在两个数据集上取得了最优的分类性能,表明所采用的

两种类型的聚合特征各具优势,具备互补性。以上实验结果充分证明了本文提出的多尺度聚合机制的有效性。

表 4 多尺度特征融合方法在公共数据集上的消融实验结果  
Table 4 Ablation experiment results of multi-scale feature fusion method on public datasets

索引	局部聚合	全局聚合	TUHEAC		NMT	
			G-mean	F1-Score	G-mean	F1-Score
1	×	×	0.608 1±0.0284	0.606 9±0.043 2	0.511 4±0.030 0	0.264 6±0.032 4
2	√	×	0.793 1±0.0198	0.785 8±0.024 1	0.726 6±0.030 1	0.556 3±0.038 2
3	×	√	0.787 1±0.021 7	0.781 0±0.024 2	0.716 1±0.042 9	0.543 5±0.054 8
4	√	√	<b>0.794 0±0.017 7</b>	<b>0.786 5±0.023 2</b>	<b>0.741 3±0.027 7</b>	<b>0.574 9±0.037 8</b>

注:加粗数字表示最优结果。

### 2.5.2 自适应结构稀疏回归对实验结果的影响

为了验证稀疏逻辑回归的不同改进策略对性能的影响,本文对新模型进行了消融研究。具体地,本小节设计了 ASLR 模型的 3 种变体进行对比实验,以证明它们在特征学习中的各自作用。模型 Model-A、Model-B 和 Model-C 分别对应将 ASLR 中的正则化项替换为  $L_1$  范数约束、仅引入加权  $L_1$  范数约束,以及在 Model-A 中引入结构正则化项。图 2 给出了不同模型在两个实际数据集上获得的平均分类结果。

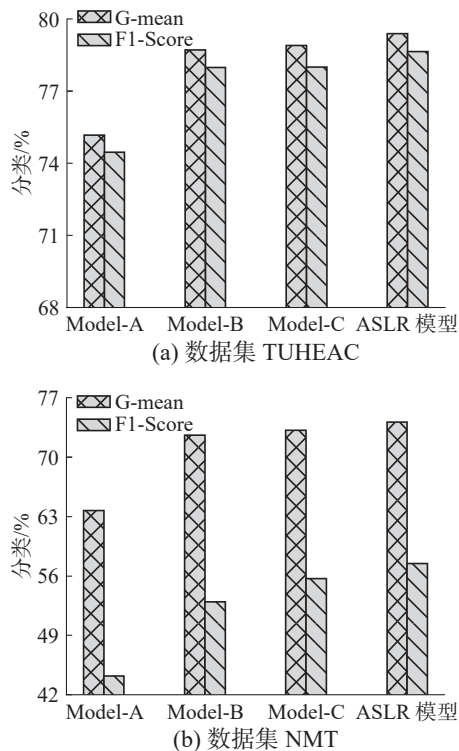


图 2 不同模型的实验结果对比

Fig. 2 Comparative analysis of experimental results obtained from different models

从图 2 可以看出,ASLR 在两个数据集上均获得了最佳性能。例如,在 NMT 数据集上,Mod-

el-B 和 Model-C 的分类结果显著优于 Model-A,表明自适应特征加权和加权成对结构正则化项在模型的特征学习过程中起着关键作用。这是因为受颅内容积传导效应影响<sup>[15]</sup>,不同脑电通道间存在高度相关性,导致脑电特征中往往包含大量冗余信息。而基于加权  $L_1$  范数约束和加权成对结构正则化项的模型可以自适应地选择组内的重要特征、剔除带有干扰信息的噪声特征,进而获得更佳分类结果。因此,两种改进策略的融合使得 ASLR 在所有方法中表现最佳。这一结论在 TUHEAC 数据集上同样得以验证。上述实验结果表明,本文新提出的正则化项对于提升异常脑电分类任务的性能是有效的。

### 2.6 参数敏感性

本节实验旨在进一步分析 ASLR 模型的参数敏感性。由式 (6) 可知,ASLR 包含 4 个预先设置的参数。这里,当评估某个特定参数时,其余 3 个参数将被固定为本文推荐的默认值。此外,在参数敏感性实验中, $\lambda$  在  $[0.01,0.06]$  内以 0.01 的步长取值, $\gamma$  在  $[0.1,0.6]$  内以 0.1 的步长取值,而  $\eta_1$  和  $\eta_2$  的取值范围限定为  $\{1,2,3,4,5,6,7,8\}$ 。图 3 给出了 ASLR 模型在不同参数值下获得的分类效果。

