

基于脑结构特征和认知行为特征的阿尔茨海默症分类诊断

温森森, 陈玉龙, 王 瑜, 霍京京*

河北工程大学数理科学与工程学院, 河北 邯郸

收稿日期: 2024年6月5日; 录用日期: 2024年6月29日; 发布日期: 2024年7月10日

摘 要

阿尔茨海默症是一种多发于老年人的神经退行性疾病, 由于病程不可逆且无法治愈, 因而及时发现该病早期阶段并采取针对性的治疗措施具有重要意义。本文基于脑结构特征和认知行为特征, 辅以人口统计学特征和生物标志物, 对阿尔茨海默症进行三分类和五分类诊断。将数据进行预处理后, 采用随机森林对特征重要性进行排序, 依据排序结果和各特征组合在机器学习模型中的表现, 分别选取出15个、21个特征, 用于训练随机森林、梯度提升树、CatBoost、LightGBM、XGBoost等机器学习模型。结果显示, LightGBM和XGBoost分别在阿尔茨海默症的三分类和五分类诊断中表现出优异性能, 分类准确率分别为99.25%和95.94%, F1-分数分别为99.25%和95.92%。经过五折交叉验证可知, 上述模型的性能和稳定性都比较出色。最后, 对比两次诊断选取出的特征, 为今后阿尔茨海默症的诊断提出部分建议。

关键词

阿尔茨海默症, 特征选择, 机器学习, 分类诊断

Classification Diagnosis of Alzheimer's Disease Based on Brain Structural Characteristics and Cognitive Behavioral Characteristics

Sensen Wen, Yulong Chen, Yu Wang, Jingjing Huo*

School of Mathematics and Physics, Hebei University of Engineering, Handan Hebei

Received: Jun. 5th, 2024; accepted: Jun. 29th, 2024; published: Jul. 10th, 2024

*通讯作者。

文章引用: 温森森, 陈玉龙, 王瑜, 霍京京. 基于脑结构特征和认知行为特征的阿尔茨海默症分类诊断[J]. 应用数学进展, 2024, 13(7): 3052-3064. DOI: 10.12677/aam.2024.137290

Abstract

Alzheimer's disease is a kind of neurodegenerative disease which mainly occurs in the elderly. Because the course of the disease is irreversible and incurable, it is of great significance to find the early stage of the disease and take targeted treatment measures. In this paper, the three-classification and five-classification diagnosis of Alzheimer's disease are performed based on brain structural characteristics and cognitive behavioral characteristics, supplemented by demographic characteristics and biomarkers. After the data is pre-processed, Random Forest is used to sort the importance of features. According to the sorting results and the performance of each feature combination in the machine learning models, 15 and 21 features are selected respectively for training machine learning models such as Random Forest, GBDT, CatBoost, LightGBM and XGBoost. The results show that LightGBM and XGBoost have excellent performance in the three-classification and five-classification diagnosis of Alzheimer's disease, with classification accuracy of 99.25% and 95.94%, and F1-score of 99.25% and 95.92%, respectively. After the five-fold cross validation, the performance and stability of the above models are excellent. Finally, the characteristics selected from the two diagnoses are compared, and some suggestions are put forward for the future diagnosis of Alzheimer's disease.

Keywords

Alzheimer's Disease, Feature Selection, Machine Learning, Classification Diagnosis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

阿尔茨海默症(Alzheimer's disease, AD)是一种起病隐匿且进展缓慢的神经退行性疾病[1]。这种疾病多发于 65 岁以上的老年人,主要表现为记忆和语言障碍、视觉听觉受损、认知功能下降和性格发生变化等。目前全球有超过 5500 万人患有 AD,因其死亡人数也在逐年增加[2]。该病的发展阶段依次如下:认知正常(Cognitively normal, CN)、主观记忆抱怨(Subjective memory complaint, SMC)、早期轻度认知障碍(Early mild cognitive impairment, EMCI)、晚期轻度认知障碍(Late mild cognitive impairment, LMCI)和 AD。轻度认知障碍(Mild cognitive impairment, MCI)是指上述 EMCI 和 LMCI 两个阶段,被认为是 AD 重要的过渡阶段,每年大约有 10%至 15%的 MCI 病人转换为 AD [3]。迄今为止,还没有药物能有效治愈 AD,因此,及时发现 AD 早期阶段并采取针对性的干预和治疗措施,有助于减缓疾病进程,降低 AD 的致死率。近年来,对于 AD 的诊断主要依靠神经心理学测验、神经影像学检查、生物标志物检测和基因检测等。综合分析以上结果需要医生具备一定的专业素养,并且耗时耗力。

随着人工智能的不断发展,计算机辅助 AD 的早期诊断技术开始兴起。机器学习因其可以从大量的医疗数据中学习特征和标签之间的复杂关系而被广泛运用到 AD 的分类中,研究人员通常使用手工方法提取特征,再采用机器学习中的各种算法进行分类。Klöppel 等人[4]将大脑的灰质密度投射到高维空间中以实现体素级别的特征提取,采用支持向量机对 CN 和 AD 患者进行分类,其诊断准确率为 89%。Li 等人[5]选取大脑皮层顶点体积和与之相关的形态学信息作为特征,采用线性支持向量机对两组数据的 CN 和 MCI 进行区分,其准确率分别为 76%和 80%。Silveira 等人[6]使用 PET 图像将大脑划分为 116 个感兴

