



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331,CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目：集成分类器对脑卒中患者脑电的分类  
作者：王方, 张雪英, 胡风云, 李凤莲  
网络首发日期：2020-10-12  
引用格式：王方, 张雪英, 胡风云, 李凤莲. 集成分类器对脑卒中患者脑电的分类. 计算机工程与应用.  
<https://kns.cnki.net/kcms/detail/11.2127.TP.20201012.0942.002.html>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 集成分类器对脑卒中患者脑电的分类

王方<sup>1</sup>, 张雪英<sup>1</sup>, 胡风云<sup>2</sup>, 李凤莲<sup>1</sup>

1.太原理工大学信息与计算机学院, 太原 030024

2.山西省人民医院神经内科, 太原 030024

**摘要:** 脑卒中患者意识障碍的检查和检测耗时耗力且非连续, 采集脑卒中患者的脑电信号, 以研究有意识障碍与无意识障碍的脑卒中患者的自动分类。对脑卒中患者的脑电图提取多达9种定量脑电特征, 构建脑网络, 将这些脑网络的连通性特征输入到分类器中, 实现对脑卒中患者是否有意识障碍的分类。为解决非平衡数据集分类时严重偏向多数类的问题, 设计集成支持向量机分类器。实验结果显示基于现有分类器的脑卒中意识障碍的分类正确率在70%左右, 敏感度在40%以下; 而基于集成支持向量机分类器的分类准确性可达96.79%, 同时敏感度和特异性分别为95.45%和100%。实验结果表明集成支持向量机分类器对非平衡数据集的脑电分类准确率显著提升, 并促进脑卒中患者意识障碍的自动识别。

**关键词:** 脑卒中; 集成分类器; 定量脑电图特征; 脑网络; 意识障碍; 机器学习

文献标志码: Adoi: 10.3778/j.issn.1002-8331.2006-0328

王方, 张雪英, 胡风云, 等. 集成分类器对脑卒中患者脑电的分类. 计算机工程与应用

WANG Fang, ZHANG Xueying, HU Fengyun, et al. Ensemble Method Classifies EEG from Stroke Patients. Computer Engineering and Applications

## Ensemble Method Classifies EEG from Stroke Patients

WANG Fang<sup>1</sup>, ZHANG Xueying<sup>1</sup>, HU Fengyun<sup>2</sup>, LI Fenglian<sup>1</sup>

1. School of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

2. Department of Neurology, Shanxi Provincial People's Hospital, Taiyuan 030024, China

**Abstract:** The detection of disorders of consciousness (DoC) based on clinical experiments is time and manpower consuming and these experiments are intervalic. The Electroencephalogram (EEG) of stroke patients is collected to study the automatic classification of stroke patients with DoC and without DoC. As many as 9 kinds of quantitative EEG features are extracted to build brain network, and then the connectivity features of the brain network are imported into classifiers to classify DoC and no DoC in stroke patients. An ensemble of support vector machine (EOSVM) is designed to solve the problem that classifiers always tend to the majority classes in the classification on imbalanced dataset. The experimental results show that the existing classifiers can classify the stroke patients with DoC and without DoC with an accuracy of about 70% and a sensitivity lower than 40%, however, the EOSVM classifier gives an accuracy of 96.79%, the sensitivity of 95.45% and specificity of 100%. These results show that the

**基金项目:** 山西省重点研发计划(社会发展)项目(No.201803D31045); 吴阶平医学基金(No.320675016129); 山西省自然科学基金(No.201801D121138); 山西省回国留学人员科研资助项目(No.201925)。

**作者简介:** 王方(1988-), 女, 博士研究生, 研究领域为脑电信号分析、机器学习; 张雪英(1964-), 女, 博士, 教授, CCF 会员, 研究领域为情绪脑电信号分析、情绪语言识别, E-mail: tyzhangxy@163.com; 胡风云(1961-), 女, 硕士, 主任医师, 研究领域为脑血管疾病、癫痫; 李凤莲(1972-), 女, 博士, 教授, 研究领域为非平衡数据分类、大数据、脑梗。

classifier EOSVM can improve the classification of imbalanced EEG dataset, and help the automatic detection of DoC in stroke patients.

**Key words:** stroke patients; ensemble classifier; quantitative EEG; brain network; disorders of consciousness; machine learning

## 1 引言

大脑是人类中枢神经系统的最高级部分,是负责个体各种行为活动的高级神经中枢。人类的大脑中有数百亿的神经元,每个神经元通过数千个突触与其它神经元连接。人体的各个功能都是在神经系统的直接或间接调控下实现的。研究表明人类的大脑的不同区域或神经元组织会相应负责不同的行为功能。因此对人类大脑的研究可以促进许多神经系统疾病尤其是脑部疾病的研究和诊断。

