DOI:10.3969/j.issn.1005-202X.2020.08.013

医学影像物理

基于随机森林算法的阿尔茨海默症医学影像分类

李长胜，王瑜，肖洪兵，邢素霞
北京工商大学计算机与信息工程学院，北京 100048

【摘要】为实现阿尔茨海默症(AD)的医学影像分类，辅助医生对患者的病情进行准确判断，本研究对采集的34名AD患者、35名轻度认知障碍患者和35名正常对照组成员的功能磁共振影像进行特征提取和分类，具体思路包括：首先利用皮尔逊相关系数计算脑区之间的功能连接，然后采用随机森林算法对被试不同脑区之间的功能连接进行重要性度量及特征选择，最后使用支持向量机分类器进行分类，利用十倍交叉验证估算分类准确率。实验结果显示，随机森林算法可以对功能连接特征进行有效分析，同时得到AD发病过程的异常脑区，基于随机森林和SVM建立的分类模型对AD、轻度认知障碍的识别具有较好的效果，分类准确率可达90.68%，相关结论可以为AD的早期临床诊断提供客观参照。

【关键词】阿尔茨海默症；功能磁共振成像；随机森林；特征选择
【中图分类号】R318；R455.2 【文献标志码】A 【文章编号】1005-202X(2020)08-1005-05

Medical image classification for Alzheimer's disease diagnosis based on random forest algorithm

LI Changsheng, WANG Yu, XIAO Hongbing, XING Suxia
School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

Abstract: For accurately classifying the medical images of Alzheimer's disease (AD) and assisting the doctors in making an accurate diagnosis of the patient's condition, a computer-aided diagnosis method is proposed based on random forest algorithm. The functional magnetic resonance imaging (fMRI) data of 34 AD patients, 35 patients with mild cognitive impairment (MCI) and 35 normal controls are collected for feature extraction and classification. Firstly, the functional connections between different brain regions are calculated using Pearson correlation coefficient. Then the importance of the functional connections between different brain regions is assessed and important features are selected by random forest algorithm. Finally, support vector machine classifier is used for classification, and ten-fold cross-validation for estimating the classification accuracy. The experimental results show that random forest algorithm can be use to effectively analyze the functional connection characteristics and obtain the abnormal brain regions of AD pathogenesis. The classification model based on random forest and support vector machine has a good effect on the recognition of AD and MCI, with a classification accuracy of 90.68%. The related experimental results provide an objective reference for the early clinical diagnosis of AD.

Keywords: Alzheimer's disease; functional magnetic resonance imaging; random forest; feature selection

前 言

阿尔茨海默症(Alzheimer's Disease, AD)作为一种常见的大脑中枢神经系统退行性疾病，其发病率随着年龄的增长不断上升。随着当前社会人口老龄化的程度不断加深，AD的诊断、预防和治疗也引起了社会的广泛关注[1]。AD的典型症状包括记忆丧失、定向障碍、语言和行为障碍等，其发病过程是不可逆的，但是目前在AD早期阶段开展该疾病的预防和治疗是最有效的，因此AD的早期诊断在临床实践中具有很高的价值[2]。

现代成像技术为探索大脑区域之间的功能性相互作用提供了有效的方法，增加了对精神疾病病理基础的理解。神经影像技术，如磁共振成像(Magnetic Resonance Imaging, MRI)、功能磁共振成像(functional MRI, fMRI)、弥散张量成像(Diffusion Tensor Imaging, DTI)等已经广泛应用于轻度认知障碍(Mild Cognitive Impairment, MCI)和AD的研究[3]。脑功能网络方法提供了大脑相互作用模式的简化表示，在神经认知理论中的应用引起了普遍关注和认可，并广泛应用于脑疾

【收稿日期】2020-02-01
【基金项目】国家自然科学基金(61671028)；国家重大科技研发子课题(ZLJC6 03-5-1)；北京工商大学校级两科培育基金项目(19008001270)
【作者简介】李长胜，硕士研究生，研究方向：计算机视觉、医学图像处理、模式识别，E-mail: 516795305@qq.com
【通信作者】王瑜，博士，副教授，研究方向：计算机视觉、医学图像处理、模式识别，E-mail: wangyu@btbu.edu.cn

