DOI:10.16016/j.1000-5404.201509138

阿尔茨海默病和轻度认知损伤诊断的逐步 Fisher 判别模型

马 翠[1],周先东[2],罗万春[1] (400038 重庆,第三军医大学生物医学工程系数学与生物数学教研室[1];401147 重庆,重庆市统计局[2])

[摘要] **目的** 建立正确率较高的阿尔茨海默病和轻度认知损伤的诊断模型。**方法** 通过对391 例包括正常人、阿尔茨海默病患者以及轻度认知损伤患者的 33 项临床指标的研究,运用逐步 Fisher 判别进行诊断模拟,选出正确率最高的判别函数。**结果** 筛选出 6 项指标建立了阿尔茨海默病以及轻度认知损伤诊断模型。经过 100 次随机抽样诊断模拟,对于正常人、轻度认知损伤以及阿尔茨海默病患者诊断的平均正确率达到75.50%,最高正确率达到80.63%。**结论** 逐步 Fisher 判别诊断模型能提高阿尔茨海默病以及轻度认知损伤的诊断正确率,为科研和临床诊断提供了一种新途径。

[关键词] 阿尔茨海默病;轻度认知损伤;逐步 Fisher 判别
[中图法分类号] R311;R745.704; R749.04 [文献标志码] A

Stepwise Fisher discriminant model of Alzheimer's disease and mild cognitive impairment

Ma Cui[1], Zhou Xiandong[2], Luo Wanchun[1] ([1]Department of Mathematics and Biomathematics, College of Biomedical Engineering, Third Military Medical University, Chongqing, 400038; [2]Chongqing Municipal Statistics Bureau, Chongqing, 401147, China)

[**Abstract**] Objective To establish a diagnostic model of Alzheimer's disease and mild cognitive impairment with higher diagnosis accuracy. Methods Based on the researches of 33 clinical indices of 391 cases including healthy controls, Alzheimer's disease patients and mild cognitive impairment patients, a stepwise Fisher discriminant model was carried out for diagnosis simulation, and the discriminant function with the highest accuracy was selected. Results Six indices were selected to establish the discriminant model of Alzheimer's disease and mild cognitive impairment. After 100 times of random sampling for diagnosis simulation in the objects, including healthy controls, Alzheimer's disease patients and mild cognitive impairment patients, the average diagnosis accuracy was 75.50%, and the highest accuracy reached 80.63%. Conclusion Our established stepwise Fisher discriminant model can be used to improve the diagnostic accuracy for Alzheimer's disease and mild cognitive impairment, which can provide a new method for their discriminant for scientific research and clinical practice.

[**Key words**] Alzheimer's disease; mild cognitive impairment; stepwise Fisher discriminant

Supported by the Natural Science Foundation of Chongqing (CSTC2013jcyjA10041). Corresponding author: Luo Wanchun, E-mail: wanchunluo@163.com

阿尔茨海默病(Alzheimer's disease, AD)是老年人常见的一种神经系统变性疾病,临床上以进行性的记忆和认知功能障碍为特征,是一种渐进性、获得性神经功能退化性疾病。AD 患者的早期诊断对于患者的评估、控制、护理和未来规划具有重要意义,但轻度认知损伤(mild cognitive impairment, MCI)患者与早期 AD 患者十分相似的临床症状却让 AD 的早期诊断变得异常困难[1-3]。随着医疗技术的进步,神经心理学、遗传学、生化学、影像学等检测手段广泛应用于 AD 诊断中,使得 AD 早期诊断有了多项参考指标[4-9]。研究表明,这些参考指标均为 AD 诊断提供了非常重要的信息,但目前大多研究人员都仅仅局限于对某一方面指标的研究[10-15]。文献[16]从数学建模和统计学的角度,综合运用各类指标进行研究,建立了诊断 AD 和 MCI 的数学模型。本研究在文献[16]的基础上,进一步探索更加准确有效的诊断模型,为 AD 和 MCI 的

