DOI: 10.19650/j.cnki.cjsi.J2210263

# 阿尔茨海默病多病程自适应筛查模型的构建*

李志刚[1,2], 胡德安[1,2], 纪 勇[3,4], 吴 昊[3,4], 单 鹏[1,2]

(1. 东北大学信息科学与工程学院 沈阳 110819; 2. 河北省微纳精密光学传感与检测技术重点实验室 秦皇岛 066004; 3. 天津环湖医院神经科 天津 300350; 4. 天津市脑血管与神经退行性疾病重点实验室 天津市痴呆研究所 天津 300350)

**摘 要**:人口老龄化导致更高的阿尔茨海默病患病率,造成沉重的家庭和社会负担。尽早发现是延缓和逆转病程的关键,但现有检测方法无法满足价格低廉、低侵入性、快速、可靠的筛查与检测需求。生理液体光谱技术已在医学检测方面显示出了潜力,但在阿尔茨海默痴呆症检测上所面临的挑战是血浆光谱中阿尔茨海默病相关特征信息难以提取以及多病程复杂分类的难题。本文从信息空间建立、特征挖掘以及筛查与检测体系构建的角度展开研究,建立模型特征波数驱动的阿尔茨海默病多病程自适应筛查与诊断体系。针对阿尔茨海默病早期、中期和晚期 3 个不同病程阶段,模型检测的灵敏度分别为 90.0%、87.5%、100%;特异性分别为 83.3%、93.7%、100%。实验结果表明,基于阿尔茨海默病多病程自适应筛查与诊断模型具有较好的检测灵敏度和特异性。

**关键词**: 阿尔茨海默病;血浆振动光谱;随机森林;Relief-F 算法;检测模型体系

**中图分类号**: TH741 O657.33 **文献标识码**: A **国家标准学科分类代码**: 416.6040

# Construction of a multiple disease courses adaptive screening model for Alzheimer's disease

Li Zhigang[1,2], Hu Dean[1,2], Ji Yong[3,4], Wu Hao[3,4], Shan Peng[1,2]

(*1. College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; 2. Hebei Key Laboratory of Micro-Nano Precision Optical Sensing and Measurement Technology, Qinhuangdao 066004, China; 3. Department of Neurology, Tianjin Huanhu Hospital, Tianjin 300350, China; 4. Tianjin Key Laboratory of Cerebral Vascular and Neurodegenerative Diseases, Tianjin Dementia Institute, Tianjin 300350, China*)

**Abstract**: An aging population will inevitably make a higher prevalence of Alzheimer's disease (AD), which results in a heavy burden on family and society. The early detection is the key to delaying or reversing the course of the disease. However, the current detection methods cannot meet the needs of inexpensive, low-invasive, rapid and reliable diagnosis of AD. The detection technology based on the biofluid spectra shows great potential in medical diagnosis. However, the challenge in the detection of dementia is difficult in extracting the characteristic information related to dementia in the plasma spectra and the complicated classification problem of multiple disease courses. For this reason, the research work will be carried out from the perspective of information space construction, feature information mining, and detection system design. An adaptive screening and diagnostic model for multiple disease courses of AD driven by the model feature wavenumber is constructed. For three different stages of AD, including early, middle, and late stages, the sensitivity of detection is 90.0%, 87.5%, and 100%, respectively. And the specificities are 83.3%, 93.7%, and 100%, respectively. Experimental results show that the proposed detection model provides superior classification power for AD.

**Keywords**: Alzheimer's disease; blood plasma vibration spectra; random forest; Relief-F algorithm; detection model system

收稿日期:2022-08-05 Received Date: 2022-08-05

*基金项目:天津市科技计划项目(22ZYCGSY00840)、天津市卫健委科技项目(ZC20121)、国家自然科学基金(61601104)、河北省自然科学基金(F2017501052)、中央高校基本科研经费(N2023021)项目资助

## 0　引　　言

阿尔茨海默病(Alzheimer disease, AD)是一种发病率极高的神经系统退行性疾病,其发病机制尚未完全明确[1-2]。2018 年,美国国家老龄化研究所和阿尔茨海默病协会(National Institute on Aging-Alzheimer′s Association, NIA-AA)提出了 β-淀粉样蛋白[3]、tau[4]、神经变性(amyloid-tangle-neurodegeneration, ATN)的研究框架[5-6],用于阿尔茨海默病的定义和诊断。其中 A 代表淀粉样蛋白(amyloid)、T 代表神经丝缠结(tangle), N 代表神经变性(neurodegeneration)。目前应用最广泛的神经影像检测方法[7-8]存在价格高昂、辐射暴露且对于早期 AD 敏感性低等诸多缺陷。脑脊液生物标志物检测方式操作复杂、侵入性高,不适于 AD 早期的检测[9],且上述方法均无法同时涵盖 ATN 诊断框架所需的全部信息。