由图 3 可见,ASLR 对于调整参数  $\gamma$  和  $\lambda$ ,以及平滑系数  $\eta_1$  的不同取值所得到的分类表现较为稳定,表明模型对这些参数的值不敏感。而对于参数  $\eta_2$ ,ASLR 在 NMT 数据集上的分类性能随着其取值增大而提升,随后趋于平稳。通过分析实验发现,当参数  $\eta_2$  取值较小时,模型选择的特征相对较多,从而影响了异常脑电识别。这再次证明了成对结构正则化项对提升模型的识别精度发挥着重要作用。因此,上述结果表明,所提出方法的学习性能对参数在一定程度上具有较高的稳定性和鲁棒性。

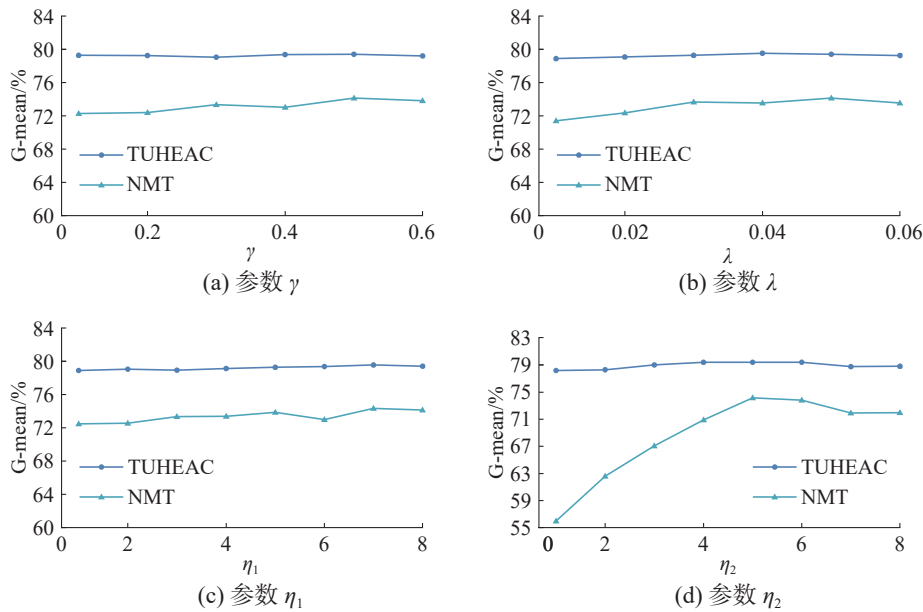


图 3 ASLR 的参数敏感性  
 Fig. 3 Parameter sensitivity of ASLR

2.7 可视化分析

为了验证 ASLR 模型选取的 EEG 特征具有高区分性, 本节引入标准的  $t$ -分布随机邻域嵌入 ( $t$ -distributed stochastic neighbor embedding,  $t$ -SNE)<sup>[33]</sup> 来可视化新方法中每个核心部分输出的特征。图 4 和图 5 给出了新方法在测试数据集上不同阶段输出特征的可视化结果。由图 4(a) 和图 5(a) 可见, 原始数据中不同类别间的特征被严重混淆, 几乎无法有效辨识样本的类别标号, 即样本数据近乎均匀分布。而观察图 4(b) 和图 5(b) 可以看出, 经离散小波变换提取的脑电特征可以形成一定的聚类效果, 不同类别之间的数据特征表现出差异性。有趣的是, 小波统计特征经自适应聚合后, 相同类别的特征具有抱团趋势, 且不同类别样本的特征之间具有良好的区分性, 如图 4(c) 所示。类似的结论在非平衡数据集 NMT 上也可以观察到, 如图 5(c) 所示。更有趣的是, 我们发现, ASLR 能够最终选取的数据特征具有更好的聚类效果, 且分类边界也更加明显、清晰, 如图 5(d) 所示。不同类别间的特征越易分割, 则对其的分类效果就越好。因此, 图 4 和图 5 给出的实验结果再一次验证了新方法能够通过多尺度特征融合和基于结构稀疏回归的特征选择来获得具有高度区分性的数据特征表示。

此外, 为了更好地理解新方法的性能, 以 TUHEAC 数据集为例, 将由 ASLR 模型学习得到的重要特征以脑地形图的方式进行可视化, 如图 6 所示。从图中可以清楚地看出, 患有脑部疾病的人群与正常人群的脑地形图具有显著差异。其

