

基于多变量逻辑回归的帕金森病认知障碍预测模型

巴梦茹, 尹晓红, 李少远

(青岛科技大学自动化与电子工程学院, 山东 青岛 266061)

摘要: 帕金森病 (Parkinson's disease, PD) 患者常伴随认知障碍, 严重影响生活质量, 因此对帕金森病认知障碍进行超前预测对于临床诊断和干预至关重要。然而, 帕金森病受多变量因素 (如年龄、性别、病程时间等) 的耦合影响, 使得认知障碍的超前预测面临严峻挑战。针对帕金森病认知障碍的多变量耦合特性, 采用多变量逻辑回归方法, 构建了一种新型列线图模型, 旨在超前预测帕金森病患者发生认知障碍 (cognitive impairment, CI) 的风险。首先, 应用最小绝对收缩选择算子 (least absolute shrinkage and selection operator, LASSO) 算法对可能影响患者认知能力的风险因素进行了分析, 筛选出相关性高的临床变量。其次, 采用多变量逻辑回归方法分析各变量之间的相关性, 构建可视化的新型列线图模型, 实现对帕金森病认知障碍的风险超前预测。最后, 模型性能评估结果表明新型认知障碍预测模型具有良好的准确性、一致性和临床实用性, 可显著提高临床医生的诊断效率。此外, 该模型还实现了对同一预测因子不同值的患者数量及分布的可视化对比和分析, 能够辅助临床医生根据每位患者的个人风险制定个性化的医疗管理和咨询方案, 有助于更早地展开对患者的干预和治疗, 具备一定的临床诊断价值。

关键词: 帕金森病; 认知障碍; 多变量逻辑回归; 列线图

中图分类号: TP181

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202413

A predictive model of cognitive impairment in Parkinson's disease based on multivariate logistic regression

BA Mengru, YIN Xiaohong, LI Shaoyuan

School of Automation & Electronic Engineering, Qingdao University of Science & Technology, Qingdao 266061, China

Abstract: Parkinson's disease (PD) patients were often accompanied by cognitive impairment, which seriously affected the quality of life, so the over-prediction of cognitive impairment in Parkinson's disease was crucial for clinical diagnosis and intervention. However, Parkinson's disease was affected by the coupling of multivariate factors, such as age, gender, and disease duration, which made the overprediction of cognitive impairment a serious challenge. Aiming at the multivariate coupling characteristics of cognitive impairment in Parkinson's disease, in this study, multivariate logistic regression was used to construct a novel column-linear graphical model aiming at over-predicting the risk of cognitive impairment (CI) in Parkinson's disease patients. First, the least absolute shrinkage and selection operator (LASSO) algorithm was applied to analyze the risk factors that may affect the cognitive ability of patients, and the clinical variables with high corre-

收稿日期: 2023-11-23; 修回日期: 2024-04-29

通信作者: 尹晓红, yxh@qust.edu.cn

基金项目: 山东省自然科学基金项目 (No. ZR2023ZD49, No. ZR2023MF028); 山东省重点研发项目 (重大科技创新工程) (No. 2020CXGC011402); 青岛市临床重点专科项目; 青岛市科技示范指导项目 (No. 20-3-4-37-NSH); 山东大学青岛齐鲁医院灵活人才项目 (No. QDKY2019RX05, No. QDKY2019RX13)

Foundation Items: Natural Science Foundation of Shandong Province (No. ZR2023ZD49, No. ZR2023MF028), Shandong Province Key Research and Development Project (Major Science and Technology Innovation Project) (No. 2020CXGC011402), Qingdao Clinical Key Specialty Project Fund, Qingdao Science and Technology Demonstration and Guidance Project (No. 20-3-37-NSH), Shandong University Qingdao Qilu Hospital Flexible Talent Project (No. QDKY2019RX05, No. QDKY2019RX13)

lation were screened out. Second, multivariate logistic regression was used to analyze the correlation between variables and construct a visualized novel column-line diagram model to achieve the risk over prediction of cognitive impairment in Parkinson's disease. Finally, the results of model performance evaluation show that the novel cognitive impairment prediction model proposed in this paper has good accuracy, consistency and clinical practicability, which can significantly improve the diagnostic efficiency of clinicians; in addition, the model also realizes the visual comparison and analysis of the number and distribution of patients with different values of the same predictor, which can assist clinicians in formulating personalized healthcare management and consulting programs according to the individual risk of each patient, and help to start the intervention and treatment of the patients at an early stage, and it has a certain clinical diagnostic value.