研究已经表明外周血血浆是理想的 AD 相关特征信息的来源[10]。而振动光谱具有低成本、高通量、免试剂、易操作、可重复的优点,在医学筛查与检测的研究与应用领域已经显示出巨大的优势,尤其傅里叶变换红外衰减全反射光谱(Fourier transform infrared attenuated total reflectance, FTIR-ATR)适用于液体样本[11-12],能够同时提取 ATN 诊断框架中的相关信息,非常适合 AD 等复杂致病因素疾病的筛查与诊断。基于血液振动光谱的 AD 筛查与检测的研究已经在特征波数的选取、筛查与检测模型的建立、筛查与检测模型的融合等方面取得了一定的进展[13-14],但是国内在基于血浆的 AD 筛查与检测方面的研究目前还很少,大多还是采用神经影像技术对于 AD 的分类[15]。为此,本文将探索利用血浆光谱构建 AD 多病程的自适应筛查检测模型体系的研究。

## 1　自适应切换筛查体系工作原理与分析

振动光谱通过提供光与物质相互作用的度量,能够以无损、无标记的方式检测出给定样品的独特生物分子特征信息。血浆振动光谱,尤其是傅里叶变换红外光谱仪(Fourier transform infrared, FTIR)结合人工智能进行 AD 筛查与诊断已成为极具研究价值的领域[9,11,16]。FTIR 作为一种多分子生化分析仪器,可以有效地测量蛋白质、脂质、多糖、核酸等并在分子水平上提供关于生物样品的结构和化学成分信息。更为关键的是,FTIR 不像组织病理学方法那样评估形态学差异,而是在发生形态学变化之前在分子水平上分析生物样品,以被测样本的光谱特征信息挖掘与提取的方式,更早地检测出与疾病关联的差异,这个特点对于疾病早期的筛查尤为重要。此外,本质上不同阶段的疾病诊断或筛查是与故障诊断类似的多分类问题。一般来说,单一的模型无法胜任多亚类的疾病检测,因此多模型联合应用是此类问题有效的解决途径。针对血液中 AD 的生物标志物含量微弱,患者与健康者血浆光谱混叠严重以及 AD 症状不同阶段光谱细微差异导致的利用血浆光谱进行整体诊断痴呆准确率不高的问题,受数据驱动与自适应切换控制智能机理的启发,本文创新性构建采用不同病程分类的 AD 多模型自适应切换筛查体系。其工作流程为:首先,通过被检测样本的模型特征,结合筛查体系中的切换机制完成被测样本的模型通道的选择,被测样本进入相应的检测通道,再利用其分类特征完成最终的分类筛查。自适应切换筛查体系整体架构如图 1 所示。

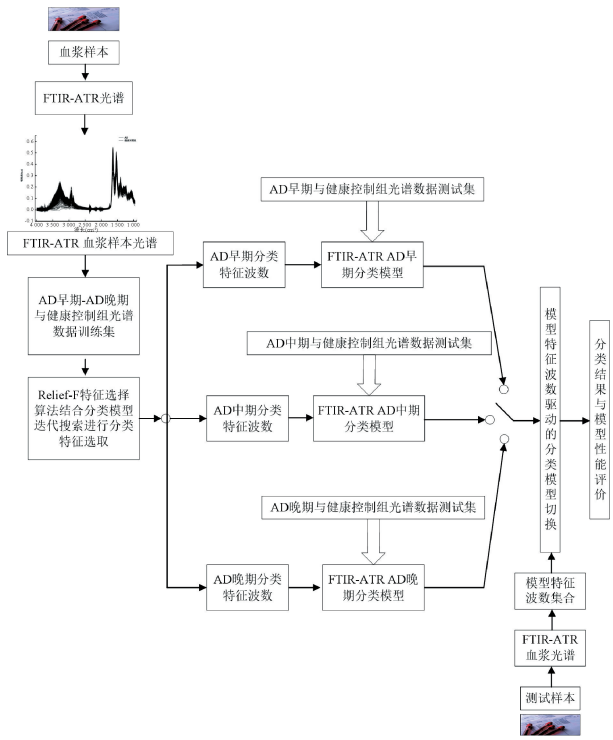



图 1　AD 多模型自适应切换筛查体系

Fig. 1　AD multi-model adaptive switching screening system

### 1.1　多模型分类体系切换机制设计

为了构建阿尔茨海默多模型分类体系,本文将文献[9]中给出的针对不同病程阶段的阿尔茨海默症具有显著性的 6 个波数定义为模型特征波数, $\nu_{\mathrm{MF}}=\{1\,547, 1\,504, 1\,369, 1\,219, 1\,080, 1\,034\}$, 整体 AD 数据集包含 AD 早期、AD 中期和 AD 晚期 3 个亚类($AD_i, i=\mathrm{early}, \mathrm{middle}, \mathrm{late}$),针对每一亚类 AD 构建对应的筛查与诊断子模型,形成多模型自适应切换筛查与诊断体系,增加模型对不同症状阶段的适应度和筛查能力。其多模型通道切换规则如下:测试样本在模型特征波数上的光谱吸收度被定义为 $A_{\mathrm{new}}=\{A_{1\,547}, A_{1\,504}, A_{1\,369}, A_{1\,219}, A_{1\,080}, A_{1\,034}\}$,