病的研究[4]。已有研究表明，大脑认知功能减退是一系列脑网络异常的体现，与大脑功能区的功能连接异常有一定关系[5]，在针对MCI患者的研究中，发现其脑默认网络（Default Mode Network, DMN）的完整性被破坏[6]。在脑功能连接网络的相关研究中，对于脑功能连接网络的建立大都基于通用的自动解剖标记（Anatomical Automatic Labeling, AAL）模板[7]和Brodmann分区模板[8]等。中国自动化所脑网络组研究中心利用MRI技术在获取大量脑影像样本的基础上，能够对脑结构和功能区进行精细划分，并制作出适用的人类脑网络组图谱（Brainnetome Atlas）[9]，包括246个精细脑区亚区，以及脑区亚区间的多模态连接模式。

计算机辅助诊断作为一种有价值的自动诊断工具，利用计算机辅助预测早期AD，协助医生进行临床诊断。基于fMRI数据结合机器学习和模式识别算法，可以有效地对不同被试进行分类[10]，Rabin等[11]从神经心理测量、患者陈述和家属告知三方面评估认知功能，并使用支持向量机（Support Vector Machine, SVM）模型对AD进行预测。李亚鹏等[12]采用脑功能网络模型研究AD患者大脑功能的变化；李慧卓等[13]采用Adaboost集成分类方法区分MCI、AD和NC的功能与结构磁共振成像数据，取得较好的分类结果。

在利用MRI数据诊断AD的过程中，其多种特征的高维度及其复杂的相互作用使得数据难以解释。随机森林具有处理高度非线性相关数据的能力，优势很多，例如对噪声具有较强的鲁棒性等[14]，在许多科学领域和其他神经系统疾病中被广泛应用，并取得良好的效果。2016年Guo等[15]通过试验证明了小脑萎缩与AD等常见神经疾病的关系。因此，本文通过对被试的fMRI图像数据提取包括小脑在内的全脑功能连接矩阵作为初始特征，并利用随机森林对初始特征进行特征选择，最后利用机器学习的方法进行分类，达到对AD的辅助诊断功能。

1 数据采集及预处理

1.1 数据采集

本研究共采集了104例被试的静息态fMRI图像数据，其中包括34名AD患者、35名MCI患者和35名正常对照（Normal Control, NC）组成员，fMRI数据全部来源于AD神经影像学（Alzheimer's Disease Neuroimaging Initiative, ADNI）数据库。数据库中每个被试的数据分别包含140幅图像，每幅图像扫描48层，所有被试样本的年龄、性别信息如表1所示，3组样本之间性别、年龄差异均无统计学意义（P>0.05）。

表1 被试数据统计信息

Tab.1 General information of subjects

组别	样本量	男/女	年龄/岁
AD	34	18/16	73.29±7.65
NC	35	20/15	77.11±6.69
MCI	35	13/22	73.34±8.43
P值	-	0.190	0.995

1.2 数据预处理

由于不同被试的脑部结构大小不一，且fMRI图像数据本身的高噪声等特点，在对图像进行特征提取之前需要进行预处理。本试验主要使用DPARSF软件，在MATLAB平台对104例被试fMRI图像数据进行标准的数据预处理操作。实验环境为个人PC机，处理器：Intel（R）Xeon（R）E5-2643，CPU@1.70 GHz，内存为24 GB，试验运行环境为MATLAB2017a。从ANDI获取的fMRI数据格式为常用的Analyze格式，对每个被试的静息态fMRI图像数据依次进行如下处理：首先以第47层图像作为参考层进行时间层校正；进行头动校正；对图像进行空间标准化；对图像进行空间平滑（高斯核半宽全高设为6 mm×6 mm×6 mm）；对平滑后的图像进行去线性漂移和0.01~0.08 Hz的低频滤波；最后去除头动校正时生成的头动参数协变量，去除全脑均值信号、白质信号和脑脊液信号的协变量，具体流程如图1所示。经过预处理之后所有fMRI图像规格统一，方便进行后续的特征提取，预处理前后的fMRI图像对比如图2所示。