Key words: Parkinson's disease, cognitive impairment, multivariate logistic regression, nomogram

0 引言

帕金森病^[1-2]是一种涉及神经系统的慢性疾病, 不仅表现为典型的运动障碍, 而且大约一半以上的帕金森病患者在疾病进展过程中会经历认知障碍。认知障碍的症状包括记忆力下降、注意力不集中、执行功能障碍(如计划和组织能力受损)、语言障碍和空间意识问题等, 不但对患者的生活质量产生显著影响, 还进一步增加了患者护理成本和护理人员负担。研究表明, 患有帕金森病认知障碍(Parkinson's disease with cognitive impairment, PD-CI)的患者中, 超过80%的患者在疾病持续发展的过程中会发展成帕金森病痴呆(Parkinson's disease dementia, PDD)^[3-5]。因此, 为确保PD-CI的早期诊断和治疗, 亟须开发准确的预测模型, 帮助医护人员在疾病早期识别风险, 及时提供精准治疗。

当前, 国内外关于PD-CI的研究主要集中在利用不同的认知评估工具和生物标志物技术进行诊断。在评估工具的研究中, Hu等^[6]采用蒙特利尔认知评估量表(montreal cognitive assessment, MoCA)和最小精神状态检查(minimum mental state examination, MMSE)对认知功能进行了评估。结果显示, 根据MoCA和MMSE的分类阈值, 在筛查时, 很大一部分患者已经患有认知障碍(CI)和痴呆范围的认知障碍。使用筛选阈值, MMSE在检测基线损伤方面似乎比MoCA更敏感。钱忠立等^[7]研究了MoCA在评估睡眠剥夺对认知功能的影响方面的应用。研究发现, 24小时睡眠剥夺对总体认知功能产生了影响。然而, 以上研究仅限于MoCA的应用, 而评估工具本身的时效性和可信度对研究结果可能产生较大的影响, 存在潜在的偏差。同时, 由于传统评估工具对认知功能的全面评估存在不足^[8], 且测试项目繁多, 其在临床应用上存在一定限制。

值得关注的是, 目前对于PD-CI的评估主要局限于已经发生认知障碍的患者, 存在滞后性评估的问题。然而, 如何超前并精准预测尚未发生认知障碍但存在高危险性的PD患者及其发生认知障碍的风险概率仍亟待突破。

生物学标志物在诊断认知障碍方面的研究逐渐引起关注。Naskar等^[9]基于无标记无偏非靶向蛋白质组学(LC-MS/MS)分析提出(A)纤维蛋白原 α 链(FGA)作为区分PD-CI的可能生物标志物, 对于PD-CI、PD和正常压力脑积水(NPH)患者, 使用酶联免疫吸附实验(enzyme linked immunosorbent assay, ELISA)技术估计了患者脑脊液(CSF)中 α -突触核蛋白的水平。结果显示FGA可能是区分PD-CI的更好标志物。Chen等^[10]评估血清外泌体中是否存在用于区分认知障碍的生物标志物, 采用基于串联质谱标签方法的非靶向蛋白质组学, 展示了认知障碍患者和健康对照之间外泌体中差异表达的蛋白。通过免疫印迹和酶联免疫吸附测定验证了这些差异蛋白, 并确认了认知能力与血清外泌体蛋白水平的相关性。刘伟林等^[11]对来自24篇相关文章的4 661名个体进行了一项综合分析, 确定了脑脊液和血液中神经粒蛋白水平是阿尔茨海默病患者和认知障碍患者的诊断生物标志物。Cosgrove等^[12]综述了帕金森病认知障碍的病理生理机制和管理策略, 详细描述了认知障碍的多种表现, 包括注意力、记忆、执行功能和视觉空间能力等方面的受损。同时, 还讨论了目前对于帕金森病认知障碍的治疗方法和未来的研究方向。然而, 使用生物标志物进行认知障碍的诊断需要对患者进行长期随访, 这使得预测相对复杂且耗时较长, 可能导致患者错过早期治疗的机会。

为了解决上述问题, 本文深入探讨预测PD-CI的新方法, 提出了一种基于多变量逻辑回归的新型

列线图预测模型,该模型耗时短、操作简单、准确性高、预测结果能够可视化,从而能够快速有效地超前预测PD患者发生认知障碍的风险,帮助临床医生尽早采取干预措施,提供更准确的诊断和诊疗方案,降低PD-CI的发生率,提高PD患者的生活质量。