$AD_{\text{early}}$ 样本中心点定义为 $A^{C_\text{early}} = \{A_{1\,547}^{\text{early}}, A_{1\,504}^{\text{early}}, A_{1\,369}^{\text{early}}, A_{1\,219}^{\text{early}}, A_{1\,080}^{\text{early}}, A_{1\,034}^{\text{early}}\}$，$AD_{\text{middle}}$ 的光谱样本集中心点表达为 $A^{C_\text{middle}} = \{A_{1\,547}^{\text{middle}}, A_{1\,504}^{\text{middle}}, A_{1\,369}^{\text{middle}}, A_{1\,219}^{\text{middle}}, A_{1\,080}^{\text{middle}}, A_{1\,034}^{\text{middle}}\}$，$AD_{\text{late}}$ 的定义为 $A^{C_\text{late}} = \{A_{1\,547}^{\text{late}}, A_{1\,504}^{\text{late}}, A_{1\,369}^{\text{late}}, A_{1\,219}^{\text{late}}, A_{1\,080}^{\text{late}}, A_{1\,034}^{\text{late}}\}$，由此定义测试样本与各个亚类样本集中心点的欧氏距离：

$$D_{AD_\text{early}} = A_{\text{new}} - A_{C_\text{early}} \tag{1}$$

$$D_{AD_\text{middle}} = A_{\text{new}} - A_{C_\text{middle}} \tag{2}$$

$$D_{AD_\text{late}} = A_{\text{new}} - A_{C_\text{late}} \tag{3}$$

则测试样本到各亚类样本中心点最短距离为 $D_{\min} = \min\{D_{AD_\text{early}}, D_{AD_\text{middle}}, D_{\text{AD_-late}}\}$，此距离决定了测试样本所归属的分类模型。这即是多病程自适应筛查与诊断模型的切换规则。

### 1.2 分类特征波数选取

AD 主要因素是蛋白质结构变异，也有氧化过激、炎症等因素[9]，为了避免潜在有用信息的遗漏，本文基于阿尔茨海默患者和健康控制组的血浆全光谱数据的自身统计分布特性，利用 Relief-F 特征选取方法获得全谱波数的分类显著性指标值；Relief-F 算法是一种著名的过滤式特征选择方法，通过单个特征对分类的影响，作为特征权重的衡量标准，从而实现对整个数据的特征按照特征权重进行排序[17-18]。Relief-F 首先在样本数据中随机选取样本 $S$，根据 $S$ 在 $S$ 的同类和异类中分别选取 $S$ 的近邻集 $M$ 和 $N$。$S$ 和 $M$ 的关于某一特征差异越大则该特征权重就越低，相反 $S$ 和 $N$ 的关于某一特征差异越大则该特征权重越高。Relief-F 的特征权重计算公式为：

$$W_f = W_f - \text{diff}(f,S,M)/k + \text{diff}(f,S,N)/k \tag{4}$$

其中，$W_f$ 为特征 $f$ 的特征权重，$k$ 为随机抽样个数；$\text{diff}(f,S,M)$ 和 $\text{diff}(f,S,N)$ 分别为 $S$ 样本与其同类近邻集 $M$ 和异类近邻集 $N$ 在特征 $f$ 下的距离和，的其计算公式分别为：

$$\text{diff}(f,S,M) = \sum_{i=1}^{k} \frac{|S_f - M_{fi}|}{\max(S_f) - \min(S_f)} \tag{5}$$

$$\text{diff}(f,S,N) = \sum_{i=1}^{k} \frac{|S_f - N_{fi}|}{\max(S_f) - \min(S_f)} \tag{6}$$

式中：$S_f$ 为样本 $S$ 在特征 $f$ 下的值，$M_{fi}$ 和 $N_{fi}$ 分别为同类样本集 $M$ 和异类样本集 $N$ 的第 $i$ 个样本在特征 $f$ 下的值。最后，按照筛查与诊断模型的性能指标和波数显著性排序，选取出不同症状阶段筛查与诊断模型所对应的用于区分阿尔茨海默症患者和健康控制组的分类特征波数。

## 2 筛查实验验证与模型评估

### 2.1 样本采集环境与病程数据划分

验证与测试实验使用带有 ATR 附件 Alpha-FTIR 光谱仪（Bruker Optics Ltd.），通过 OPUS 5.5 软件进行采集血浆样本 ATR-FTIR 光谱的操作。血液样本按照标准操作程序收集在 EDTA 管中。全血 4℃，2 000 rpm 离心 10 min，获取血浆。血浆样本保存在-80℃。测量时，样品在室温下解冻。冷冻样品解冻后，分别取 1 μL 置于 ATR 表面，通风干燥 5 min 后进行测定。光谱分辨率 4 $\text{cm}^{-1}$，波长范围为 4 000～950 $\text{cm}^{-1}$，。每次将 ATR 的金刚石晶体用蒸馏水清洗并干燥后再进行下一个样品。背景光谱也在每个样品分析之前进行，以减少环境或环境条件变化的影响。

验证与测试实验过程中，阿尔茨海默症的病程划分主要依据简易精神状态量表检查（mini-mental state examination，MSE）和蒙特利尔认知评估量表检查（Montreal cognitive assessment，MoCA）[19]。此外，针对隐匿性强的早期患者还进行了淀粉样蛋白 Aβ 的正电子发射断层显像/计算机断层扫描（positron emission tomography scans-computed tomography，PET-CT）扫描确认。MMSE 量表包括时间定向力、地点定向力、即刻记忆、注意力及计算力、延迟记忆、语言以及视空间等 7 个方面的评估，量表总分范围为 0～30 分。分数值与受教育程度有关，文盲≤17 分，小学程度≤20 分，中学或以上程度≤24 分，为有认知功能缺陷，以上为正常。具体评分参考：13～23 分为早期，5～12 分为中期，<5 分为晚期。而 MoCA 量表是用来对认知功能异常进行快速筛查的评定工具，包括注意与集中、执行功能、记忆、语言、视空间功能、抽象思维、计算和定向力等 8 个认知领域。MoCA 量表总分 30 分，得分≥26 分为正常。MoCA 量表对早期轻度患者更加灵敏，是 MMSE 量表的有效补充。