2 基于随机森林算法的特征选择

2.1 fMRI数据特征提取

为了更好的使用机器学习方法对fMRI图像进行分类，需要对预处理后的fMRI图像进行特征提取，过程如图3所示。所有被试的fMRI数据经过预处理之后，通过自动解剖标记（Anatomical Automatic Labeling, AAL）分区模板进行匹配，每个被试的大脑被分为116个脑区，然后计算每个脑区的体素平均值时间序列的皮尔逊相关系数：

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

其中，x_i和y_i分别表示两个不同脑区的体素平均值时间序列，$\bar{x}$和$\bar{y}$分别表示时间序列的均值。通过计算每个被试的所有脑区之间的皮尔逊相关系数，得到能够代表全脑功能连接状态的功能连接矩阵。由于功能连接矩阵为对称矩阵，因此取下三角矩阵作为实验获取的初始特征。

2.2 fMRI数据特征选择

由于fMRI图像的高维特点，对fMRI图像提取的

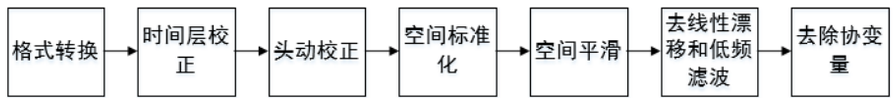
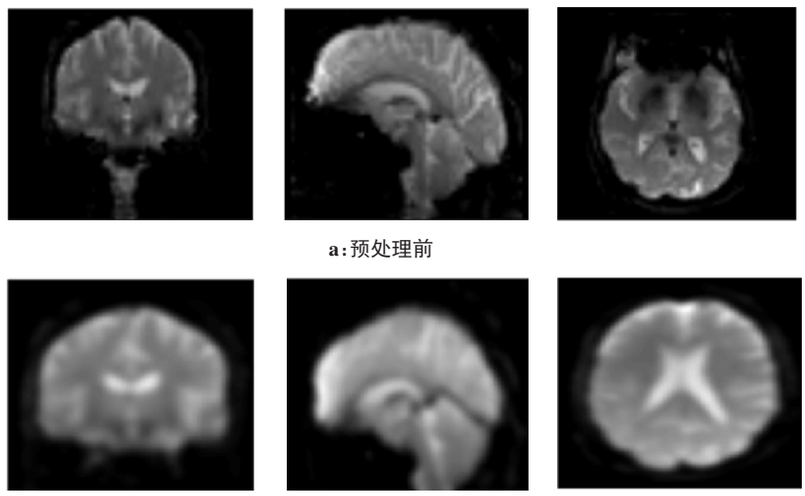


图1 fMRI预处理流程图

Fig.1 fMRI pre-processing



a:预处理前

b:预处理后

图2 fMRI预处理前后对比图

Fig.2 Comparison before and after fMRI pre-processing

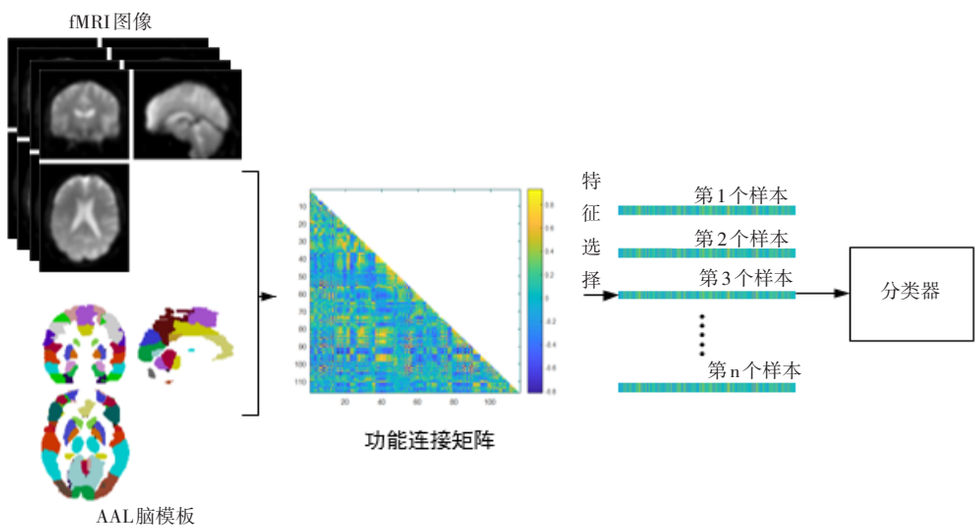


图3 fMRI图像特征提取过程

Fig.3 fMRI image feature extraction process

初始特征仍然包含大量冗余信息,无法直接进行分类。随机森林是一种广泛使用的对密集特征进行降维的有效方法[16],为了最大程度的减少无关特征,保留有效信息,本试验采取随机森林进行特征选择。