中, 患病人群的脑电特征值在部分大脑区域明显比正常对照组更大, 特别是在额叶区域。有趣的是, 这些发现与现有基于脑电信号的神经系统疾病研究相一致<sup>[34-36]</sup>。例如, 癫痫发作波最常见于患者的额叶区域, 而其他脑部区域尖峰放电的发生率小于 10%<sup>[35]</sup>。这是因为额叶在认知与情感调节方面起着至关重要的作用。此外, 从图 6 也可以观察到, 在异常脑电组中, 由于患者的脑功能紊乱导致不同导联上的平均特征值之间存在较大的差异性, 这与 Zhao 等<sup>[37]</sup> 的分析结果相一致。因此, 以上实验结果从特征层面再次验证了 ASLR 模型是一种有效的异常脑电识别方法。

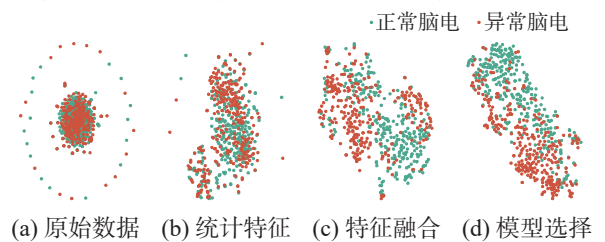


图 4 TUHEAC 数据集上不同特征的可视化结果  
 Fig. 4 Visualization results of different features on the TUHEAC dataset

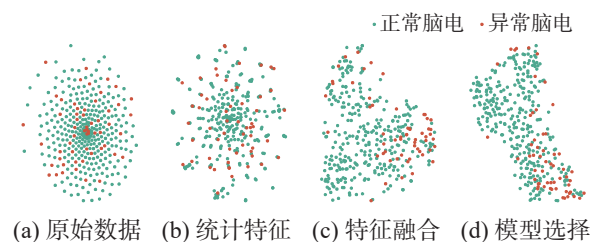


图 5 NMT 数据集上不同特征的可视化结果  
 Fig. 5 Visualization results of different features on the NMT dataset

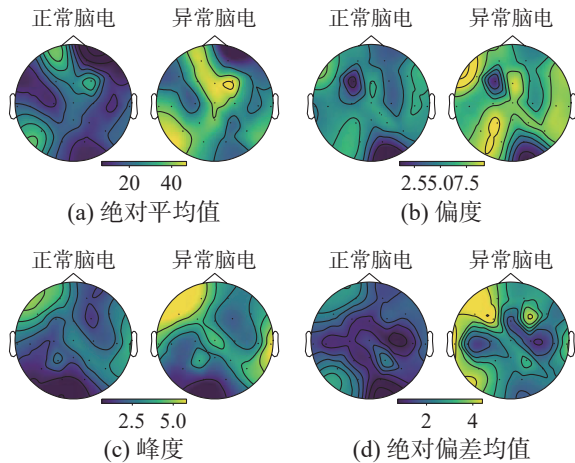


图 6 TUHEAC 数据集上不同特征的脑地形图

Fig. 6 Topographical distribution of different statistical features on the TUHEAC dataset

### 3 结束语

本文提出一种自适应的方法 ASLR, 用于对蕴含多种脑疾病信息的高维复杂脑电数据进行分类分析。通过多尺度特征融合技术, 可以以无监督方式从原始脑电数据中学习出具有高鉴别性的数据表征; 接着定义了一个新的自适应结构稀疏回归模型, 能够从特征中充分挖掘出有效鉴别模式的成组重要特征的同时, 实现精准的脑电二分类。与现有其他异常脑电识别方法相比, ASLR 提出的两种特征权重构造策略仅仅依赖于特征的概率分布, 并非是依赖数据的真实值, 因而新方法不仅能够以数据自驱动的方式选择重要特征, 还可以改善甚至避免现有特征选择模型中的有偏估计。本文所提方法充分利用了不同视图特征的互补性以及特征的群组性质, 通过在实际应用数据上进行对比实验, 结果表明了 ASLR 可以有效地进行异常脑电识别, 其分类效果优于新近提出的有关方法。下一步的工作重点是寻找能够根据数据样本分布自适应学习出模型可调参数相应最优值的方法, 及对异常值和产生较大误差的可能原因进行深入研究, 以进一步提高方法的稳定性。

### 参考文献:

[1] 王振宇, 向泽锐, 支锦亦. 离散小波变换和自编码器耦合的脑电信号异常检测方法[J]. 北京邮电大学学报, 2024, 47(2): 66-73.  
WANG Zhenyu, XIANG Zerui, ZHI Jinyi. Detection method for abnormal electroencephalographic signals coupled with discrete wavelet transform and autoencoder [J]. Journal of Beijing University of Posts and Telecommunications, 2024, 47(2): 66-73.