### 2.2 数据集划分、特征选取与模型评估

AD 不同病程的血浆光谱样本集和健康者血浆光谱组成不同筛查通道样本集，利用 KS 算法[20] 按照 7∶1.5∶1.5 的比例将数据集划分为训练集、验证集和测试集。KS 算法是一种基于欧氏距离的样本划分方法，能够较均衡的划分数据样本空间，在众多领域得到了应用[21-22]。

在综合比较了常用分类方法的基础上，本文采取随机森林分类方法构建分类模型。在科学研究中，随机森林算法得到了广泛的应用[23-26]。随机森林算法是 Breiman[27] 在 2001 年开发的一种以决策树为基分类器的组合方法。随机森林算法是决策树算法的增强，在生成决策树过程中引进随机属性选择，通过采用自助法重采样技术，有放回的从原始样本集中随机抽取部分样本构建决策树，不断重复后产生多棵决策树组成随机森林。每颗决策树的节点有 2 种分裂方式：1）基于信息增益的，信息增益越大越好；2）基于基尼指数的，指数越小越好。由于基尼系数的计算公式比较简单，不需要进行对数运算，计算速度会更快，故本次的实验选择基尼系数的结点

分裂方式。3 种病程的随机森林分类模型体系的基分类树的数目分别为 120、500 和 30。模型评价采用医学诊断领域的 2 个主要指标:灵敏度和特异性。

$$灵敏度 = \frac{TP}{TP + FN} \times 100\% \tag{7}$$

$$特异性 = \frac{TN}{TN + FP} \times 100\% \tag{8}$$

其中,*TP* 表示被判定为正,实际为正(真阳性);*FN* 表示被判定为负,实际为正(假阴性);*FP* 表示被判定为正,实际为负(假阳性);*TN* 表示被判定为负,实际为负(真阴性)。应用这两个指标就可以很好的评定筛查与诊断模型的实际应用性能。

最后,基于 Relief-F 获取的波数显著性排序,利用随机森林分类模型进行不同筛查通道分类波数的迭代选取。迭代指标设置为验证集的正确率(正确检测 AD 患者与健康者数目占验证集数据量的比例)。

## 3　自适应切换筛查体系实验结果与分析

验证与应用测试实验中的数据集有 196 个阿尔兹海默症 AD 样本(包含早期 67 个,中期 111 个以及晚期 18 个)和 252 个健康样本作为健康对照组,来自南开大学附属环湖医院。AD 与健康对照组样本的平均光谱如图 2 所示。

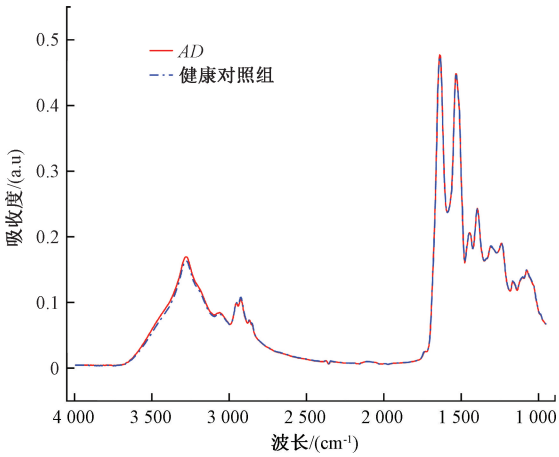



图 2　*AD* 与健康对照组血浆样本平均光谱图

Fig. 2　Mean plasma spectra of *AD* and healthy controls samples

进一步,AD 患者数据集包含早期、中期和晚期 3 个不同的病程,各阶段样本数目分布如表 1 所示。

针对不同 AD 病程的血浆平均光谱如图 3 所示,可以看出不同 AD 光谱高波数 3 500~2 700 $cm^{-1}$ 特征区域和 1 600~1 400 $cm^{-1}$ 指纹区域的吸收度不同,其中 AD 早期吸收度最高,AD 中期次之,AD 晚期光谱吸收度最低。依据 KS 样本数据集划分方法,实验样本划分结果如表 2 所示。