集成学习是当前非常流行的机器学习算法,通过在数据集上构建多个模型,集成所有模型的建模结果。随机森林是非常具有代表性的集成算法,它是一组随机生成的决策树组合,通过组合多个弱分类器,使整体模型的结果具有较高的精确度和泛化性能。随机森林的一个重要特点是使用"袋外数据"(Out-Of-Bag, OOB)进行泛化误差的估计,在构建决策树时,采用随机有放回的抽取,OOB是在当前树的训练中未使用的样本集,这种对泛化误差的内部估计增强了决策树分类的准确性,同时有助于对特征重要性的量化。OOB估计就是对每个样本计算它作为OOB样本时决策树对它的分类情况,多数投票作为该样本的分类结果,用误分个数占样本总数的比率作为随机森林的OOB误分率,是随机森林泛化误差的一个无偏估计,它可以在内部进行评估,也就是在生成的过程中就可以对误差建立一个无偏估计。因此随机森林进行特征重要性的度量时,首先对

于每一棵决策树，计算其OOB误分率（err（X^j）），选取一个特征，随机对特征加入噪声干扰，再次计算OOB误分率（err（X^j_{oob}）），则特征的重要性VI（X^j）[17]如式（2）所示：

$$VI(X^j)=\frac{1}{n}\sum_j(|err(X^j)-err(X^j_{oob})|) \quad (2)$$

在本文的试验中，利用随机森林的特征重要性选择特征。首先，随机森林对初始特征的训练集进行训练，然后计算特征重要性并以降序存储。最后，选择一组特征来训练分类器，分别利用K最近邻（k-Nearest Neighbor, KNN）分类器和SVM分类器进行分类。

2.3 模型评价标准

本研究采用10折交叉验证的方法得到准确率，来验证分类器的性能。除此之外，为了更准确的评价分类模型的效果，实验选取分类精确率（P）、召回率（R）和$F1$值作为评价标准，其计算公式分别如下：

$$P=\frac{TP}{TP+FP} \quad (3)$$

$$R=\frac{TP}{TP+FN} \quad (4)$$

$$F1=\frac{2PR}{P+R} \quad (5)$$

其中，TP为分类器将正样本判定为正的个数，FP为分类器将负样本判定为正的个数，FN为分类器将正样本判定为负的个数。在结果的分析过程中，精确率和召回率的含义也可以被理解为查准率和查全率，而$F1$值可以被看作是模型准确率和召回率的一种加权平均，最大值为1，最小值为0。$F1$值作为统一准确率和召回率的评估标准，来衡量模型分类的性能，$F1$的值越高说明分类模型的分类效果越好。

3 实验结果与分析

为了证明随机森林算法在AD分类模型中的有效性，本文首先根据AAL模板计算初始特征，通过随机森林算法对初始特征进行特征选择，根据特征的权重进行排序，并在AAL模板找到权重较大的特征对应的脑区。经过计算特征权重，特征重要性排名前十的特征子集对应脑区如表2所示，由于本研究所选初始特征为脑区的功能连接，因此每一个特征对应两个脑区。其中，中央前回、尾状核、颞中回、岛盖部额下回、中央后回、楔前叶、枕中回与文献[18]得出的AD患者异常脑区结果一致，说明随机森林算法可以很好的对脑功能连接数据进行特征选择。

表2 特征子集对应脑区列表

Tab.2 List of brain regions corresponding to feature subsets

排名	AAL编号	脑区名称	排名	AAL编号	脑区名称
1	52	枕中回	6	57	中央后回
	1	中央前回		52	枕中回
2	107	小脑半球左小叶	7	98	小脑半球右小叶
	1	中央前回		26	眶内额上回
3	72	尾状核	8	40	海马旁回
	49	枕上回		13	三角部额下回
4	87	颞极:颞中回	9	67	楔前叶
	55	梭状回		52	枕中回
5	111	小脑蚓体	10	44	距状裂周围皮层
	11	岛盖部额下回		30	脑岛