感兴趣区域(Regions of interest, ROI), 从中提取特征, 采用提升法组合多个分类器, 在 AD 和 MCI 的检测中得到的准确率分别为 90.97%和 79.63%。Wang 等人[7]首先计算 ROI 之间的相关系数, 然后采用正则化的线性判别分析对数据进行降维, 最后构建 AdaBoost 模型实现 AD 和 MCI 诊断。刘茜等人[8]运用支持向量机递归特征消除和线性判别分析相结合的方式提取 90 个脑区的灰质体积作为特征, 并利用支持向量机进行 AD 分类诊断, 在 CN 和 AD、AD 和 SMC、CN 和 SMC 三种情况下的平均准确率分别为 94%、100%和 93.6%。杨邦坤等[9]使用内核局部 Fisher 判别分析算法对 sMRI 图像中部分与大脑有关数据进行特征提取, 并利用基于核支持向量机算法进行分类, 该方法对 CN 和 AD、CN 和 MCI 两种情况的判别准确率均达到 95%以上。李彩和范炤[10]将 L1 正则化逻辑回归方法与人口统计学特征、简易智力状态检查量表相结合分析 sMRI 图像, 提取面积和体积等形态学指标作为特征, 并运用支持向量机进行分类, 该方法对 CN 和 AD 的分类准确率为 97.66%。刘德华等人[11]选用阿尔茨海默病影像学倡议数据库中 CN、EMCI、LMCI 和 AD 的数据, 结合随机森林的特征重要性排序和基于分类精度的序列前向选择方法, 进行特征选择, 将选出的特征代入 L1 正则化逻辑回归、支持向量机、BP 神经网络、随机森林中自动化识别出最优分类模型, 给出其对 AD 病程进行两两分类的效果。

综上所述, 对于运用机器学习实现 AD 分类诊断, 研究人员提取的特征大多为脑结构特征, 进行的诊断方向也基本为二分类诊断, 但这种情况所提取的特征较为单一, 忽略了患者其他诊断信息; 且对 AD 的分类诊断不够细致, 不利于临床精准用药治疗。本文将基于脑结构特征和认知行为特征, 辅以人口统计学信息及生物标志物等特征构建 AD 的分类诊断模型, 实现 AD 的三分类诊断和五分类诊断。

2. 实验数据

2.1. 数据来源与分析

本文所使用的数据来自 2022 年数维杯国际赛 C 题。该数据集包含 2425 例受检者, 其中男性有 1155 例, 女性有 1270 例, 年龄在 50 至 92 岁之间。超九成的受检者并非西班牙裔或拉丁裔, 并且绝大多数受检者为白人。该数据集主要包括 51 个特征以及两种诊断结果, 将这些特征进行分类, 主要分为脑结构特征、认知行为特征、人口统计学特征和生物标志物四类, 结果如表 1 所示。

Table 1. Feature classification result
表 1. 特征分类结果

	脑结构特征	认知行为特征	人口统计学特征	生物标志物
数目	7	30	6	8
来源	MRI 图像	量表测试	问询或填表	基因诊断、PET 图像、脑脊液检查

其中, 脑结构特征主要包括颅内容积(ICV)、脑室(Ventricles)、内嗅皮层(Entorhinal)等 7 个区域结构的量化值; 认知行为特征主要包括临床痴呆评分总和(CDRSB)、阿尔茨海默症认知功能评价量表(ADAS-cog)得分、简易智力状态检查量表(MMSE)得分等 30 个量表得分; 人口统计学特征包括年龄(AGE)、种族(PTETHCAT)、血统(PTRACCAT)等 6 个基本信息; 生物标志物主要包括载脂蛋白 E ϵ 4 等位基因数目(APOE4)以及 β -淀粉样蛋白(ABETA)、氟代脱氧葡萄糖(FDG)等 7 种检测物含量。

数据集中两种诊断结果是指基线诊断结果(DX-bl)和随访诊断结果(DX)。基线诊断结果是患者首次来访进行诊断的结果, 包括 CN、SMC、EMCI、LMCI 和 AD 五种类型, 共计 2425 条数据; 随访诊断结果是距离患者首次或上一次来访至少六个月后的诊断结果, 包括 CN、MCI 和 AD 三种类型, 共计 16,222 条数据。本文将基于以上两种诊断结果的诊断记录进行 AD 的三分类和五分类诊断。

2.2. 数据预处理

在进行分类诊断之前，需要对数据进行预处理，具体过程如图 1 所示。

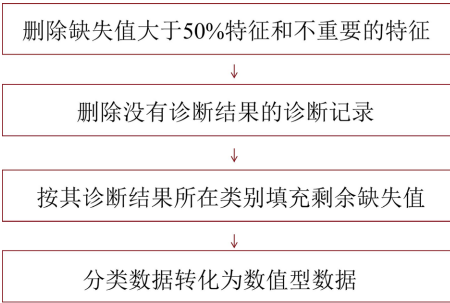


Figure 1. Data preprocessing procedure
图 1. 数据预处理过程

首先,对列进行处理,删除缺失值大于 50%特征和不重要的特征(如患者编号、数据收集研究协议等)。其次,对行进行处理,删除没有诊断结果的诊断记录。再次,对具有少量缺失值的数据按其诊断结果所在类别进行填充,对于数值型数据使用其所在类别的均值进行填充,对于分类数据使用其所在类别的众数进行填充。最后,将数据集中的分类数据转化为数值型数据,以便于后续的特征选择。经过数据预处理后的各类特征情况如表 2 所示。

Table 2. Characteristics before and after preprocessing
表 2. 预处理前后特征情况

	三分类	五分类	原始数据
脑结构特征	7	7	7
认知行为特征	14	29	30
人口统计学特征	6	6	6
生物标志物	1	5	8
总数	28	47	51

原始数据集主要包含 51 个特征,在经过数据预处理后,用于三分类的随访诊断记录保留了 28 个特征,而用于五分类的基线诊断记录保留了 47 个特征,特征个数的差异主要体现在认知行为特征和生物标志物这两类特征。