**表 1　AD 不同病程阶段的样本数**

**Table 1　The number of samples in different stages of AD**

| AD 病程 | AD 早期 | AD 中期 | AD 晚期 |
|---|---|---|---|
| AD 数目 | 67 | 111 | 18 |

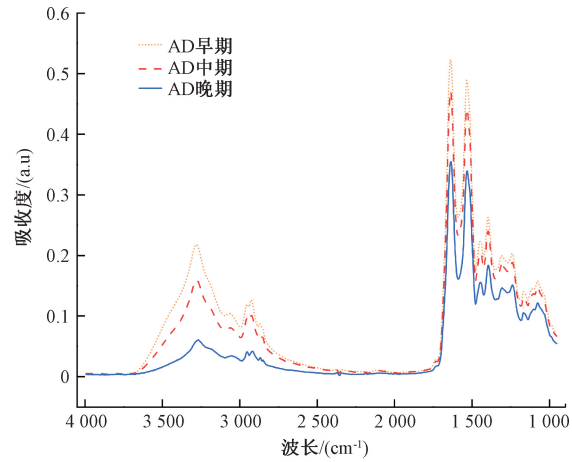



图 3　不同 AD 病程血浆样本平均光谱图

Fig. 3　The average spectrogram of plasma samples of different AD courses

**表 2　AD 数据集划分结果**

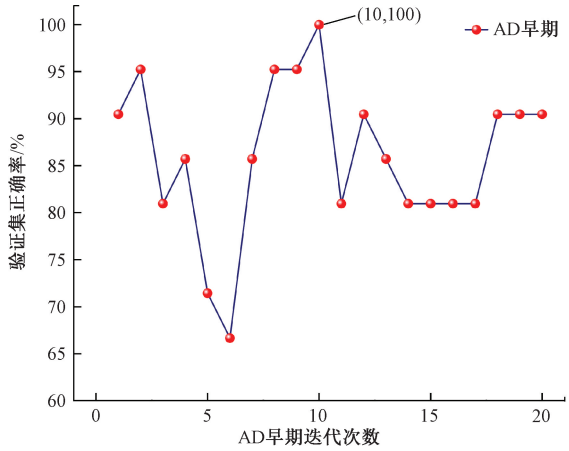
**Table 2　The results of AD data set division**

| 模型数据集 | 训练集数目 | | 验证集数目 | | 测试集数目 | |
|---|---|---|---|---|---|---|
| | AD | 健康 | AD | 健康 | AD | 健康 |
| AD 早期与健康对照组 | 47 | 61 | 10 | 11 | 10 | 12 |
| AD 中期与健康对照组 | 78 | 100 | 17 | 21 | 16 | 22 |
| AD 晚期与健康对照组 | 12 | 17 | 3 | 4 | 3 | 4 |

对于分类特征选取的结果,由各自检测通道中随机森林分类模型结合波数显著性进行分类特征迭代选取的迭代图可知,不同 AD 病程筛查模型通道对应的分类特征波数个数分别为 10、8 和 12。如图 4 所示。

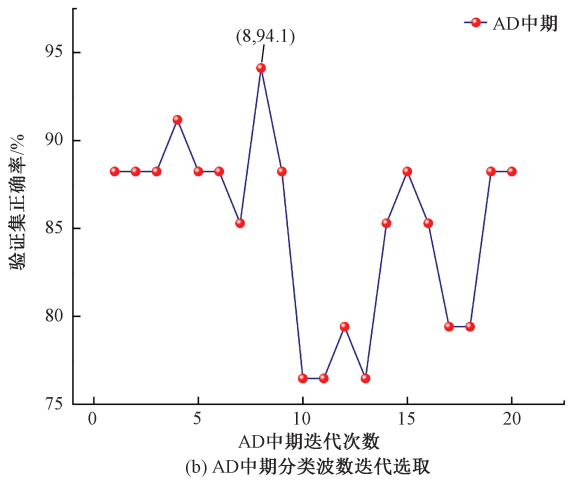
具体不同病程阶段 AD 筛查模型通道所对应的入选分类波数如表 3 所示。

亚类 AD 早期中的 1 547 $cm^{-1}$、1 551 $cm^{-1}$、1 543 $cm^{-1}$、1 539 $cm^{-1}$、1 535 $cm^{-1}$、1 555 $cm^{-1}$、1 531 $cm^{-1}$波数来自于酰胺Ⅱ区(1 500~1 560 $cm^{-1}$,肽骨架的 N—H 弯曲和 C—N 伸展振动),1 647 $cm^{-1}$ 和 1 651 $cm^{-1}$ 波数来自于酰胺Ⅰ区(1 600~1 700 $cm^{-1}$,C ═O 伸缩振动)[28-29]。这两个区域是典型的蛋白质显著区域,因而成为评估神经退