头皮脑电图(Electroencephalogram, EEG)价格低廉无创,方便长时间监测,已成为许多研究机构及临床神经系统疾病监测的神经成像方式。定量脑电特征(Quantitative, QEEG)是脑电信号分析的一个重要方法,它借助计算机提取脑电信号频域和时域等多方面的定量特征,大量的研究表明QEEG在各种临床研究中有重要应用价值。例如脑功率谱对称指数(Brain Symmetry Index, BSI)与脑卒中患者的卒中程度指标之间存在正相关关<sup>[1]</sup>。相位滞后指数(Phase lag index, PLI)被用于研究中风患者<sup>[2]</sup>、痴呆患者<sup>[3]</sup>及帕金森<sup>[4]</sup>等患者的临床状况。脑电慢波功率比(Delta/Alpha, DAR)被证明能够有效区分缺血性脑卒中组和健康控制<sup>[5]</sup>。另外,加权相位滞后指数(Weighted phase lag index, WPLI)<sup>[6-8]</sup>、相位同步(Phase synchronization, PS)<sup>[9-12]</sup>和一致性的虚部(Imaginary component of coherence, IC)<sup>[13]</sup>等关于脑电相位的定量特征也被用于临床研究中。

脑卒中患者发病后经常会出现昏迷、昏睡、嗜睡等意识障碍,需要给予及时的治疗和干预,长期处于昏迷会给大脑带来不可恢复的损伤<sup>[14]</sup>。已经有多种临床意识量表用于评估意识损失和昏迷的深度及持续时间等。但这些神经学评估量表的评估是依靠医生通过一系列的临床检查和观察来完成的,这种方法需要花费大量人力和时间成本,并且依靠这种间歇性临床检查很难实时的了解患者的意识水平

的变化。意识状态的实时监测和意识障碍的及时发现对脑卒中患者的后期康复意义重大<sup>[15]</sup>。

本文提取了脑卒中被试者脑电图中的9种定量脑电图特征,将这9种定量脑电特征作为脑网络的边的权重构成脑网络。脑网络的连通性特征被输入到分类器中,对脑卒中患者是否有意识障碍进行自动二分类。

现已有多种机器学习方法应用于脑电信号的分类,如支持向量机(Support vector machine, SVM)<sup>[16]</sup>,邻近算法(K-Nearest Neighbor, KNN)<sup>[17]</sup>,集成装袋树(Bagging tree)<sup>[18]</sup>,集成提升树(Adaptive Boosting, AdaBoost)<sup>[19]</sup>等,但大多数并不适用于非平衡数据集的分类。

为提高非平衡数据集的分类准确性和敏感度,本文设计了集成支持向量机分类器(ensemble of support vector machine, EOSVM),并将其分类结果与其它分类器(SVM、KNN、AdaBoost及Bagging tree)的分类结果进行了对比。

## 2 脑电特征提取

### 2.1 脑电信号的采集及预处理

147例脑卒中患者(平均年龄 $\pm$ 标准偏差=64.76 $\pm$ 5.63)的脑电采集是在山西省人民医院神经内科完成的。用脑电图监测系统(太阳能2000 N,太阳能电子技术有限公司,北京,中国北京)记录大脑皮层自发的脑电图。根据国际10-20标准,电极点放置于FP1、FP2、C3、C4、O1、O2、T7、T8、A1和A2来采集脑电图信号。采样频率为100 Hz,阻抗小于10 k $\Omega$ 。脑电图记录与意识评估时间最大间隔是30 min。患者的意识状态是通过一系列的临床行为评估来确定的<sup>[20]</sup>。

表1概括了参与者的人口统计特征和意识状态。参与者的入选标准如下:(1)MRI或CT诊断为缺血性脑卒中,(2)有脑电图记录和相应的意识水平记录(3)18岁以上。排除标准如下:(1)脑出血(2)孕妇或哺乳期妇女。

表 1 参与者的人口统计和临床特征

Table 1 Demographics and clinical characteristics of subjects

意识状态 (分值)	参与者 数量	女性 数量	年龄
清醒 (1)	93	24	63.05±15.31
嗜睡 (2)	30	9	69.17±18.17
昏睡 (3)	11	5	65.29±11.22
浅昏迷 (4)	7	2	70.71±16.64
中昏迷 (5)	5	2	60.33±13.94
深昏迷 (6)	1	0	76.00±0.00

所有的脑电图预处理及分析都是在 EEGLAB 及 MATLAB 中进行的。预处理的步骤包括平均重参考、基本的有限脉冲响应对数据分别进行高通滤波 (0.5 Hz) 和低通滤波 (30 Hz) 及去伪迹处理。去伪迹处理采用的工具为 EEGLAB 中自动删除连续信号的伪迹工具。最终提前的定量脑电特征是在每个参与者的前 10 分钟长度的脑电信号上提取的。

## 2.2 定量脑电特征计算

本文提取的 11 种定量脑电特征包括: 脑功率谱对称指数 BSI, 左右脑相位同步指数 (Phase Synchrony Index of Left and Right Hemisphere, PSI-LR)、一致性 (Coherence, C)、一致性的实部特征 (Real Part of Coherence, RC) 和虚部 IC, 及四种不同频段的快慢波比例特征。9 种特征的具体计算如下文所示。