3. 实验方法

3.1. 特征选择

特征选择作为机器学习中的关键环节,旨在通过剔除与学习任务不相关或冗余的特征,从原始数据集中选择出最具代表性和预测能力的特征,以降低模型的复杂性,提高模型的性能和可解释性。本文使用随机森林算法,分别对预处理后基线诊断记录和随访诊断记录中的特征进行重要性排序,结合特征类别和相关医学知识找到最适合进行 AD 三分类和五分类诊断的特征。

3.2. 分类模型

3.2.1. 随机森林

随机森林(Random forest, RF)是 2001 年 Leo Breiman 和 Adele Cutler 提出的一种基于决策树的集成学习方法,它通过组合多个决策树来进行分类和预测,每个决策树之间没有关联。在分类问题中,随机森林以决策树作为基学习器,并通过投票的方式选取最终的分类结果[12],即所有决策树预测最多的分类结果作为随机森林的最终结果。随机森林的构建过程如下:

- 1) 从训练集中有放回地随机抽取样本,构建每个决策树的训练集。
- 2) 对于每个决策树的每个节点,在一个随机的特征子集中选择最优的特征来进行节点划分。
- 3) 根据选择的特征,递归地构建决策树,直到满足停止的条件。
- 4) 重复以上步骤,创建多个决策树。
- 5) 根据这些决策树的预测结果进行投票,得到最终分类结果。

3.2.2. 梯度提升树

梯度提升树(Gradient boosting decision tree, GBDT)是 1999 年 Friedman 提出的一种迭代的决策树集成算法,它由多个决策树组成,是被公认的泛化能力较强的算法。它对每棵决策树分配不同的权重以训练模型,并在每一步迭代中,通过计算上一次模型的负梯度来改进模型,然后在残差减少的梯度方向上建立新的决策树[13]。GBDT 的迭代思路是通过计算残差的梯度来弥补模型的不足。GBDT 的核心在于累加所有决策树的结果作为最终结果,所以 GBDT 中的决策树都是回归树,而不是分类树。回归树在分枝时会穷举每个特征的每个阈值来找到最好的分割点,以最小化均方误差作为衡量标准。

3.2.3. XGBoost

XGBoost (Extreme gradient boosting)是 2014 年华盛顿大学的博士生陈天奇提出的一种基于 GBDT 的集成学习算法,该算法不仅学习效果优良,而且训练速度高效,广泛应用于多个领域且具有显著效果。XGBoost 是大规模并行的开源提升树工具包,比常见的工具包快 10 倍以上。与 GBDT 最大的不同,XGBoost 的目标函数由损失函数和正则化项两部分构成。在训练过程中,XGBoost 采用二阶泰勒展开式近似损失函数,并通过最小化损失函数来求解最优的树结构和叶子节点的值,引入正则化项控制模型的复杂度,有效防止模型过拟合[14]。在寻找最佳分割节点时,XGBoost 采用完全搜索式的精确贪心算法,选择增益最大的特征来分裂,开启多线程进行,大大提高了运行速度。XGBoost 的并行,不是每棵树可以并行训练,而是特征维度的并行。

3.2.4. LightGBM

LightGBM (Light gradient boosting machine)是 2017 年微软亚洲研究院(MSRA)提出的一种基于 GBDT 的高效、可扩展的集成学习算法,主要用于解决 GBDT 算法在处理海量数据时计算效率低下的问题[15]。与 XGBoost 相比,LightGBM 的优势在于训练速度更快、内存消耗更低、准确率更好以及分布式支持。LightGBM 主要有以下 4 个方面的改进:

- 1) 基于梯度的单边采样算法:仅保留梯度大的样本,并且对梯度小的样本进行随机抽样,在计算增益时给梯度小的样本引入常数进行平衡,减少复杂度。
- 2) 互斥特征捆绑算法:将互斥特征进行融合绑定,在尽可能小的特征损失下降低特征数量。
- 3) 直方图算法:将连续的特征离散化,构造直方图统计信息,以有限的遍历次数找到最佳分裂点。
- 4) 基于最大深度的 Leaf-wise 的垂直生长算法:设置最大分裂深度,只对增益最大的叶子节点进行分裂以减少同层叶子一起分裂的误差。

3.2.5. CatBoost

CatBoost (Categorical Boosting)是 2017 年俄罗斯的搜索巨头 Yandex 开发的一种基于 GBDT 的自适应增强学习算法,能够高效合理地处理类别型特征,通过其独特的策略和优化技术,提高算法的准确性和泛化能力[16]。该算法对数据集进行随机排列,计算相同类别值的样本的平均标签值时,将这个样本之前的样本的标签值纳入计算,既降低过拟合又保证全部数据集都可用于学习。该算法包含了对类别型特征的空值处理、ordered TS 编码、特征组合处理方式,避免了传统方法中需要进行大量的数据预处理和特征工程的问题。此外,在 CatBoost 算法中,梯度的计算是基于整个训练数据集的,而不是基于单个样本的,有助于减少梯度估计的偏差和解决预测偏移的问题。

3.3. 模型评估

K 折交叉验证是一种常用的评估机器学习模型性能和泛化能力的方法。在 K 折交叉验证中,原始数据集随机地被平分成 K 个不重叠子集,然后轮流使用其中的 K - 1 个子集作为训练集用以训练模型,剩余的一个子集作为测试集用于评估上述模型性能。迭代结束后,计算所有测试集评估指标的平均值作为该模型性能的最终估计[17]。K 折交叉验证的优势在于充分利用数据集集中的所有样本,提高了模型稳定性和可靠性,通过多次迭代可以减少单次划分数据集所带来的随机误差。在实际应用中,五折交叉验证和十折交叉验证较为常见。本文采用五折交叉验证,通过计算平均准确率和方差来评估模型性能。