由于Brainnetome Atlas包括246个精细脑区亚区，比传统的脑区图谱精细2倍以上，因此有更加精准的边界定位。为对比基于不同脑模板计算的全脑功能连接对分类结果的影响，试验分别采用传统的AAL和Brainnetome Atlas两种脑模板进行初始特征提取，并利用随机森林算法进行特征选择，最后通过SVM分类器进行分类，得到AD、MCI和NC分类的准确率。同时在进行特征选择过程中，通过设置不同的阈值，观察最终分类结果的影响，为保证训练结果的可靠性，分类准确率采用十折交叉验证的结果表示，分类结果变化趋势如图4所示。

根据图4可以看出，在选择合适阈值的情况下，基于Brainnetome Atlas模板的分类效果优于传统的AAL模板。其中使用AAL模板进行特征选择时，当阈值设置为0.000 5时，保留159维特征，分类结果最高（75.88%）；而基于Brainnetome Atlas模板的特征选择，当阈值设置为0.000 15时，保留1 034维特征，分类结果最高（90.68%）。说明对全脑进行更加精细的划分后，

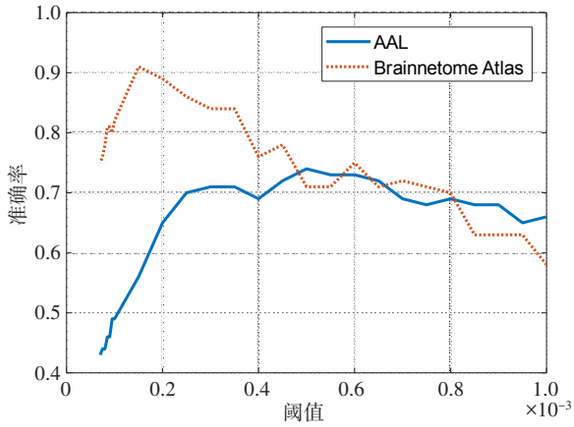


图4 准确率变化趋势图

Fig.4 Accuracy changing with thresholds

其功能连接作为特征更容易被分类，对AD的辅助诊断更有帮助。本试验将基于Brainnetome Atlas模板提取的初始特征，利用随机森林算法进行特征选择后，分别使用KNN和SVM进行分类，结果如表3所示。

表3 分类结果

Tab.3 Classification results

分类器	类别	精确率/%	召回率/%	*F*1值
KNN	AD	66.67	83.33	0.74
	NC	77.78	77.78	0.78
	MCI	75.00	54.55	0.63
SVM	AD	90.00	100	0.94
	NC	100	83.33	0.91
	MCI	91.67	100	0.96

根据表3的结果可以看出，SVM的分类效果明显优于KNN，*F*1值最高为0.96，说明全脑功能连接矩阵可以作为区分AD、NC和MCI的特征，同时随机森林算法对功能连接矩阵可以有效的进行特征选择，结合SVM分类算法可以有效的区分AD、MCI、NC，以达到计算机辅助诊断的目的。

4 总 结

本文提出一种基于随机森林算法的AD计算机辅助诊断方法，以被试的fMRI图像数据为研究对象，利用皮尔逊相关系数计算其全脑功能连接网络，并利用随机森林算法进行特征选择，最后分别采用KNN和SVM方法对AD、NC、MCI进行分类，试验结果表明本文提出的方法可以对fMRI图像数据进行准确的分类。由于Brainnetome Atlas模板对脑区进行了更加精细的划分，因此基于Brainnetome Atlas模板的分类准确率明显优于基于传统的AAL模板得到的结果，同时也说明全脑功能连接矩阵能够反应出AD、MCI和NC的差异，通过特征选择后寻找到重要特征，得出AD异常脑区，为阿尔茨海默症的诊断提供有效的判断依据。

【参考文献】

[1] OBOUDIYAT C, GLAZER H, SEIFAN A, et al. Alzheimer's disease [J]. Semin Neurol, 2013, 33(4): 313-329.

[2] ZHANG J, GAO Y, GAO Y Z, et al. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis[J]. IEEE Trans Med Imaging, 2016, 35(12): 2524-2533.