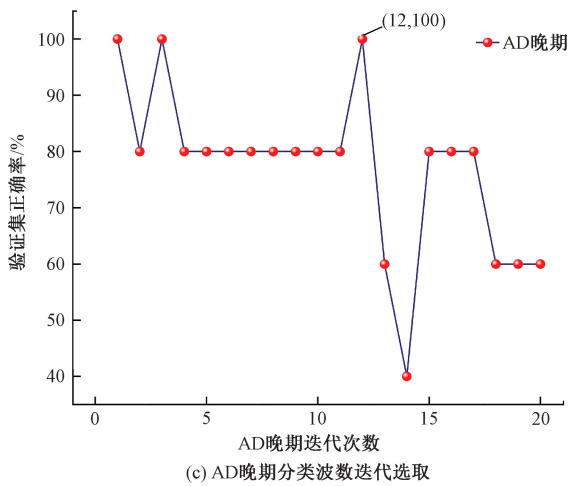

(a) AD早期分类波数迭代选取

(a) Iterative selection of classification wave-numbers in $AD_{early}$




(b) AD中期分类波数迭代选取

(b) Iterative selection of classification wave-numbers in $AD_{middle}$




(c) AD晚期分类波数迭代选取

(c) Iterative selection of classification wave-numbers in $AD_{late}$

图 4 分类波数迭代数目

Fig. 4 Number of classification wavenumber iterations

行性疾病中蛋白质聚集的重要位置,而波数 3 280 $cm^{-1}$ 可能归功于 N—H 的振动(蛋白质)的贡献[30-31];AD 中期的波数同样也在酰胺Ⅱ区中;AD 晚期的波数除了 1 555 $cm^{-1}$ 在酰胺Ⅱ区,其他的主要集中在 1 040~1 080 $cm^{-1}$ 之间。这一区域主要是核酸中 $PO_2$ 基团的对称伸展和一系列复杂的峰序列主导[28]。在上述实验参数选定,多模型自适应切换 AD 筛查体系构建完成的前提下,本文在验证集和测试集上进行了整体 AD 筛查与多模型自适应切换 AD 筛查分类对比实验,其筛查分类结果如表 4 所示。

**表 3 不同病程阿尔茨海默症筛查通道分类波数**

**Table 3 Classification wavenumber of different AD**

| 类型 | 分类特征波数/($cm^{-1}$) |
|---|---|
| AD | 1 072 1 128 1 296 1 315 1 343 1 393 1 396 1 402 1 455 1 512 1 601 1 666 1 690 3 256 3 287 |
| AD 早期 | 1 048 1 342 1 455 1 531 1 535 1 539 1 630 1 651 1 656 3 256 |
| AD 中期 | 1 048 1 315 1 501 1 535 1 547 1 531 1 547 3 325 |
| AD 晚期 | 1 034 1 045 1 053 1 057 1 069 1 076 1 080 1 090 1 555 1 686 2 997 3 310 |

**表 4 AD 筛查分类结果**

**Table 4 The screening classification results of AD** %

| 类型 | 验证集 | | 测试集 | |
|---|---|---|---|---|
| | 灵敏度 | 特异性 | 灵敏度 | 特异性 |
| AD | 82.8 | 93.3 | 70.0 | 67.7 |
| AD 早期 | 100.0 | 100 | 90.0 | 83.3 |
| AD 中期 | 94.1 | 94.1 | 87.5 | 93.7 |
| AD 晚期 | 100 | 100 | 100 | 100 |

实验表明,针对不同病程 AD 设计的多模型自适应切换 AD 筛查体系相比整体 AD 直接筛查,性能显著提升。独立测试实验中,AD 早期和 AD 中期灵敏度和特异性均超过了 80%,达到了临床筛查 70%的指标要求。而对于阿尔茨海默症早期 AD 准确的筛查是具有巨大的意义和价值。经过细分筛查通道后,独立测试实验中整体阿尔茨海默症的筛查灵敏度和特异性都分别提升了 10%以上,充分表明了所提出的多模型自适应切换 AD 筛查体系具有显著优势。为了进一步论证体系的筛查分类性能,多模型自适应切换 AD 筛查体系测试实验的 ROC 曲线[32]以及整体 AD 单一筛查测试实验的 ROC 曲线如图 5 所示。

可以看出 AD 早期-AD 晚期筛查与检测模型中 *AUC* 分别为 0.925 0、0.935 5 和 1,而单一整体筛查模型 AD 中的 *AUC* 为 0.686 0。可见细分的筛查模型的筛查性能均显著高于单一整体筛查模型。

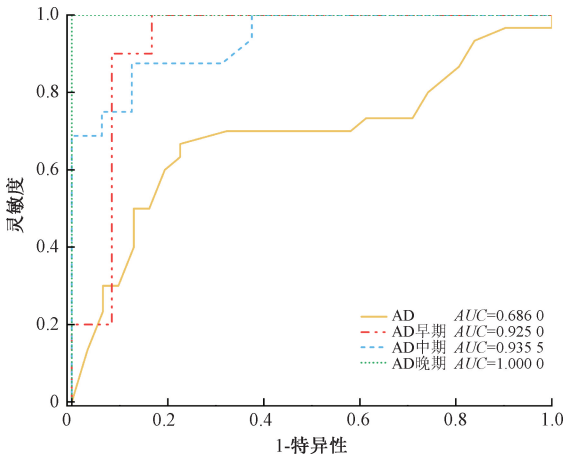



图 5　测试集 ROC 曲线图

Fig. 5　ROC curves of test set

## 4　结　　论

本文针对 AD 的复杂病程引起的筛查困难,提出基于血浆光谱模型特征波数驱动的阿尔茨海默症多病程的自适应筛查与诊断模型体系。将 AD 的样本进行细分,改善了样本空间信息分布,兼顾了 AD 与健康者的宏观区别以及 AD 患者不同病程阶段的细微区别。实验结果也验证了提出的基于血浆光谱模型特征波数驱动的阿尔茨海默症多病程的自适应筛查与诊断模型体系能够有效解决复杂疾病筛查与检测的难题,具有较强的灵活性、可靠性。为未来国内实现 AD 筛查与检测模型体系的临床应用提供了一定的研究基础和探索。

## 参考文献


[1] LIU P P, XIE Y, MENG X Y, et al. History and progress of hypotheses and clinical trials for Alzheimer's disease[J]. Signal Transduction and Targeted Therapy, 2019, 4(1): 1-22.