[2] 张军鹏, 施玉杰, 蒋睿, 等. 基于脑电信号的认知功能障碍识别与分类进展综述[J]. 计算机应用, 2023, 43(10): 3297-3308.

ZHANG Junpeng, SHI Yujie, JIANG Rui, et al. Review on advances in recognition and classification of cognitive impairment based on EEG signals[J]. Journal of computer applications, 2023, 43(10): 3297-3308.

[3] LÓPEZ DE DIEGO S. Automated interpretation of abnormal adult electroencephalograms[D]. Philadelphia: Temple University, 2017.

[4] MEI Ning, GROSSBERG M D, NG K, et al. Identifying sleep spindles with multichannel EEG and classification optimization[J]. Computers in biology and medicine, 2017, 89: 441-453.

[5] GEMEIN L A W, SCHIRRMESTER R T, CHRA-BAŚSZCZ P, et al. Machine-learning-based diagnostics of EEG pathology[J]. NeuroImage, 2020, 220: 117021.

[6] TANG Fengzhen, ADAM L, SI Bailu. Group feature selection with multiclass support vector machine[J]. Neuro-computing, 2018, 317: 42-49.

[7] WANG Min, YIN Xuefei, ZHU Yanming, et al. Representation learning and pattern recognition in cognitive biometrics: a survey[J]. Sensors, 2022, 22(14): 5111.

[8] ALBAQAMI H, HASSAN G M, SUBASI A, et al. Automatic detection of abnormal EEG signals using wavelet feature extraction and gradient boosting decision tree[J]. Biomedical signal processing and control, 2021, 70: 102957.

[9] GOWTHAM R N, HAIT S R, GUHA D, et al. Classification of epileptic EEG signals with the utilization of Bonferroni mean based fuzzy pattern tree[J]. Expert systems with applications, 2024, 239: 122424.

[10] SHARMA M, PATEL S, ACHARYA U R. Automated detection of abnormal EEG signals using localized wavelet filter banks[J]. Pattern recognition letters, 2020, 133: 188-194.

[11] IDOWU O P, ADELOPO O, ILESANMI A E, et al. Neuro-evolutionary approach for optimal selection of EEG channels in motor imagery based BCI application[J]. Biomedical signal processing and control, 2021, 68: 102621.

[12] ZHONG Yunning, WEI Hongyu, CHEN Lifei, et al. Automated EEG pathology detection based on significant feature extraction and selection[J]. Mathematics, 2023, 11(7): 1619.

[13] 刘建伟, 崔立鹏, 罗雄麟. 结构稀疏模型[J]. 计算机学报, 2017, 40(6): 1309-1337.  
LIU Jianwei, CUI Lipeng, LUO Xionglin. Structured sparse models[J]. Chinese journal of computers, 2017, 40(6): 1309-1337.

[14] ZHENG Wenming. Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis [J]. IEEE transactions on cognitive and developmental systems, 2016, 9(3): 281-290.

[15] 徐雪远, 刘建红, 李子遇, 等. 基于正交回归和特征加权的脑电情感特征选择方法[J]. 中国科学: 信息科学, 2023, 53(1): 33-45.  
XU Xueyuan, LIU Jianhong, LI Ziyu, et al. EEG emotional feature selection method based on orthogonal regression and feature weighting[J]. Scientia sinica (informationis), 2023, 53(1): 33-45.