混淆矩阵是机器学习中用于总结分类模型预测结果的表格,它可以提供模型在各个类别上的性能概览。混淆矩阵中的元素可以用于计算评估指标,如准确率、精确率、召回率、F1-分数等,这些指标可以进一步了解模型在各个类别上的性能表现,并且根据评估结果可以对模型进行改进和优化。

4. 实验结果

4.1. 特征选择结果

4.1.1. 三分类诊断的特征选择结果

随访诊断结果包括 CN、MCI 和 AD 三种类型,其诊断记录用于 AD 的三分类诊断任务中。在对随访诊断记录进行数据预处理后,利用随机森林算法将剩余 28 个特征进行重要性排序,结果如图 2 所示。

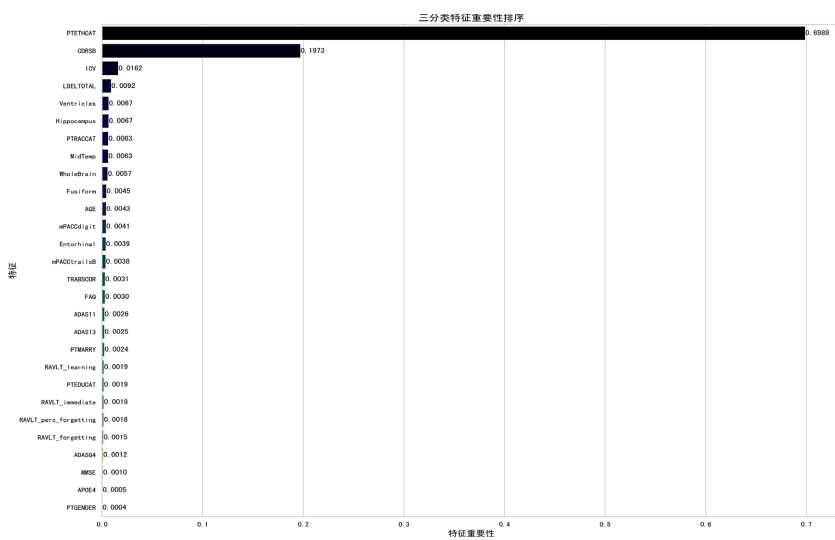


Figure 2. Three-classification feature importance ranking
图 2. 三分类特征重要性排序

在 AD 的三分类诊断中，种族(PTEHAT)在 28 个特征中重要性得分最高，达到了 0.6989。与种族密切相关的特征——血统(PTRACCAT)，其重要性排序位次也比较高，居于第七位。脑结构特征在 AD 三分类特征重要性排序中位次普遍较高，与智力和记忆等认知能力相关的颅内容积(ICV)位于第 3 位，能够反映大脑老化过程的脑室(Ventricles)位于第 5 位，负责短时记忆的海马体(Hippocampus)位于第 6 位，命名性语言中枢的颞中回(MidTemp)位于第 8 位，全脑(WholeBrain)位于第 9 位，与面孔等识别有关的梭状回(Fusiform)位于第 10 位，作为刺激记忆开关的内嗅皮层(Entorhinal)位于第 13 位。此外，认知行为特征的重要性得分有高有低，重要性较高的是临床痴呆评分总和(CDRSB)和逻辑记忆延迟回忆(LDELTOTAL)测试得分，分别排在第 2 位和第 4 位。

年龄与 AD 的发病密切相关，有研究表明，随着年龄的不断增长，患者得 AD 风险会逐年上升[18]。鉴于此，将年龄(AGE)所在位次设定为起始特征个数，并在此基础上依次增加其他特征。通过不断训练机器学习模型，观察各模型在测试集上的准确率和 F1-分数以比较模型性能。表 3 展示了以 8:2 划分训练集和测试集最优模型情况(部分)。

Table 3. Three-classification optimal model (part)
表 3. 三分类最优模型(部分)

特征数	最优机器学习模型	准确率	F1-分数
12	LightGBM	0.9912	0.9912
13	CatBoost	0.9912	0.9912
15	XGBoost	0.9912	0.9912

在特征数为 12、13 和 15 时，LightGBM、CatBoost 和 XGBoost 模型分别表现出了较为优异的性能，它们的准确率和 F1-分数都可以达到 99.12%。但观察特征重要性排序(图 2)，以上数目的特征仅包含脑结构特征、认知行为特征和人口统计学特征，缺乏生物标志物。为平衡特征类别，将预处理后用于三分类的随访诊断记录中唯一的生物标志物——APOE4 加入，以 8:2 划分训练集和测试集，重新逐次训练机器学习模型，通过准确率和 F1-分数比较模型性能，部分结果见表 4。

Table 4. Adding APOE4 three-classification optimal model (part)
表 4. 加入 APOE4 三分类最优模型(部分)

特征数	最优机器学习模型	准确率	F1-分数
11 + 1	XGBoost	0.9916	0.9916
12 + 1	LightGBM	0.9916	0.9916
13 + 1	LightGBM	0.9921	0.9921
14 + 1	LightGBM	0.9925	0.9925
15 + 1	XGBoost	0.9912	0.9912

加入 APOE4 这一特征后，不仅弥补了生物标志物的缺失，并且最优机器学习模型对 AD 三分类诊断的准确率和 F1-分数都有一定的提升。其中，特征重要性得分排在前 14 位的特征，在添加 APOE4 后，其最优机器学习模型 LightGBM 的准确率达到 99.25%，F1-分数也达到 99.25%。较未添加 APOE4，特征数目为 12、13 和 15 所训练出的最优机器学习模型，分类准确率和 F1-分数提高了 0.13%。综上所述，选取特征重要性排序前 14 位的特征和 APOE4 进行进一步的分析，相关特征类别情况见表 5。