[基金项目] 重庆市自然科学基金(CSTC2013jcyjA10041)
[通信作者] 罗万春, E-mail: wanchunluo@163.com
[优先出版] http://www.cnki.net/kcms/detail/51.1095.R.20151230.1138.004.html(2015-12-30)

临床诊断提供更为有效的参考依据。

1 资料与方法

1.1 临床资料

数据源于 The LONI Image Data Archive(https://ida.loni.usc.edu/login.jsp),经删除缺失信息处理后,选出所有指标数据齐全的样本 391 例,包括正常人(healthy controls, HC)101 例,MCI 患者 200 例,AD 患者 90 例。数据矩阵维度为 391×34,每个样本都包含年龄、受教育程度、简易智力状况检查量表得分(mini-mental state examination score, MMSE)等 33 项指标,分别记为 x_1, x_2,…, x_{33}(表 1),分别将 HC、MCI 患者、AD 患者赋值为 1、2、3。

表 1 样本指标对照表

符号	指标	含义
x_1	Age	年龄
x_2	Education	受教育程度
x_3	MMSE	简易智力状况检查量表得分
x_4	APGEN1	载脂蛋白 ApoE4 基因检测 1
x_5	APGEN2	载脂蛋白 ApoE4 基因检测 2
x_6	APOE4load	载脂蛋白 ApoE4 综合基因检测
x_7	TOTALMOD	ADAS-cog 修正值
x_8	BSI_BRAINVOL	大脑容积
x_9	BSI_VENTVOL	心室容积
x_{10}	BSI_relVENTVOL	心室与大脑容积之比
x_{11}	FSXvol_LEFTHIPPO	左海马体积
x_{12}	FSXvol_RIGHTHIPPO	右海马体积
x_{13}	FSXvol_relHIPPOk	海马体积与脑容量之比的 1 000 倍
x_{14}	PRT	左脑的 PRT 值
x_{15}	TMP	左右半脑的 TMP 均值
x_{16}	FRT	左脑的 FRT 值
x_{17}	OCT	左右半脑的 OCT 均值
x_{18}	PCING	左右半脑的 PCING 均值
x_{19}	ACING	左右半脑的 ACING 均值
x_{20}	MFRT	左右半脑的 MFRT 均值
x_{21}	MPRT	左右半脑的 MPRT 均值
x_{22}	SMC	左右半脑的 SMC 均值
x_{23}	VC	左右半脑的 VC 均值
x_{24}	CAD	左右半脑的 CAD 均值
x_{25}	CBL	左右半脑的 CBL 均值
x_{26}	VER	左右半脑的 VER 均值
x_{27}	HEMI	左右半脑的 HEMI 均值
x_{28}	PNS	左右半脑的 PNS 均值
x_{29}	CTX	左右半脑的 CTX 均值
x_{30}	PTM	左右半脑的 PTM 均值
x_{31}	PARH	左右半脑的 PARH 均值
x_{32}	AMG	左右半脑的 AMG 均值
x_{33}	THL	左右半脑的 THL 均值

x_{14}至x_{33}均为 PET 扫描图像指标

1.2 建模思路

为便于数据分析和建立诊断模型,将所有样本随机分为训练样本和测试样本,共产生 100 组训练样本集(T_1, T_2,…,T_{100})和对应的测试样本集(C_1, C_2,…, C_{100}),每个 T_i 中包含 HC 50 例、MCI 患者 100 例、AD 患者 50 例,剩余样本所组成的集合作为测试样本集 C_i(即每个 C_i 包含 HC 51 例、MCI 患者 100 例、AD 患者 40 例)。

鉴于样本数据的 33 项指标中既有离散指标又有连续指标,而 Fisher 判别不拘于对样本条件的限制,故先用逐步判别分析筛选出对诊断敏感的指标后,再用 Fisher 判别进行判别诊断,简称逐步 Fisher 判别。具体如下:

对第 i 组训练样本集 T_i,运用逐步判别筛选出敏感指标,选取 T_i 中敏感指标及病例类型构成新的训练样本集,运用 Fisher 判别对新训练样本集进行训练,即可得到第 i 组训练样本集 T_i 的判别函数。用该判别函数对测试样本集 C_i 进行判别,并计算出 HC、MCI 患者、AD 患者的平均判别正确率。按照上述方法,模拟训练 100 次,即可得到每组相应的敏感指标、判别函数以及平均判别正确率。最后,选择平均判别正确率最大时所对应的判别函数作为最终诊断模型的判别函数。

1.3 统计学方法

采用 EXCEL 2007 删除带有缺失值的原始数据;采用 MATLAB R2013a 编程实现训练样本与测试样本的抽样、逐步判别分析筛选敏感指标和 Fisher 判别分类的过程[17]。利用 SPSS 13.0 绘制敏感指标的 ROC 曲线(ROC 曲线下面积的比较采用 MedCalc 11.4 中 Delong 法完成),对逐步 Fisher 判别、全指标 Fisher 判别与非线性判别模型所得正确率采用配对样本 t 检验进行差异性比较。

2 结果

2.1 判别函数

运用软件编程实现,得到平均判别正确率最大时,逐步 Fisher 判别所对应的敏感指标为:$y=(x_3, x_4, x_7, x_{11}, x_{14}, x_{33})^T$,判别函数为:

$$f_{HC}=(V_1y+3.36)^2+(V_2y-3.54)^2,$$

$$f_{MCI}=(V_1y+3.12)^2+(V_2y-3.65)^2,$$

$$f_{AD}=(V_1y+2.63)^2+(V_2y-3.57)^2,$$

$V_1=(-0.06, -0.002, 0.01, -0.00007, -0.69, -0.72)$,

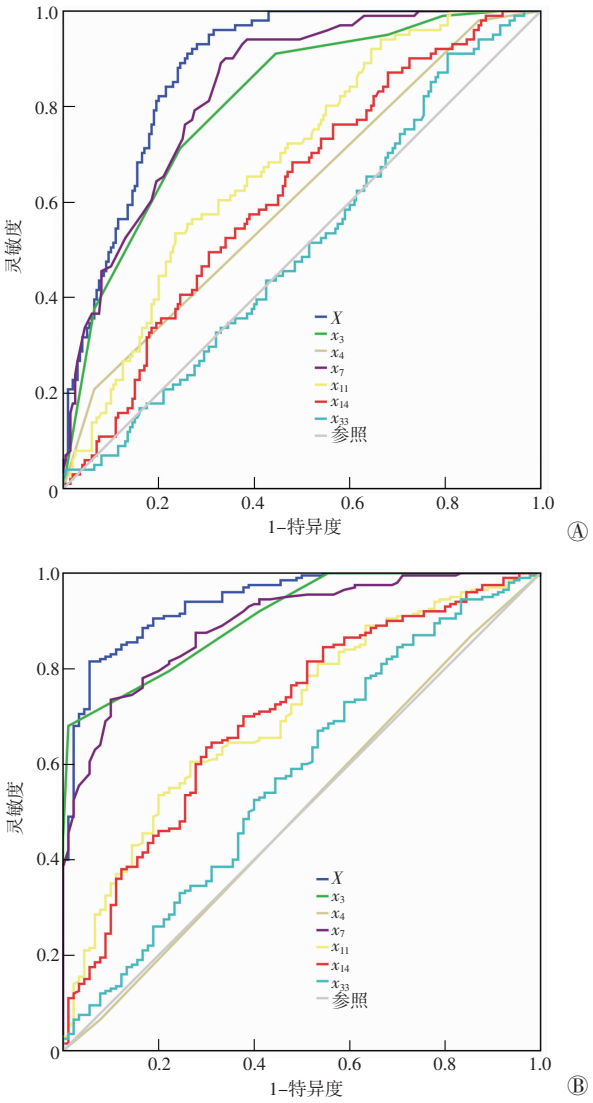
$V_2=(0.05, 0.46, 0.01, -0.0001, 0.86, 0.20)$.