左右半脑相位同步指数 PSI-LR<sup>[20]</sup>评估了左半球和右半球之间的相位差异。相位同步指数 (Phase Synchrony Index, PSI) 是指来自同源电极对的两个脑电图信号之间的相位差异。要计算 PSI-LRs 和 PSIs, 首先要知道每个通道的瞬时相位。本文用希尔伯特变换<sup>[21]</sup>来计算脑电图信号的瞬时相位。PSI-LR 计算公式如下:

$$PSI-LR = \frac{1}{N} \sum_{k=0}^{N-1} \left| \exp(i(\phi_L(t_k) - \phi_R(t_k))) \right| \quad (1)$$

其中  $\phi_L(t_k)$  和  $\phi_R(t_k)$  分别表示左右一对对称通道的瞬时相位,  $i$  是虚数单位,  $t_k$  代表离散时间。

脑功率谱对称指数 (BSI) 是一种在功率谱上计算左右半球对称的脑电特征, 它是在 1~25 Hz 的频

率范围内计算得到的<sup>[1]</sup>。BSI 是一种被广泛应用的定量脑电特征, 它的计算公式如下:

$$BSI = \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{M} \sum_{j=1}^M \frac{R_{ij} - L_{ij}}{R_{ij} + L_{ij}} \right\| \quad (2)$$

其中  $R_{ij}$  ( $L_{ij}$ ) 是左 (右) 通道的脑电信号的傅里叶变换值,  $i$  代表通道对数,  $j$  是傅里叶系数。

一致性 (Coherence, C)<sup>[22]</sup> 是一种度量各通道信号线性相关性的定量脑电特征。它是在互功率谱密度基础上计算得到的。C 的定义如下:

$$C = \frac{|S_{uv}(f)|^2}{S_{uu}(f)S_{vv}(f)} \quad (3)$$

其中  $S_{uv}(f)$  是两个通道 EEG 信号  $u(t)$  和  $v(t)$  的互功率谱密度,  $S_{uu}(f)$  和  $S_{vv}(f)$  分别是信号  $u(t)$  和  $v(t)$  的自功率谱密度。

本文还计算了一致性 C 的实部和虚部, IC 代表 C 的虚部的绝对值, RC 代表 C 的实部的绝对值<sup>[22]</sup>。

$$IC = \text{Im } ag(C) \quad (4)$$

$$RC = \text{Real}(C) \quad (5)$$

脑电信号的快波与慢波所占比例是应用非常广泛的脑电频域特性分析方法。本文计算了四个不同频段的功率比值, 即  $\delta$  (1~3 Hz),  $\theta$  (4~7 Hz),  $\alpha$  (8~12 Hz) 和  $\beta$  (13~30 Hz) 频段。

DAR 是  $\delta/\alpha$  频段的功率比值, DTAR 是  $(\delta+\theta)/\alpha$  频段的功率比值, DTABR 是  $(\alpha+\beta)/(\delta+\theta)$  频段的功率比值, PRI 是  $(\delta+\theta)/(\alpha+\beta)$  频段的功率比值<sup>[5]</sup>。

## 2.3 脑网络连通性特征提取

大脑区域之间的关系可以描述为一个大脑网络, 其顶点和边分别对应于大脑区域及其连接。如果对边进行加权, 则可以用脑功能连接度量指标来表示各区域连接的强度。图 1 显示了来自 8 个 (顶点) 电极 (Fp1、Fp2、C3、C4、O1、O2、T7 和 T8) 的 28 条边 (图 1)。对这些边, 我们计算了 9 种定量脑电特征来作为这些边的权重来作为特征。



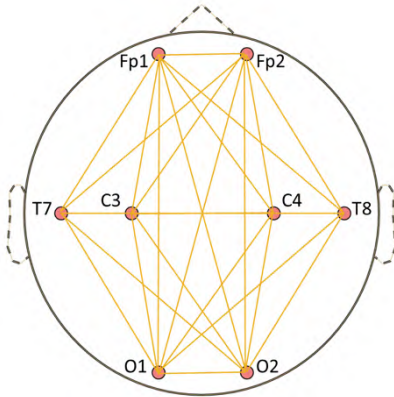


图1 脑网络

Fig.1 Brain network

定量特征 PSI-LR 和 BSI 是左右脑对称的指数，因此计算了它们的 4 对左右脑半球对称的指标做为连通性特征 (FP1-FP2、C3-C4、O1-O2、T7-T8)。其它 7 个定量脑电特征都分别计算其 28 个电极对对应的定量脑电特征，然后求 28 个电极对特征的平均值作为输入分类器的特征。因此输入下文分类器中的脑网络连通性特征是  $2 \times 4 + 7 = 15$  维的。