1 对象和方法

基于多变量逻辑回归的帕金森病认知障碍预测主要包括以下步骤:数据预处理、高风险影响因素选择、基于逻辑回归的列线图模型、分析评估,预测工作流程如图1所示。

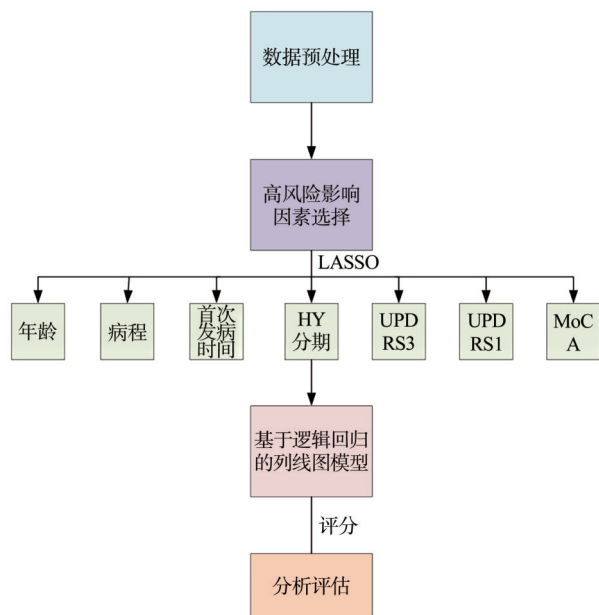


图1 预测工作流程

1.1 数据预处理

1.1.1 研究对象

本文中使用的数据来自帕金森病进展标志物倡议(PPMI)数据库^[13]。

研究对象筛选流程如图2所示,在PPMI数据库中随机选择268名患者,在排除具有未知HY(Hoehn-Yahr)分期和统一PD评分量表(unified Parkinson's disease rating scale, UPDRS),患有严重系统性疾病的患者,有严重焦虑、抑郁和精神分裂的患者后,将其余患者纳入数据集并且按照7:3的比例划分为训练集($n=122$)和验证集($n=52$)。

1.1.2 预处理

为了提高数据的准确性和可靠性,本文对数据

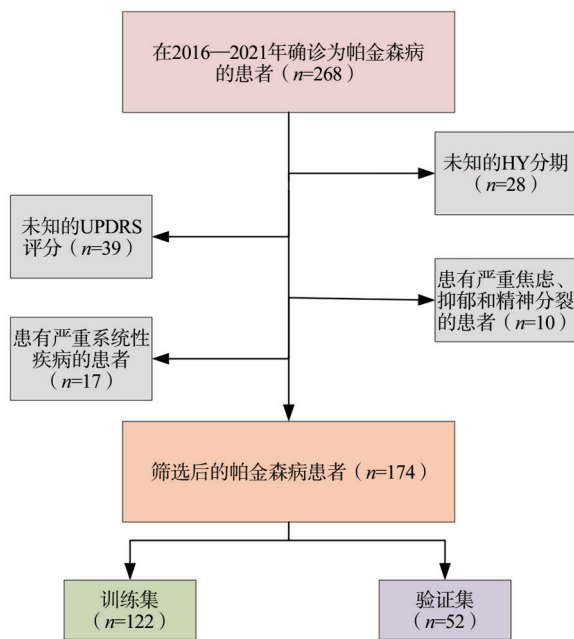


图2 研究对象筛选流程

进行预处理,包括数据的清洗和标准化过程。首先,在数据清洗阶段剔除异常值和无效值;其次,在标准化阶段采用Z-Score标准化方法^[14],提高数据的可比性。Z-Score标准化如式(1)所示:

$$Z = \frac{x - \mu_1}{\sigma} \quad (1)$$

其中, x 为个体的观测值, μ_1 为总体数据的均值, σ 为总体数据的标准差,计算方法如式(2)所示:

$$\sigma = \sqrt{\sum_{i=1}^n p_i (x_i - \mu_2)^2} \quad (2)$$

其中, μ_2 为随机变量 x 的均值, p_i 为随机变量 x_i 的出现概率。

1.2 特征选择

1.2.1 相关因素的选择

相关研究^[15-16]证实,年龄、首次发病时间和病程是影响帕金森病的重要因素。病程的长短与病情的严重程度和进展速度密切相关。一般来说,病程越长,患者的病情往往越严重,运动症状和非运动症状也更加明显。另外,首次发病年龄越大,帕金森病的患病率越高。首次发病时间还可能影响患者的症状表现和疾病进展速度。此外,HY分期、UPDRS、嗜睡量表(epworth sleepiness scale, ESS)和蒙特利尔认知评估量表(Montreal cognitive assessment, MoCA)都是评估帕金森病严重程度的重要工具。其中,HY分期是根据患者的症状和

体征，将帕金森病分为1到5期。UPDRS是一个全面的评估工具，包括4个部分（分别用UPDRS1~UPDRS4表示），分别用于评估精神行为、日常生活活动、运动功能和并发症。ESS用于评估患者的日间嗜睡程度，MoCA是一个用于评估轻度认知功能障碍的敏感工具。因此，本文选择了年龄、病程、首次发病时间、HY分期、UPDRS1、UPDRS2、UPDRS3、UPDRS4、ESS和MoCA作为相关风险因素。

1.2.2 高风险因素筛选

LASSO算法是Tibshirani R^[17]在1996年提出的一种用于线性回归和模型的特征选择方法。LASSO算法在优化模型时，通过L1范数形成的惩罚项来约束模型系数，推动部分系数变为零，从而实现特征的稀疏性，即自动进行特征选择。LASSO的优化目标如式（3）所示：

$$\arg \min_{\beta} \sum_{n=1}^N (y^n - \sum_{p=1}^m \beta_p x_p^n)^2 + \lambda \sum_{p=1}^m |\beta_p| \quad (3)$$