y 为任意 HC、MCI、AD 病例的指标 x_3, x_4, x_7, x_{11}, x_{14}, x_{33}所组成的列向量。

一旦获得了某人的这6项指标,只要代入上述3个函数比较大小即可。如果f_{HC}最小,则为HC;如果f_{MCI}最小,则为MCI患者;如果f_{AD}最小,则为AD患者。

2.2　敏感指标的灵敏度与特异度分析

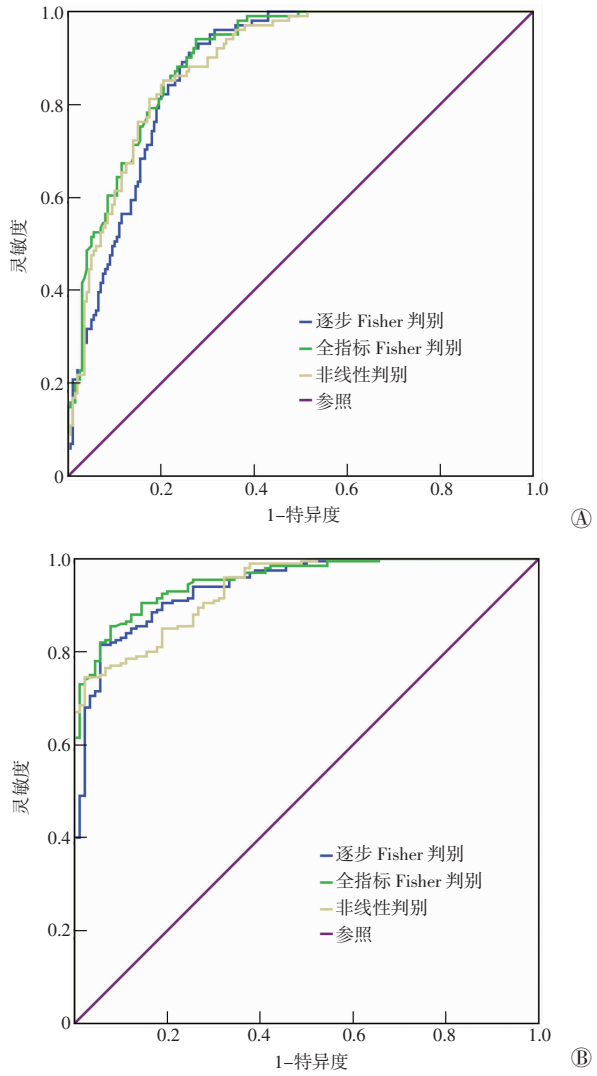
根据2.1所得的判别函数计算出诊断HC、MCI、AD的综合指标X(诊断HC和MCI时,$X=f_{HC}-f_{MCI}$;诊断MCI和AD时,$X=f_{MCI}-f_{AD}$),对X及筛选出的6个敏感指标分别作ROC曲线分析(图1)。可见,无论是诊断HC和MCI,还是诊断MCI和AD,经判别函数计算所得的综合指标X的ROC曲线下面积(AUROC)与单个敏感指标相比,均具有显著性差异。



A:正常人-MCI患者;B:MCI患者-AD患者

图1　逐步判别分析所选敏感指标的ROC曲线分析

此外,虽然逐步Fisher判别与直接采用33个指标的全指标Fisher判别方法以及文献[16]中的非线性判别方法的ROC曲线下面积均无显著差异(图2),但从3种方法诊断HC、MCI患者、AD患者的灵敏度和特异度(表2~4)来看:①逐步Fisher判别在诊断HC和MCI时的灵敏度相对较低、特异度相对较高,说明其正确诊断HC的能力相对低一些,而正确诊断MCI患者的能力较强;②逐步Fisher判别在诊断MCI和AD时的灵敏度相对较高,而其特异度仅次于全指标Fisher判别,这也说明其正确诊断MCI的能力相对较强,且正确诊断AD的能力较非线性判别强。③若不考虑MCI患者,逐步Fisher判别较全指标Fisher判别具有很强的诊断HC和AD的能力。