Table 5. Three-classification features category situation
表 5. 三分类特征类别情况

	原始数据	预处理后	特征选择
脑结构特征	7	7	7
认知行为特征	30	14	4
人口统计学特征	6	6	3
生物标志物	8	1	1
总数	51	28	15

4.1.2. 五分类诊断的特征选择结果

基线诊断结果包括 CN、SMC、EMCI、LMCI 和 AD 五种类型，其诊断记录用于 AD 的五分类诊断任务中。在对基线诊断记录进行数据预处理后，利用随机森林算法将剩余 47 个特征进行重要性排序，结果如图 3 所示。

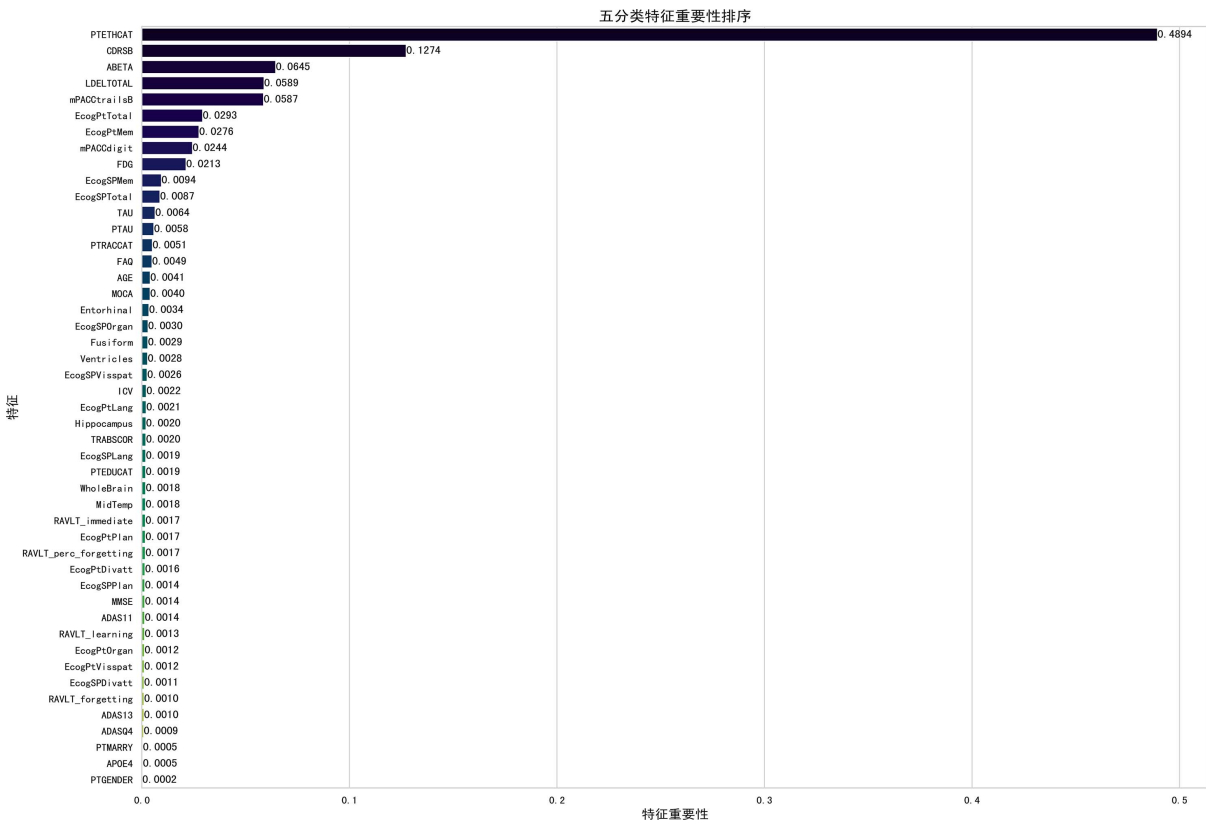


Figure 3. Five-classification features importance ranking
图 3. 五分类特征重要性排序

在 AD 的五分类诊断中，种族(PTETHCAT)在 47 个特征中重要性得分最高，达到了 0.4894，说明它的重要程度较高，这与 AD 三分类诊断的特征重要性排序显现出的结果一致。血统(PTRACCAT)在重要性排序中的位次下降至 14 位，但其重要性得分与三分类相差不大。脑结构特征的重要性得分有所下降，位次也有所降低。内嗅皮层(Entorhinal)位次最高，居于第 18 位，颞中回(MidTemp)位次最低，居于第 30 位。认知行为特征在 AD 五分类诊断中的重要性有所提高，特征重要性排序前 10 位中有 7 位是认知行为

特征。此外，由于基线诊断记录的缺失值较少，生物标志物在 AD 分类诊断中的重要性开始显现。在 AD 五分类特征重要性排序中， β -淀粉样蛋白(ABETA)、氟代脱氧葡萄糖(FDG)、tau 蛋白(TAU)及高度磷酸化的 tau 蛋白(PTAU)这些生物标志物的位次较高，分别位于第 3、9、12 和 13 位。

以年龄所在位次作为起始特征数量，逐步增加一个特征，按照 8:2 划分训练集和测试集，不断训练机器学习模型，通过比较各模型在测试集上的准确率和 F1-分数来选定最优分类模型。表 6 展示了五分类最优模型情况(部分)。

Table 6. Five-classification optimal model (part)
表 6. 五分类最优模型(部分)

特征数	最优机器学习模型	准确率	F1-分数
19	XGBoost	0.9549	0.9548
20	XGBoost	0.9571	0.957
21	XGBoost	0.9594	0.9592
22	XGBoost	0.9594	0.9592
23	XGBoost	0.9571	0.957