[3] 李均, 杨澄, 王远军, 等. 基于弥散张量成像构建阿尔茨海默病患者脑网络的研究进展[J]. 中国医学物理学杂志, 2017, 34(2): 204-210.
LI J, YANG C, WANG Y J, et al. Progress in brain network construction for patients with Alzheimer's disease based on diffusion tensor imaging[J]. Chinese Journal of Medical Physics, 2017, 34(2): 204-210.

[4] 李敏, 曾卫明. 基于自适应区域增长的fMRI脑功能激活区检测[J]. 计算机应用与软件, 2017, 34(3): 165-169.
LI M, ZENG W M. Activate region detection of brain function by fMRI based on self-adaptive region growing [J]. Computer Applications and Software, 2017, 34(3): 165-169.

[5] GEERLIGS L, MAURITS N M, RENKEN R J, et al. Reduced specificity of functional connectivity in the aging brain during task performance[J]. Hum Brain Mapp, 2014, 35(1): 319-330.

[6] MCKINNON A C, DUFFY S L, CROSS N E, et al. Functional connectivity in the default mode network is reduced in association with nocturnal awakening in mild cognitive impairment [J]. J Alzheimers Dis, 2017, 56 (4): 1373-1384.

[7] TZOURIO-MAZOYER N, LANDEAU B, PAPAANASSIOU D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain[J]. Neuroimage, 2002, 15(1): 273-289.

[8] 胡颖, 王丽嘉, 聂生东. 静息态功能磁共振成像的脑功能分区综述[J]. 中国图象图形学报, 2017, 22(10): 1325-1334.
HU Y, WANG L J, NIE S D. Review on brain functional parcellation based on resting-state functional magnetic resonance imaging data[J]. Journal of Image and Graphics, 2017, 22(10): 1325-1334.

[9] FAN L Z, HAI L, ZHUO J J, et al. The human brainnetome atlas: a new brain atlas based on connectional architecture[J]. Cereb Cortex, 2016, 26(8): 3508-3526.

[10] KHAZAEE A, EBRAHIMZADEH A, BABAJANI-FEREMI A, et al . Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI[J]. Behav Brain Res, 2017, 322(Pt B): 339-350.

[11] RABIN L A, WANG C, KATZ M J, et al. Predicting Alzheimer's disease: neuropsychological tests, self-reports, and informant reports of cognitive difficulties[J]. J Am Geriatr Soc, 2012, 60(6): 1128-1134.

[12] 李亚鹏, 覃媛媛, 李炜. 阿尔茨海默病患者大脑功能网络的改变[J]. 中国医学物理学杂志, 2013, 30(6): 4510-4514.
LI Y P, TAN Y Y, LI W. The functional brain network changes of Alzheimer's disease[J]. Chinese Journal of Medical Physics, 2013, 30 (6): 4510-4514.

[13] 李慧卓, 相洁, 秦嘉玮, 等. 基于Adaboost的轻度认知障碍和阿尔茨海默病分类[J]. 中国医学影像技术, 2016, 32(4): 623-627.
LI H Z, XIANG J, QIN J W, et al. Classification of mild cognitive impairment and Alzheimer disease based on adaboost algorithm[J]. Chinese Journal of Medical Imaging Technology, 2016, 32(4): 623-627.

[14] SARICA A , CERASA A, QUATTRONE A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review[J]. Front Aging Neurosci, 2017, 9: 329.

[15] GUO C C, TAN R, HODGES J R, et al. Network-selective vulnerability of the human cerebellum to Alzheimer's disease and frontotemporal dementia[J]. Brain, 2016, 139(Pt 5): 1527-1538.

[16] SANDRI M, ZUCCOLOTTO P. Variable selection using random forests [M]. Data Analysis, Classification and the Forward Search. Berlin: Springer, 2006: 263-270.

[17] GHARSALLI S, EMILE B, DE SOUZA DURAN F, et al. Random forest-based feature selection for emotion recognition [C]. International Conference on Image Processing Theory, Tools and Applications (IPTA), Orleans, France, 2015.

[18] RONDINA J M, FERREIRA L K, LUIS D S , et al. Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: a comparison across functional and structural imaging modalities and atlases[J]. Neuroimage Clin, 2018, 17: 628-641.

（编辑：薛泽玲）