### 3 分类

现实世界中获取的数据经常是不平衡的，例如医院患者的疾病种类所占比例不平衡。本文试图分类脑电数据集也是严重不平衡的。147 名被试的脑电数据集中，有 93 例是清醒患者，54 例是有意识障碍的患者，前者几乎是后者的两倍。因此，利用普通的分类器会导致分类预测结果严重偏向数据集中的大多数（清醒类），从而导致整个分类预测的失败。

#### 3.1 集成分类器构建

为了能有效的对此非平衡数据集进行分类，本文设计了一个集成支持向量机 (Ensemble of support vector machine, EOSVM)，它是由  $N$  个普通支持向量机 (Support vector machine, SVM) 组成的。每个 SVM 会输入一个平衡的子训练集，SVM 的个数  $N$  可根据数据集的不平衡程度进行调整，数据集越不平衡， $N$  的数值需要设定的越大。

EOSVM 采用装袋 (Bagging) 方法将多个支持向量机组组合，构成一个强学习器，从而提高非平衡数据集的分类准确性。EOSVM 与普通装袋方法的数据分割方法不同，即生成子训练样本的方法不同。EOSVM 产生的所有子训练样本包含所有少数类样本，同时

随机选择同样数量的多数类样本，目标是解决非平衡数据集的分类问题。

图 2 展示了集成支持向量机 EOSVM 的结构框架，EOSVM 的分类分为三步：数据准备、分类器的训练和预测。

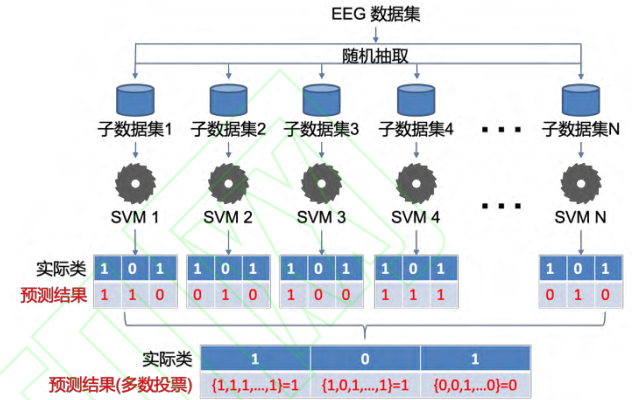


图2 分类器结果框架

Fig.2 Framework of EOSVM classifier

#### (1) 数据准备

数据准备阶段需要将样本集分离为多个子样本集。首先将整个数据集分离为训练数据集和测试数据集。本实验中训练集和测试集中数据数量的比例设为 80% 和 20%。之后需要将训练集分为多个子训练集，构成每个子样本集中的训练集。而每个子样本集中的测试集是相同的。

图 3 为子样本集生成的示意图。蓝色与橙色为两类样本，蓝色为样本集中的多数类，橙色为样本集中的少数类。每个子样本集中的少数类（橙色）样本相同，为训练集中的所有少数类（橙色）样本。每个子样本集中的多数类（蓝色）不相同，子训练集中的蓝色类是从总训练集中的蓝色类中随机选择的，且数量与子训练集中的橙色类相同。这样构成的每个子训练集是不同的，但又充分利用了少数类（橙色）样本，并且每个子训练集是平衡的（橙色与蓝色类数量相同）。

#### (2) 分类器训练

每个 SVM 会输入一个平衡的子样本集，每个子样本集中包含相同数量的无意识障碍患者和有意识障碍患者。

分类器训练阶段，SVM 的核函数 (Kernel function) 选取了高斯核函数 (Gaussian kernel function)。超参数是通过使间隔最大化来得到的。

### (3) 预测

在分类器的预测阶段, 每个 SVM 会对相同的测试集给出一个预测结果。如果 EOSVM 中包含 100 个 SVM, 则测试数据集中每个被试会得到 100 个预测结果(分别来自 100 个 SVM)。根据这 100 个预测结果, EOSVM 给测试集中的被试一个最终的预测结果。

最终的预测结果会因为 EOSVM 中多数投票 (Majority voting) 的阈值不同而改变。如果投票规则的阈值设定为 50%, 则测试集中的被试者的 100 个预测结果中大于 50% 的那一类预测结果将作为其最终预测结果。

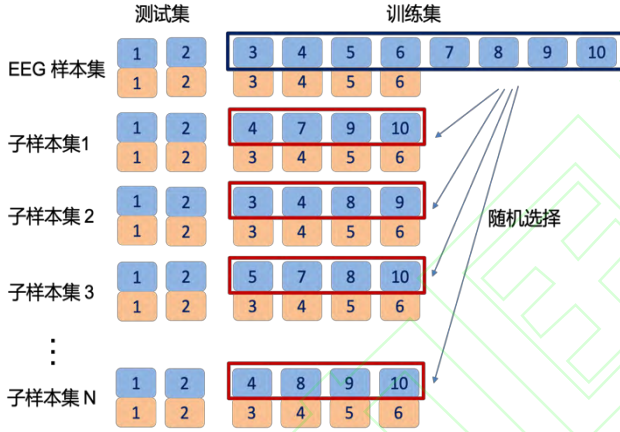


图3 数据准备

Fig.3 Preparefor dataset

## 3.2 分类效果评价

本文采用了三个参数来对分类学习算法效果进行评价, 即准确性 (Accuracy)、敏感度 (Sensitivity) 和特异性 (Specificity)。准确性代表所有参与者的预测总准确率, 敏感性代表有意识障碍参与者的预测准确率, 特异性代表无意识障碍参与者预测的准确性。三个评价参数的计算方式如下:

$$Accuracy = \frac{T_{awa} + T_{DoC}}{N_{awa} + N_{DoC}} \times 100\% \quad (6)$$

$$Sensitivity = \frac{T_{DoC}}{N_{DoC}} \times 100\% \quad (7)$$

$$Specificity = \frac{T_{awa}}{N_{awa}} \times 100\% \quad (8)$$

其中,  $N_{awa}$  为测试集中清醒被试者的数量,  $N_{DoC}$  为测试集中存在意识障碍的被试者的数量,  $T_{awa}$  为测试集中被正确分类的清醒被试者的数量,  $T_{DoC}$  为测试集中被正确分类的患有意识障碍的被试者的数量。

## 4 实验结果与分析

### 4.1 统计分析

统计分析是在 IBM SPSS 软件及 R-Studio 中进行的。本文用皮尔逊相关性系数分析了各个定量脑电特征和患者意识水平的相关性水平。

为了分析临床状态和定量脑电特征之间的关系, 我们给中风患者的意识水平 (Level of consciousness, LOC) 赋值为 1 到 6。清醒患者意识水平设定为 1, 嗜睡患者意识水平为 2, 昏睡为 3, 浅昏迷为 4, 中昏迷为 5, 深昏迷为 6。意识水平值越高, 代表意识障碍越严重。下面的分析是在如上的赋值基础上进行。

图 4 和图 5 显示了脑卒中患者的意识水平 (Level of Consciousness, LOC) 与各个定量脑电特征之间的相关关系水平。图中显示的数字为皮尔逊相关系数。皮尔逊相关系数右上角的红色星号代表显著水平。一颗红色星号代表显著水平  $P=0.05$ ; 两颗红色星号代表显著水平  $P=0.01$ ; 三颗星代表显著水平  $P=0.001$ 。

图 4 第一行显示了 LOC 与定量脑电特征 PSI-LR (三个频段上提取:  $\alpha$  (8~12 Hz)、 $\beta$  (13~30 Hz) 和 1~30 Hz)、C、IC 和 RC 之间的相关关系。观察图可发现 PSI-LR 在  $\alpha$ 、 $\beta$  和 1~30 Hz 频段都与意识水平存在显著正相关性, 但  $\beta$  频段上提取的 PSI-LR

与意识障碍水平相关性最显著 (PSI-LR(c):  $\rho = 0.76$ ,  $P=0.001$ ; PSI-LR( $\alpha$ ):  $\rho = 0.54$ ,  $P=0.01$ ; PSI-LR(1~30Hz):  $\rho = 0.58$ ,  $P=0.01$ )。因此下文分类实验中的 PSI-LR 是在  $\beta$  频段上提取的。C 与意识水平 LOC 的相关关系同样显著 ( $\rho = 0.60$ ,  $P=0.01$ )。然而, IC 及 RC 与患者的意识状态相关性不显著。