当特征数为 21 和 22 时，XGBoost 模型在 AD 五分类诊断中表现出优异的性能，其分类准确率达到 了 95.94%，F1-分数达到了 95.92%。通过查阅在五分类特征重要性排序中位于第 22 位的 EcogSPVisspat 问卷，能够发现该问卷内容与位于第 10 位、第 11 位和第 19 位的 EcogSPMem、EcogSPTotal 和 EcogSPOrgan 等问卷内容具有重合的情况。因此，最终选定位于特征重要性排序前 21 位的特征进行下一步的分析，相关特征类别情况见表 7。

Table 7. Five-classification features category situation
表 7. 五分类特征类别情况

	原始数据	预处理后	特征选择
脑结构特征	7	7	3
认知行为特征	30	29	11
人口统计学特征	6	6	3
生物标志物	8	5	4
总数	51	47	21

4.2. 不同机器学习算法分类结果

4.2.1. 三分类模型诊断结果

随访诊断记录进行特征选择后，保留了 15 个特征，将这 15 个特征和随访诊断结果组成新的数据集，按照 8:2 划分训练集和测试集，选用随机森林、梯度提升树、CatBoost、LightGBM、XGBoost 等机器学习算法进行 CN、MCI 和 AD 之间的三分类诊断，结果如表 8 所示。

Table 8. Three-classification model diagnosis results
表 8. 三分类模型诊断结果

	准确率	召回率	精确率	F1-分数	交叉验证的平均准确率	交叉验证的方差
随机森林	0.9894	0.9894	0.9895	0.9894	0.9794	0.009
梯度提升树	0.9833	0.9833	0.9833	0.9833	0.9776	0.009

续表

CatBoost	0.9912	0.9912	0.9912	0.9912	0.9776	0.0101
LightGBM	0.9925	0.9925	0.9925	0.9925	0.9789	0.0092
XGBoost	0.9908	0.9908	0.9908	0.9908	0.978	0.0099

观察各模型的评价指标,可以发现上述模型的准确率、召回率、精确率和 F1-分数均在 98%以上,说明基于随机森林重要性排序并结合特征类型所选取的 15 个特征在 AD 三分类诊断中的效果较好。对比以上模型,不难看出 LightGBM 的分类效果最好,其准确率、召回率、精确率和 F1-分数均达到 99.25%,进一步采用五折交叉验证评估模型性能,发现各模型交叉验证的平均准确率都在 97%以上,方差在 0.01 左右,说明各模型性能较好且比较稳定。

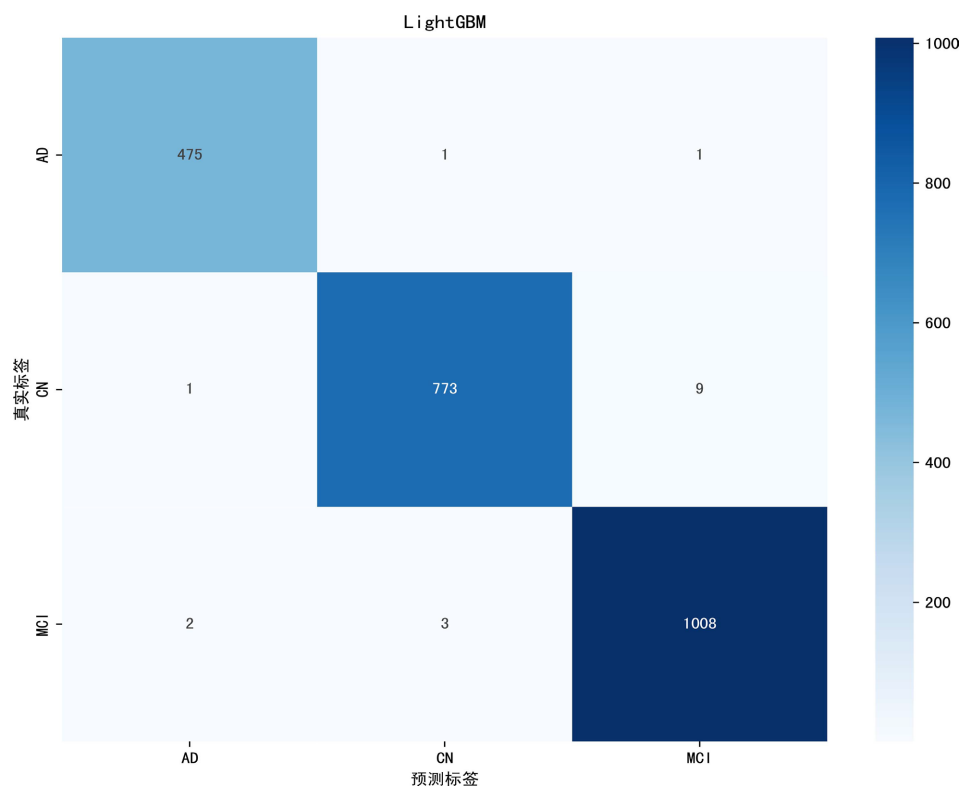


Figure 4. LightGBM confusion matrix

图 4. LightGBM 混淆矩阵

选取最优的 AD 三分类诊断模型 LightGBM, 绘制混淆矩阵(见图 4)。通过计算,可以得到该模型对各类别诊断的精确率分别是: AD 的诊断精确率为 99.37%; CN 的诊断精确率为 99.49%; MCI 的诊断精确率为 99.02%。也就是说,该模型能够通过选取的 15 个特征,实现较为精确的 AD 的三分类诊断。