其中， y^n 为第 n 个样本的类别标签； N 为样本数量； x_p^n 为第 n 个样本的第 p 个临床影响因子； β_p 为第 p 个影响因子的回归系数； λ 代表正则化参数。 λ 越大，模型越稀疏，意味着更多的 β_p 被置为0。

本文基于LASSO回归算法对临床数据中的10个影响因子进行分析和筛选，将特征的系数进行压缩并使某些回归系数变为0，进而达到特征筛选的目的^[18]。

1.3 基于逻辑回归分析的列线图模型

帕金森病认知障碍受多变量因素的耦合影响，如年龄、性别、病程时间等，逻辑回归分析^[19]是统计学中用于研究两个或多个变量之间关系的模型。本文基于逻辑回归分析方法，探讨与PD相关的各种影响因素，构建可视化的新型列线图模型，旨在超前预测PD患者出现认知功能障碍的风险。本文采用Sigmoid函数，将逻辑回归模型的输出值转化为二分类的变量，通常表示为0或1。Sigmoid函数是对数几率函数，其函数曲线如图3所示。

基于二分类任务的逻辑回归模型可以通过事先设定阈值或通过给定的输入 x ，比较两个概率值的大小，从而对样本进行分类。具体如下：

$$P(Y=1|x) = \frac{\exp(\omega^T x + b)}{1 + \exp(\omega^T x + b)} \quad (4)$$

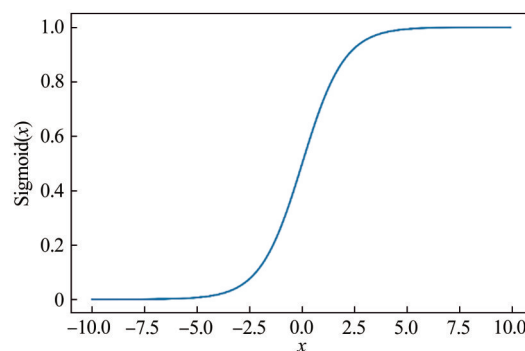


图3 Sigmoid函数曲线

$$P(Y=0|x) = \frac{1}{1 + \exp(\omega^T x + b)} \quad (5)$$

其中， $x \in \mathbf{R}^n$ 是输入， $Y \in \{0, 1\}$ 是输出， $\omega \in \mathbf{R}^n$ 和 $b \in \mathbf{R}$ 是参数，属于权值向量。

本文通过可视化多因素逻辑回归分析结果，构建了新型PD-CI的新型列线图模型。与传统的列线图相比，这一新型模型不仅能够有效预测PD-CI的风险，而且在可视化方面具备独特的优势。该模型能够清晰地展示同一预测因子在不同数值下患者数量的分布情况，从而更全面地呈现数据集中患者的分布特征。这为医生提供了更丰富的信息，有助于深入分析和研究特定范围内患者的分布情况。

1.4 模型评估

为了评估模型的准确性、一致性和临床实用性，本文基于ROC曲线、校准曲线和决策曲线（decision curve analysis, DCA）等指标来验证模型的性能。

1.4.1 ROC曲线

ROC曲线是用于量化诊断测试或预测模型鉴别准确性的重要工具，其图形化呈现了诊断测试的真正阳性率（Y轴）与假阳性率（X轴）之间的关系。对角线是从（0，0）到（1，1）的线段，代表了纯粹猜测的情况，即模型的鉴别准确性仅为随机水平，因此其AUC值为0.5，表明模型无效。

通过计算不同阈值下的真正阳性率和假阳性率，绘制出曲线，从而更直观地展示了诊断测试在不同阈值下的性能表现。曲线下面积（area under the curve, AUC）则为ROC曲线的一个综合性指标，用于全面评估测试的准确性。在理想情况下，AUC值为1.0，表示测试完全能够区分阳性和假阳性两种情况^[20]；而AUC值为0.5则说明测试没有区分能力，相当于随机猜测。

因此,通过ROC曲线分析,可以直观地了解诊断测试在不同阈值下的性能,并通过AUC值进行定量评估。这一方法在评估医学诊断测试、模型预测性能等方面被广泛应用,为研究者和临床医生提供了有效的工具,用于评估和比较不同测试或模型的准确性和鉴别能力。

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

其中,TPR代表真阳性率,FPR代表假阳性率。ROC曲线的水平坐标表示FPR,垂直坐标表示TPR。

1.4.2 校准曲线

校准曲线用于评估模型预测概率与实际概率之间的一致性。在坐标系中,模型预测概率为横坐标,实际概率为纵坐标。一条45°向上且经过原点的直线(理想线)表示模型预测概率与实际概率完全吻合,即模型具有完全一致性。Apparent线表示模型在自身训练中的表现,而Bais-corrected线表示经过反复自助抽样后训练出的模型表现,它校正了过拟合情况。预测模型的预测输出与数据真实输出之间的距离大小反映了该预测模型与实际数据的一致性程度,距离越小,一致性越高。