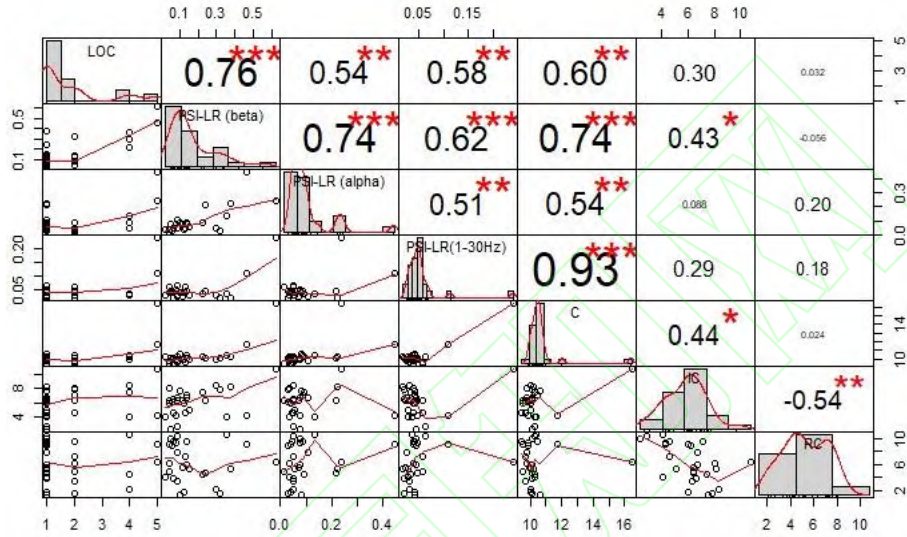


图 4 LOC 与定量脑电特征的相关矩阵图

Fig.4 Visualize correlation matrix of LOC and QEEGs

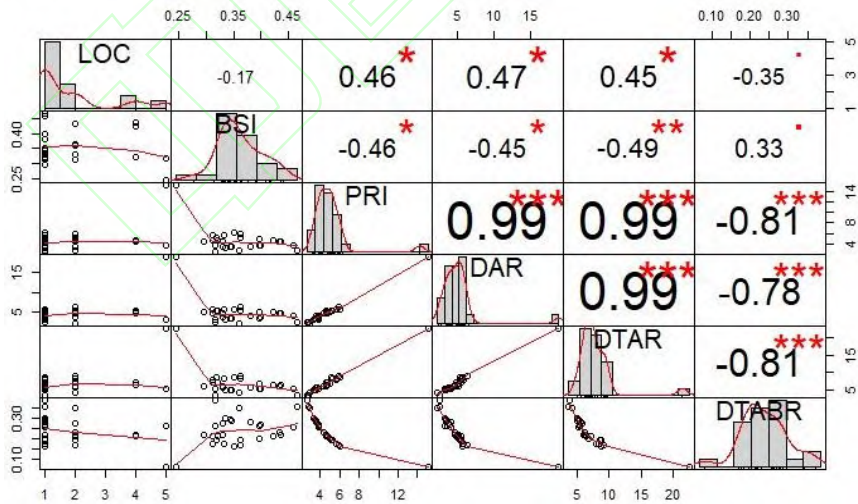


图 5 LOC 与定量脑电特征的相关矩阵图

Fig.5 Visualize correlation matrix of LOC and QEEGs

图 5 第一行显示了 LOC 与定量脑电特征 BSI 及慢波比例特征 PRI、DAR、DTAR、DTABR 之间的相关关系。皮尔逊相关系数显示 PRI、DAR、DTAR 与 LOC 呈现显著正相关 (PRI:  $\rho = 0.46$ ,  $P=0.05$ ; DAR:  $\rho = 0.47$ ,  $P=0.05$ ; DTAR:  $\rho = 0.45$ ,  $P=0.05$ )。而 DTABR

与 LOC 呈负相关, BSI 与 LOC 无显著相关关系。虽然 BSI 与脑卒中患者的卒中程度 (NIHSS 指数) 相关<sup>[1]</sup>, 但与患者的意识水平并无显著相关关系。



## 4.2 分类结果与分析

分类实验针对有意识障碍（嗜睡、昏睡、浅昏迷、中昏迷、深昏迷）和无意识障碍（清醒）两类进行二分类。实验中，15 维的脑网络连通性特征被输入到分类器中。

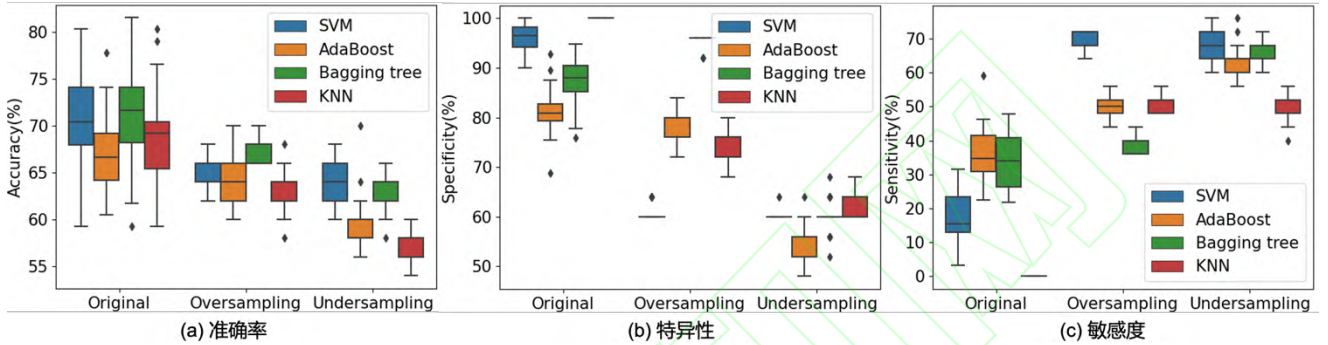


图 6 四种分类器结果对比

Fig.6 Classification results from four classifiers

由图 6 中原始数据集 (Original) 的结果可看出，四个分类器的准确率 (Accuracy) 在 70% 左右，敏感度 (Sensitivity) 在 40% 以下，特异性 (Specificity) 在 80% 以上。四个分类器的分类结果普遍存在敏感度过低的问题。敏感度过低的原因是这些分类器将测试集中的大部分被试者预测为了清醒类，即在原始数据集中占大多数的这一类。从而导致分类结果特异性很高，但敏感度很低。这将导致大量有意识障碍的患者被预测为无意识障碍，从而导致治疗时机的延误。