4.2.2. 五分类模型诊断结果

基线诊断记录经过特征选择后,保留了 21 个特征,将其与基线诊断结果组成新的数据集,以 8:2 划分训练集和测试集,选用随机森林、梯度提升树、CatBoost、LightGBM、XGBoost 等机器学习算法进行 CN、SMC、EMCI、LMCI 和 AD 之间的五分类诊断,结果如表 9 所示。

Table 9. Five-classification model diagnosis results
表 9. 五分类模型诊断结果

	准确率	召回率	精确率	F1-分数	交叉验证的准确率	交叉验证的方差
随机森林	0.9549	0.9549	0.9558	0.9547	0.9295	0.0342
梯度提升树	0.9458	0.9458	0.9477	0.9458	0.9363	0.0314
CatBoost	0.9526	0.9526	0.9543	0.9525	0.9345	0.0392
LightGBM	0.9503	0.9503	0.9511	0.9501	0.9386	0.0287
XGBoost	0.9594	0.9594	0.9604	0.9592	0.9363	0.0305

由表 9 可知,上述模型的准确率、召回率、精确率和 F1-分数均在 94%以上。对于 AD 五分类诊断而言,XGBoost 的分类效果最好,其准确率、召回率、精确率和 F1-分数都是最高的,分别为 95.94%、95.94%、96.04%和 95.92%。进一步采用五折交叉验证评估模型性能,发现各模型交叉验证的平均准确率都在 93%左右,方差在 0.03 左右,可以看出模型的性能和稳定性具有一定保证。与 AD 三分类诊断相比,各模型评估指标略有下降,推断是诊断类别增多导致的。

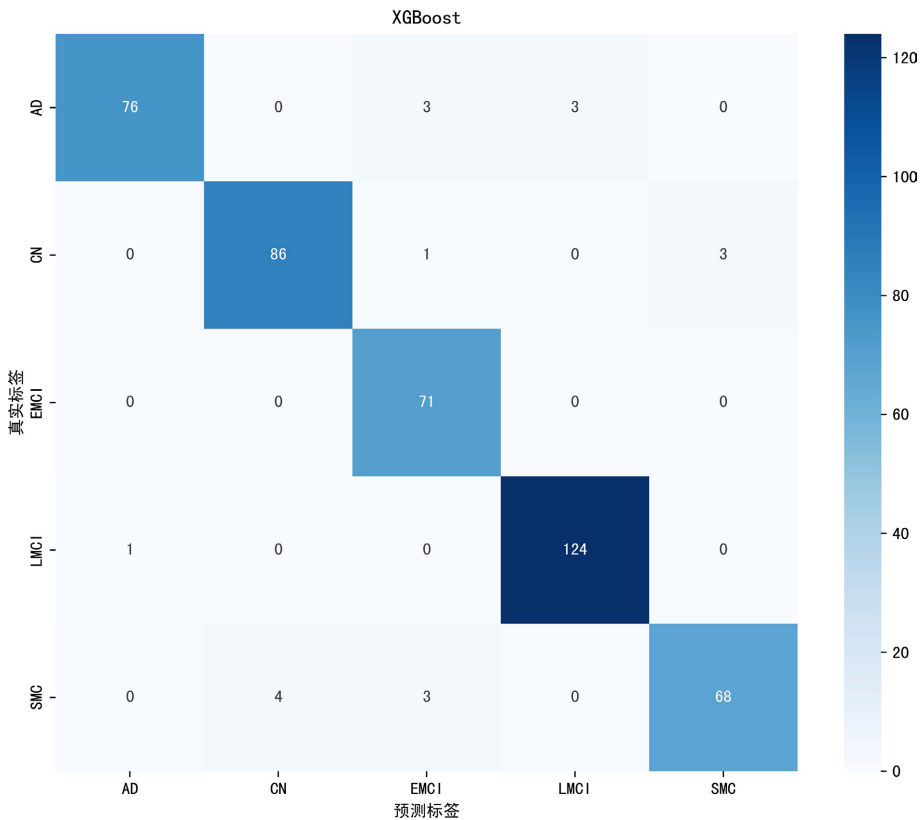


Figure 5. XGBoost confusion matrix
图 5. XGBoost 混淆矩阵

选取最优的 AD 五分类诊断模型 XGBoost, 绘制混淆矩阵(见图 5)。通过计算,可以得出该模型对各类别诊断的精确率分别是: AD 的诊断精确率为 98.70%; CN 的诊断精确率为 95.56%; EMCI 的诊断精确率为 91.03%; LMCI 的诊断精确率为 97.64%; SMC 的诊断精确率为 95.77%。说明基于随机森林特征

重要性排序并结合医学知识选取的 21 个特征,对于症状相对明显的 LMCI 和 AD 阶段诊断较为精确;对于症状较轻的 CN 和 SMC 阶段诊断精确率可以达到平均水平;而对于 EMCI 阶段,会常将其误诊成 SMC、AD 和 CN,导致诊断精确率的降低,这与 EMCI 阶段发病症状的不确定性有关。