- [16] ALGAMAL Z Y, LEE M H. Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification[J]. *Expert systems with applications*, 2015, 42(23): 9326–9332.
- [17] LIU Cheng, WONG H S. Structured penalized logistic regression for gene selection in gene expression data analysis[J]. *IEEE/ACM transactions on computational biology and bioinformatics*, 2019, 16(1): 312–321.
- [18] WANG Yadi, ZHANG Wenbo, FAN Minghu, et al. Regression with adaptive lasso and correlation based penalty[J]. *Applied mathematical modelling*, 2022, 105: 179–196.
- [19] BLANCO DÍAZ C F, RUIZ OLAYA A F. A novel method based on regularized logistic regression and CCA for P300 detection using a reduced number of EEG trials[J]. *IEEE Latin America transactions*, 2020, 18(12): 2147–2154.
- [20] GARCÉS P, BAUMEISTER S, MASON L, et al. Resting state EEG power spectrum and functional connectivity in autism: a cross-sectional analysis[J]. *Molecular autism*, 2022, 13(1): 22.
- [21] WANG Yadi, YANG Xinguang, LU Yongjin. Informative gene selection for microarray classification via adaptive elastic net with conditional mutual information[J]. *Applied mathematical modelling*, 2019, 71: 286–297.
- [22] PATIL A R, PARK B K, KIM S. Adaptive lasso with weights based on normalized filtering scores in molecular big data[J]. *Journal of theoretical and computational chemistry*, 2020, 19(4): 2040010.
- [23] PARK H, SHIRAIISHI Y, IMOTO S, et al. A novel adaptive penalized logistic regression for uncovering biomarker associated with anti-cancer drug sensitivity[J]. *IEEE/ACM transactions on computational biology and bioinformatics*, 2017, 14(4): 771–782.
- [24] LI Juntao, WANG Yadi, XIAO Huimin, et al. Gene selection of rat hepatocyte proliferation using adaptive sparse group lasso with weighted gene co-expression network analysis[J]. *Computational biology and chemistry*, 2019, 80: 364–373.
- [25] WANG Yadi, LI Xiaoping, RUIZ R. Feature selection with maximal relevance and minimal supervised redundancy[J]. *IEEE transactions on cybernetics*, 2023, 53(2): 707–717.
- [26] BOU A E, NGUYEN D K, RIHANA S, et al. Towards accurate prediction of epileptic seizures: a review[J]. *Bio-medical signal processing and control*, 2017, 34: 144–157.
- [27] YAGER R R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking[J]. *IEEE transactions on systems, man, and cybernetics*, 1988, 18(1): 183–190.
- [28] WANG Lianxi, JIANG Shengyi, JIANG Siyu. A feature selection method via analysis of relevance, redundancy, and interaction[J]. *Expert systems with applications*, 2021, 183: 115365.
- [29] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization paths for generalized linear models via coordinate descent[J]. *Journal of statistical software*, 2010, 33(1): 1–22.
- [30] KHAN H A, AIN R U, KAMBOH A M, et al. The NMT scalp EEG dataset: an open-source annotated dataset of healthy and pathological EEG recordings for predictive modeling[J]. *Frontiers in neuroscience*, 2022, 15: 755817.
- [31] ROY S, KIRAL-KORNEK I, HARRER S. ChronoNet: a deep recurrent neural network for abnormal EEG identification[M]//Artificial Intelligence in Medicine. Cham: Springer International Publishing, 2019: 47–56.
- [32] ZHANG Zhongheng, TREVINO V, HOSEINI S S, et al. Variable selection in Logistic regression model with genetic algorithm[J]. *Annals of translational medicine*, 2018, 6(3): 45.
- [33] MAATEN L V D, HINTON G. Visualizing data using t-SNE[J]. *Journal of machine learning research*, 2008, 9(86): 2579–2605.
- [34] KHARE S K, ACHARYA U R. Adazd-Net: Automated adaptive and explainable Alzheimer’s disease detection system using EEG signals[J]. *Knowledge-based systems*, 2023, 278: 110858.
- [35] YASUHARA A. Correlation between EEG abnormalities and symptoms of autism spectrum disorder (ASD)[J]. *Brain and development*, 2010, 32(10): 791–798.
- [36] ALDEMIR R, DEMIRCI E, PER H, et al. Investigation of attention deficit hyperactivity disorder (ADHD) sub-types in children via EEG frequency domain analysis[J]. *The international journal of neuroscience*, 2018, 128(4): 349–360.
- [37] ZHAO Qinglin, JIANG Hua, HU Bin, et al. Nonlinear dynamic complexity and sources of resting-state EEG in abstinent heroin addicts[J]. *IEEE transactions on nanobioscience*, 2017, 16(5): 349–355.

#### 作者简介:



吴涛, 讲师, 博士, 主要研究方向为机器学习、异常脑电分析以及模式识别。参与国家和省部级科研项目 3 项。发表学术论文 15 余篇。E-mail: [wtao@fafu.edu.cn](mailto:wtao@fafu.edu.cn)。



刘夏, 硕士研究生, 主要研究方向为脑电信号处理、数据挖掘。E-mail: [52362047002@fafu.edu.cn](mailto:52362047002@fafu.edu.cn)。



孔祥增, 教授, 博士生导师, 国家级人才, 主要研究方向为机器学习、脑机接口系统。主持和参与国家和省部级科研项目 10 项。发表学术论文 40 余篇。E-mail: [xzkong@fafu.edu.cn](mailto:xzkong@fafu.edu.cn)。