1.4.3 DCA曲线

在帕金森疾病的临床研究中,干预后患者的受益情况是判断模型是否有效的重要指标之一。本文利用DCA曲线分析方法,评估了帕金森患者在接受治疗干预后的受益情况。

DCA曲线是一种基于不同影响因素预测临床结果的方法^[21],其本质是根据假阳性和假阴性的相对损失值,确定一个阈值概率来判断是否进行干预。如果受试者的概率在可接受范围内,则建议进行干预或治疗。净收益是真阳性减去假阳性的权衡,通过式(8)计算:

$$NB = \frac{TP}{N} - \frac{FP}{N} \times \frac{P_t}{1 - P_t} \quad (8)$$

其中, N 是样本大小, P_t 是阈值概率,TP表示真阳性,FP表示假阳性值。

1.4.4 交叉验证

为了验证模型在现有样本上的稳定性和泛化能力。本文使用 K 折交叉验证^[22]对模型进行训练。 K 折交叉验证适用于不同大小的数据集,通过选择合适的 K 值,可以有效避免模型过于复杂而引起的过拟合,以及模型过于简单而引起的欠拟合。通过 K 折交叉验证方法对样本数据进行分组训练,选定 $K=10$,本文将174个样本分为10组进行10次训练和验证,在每次迭代中,模型在选定的9组上进行训练,并在剩余的1组中进行验证,十折交叉验证示意如图4所示。

2 实验结果与分析

2.1 研究对象的临床数据分析

在PPMI数据库中选取的2016年至2021年的268名PD患者中,总共有174名PD患者满足了入组的条件。依据2012年MDS工作组所制定的PD-CI诊断准则^[23],在174名帕金森病患者中,有42名

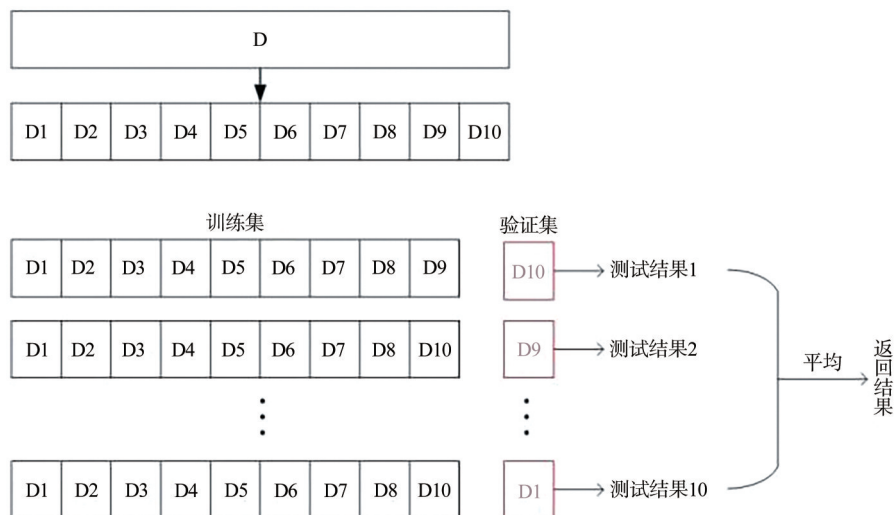


图4 十折交叉验证示意

患者（18分 $\leq$ MoCA评分 $\leq$ 26分）被诊断为PD-CI，而132名患者（MoCA评分 $>$ 26分）被诊断为帕金森病非认知功能障碍（PD-nCI），PD-CI的发病率达到了24.14%。表1展示了包括患者特定特征参数（如人口统计数据和临床表现）在内的数据，其中 $\bar{x}$ 表示各影响因素的平均值， s 表示上下浮动范围。

表1 CI与非CI患者特定特征参数的差异

影响因素	非CI(%) $n=132$	CI(%) $n=42$	总计 $n=174$
年龄($\bar{x}\pm s$)	59.48 $\pm$ 10.69	66.43 $\pm$ 7.87	61.35 $\pm$ 10.46
病程($\bar{x}\pm s$)	7.00 $\pm$ 7.41	7.75 $\pm$ 11.01	7.08 $\pm$ 6.54
首次发病时间($\bar{x}\pm s$)	57.34 $\pm$ 10.19	64.42 $\pm$ 7.85	59.86 $\pm$ 9.99
HY分期($\bar{x}\pm s$)	1.09 $\pm$ 1.03	2.01 $\pm$ 0.84	1.34 $\pm$ 1.06
UPDRS3($\bar{x}\pm s$)	17.41 $\pm$ 16.85	33.05 $\pm$ 16.34	23.48 $\pm$ 19.32
UPDRS1($\bar{x}\pm s$)	6.43 $\pm$ 5.02	10.04 $\pm$ 5.86	7.41 $\pm$ 5.48
UPDRS2($\bar{x}\pm s$)	5.25 $\pm$ 6.04	11.04 $\pm$ 7.32	6.82 $\pm$ 3.28
UPDRS4($\bar{x}\pm s$)	2.77 $\pm$ 3.03	2.36 $\pm$ 3.04	2.61 $\pm$ 0.44
ESS($\bar{x}\pm s$)	6.44 $\pm$ 4.19	8.81 $\pm$ 4.95	7.08 $\pm$ 4.52
MoCA($\bar{x}\pm s$)	27.35 $\pm$ 2.55	25.39 $\pm$ 3.77	26.69 $\pm$ 3.28