针对于非平衡数据集分类中偏向多数类的问题，主要采用的解决方法是重采样，即利用过采样 (Oversampling) 或欠采样 (Undersampling) 等方法对原始数据集进行重采样，从而构造一个平衡的数据集。之后，再利用分类器在这个重构的平衡的数据集上进行分类。

本实验也利用过采样和欠采样的方式对原始数据集进行了重构，之后用上述四种分类器进行了二分类实验，实验结果见图 6 中的 Oversampling 和 Undersampling 结果。观察图 6 (c) 可发现，重采样提高了分类的敏感度 (Sensitivity)，尤其是 SVM 的分类敏感度的提高到接近 70%。但图 6 (a) 显示，重采样后的分类准确率普遍下降，并且大多数下降到了 65% 以下。

因此，只利用重采样并不能解决本文非平衡数据集分类存在的敏感度过低和准确率偏低的问题。

### 4.2.1 现有分类器分类

首先利用四种分类器 (SVM、KNN、AdaBoost 及 Bagging tree) 对 147 例患者构成的脑电数据集进行二分类，分类结果见图 6 (20 次训练结果的统计)。

### 4.2.2 集成分类器 EOSVM 分类

实验利用随机欠采样方式 (RandomUndersampling) 将随机重构的多个平衡数据集输入到多个独立的分类器中，再用多数投票的方式将它们组成集成分类器。鉴于图 6 (c) 显示 SVM 在重采用后的平衡数据集上的敏感度最高，且 SVM 是小数据集分类的经典算法，本文利用 SVM 构建了集成分类器 EOSVM。

实验中 EOSVM 中 SVM 的个数设定为 40，EOSVM 的多数类投票阈值设定为 90% 到 100% 之间。此时分类器的分类结果如图 7 所示 (20 次训练结果的平均值和方差统计)。

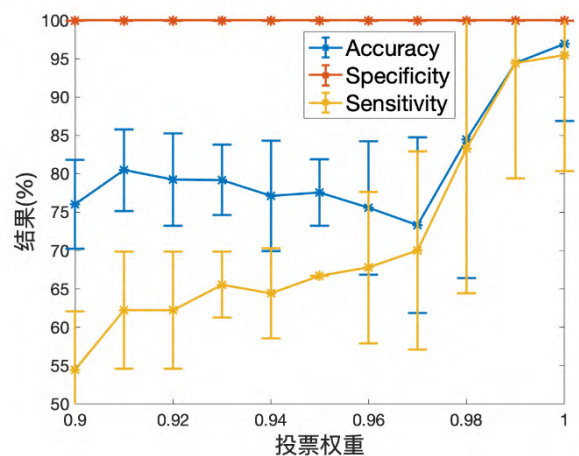


图 7 集成分类器 EOSVM 分类结果



Fig.7 Classification results from classifier EOSVM

从图 7 中的结果可观察到, 随着 EOSVM 投票规则中投票权重的增大, 分类效果 (包括准确率、敏感度、特异性) 逐渐提高。当 EOSVM 的投票权重为 100% 时, 分类预测的效果最好, 总准确率达到 95.45%, 敏感度 96.79%, 特异性 100%, 此时可预测患者比例为 23%。

EOSVM 的分类结果比上文中分类器 SVM、KNN、AdaBoost 及 Bagging tree 的分类准确性和敏感度有显著提升。EOSVM 对脑卒中患者意识障碍的识别分类将为医生的诊断提供辅助信息。

## 5 结束语

不平衡数据集的分类经常出现的问题是, 在预测时将所有被试者预测为数据集中的大多数类别 (即本实验中的无意识障碍类), 从而导致整个分类预测的失败。本文设计的集成支持向量机 EOSVM 有效提高了非平衡数据集分类的准确性。

本文利用定量脑电图特征构建的脑网络及 EOSVM 分类器可以以 96.97% 的准确率, 95.45% 的敏感度和 96.97% 的特异性预测出 23% 的脑卒中患者的意识障碍。利用脑电图监测评估脑卒中患者意识障碍的方法, 可为医生的诊断提供辅助信息。