5. 结论与建议

以往 AD 分类诊断大多集中于二分类研究,并且侧重于脑结构特征。本研究主要基于脑结构特征和认知行为特征,辅以人口统计学特征和生物标志物实现对 AD 的三分类和五分类诊断。采用随机森林算法对特征重要性进行评估,根据评估结果和各特征组合在机器学习模型中表现性能,结合医学知识及特征类型,分别筛选出 15 个和 21 个特征进行 AD 的三分类和五分类诊断,分类准确率分别最高可达 99.25%、95.94%, F1-分数可以达到 99.25%、95.92%。在 AD 三分类诊断中,所选取的特征主要集中在脑结构特征,约占所选取特征总数的 46.7%,认知行为特征约占选取特征的 26.7%。但在五分类中,认知行为特征更为重要,约占选取特征的 52.4%,脑结构特征所占比例下降至 19.0%。这说明当对 AD 进行简单分类诊断时,可以主要以 MRI 图像所显示的结果(指文中所提到的七个区域的量化值)为主要诊断依据,对于认知行为方面的问卷评估应尽量精简,以临床痴呆评分总和(CDRSB)和逻辑记忆延迟回忆(LDELTOTAL)测试为主。如果进一步对 AD 进行更为细化的诊断时,则需要增加对认知行为方面的评估,主要包括日常认知功能问卷(Ecog)、蒙特利尔认知评估量表(MOCA)、社会功能活动问卷(FAQ)等。不容忽视的是,人口统计学特征中的种族、血统和年龄在两次分类诊断中都起到了不小的作用。此外,生物标志物的存在有助于提高 AD 诊断的准确率,并且随着分类的细化, β -淀粉样蛋白、氟代脱氧葡萄糖、tau 蛋白及高度磷酸化的 tau 蛋白等生物标志物在 AD 的分类诊断中更为重要。

基于上述研究可以发现,随着 AD 分类诊断的细化,需要通过更多的问卷对认知行为进行评估。但由于目前现有各类问卷的制定机构、考量方式、量化准则等不同,不同问卷中的个别问题难免存在着相似或重复的现象,这将大大延长患者的就诊时间。此外,在实际诊疗中医生核对多份问卷工作量相对较大,难免发生疏漏。如果能够根据以上特征选择结果,将多份问卷内容整合成一份,分别设计适用于考量三分类和五分类认知行为的问卷,这将大大减少患者就诊时间,减轻医生工作量,并且对 AD 的分类诊断具有一定的现实意义。

基金项目

河北省自然科学基金面上项目(A2020402006)。

参考文献

- [1] Hahn, K., Myers, N., Prigarin, S., Rodenacker, K., Kurz, A., Förstl, H., *et al.* (2013) Selectively and Progressively Disrupted Structural Connectivity of Functional Brain Networks in Alzheimer's Disease—Revealed by a Novel Framework to Analyze Edge Distributions of Networks Detecting Disruptions with Strong Statistical Evidence. *NeuroImage*, **81**, 96-109. <https://doi.org/10.1016/j.neuroimage.2013.05.011>
- [2] Thies, W. and Bleiler, L. (2012) 2012 Alzheimer's Disease Facts and Figures Alzheimer's Association. *Alzheimer's & Dementia*, **8**, 131-168. <https://doi.org/10.1016/j.jalz.2012.02.001>
- [3] Mitchell, A.J. and Shiri-Feshki, M. (2009) Rate of Progression of Mild Cognitive Impairment to Dementia—Meta-analysis of 41 Robust Inception Cohort Studies. *Acta Psychiatrica Scandinavica*, **119**, 252-265. <https://doi.org/10.1111/j.1600-0447.2008.01326.x>
- [4] Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., *et al.* (2008) Automatic Classification of MR Scans in Alzheimer's Disease. *Brain*, **131**, 681-689. <https://doi.org/10.1093/brain/awm319>
- [5] Li, S., Yuan, X., Pu, F., Li, D., Fan, Y., Wu, L., *et al.* (2014) Abnormal Changes of Multidimensional Surface Features Using Multivariate Pattern Classification in Amnesic Mild Cognitive Impairment Patients. *Journal of Neuroscience*, **34**, 10541-10553. <https://doi.org/10.1523/jneurosci.4356-13.2014>

-
- [6] Silveira, M. and Marques, J. (2010) Boosting Alzheimer Disease Diagnosis Using PET Images. 2010 20th *International Conference on Pattern Recognition*, Istanbul, 23-26 August 2010, 2556-2559. <https://doi.org/10.1109/icpr.2010.626>
 - [7] Wang, Z., Zheng, Y., Zhu, D.C., Bozoki, A.C. and Li, T. (2018) Classification of Alzheimer's Disease, Mild Cognitive Impairment and Normal Control Subjects Using Resting-State Fmri Based Network Connectivity Analysis. *IEEE Journal of Translational Engineering in Health and Medicine*, **6**, 1-9. <https://doi.org/10.1109/jtehm.2018.2874887>
 - [8] 刘茜, 王瑜, 付常洋, 等. 基于特征选择的阿尔茨海默症辅助诊断[J]. 中国医学物理学杂志, 2020, 37(5): 656-660.
 - [9] 杨邦坤, 汪乐生, 聂颖, 等. 基于机器学习的阿尔兹海默症初期行为辨识方法[J]. 生物医学工程研究, 2021, 40(2): 121-125.
 - [10] 李彩, 范焰. 基于机器学习的阿尔兹海默症分类预测[J]. 中国医学物理学杂志, 2020, 37(3): 379-384.
 - [11] 刘德华, 殷国盛, 范焰. 基于特征选择下机器学习对阿尔茨海默病的分类[J]. 中国医学影像学杂志, 2023, 31(2): 167-174.
 - [12] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
 - [13] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, **29**, 1189-1232. <https://doi.org/10.1214/aos/1013203451>
 - [14] Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
 - [15] Ke, G., Meng, Q., Finley, T., et al. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 3149-3157.
 - [16] Prokhorenkova, L., Gusev, G., Vorobev, A., et al. (2018) CatBoost: Unbiased Boosting with Categorical Features. *32nd Conference on Neural Information Processing Systems*, Montréal, 3-8 December 2018, 6639-6649.
 - [17] Rodriguez, J.D., Perez, A. and Lozano, J.A. (2010) Sensitivity Analysis of K-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 569-575. <https://doi.org/10.1109/tpami.2009.187>
 - [18] Citron, M. (2002) Alzheimer's Disease: Treatments in Discovery and Development. *Nature Neuroscience*, **5**, 1055-1057. <https://doi.org/10.1038/nn940>