2.2 LASSO特征筛选实验结果

本文将认知障碍是否发生作为输出变量，将174例帕金森病患者的10个高风险指标作为自变量，采用LASSO回归方法进行降维，将系数向量中小的权重变为0，从而实现特征选择和模型稀疏性，结果如图5和图6所示。

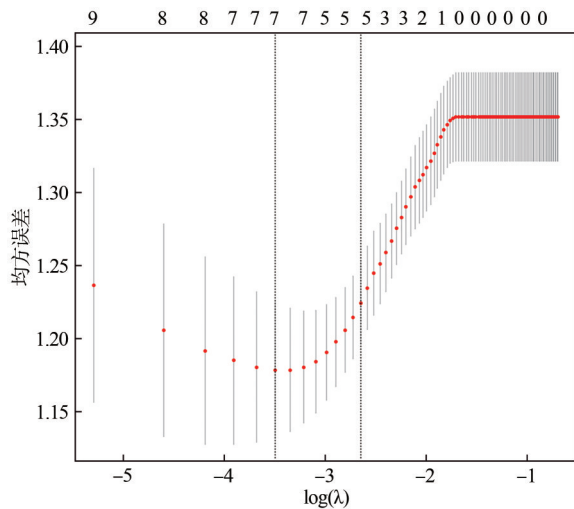


图5 LASSO拟合的10倍交叉验证选择参数

LASSO拟合的10倍交叉验证选择参数如图5所示，纵坐标为交叉验证的均方误差，纵坐标越小说明LASSO拟合效果越好，上横坐标为不同 λ 值对应的特征个数，下横坐标为 $\log \lambda$ ，也叫 λ 惩罚系

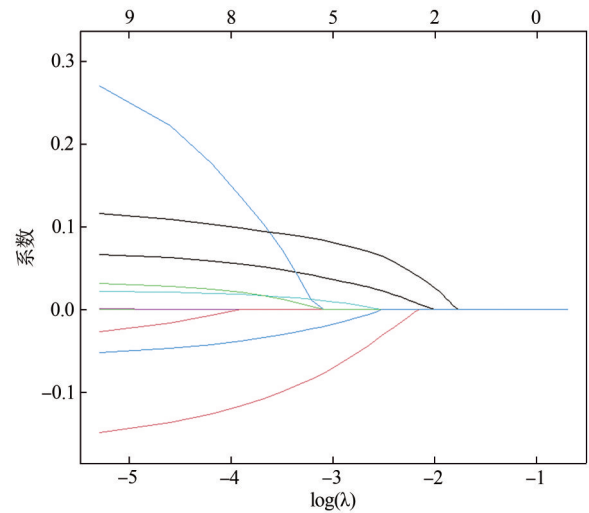


图6 非零系数特征的LASSO系数分布

数的对数。当误差最小，即 $\log \lambda$ 为-3.5时，垂直线对应的特征数量为7，即筛选出非0系数的特征个数为7。非零系数特征的LASSO系数分布如图6所示，其中10条不同颜色的线代表10个高风险指标，每一条曲线代表每一个自变量系数的变化轨迹，上横坐标为模型中非0系数的个数，纵坐标为系数，下横坐标为 $\log \lambda$ 也叫 λ 惩罚系数的对数，当 $\log \lambda$ 为-3.5时，筛选出相应的特征分别是年龄、病程、首次发病时间、UPDRS1评分、UPDRS3评分、MoCA以及HY分期。

最大相关最小冗余（max-relevance and min-redundancy, mRMR）算法^[24-25]是特征选择领域中的一种先进方法，旨在找到与目标变量最相关且彼此之间冗余性最小的特征子集。

本文对LASSO算法和mRMR算法进行了实验对比和验证，通过LASSO算法筛选出以下高风险因素：年龄、病程、首次发病时间、UPDRS1评分、UPDRS3评分、MoCA以及HY分期。采用mRMR算法得到的高风险因素包括：年龄、病程、UPDRS1评分、ESS、MoCA以及HY分期。为了客观评估这两种算法在帕金森病认知障碍预测中的性能，进行了详细的实验对比，结果如表2所示。

表2 不同特征筛选方法的对比和评估

方法	数据集	准确率	精确率	召回率	F1分数
LASSO	训练集	0.884	0.875	0.860	0.868
	验证集	0.880	0.870	0.850	0.860
mRMR	训练集	0.764	0.750	0.710	0.730
	验证集	0.738	0.718	0.705	0.712