[2] ZHAO Y L, OU Y N, MA Y H, et al. Association between Life's simple 7 and cerebrospinal fluid biomarkers of Alzheimer's disease pathology in cognitively intact adults: The CABLE study [J]. Alzheimer's Research & Therapy, 2022, 14(1): 74-82.

[3] SARROUKH R, GOORMAGHTIGH E, RUYSSCHAERT J M, et al. ATR-FTIR: A "rejuvenated" tool to investigate amyloid proteins[J]. Biochimica et Biophysica Acta, (BBA)-Biomembranes, 2013, 1828(10): 2328-2338.

[4] MORDECHAI S, SHUFAN E, KATZ B P, et al. Early diagnosis of Alzheimer's disease using infrared spectroscopy of isolated blood samples followed by multivariate analyses [J]. Analyst, 2017, 142(8): 1276-1284.

[5] O'BRYANT S E, ZGANG F, PETERSEN M, et al. Neurodegeneration from the AT(N) framework is different among mexican americans compared to non-Hispanic whites: A health & aging brain among latino elders (hable) study[J]. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2022, 14(1): 12267-12277.

[6] JACK C R, BENNETT D A, BLENNOW K, et al. NIA-AA research framework: Toward a biological definition of Alzheimer's disease [J]. Alzheimer's & Dementia, 2018, 14(4): 535-562.

[7] ANDERSEN E, CASTEIGNE B, CHAPMAN W D, et al. Diagnostic biomarkers in Alzheimer's disease[J]. Biomarkers in Neuropsychiatry, 2021, 5(1): 100041-100051.

[8] BLOCH L, FRIEDRICH C M. Data analysis with Shapley values for automatic subject selection in Alzheimer's disease data sets using interpretable machine learning[J]. Alzheimer's Research & Therapy, 2021, 13(1): 1-30.

[9] PARASKEVAIDI M, MORAIS C L, LIMA K M, et al. Differential diagnosis of Alzheimer's disease using spectrochemical analysis of blood[J]. Proceedings of the National Academy of Sciences, 2017, 114(38): 7929-7938.

[10] LEUZY A, MATTSSON-CARLGREN N, PALMQVIST S, et al. Blood-based biomarkers for Alzheimer's disease[J]. EMBO Molecular Medicine, 2022, 14(1): 14408-14422.

[11] CAMERON J M, BUTLER H J, ANDERSON D J, et al. Exploring pre-analytical factors for the optimization of serum diagnostics: Progressing the clinical utility of ATR-FTIR spectroscopy[J]. Vibrational Spectroscopy, 2020, 109: 103092-103104.

[12] 高克铉, 李志刚, 徐长明, 等. 混合整体趋势扩散的虚拟样本构建及其血液光谱分析应用[J]. 仪器仪表学报, 2019, 40(8): 167-175.
GAO K X, LI ZH G, XUN CH M, et al. Virtual sample establishment of Hybrid-MTD and its application in blood spectrum analysis [J]. Chinese Journal of Scientific

Instrument, 2019, 40(8): 167-175.

[13] PARASKEVAIDI M, MORAIS C L, FREITAS D L, et al. Blood-based near-infrared spectroscopy for the rapid low-cost detection of Alzheimer′s disease[J]. Analyst, 2018, 143(24): 5959-5964.

[14] NASEER K, ALI S, QAZI J. ATR-FTIR spectroscopy as the future of diagnostics: A systematic review of the approach using bio-fluids[J]. Applied Spectroscopy Reviews, 2021, 56: 85-97.

[15] 林伟铭, 袁江南, 冯陈伟, 等. 基于极限学习机的阿尔兹海默病辅助诊断[J]. 中国生物医学工程学报, 2020, 39(3):288-294.
LIN W M, YUAN J N, FENG CH W, et al. Computer-aided diagnosis of Alzheimer′s diseases based on extreme learning machine[J]. Chinese Journal of Biomedical Engineering, 2020, 39(3): 288-294.

[16] CARMONA P, MOLINA M, CALERO M, et al. Discrimination analysis of blood plasma associated with Alzheimer′s disease using vibrational spectroscopy[J]. Journal of Alzheimer′s Disease, 2013, 34: 911-920.

[17] YE L, SHI J F, FERDINANDO H, et al. A multi-sensor school violence detecting method based on improved Relief-F and D-S algorithms[J]. Mobile Networks and Applications, 2020, 25(5): 1655-1662.

[18] SHUKL A K, PIPPAL S K, GUPTA S, et al. Knowledge discovery in medical and biological datasets by integration of Relief-F and correlation feature selection techniques[J]. Journal of Intelligent & Fuzzy Systems, 2020, 38(5): 6637-6648.

[19] 裴芳, 孟涛, 张凯旋, 等. 简易智能状态检查量表和蒙特利尔认知评估量表在老年人认知功能障碍筛查中的比较[J]. 中国药物与临床, 2020, 20(11): 1771-1774.
PEI F, MENG T, ZHANG K X, et al. Value of MMSE vs MoCA in screening cognitive dysfunction in elderly[J]. Chinese Remedies & Clinics, 2020, 20(11): 1771-1774.