A:正常人-MCI患者;B:MCI患者-AD患者

图2　3种诊断方法的ROC曲线分析

2.3　诊断模拟

按照1.2中训练样本集和测试样本集的组成方法,进行100次逐步Fisher判别测试,得到了测试样本判别3类病例的平均、最大、最小正确率以及HC、MCI和AD各自的平均、最大、最小正确率,并对同样的测试样本,利用全指标Fisher判别所得结果与文献[16]非线性判别进行比较,结果如图3所示。可见,逐步

Fisher 判别和全指标 Fisher 判别的诊断效果明显好于非线性判别,在 3 类病例平均诊断正确率和 AD 诊断正确率方面,逐步 Fisher 判别略低于全指标 Fisher 判别,但在诊断 MCI 方面,逐步 Fisher 判别却高于全指标 Fisher 判别,这对于提前确诊 MCI 具有一定的意义。

表 2 正常人与轻度认知损伤患者的灵敏度和特异度分析(HC:$n=51$,MCI:$n=100$)

诊断方法	类型	判断为 HC(平均例数)	判断为 MCI(平均例数)	灵敏度(%)	特异度(%)
逐步 Fisher 判别	HC	42.63	8.37	83.59	70.89
	MCI	20.61	70.89		
全指标 Fisher 判别	HC	43.24	7.19	85.74	68.35
	MCI	21.98	68.35		
非线性判别	HC	46.97	4.03	92.10	66.86
	MCI	32.62	66.86		

表 3 轻度认知损伤患者与阿尔茨海默病患者的灵敏度和特异度分析(MCI:$n=100$,AD:$n=40$)

诊断方法	类型	判断为 MCI(平均例数)	判断为 AD(平均例数)	灵敏度(%)	特异度(%)
逐步 Fisher 判别	MCI	70.89	8.50	70.89	76.70
	AD	9.32	30.68		
全指标 Fisher 判别	MCI	68.35	9.67	68.35	85.58
	AD	5.77	34.23		
非线性判别	MCI	66.86	0.52	66.86	47.92
	AD	20.83	19.17		

表 4 正常人与阿尔茨海默病患者的灵敏度和特异度分析(HC:$n=51$,AD:$n=40$)

诊断方法	类型	判断为 HC(平均例数)	判断为 AD(平均例数)	灵敏度(%)	特异度(%)
逐步 Fisher 判别	HC	42.63	0.00	100.00	100.00
	AD	0.00	30.68		
全指标 Fisher 判别	HC	43.24	0.57	98.70	100.00
	AD	0.00	34.23		
非线性判别	HC	46.97	0.00	100.00	100.00
	AD	0	19.17		

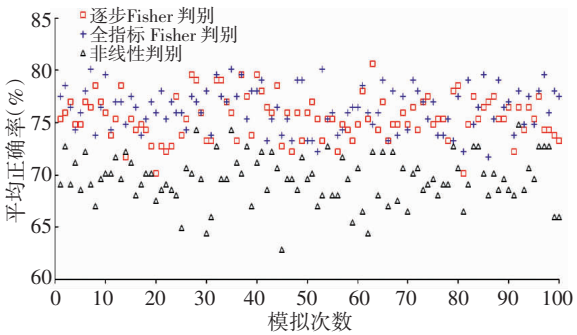
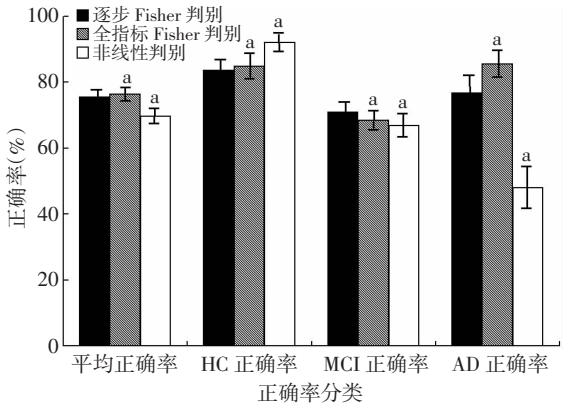