## 参考文献:

- [1] deVos CC, van Maarseveen SM, Brouwers PJ, et al. Continuous EEG monitoring during thrombolysis in acute hemispheric stroke patients using the brain symmetry index[J]. *Journal of Clinical Neurophysiology*, 2008, 25(2): 77-82.
- [2] T. Kawano, N. Hattori, M. Hatakenaka, et al. Comparison of EEG Synchrony Measures for Post-Stroke Neurorehabilitation[C]//IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2018, pp. 35-38.
- [3] M. Dauwan, M.M. Linszen, A.W. Lemstra, et al. EEG-based neurophysiological indicators of hallucinations in Alzheimer's disease: Comparison with dementia with Lewy bodies[J]. *Neurobiology of Aging*, 2018, 67: 75-83.
- [4] M. Chaturvedi, J. Bogaarts, F. Hatz, et al. F67. Distinguishing Parkinson's Disease Dementia (PDD) patients from Parkinson's Disease (PD) patients using EEG frequency and connectivity measures[J]. *Clinical Neurophysiology*, 2018, 129: e92.
- [5] Finnigan S, Wong A, Read SJCN. Defining abnormal slow EEG activity in acute ischaemic stroke: Delta/alpha ratio as an optimal QEEG index[J]. *Clinical Neurophysiology*, 2016, 127(2): 1452-1459.
- [6] J. Marquetand, S. Vannoni, M. Carboni, et al. Reliability of MEG and hd-EEG resting-state functional connectivity metrics[J]. *Brain Connectivity*, 2019, 9(7).
- [7] C. Tramonti, L.S. Imperatori, C. Fanciullacci, et al. Predictive value of EEG connectivity measures for motor training outcome in multiple sclerosis: an observational longitudinal study[J]. *European Journal of Physical and Rehabilitation Medicine*, 2018, 10: 30370753.
- [8] L.S. Imperatori, M. Betta, L. Cecchetti, et al. EEG functional connectivity metrics wPLI and wSMI account for distinct types of brain functional interactions[J]. *Scientific Reports*, 2019, 9(1): 8894.
- [9] M. Fujiwara, T. Hashimoto, G. Li, et al. Changes in Phase Synchronization of EEG During Development of Symbolic Communication Systems[M]. Springer, 2018: 327-333.
- [10] A. Takamiya, J. Hirano, B. Yamagata, et al. Electroconvulsive therapy modulates resting-state EEG oscillatory pattern and phase synchronization in nodes of the default mode network in patients with depressive disorder[J]. *Frontiers in Human Neuroscience*, 2019, 13: 1.
- [11] H. Yokoyama, I. Nambu, J. Izawa, Y. Wada, et al. Alpha Phase Synchronization of Parietal Areas Reflects Switch-Specific Activity During Mental Rotation: An EEG Study[J]. *Frontiers in Human Neuroscience*, 2018, 12: 259.
- [12] S. Ahn, H. Cho, M. Kwon, et al. Interbrain phase synchronization during turn-taking verbal interaction—a hyperscanning study using simultaneous EEG/MEG[J]. *Human Brain Mapping*, 2018, 39(1): 171-188.
- [13] B. Scally, M.R. Burke, D. Bunce, et al. Resting-state EEG power and connectivity are associated with alpha peak frequency slowing in healthy aging[J]. *Neurobiology of Aging*, 2018, 71: 149-155.
- [14] Suwatcharakoon S, Meyers E, Falo C, et al. Loss of Consciousness at Onset of Subarachnoid Hemorrhage as an Important Marker of Early Brain Injury[J]. *JAMA Neurology*, 2016, 73(1): 28-35.
- [15] Claassen J, Velazquez A, Meyers E, et al. Bedside quantitative electroencephalography improves assessment of consciousness in comatose subarachnoid hemorrhage patients[J]. *Annals of Neurology*, 2016, 80(4): 541-553.
- [16] Kang J, Han X, Song J, et al. The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data[J]. *Computers in Biology and Medicine*, 2020, 120: 103722.
- [17] Singh N, Dehuri S. Multiclass classification of EEG signal for epilepsy detection using DWT based SVD and fuzzy kNN classifier[J]. *Intelligent Decision Technologies*, 2020(2): 1-14.

- [18] 陈景霞, 郑茹, 贾小云. 共空间模式结合小波包分解的脑电情感分类 [J]. 计算机工程与应用, 2019, 55(1):149-153.  
Chen J, Zheng R, Jia X, et al. EEG Emotion Classification Based on Common Spatial Patterns and Wavelet Packet Decomposition[J]. Computer Engineering and Applications, 2019, 55(1):149-153.
- [19] 韩敏, 孙卓然. 基于小波变换和 AdaBoost 极限学习机的癫痫脑电信号分类 [J]. 计算机应用, 2015, 35(9):2701-2705.  
Han M, Sun Z. Epileptic EEG signals classification based on wavelet transform and AdaBoost extreme learning machine[J]. Journal of Computer Application, 2015, 35(9):2701-2705.
- [20] Wang F, Zhang X, Hu F, et al. Using phase synchrony index for improved assessment of consciousness in ischemic stroke patients[J]. IEEE Access, 2019, 7: 30252-30260.
- [21] BrunsAJJonm. Fourier- Hilbert-and wavelet-based signal analysis: are they really different approaches?[J]. Journal of Neuroscience Methods, 2004, 137(2): 321-332.
- [22] Nicolo P, Rizk S, Magnin C, et al. Coherent neural oscillations predict future motor and language improvement after stroke[J]. Brain, 2015, 138(10): 3048-3060.