表2的实验结果显示, 基于LASSO算法筛选出的高风险因素所构建的逻辑回归模型, 在训练集和验证集上分别取得了0.884和0.880的准确率。相比之下, 采用mRMR算法筛选出的因素所构建的模型, 在训练集和验证集上的准确率则分别为0.764和0.738。此外, mRMR算法在精确率、召回率以及F1分数3个指标上也均低于LASSO算法。主要原因在于, LASSO算法通过引入L1正则化项, 更偏向于构建稀疏模型并优化帕金森病认知障碍模型整体的预测性能; mRMR算法则更侧重于探寻与帕金森病紧密相关且相互间冗余度极低的认知障碍特征集合, 虽然这有助于理解认知障碍特征与帕金森病之间的内在联系, 但牺牲了部分预测性能。因此, 在本实验中, 使用LASSO算法进行特征选择构建帕金森病认知障碍模型。

本文基于LASSO方法筛选的7个预测因子, 采用多变量逻辑回归分析, 结果如表3所示, 其中P值代表自变量对因变量的影响程度, OR值代表自变量与因变量的关联程度, $P<0.05$ 代表有统计学差异。另外, 考虑到预测因子之间可能存在一定的相关性^[26], 本文对这7个预测因子进行了相关性热力图分析, 结果如图7所示。可以看出, 在临床高风险影响因素中, 年龄和首次发病时间以及HY分期和UPDRS3评分之间具有较高的相关性。值得注意的是, 本文筛选的7个高风险变量在临床诊断和治疗过程中易于获取且具有高准确度, 因此有利于模型的准确建立, 并对于深入探讨帕金森病认知功能障碍的发生机制、制定针对性的预防和治疗策略具有重要意义。

2.3 多变量逻辑回归模型验证

基于上述LASSO特征筛选出的7个高风险变量, 本文构建了帕金森病认知障碍多变量逻辑回归模型, 并通过新型列线图模型(如图8所示), 对

PD患者的CI风险进行可视化。

根据本文所提出的列线图模型, 可以有效地计算每例PD患者发生认知障碍的风险系数。例如, 某PD患者的HY分期为1(得分49分)、年龄30岁(得分52分)、UPDRS1评分26(得分60分)、MoCA评分14(得分64)、首次发病时间70岁(得分60)、病程5年(得分54分)、UPDRS3评分10(得分43分), 则该患者的列线图模型评分总得分=49+52+60+64+60+54+43=382分。根据该患者的风险系数计算结果, 可以判断该患者有较高的风险发生认知障碍。在临床诊断和治疗过程中, 对于具有较高认知障碍风险的患者, 我们应高度重视并持续进行随访观察, 确保细致观察患者的病情变化和进展情况, 并定期检查患者的认知程度。通过及时调整治疗策略, 有助于降低认知障碍的发生风险, 提高患者的生活质量。

预测PD患者CI风险的传统列线图模型(如图9所示), 仅用于评估发生CI的风险, 无法直观地展示同一影响因素不同值的患者分布情况。然而, 本文构建的新型列线图模型不仅可以对PD-CI进行风险预测, 还能够以可视化的方式展示同一预测因子不同值的患者数量, 并对数据集中的患者分布进行比较。这一改进后的方法使研究结果更为清晰和明了, 通过可视化患者分布, 医生可以更全面地了解哪些患者群体面临更高的认知障碍风险。这为制定更为精准的预防和治疗策略提供了有力支持。因此, 该新型列线图模型不仅在帕金森病认知障碍的风险预测方面具有创新性, 而且为医疗决策提供了更为全面的信息, 有望在临床实践中取得显著的应用价值。

2.4 模型评估结果分析

为了对本文提出的列线图模型的预测性能进行准确评估, 首先采用了ROC曲线分析方法。ROC

表3 PD患者认知障碍的多变量Logistic回归分析

影响因素	95% 置信区间			OR	P
	绝对值	2.5%	97.5%		
年龄	0.06938515	1.0189332969	1.135968	1.07184895	0.011549635
病程	0.1039191	0.8037749	0.9876182	0.9012982	0.04713619
首次发病时间	0.06808892	1.0176028547	1.1347004	1.07046049	0.013384583
UPDRS1	0.09110524	1.0022862	1.2045319	1.0953843	0.04948798
UPDRS3	0.03527075	1.00250364	1.0749078	1.0359001	0.04508169
HY 分期	0.058362793	1.03947293	1.2804874	0.8293749	0.00482048
MoCA	0.1411373	0.7487251	0.9935284	0.8683701	0.04753511