图 3 随机模拟的正确率对比分析

进一步比较逐步 Fisher 判别与全指标 Fisher 判别、非线性判别的差异,将 100 次模拟获得的逐步 Fisher 判别与全指标 Fisher 判别、非线性判别正确率分别作配对样本 t 检验(图 4)发现,逐步 Fisher 判别的平均正确率、MCI 正确率和 AD 正确率均显著高于非线性判别,而 HC 正确率较非线性判别稍低。与全指标 Fisher 判别相比,除 MCI 正确率较高外,其他均略低。



a:$P<0.05$,与逐步 Fisher 判别结果比较($n=100$,$\bar{x}\pm s$)

图 4 不同正确率的差异比较

3 讨论

本研究的研究对象包括 HC、MCI 患者和 AD 患者,这更符合临床实际情况。在 MCI 的干扰下,患者前来就诊时,不大可能马上就可以排除某种疾病。因此,尽管在本研究中得到的逐步 Fisher 判别对测试样本的平均诊断正确率只有 80.63%(154/191),但对于临床诊断极为困难的 AD 诊断方法提供了有益的探索,在临床上也可作为辅助诊断的一种参考依据。

从模拟测试结果分析来看,逐步 Fisher 判别的平均正确率、MCI 正确率和 AD 正确率均显著高于非线性判别,但 HC 正确率较非线性判别稍低。从临床的角度考虑,将 MCI 患者误判为 HC,有可能延误病情、加速病情的恶化,其后果比将 HC 误判为 MCI 患者严重得多。同时,对于 HC、MCI 患者和 AD 患者的判别诊断,文献[16]所用非线性判别函数的构造方法比较巧妙,但型如 x_ix_j 交叉项在临床上不好解释,且通用性受到限制。因此,逐步 Fisher 判别的诊断效果比非线性判别模型更适用于临床诊断。

虽然模拟结果分析显示逐步 Fisher 判别的正确率略低于全指标 Fisher 判别,但全指标 Fisher 判别诊断所需要的指标信息过多,在临床诊断中往往难以实现,而逐步 Fisher 判别将逐步判别与 Fisher 判别有效结

合,利用逐步判别从大量临床指标中剔除了对诊断意义不大的指标,降低了诊断成本,且效果也与全指标Fisher判别相差无几,在临床辅助诊断中具有较强的操作性。另外,逐步Fisher判别对指标的类型(离散或者连续)没有要求,故本研究的方法对于类似的大量临床辅助诊断问题具有一定的通用性。当然,逐步Fisher判别所得灵敏度和特异度仍不够高,尤其是在正确判别AD和MCI方面,效果不是非常理想。今后考虑从两方面进行研究改进:一是收集更多样本,同时注意剔除样本中的异常值数据;二是研究探索效果更好的多分类判别方法,力争进一步提高MCI和AD的分类正确率。

此外,本研究表明,除文献[16]利用ROC曲线所选指标外,APGEN1、THL在AD的临床诊断中也较重要。尤其是APGEN1是ApoE4的重要度量指标,研究发现,ApoE4似乎可提高磷酸化tau的浓度,而磷酸化tau水平增高将最终杀死GABAergic神经元,导致学习和记忆障碍[6]。因此,本文筛选出的6项指标具有较好的生物学意义。

参考文献:

[1] Roman G, Pascual B. Contribution of neuroimaging to the diagnosis of Alzheimer's disease and vascular dementia[J]. Arch Med Res, 2012, 43(8): 671-676. DOI: 10.1016/j.arcmed.2012.10.018