[20] KENNARD R, STONE W L. Computer aided design of experiments[J]. Technometrics, 1969, 11(1): 137-148.

[21] LI Q, HUANG Y, ZHANG J, et al. A fast determination of insecticide deltamethrin by spectral data fusion of UV-vis and NIR based on extreme learning machine[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 247: 119119-119126.

[22] ZHANG L, LI G, SUN M, et al. Kennard-Stone combined with least square support vector machine method for noncontact discriminating human blood species[J]. Infrared Physics & Technology, 2017, 86: 116-119.

[23] MOORE P J, LYONS T J, GALLACHER J, et al. Random forest prediction of Alzheimer′s disease using pairwise selection from time series data[J]. PLOS One, 2019,14(2): 211558-211571.

[24] POTT L P, AMADO T J, SCHWALBERT R A, et al. Pre-planting weed detection based on ground field spectral data[J]. Pest Management Science, 2020, 76(3): 1173-1182.

[25] GRAY K R, ALJABAR P, HECKEMANN R A, et al. Random forest-based similarity measures for multi-modal classification of Alzheimer′s disease[J]. Neuroimage, 2013, 65: 167-175.

[26] DAUWAN M, VAN DER ZANDE J J, VAN DELLEN E, et al. Random forest to differentiate dementia with Lewy bodies from Alzheimer′s disease[J]. Alzheimer′s & Dementia: Diagnosis, Assessment & Disease Monitoring, 2016, 4: 99-106.

[27] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[28] CORREIA M, LOPES J, SILVA R, et al. FTIR spectroscopy-A potential tool to identify metabolic changes in dementia patients[J]. Alzheimer′s & Neurodegenerative Diseases, 2016, 2(2): 1-9.

[29] LOPES J, CORREIA M, MARTINS I, et al. FTIR and Raman spectroscopy applied to dementia diagnosis through analysis of biological fluids[J]. Journal of Alzheimer′s Disease, 2016, 52(3): 801-812.

[30] TALARI A C S, MARTINEZ M A G, MOVASAGHIN Z, et al. Advances in Fourier transform infrared (FTIR) spectroscopy of biological tissues[J]. Applied Spectroscopy Reviews, 2016, 52(5): 456-506.

[31] KUMAR S, CHAYDHARY S, JAIN D C. Vibrational studies of different human body disorders using ftir spectroscopy[J]. Open Journal of Applied Sciences, 2014, 4(3): 103-129.

[32] LINDEN A. Measuring diagnostic and predictive

accuracy in disease management: An introduction to receiver operating characteristic (ROC) analysis[J]. Journal of Evaluation in Clinical Practice, 2006, 12(2): 132-139.

## 作者简介

**李志刚**(通信作者),1999 年于燕山大学获得学士学位,2002 年于燕山大学获得硕士学位,2005 年于天津大学获得博士学位,现为东北大学副教授,主要研究方向为光谱分析与人工智能。

E-mail:lizhigang@ neuq. edu. cn

**Li Zhigang** (Corresponding author) received his B. Sc. degree in 1999 from Yanshan University, received his M. Sc. degree in 2002 from Yanshan University, received his Ph. D. degree in 2005 from Tianjin University, now he is associate professor in Northeastern University. His main research interests include spectral analysis and artificial intelligence.

**胡德安**,2020 年于安徽工程大学获得学士学位,现为东北大学在读硕士研究生,主要研究方向为光谱分析与信号处理。

E-mail:hda5017@ 163. com

**Hu Dean** received his B. Sc. degree in 2020 from Anhui Polytechnic University, now he is current a master student in Northeastern University. His main research interests include MIR and signal processing.

**纪勇**,1987 年于中国医科大学获得学士学位,1999 年于日本国立鸟取大学神经中心博士学位,现为天津环湖医院神经内科主任医师,主要研究方向认知障碍与痴呆症。

E-mail: jiyongusa@ 126. com

**Ji Yong** received his B. Sc. degree in 1987 in China Medical University, receive his Ph. D. degree in 1999 from Neurological Center, Tottori University in Japan. Now he is the chief physician of the Department of Neurology at Tianjin Huanhu Hospital His research interests are cognitive impairment and dementia.

**吴昊**, 2012 于河北医科大学获得学士学位,2015 年于首都医科大学大学获得硕士学位,2020 年至今为天津医科大学在读博士,现为天津环湖医院主治医师,主要研究方向为认知障碍与痴呆症。

E-mai: xiaowu2pm@ 163. com

**Wu Hao** received her bachelor's degree from Hebei Medical University in 2012 and her master's degree from Capital Medical University in 2015. Now she is an attending physician in Tianjin Huanhu Hospital, and her research interests are cognitive impairment and dementia.

**单鹏**,2008 年于西北大学获得学士学位,2011 年于西北大学获得硕士学位,2015 年于中国科学院大学自动化研究所获得博士学位,现为东北大学副研究员,主要研究方向为光谱分析与模式识别。

E-mail:peng. shan@ neuq. edu. cn

**Shan Peng** received his B. Sc. degree from Northwest University in 2008, his M. Sc. degree from Northwest University in 2011, and his Ph. D. degree from Institute of Automation, University of Chinese Academy of Sciences in 2015. Now he is an associate professor at Northeastern University with his research interest in spectral analysis and pattern recognition.