[2] Naylor M D, Karlawish J H, Arnold S E, *et al*. Advancing Alzheimer's disease diagnosis, treatment, and care: recommendations from the Ware Invitational Summit[J]. Alzheimers Dement, 2012, 8(5): 445-452. DOI: 10.1016/j.jalz.2012.08.001

[3] Park H, Yang J J, Seo J, *et al*. Dimensionality reduced cortical features and their use in the classification of Alzheimer's disease and mild cognitive impairment[J]. Neurosci Lett, 2012, 529(2): 123-127. DOI: 10.1016/j.neulet.2012.09.011

[4] Segovia F, Gorriz J M, Ramirez J, *et al*. Early diagnosis of Alzheimer's disease based on Partial Least Squares and Support Vector Machine[J]. Expert Syst Appl, 2013, 40(2): 677-683. DOI: 10.1016/j.eswa.2012.07.071

[5] Illan I A, Gorriz J M, Lopez M M, *et al*. Computer aided diagnosis of Alzheimer's disease using component based SVM[J]. Appl Soft Comput, 2011, 11(2): 2376-2382. DOI: 10.1016/j.asoc.2010.08.019

[6] Apostolova L G, Hwang K S, Kohannim O, *et al*. ApoE4 effects on automated diagnostic classifiers for mild cognitive impairment and Alzheimer's disease[J]. Neuroimage Clin, 2014, 4: 461-472. DOI: 10.1016/j.nicl.2013.12.012

[7] Savio A, Grana M. Deformation based feature selection for Computer Aided Diagnosis of Alzheimer's Disease[J]. Expert Syst Appl, 2013, 40(5): 1619-1628. DOI: 10.1016/j.eswa.2012.09.009

[8] Dukart J, Mueller K, Barthel H, *et al*. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI[J]. Psychiatry Res, 2013, 212(3): 230-236. DOI: 10.1016/j.pscychresns.2012.04.007

[9] Martinez-Murcia F J, Gorriz J M, Ramirez J, *et al*. Computer Aided Diagnosis tool for Alzheimer's Disease based on Mann-Whitney-Wilcoxon U-Test[J]. Expert Syst Appl, 2012, 39(10): 9676-9685. DOI: 10.1016/j.eswa.2012.02.153

[10] Cho Y, Seong J K, Jeong Y, *et al*. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data[J]. Neuroimage, 2012, 59(3): 2217-2230. DOI: 10.1016/j.neuroimage.2011.09.085

[11] Zhang D, Wang Y, Zhou L, *et al*. Multimodal classification of Alzheimer's disease and mild cognitive impairment[J]. Neuroimage, 2011, 55(3): 856-867. DOI: 10.1016/j.neuroimage.2011.01.008

[12] Illan I A, Gorriz J M, Ramirez J, *et al*. ^{18}F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis[J]. Information Sciences, 2011, 181(4): 903-916. DOI: 10.1016/j.ins.2010.10.027

[13] Grana M, Termenon M, Savio A, *et al*. Computer aided diagnosis system for Alzheimer disease using brain diffusion tensor imaging features selected by Pearson's correlation[J]. Neurosci Lett, 2011, 502(3): 225-229. DOI: 10.1016/j.neulet.2011.07.049

[14] Abdulkadir A, Mortamet B, Vemuri P, *et al*. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier[J]. Neuroimage, 2011, 58(3): 785-792. DOI: 10.1016/j.neuroimage.2011.06.029

[15] Polikar R, Topalis A, Green D, *et al*. Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease[J]. Comput Biol Med, 2007, 37(4): 542-558.

[16] 罗万春. 阿尔茨海默病和轻度认知损伤诊断的建模分析[J]. 第三军医大学学报, 2013, 35(15): 1613-1615. DOI: 10.16016/j.1000-5404.2013.15.027

[17] 谢中华. MATLAB统计分析与应用: 40个案例分析[M]. 北京: 北京航空航天大学出版社, 2010.

(**收稿**:2015-09-25;**修回**:2015-11-27)

(编辑 邓强庭)