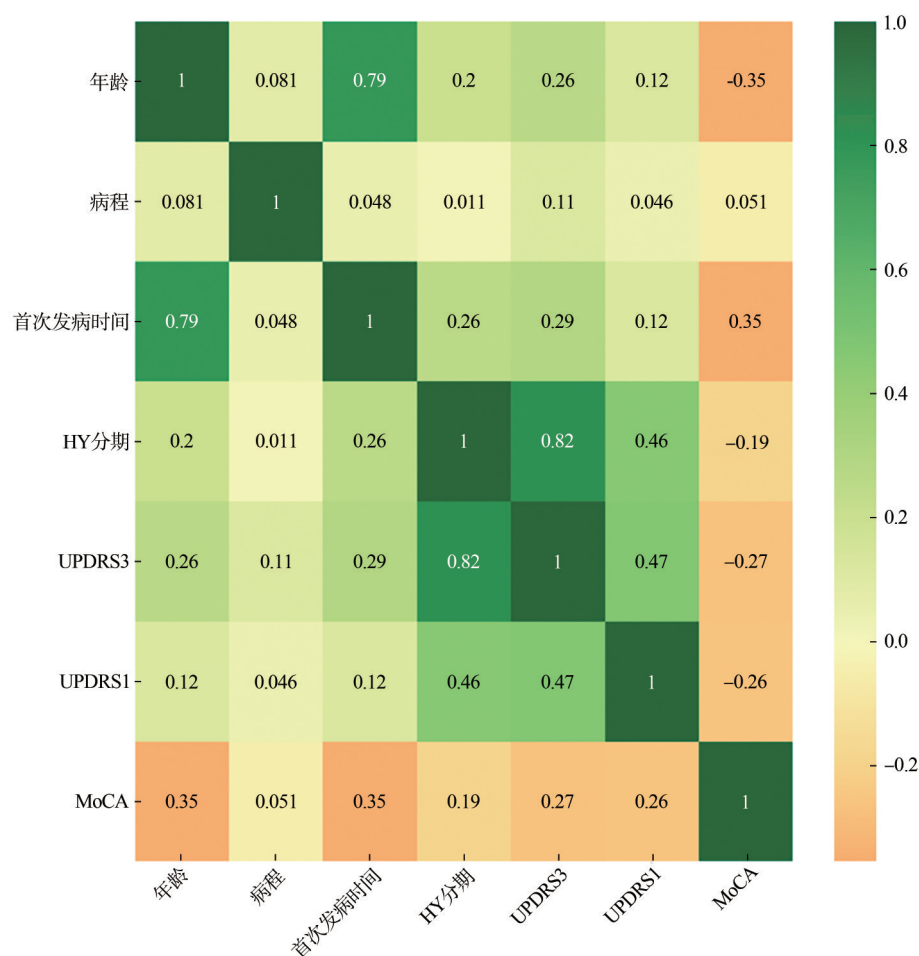


图7 CI影响因子相关性热力图

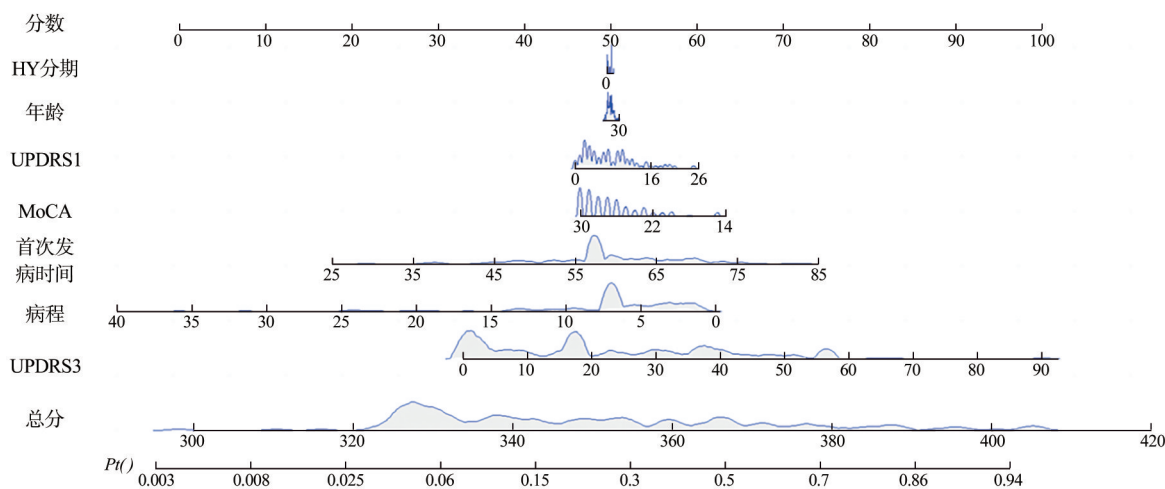


图8 预测PD患者CI风险的新型列线图模型

曲线根据 AUC 值来评估模型的预测准确性，它能够直观地展示模型在不同分类阈值下的效能。通过使用模型校准曲线，本文对列线图模型的一致性进行了评估。值得注意的是，虽然上述评价指标可以

验证所创建的列线图模型具有良好的一致性，但这并不意味着所有患者都能从中受益。为了明确列线图模型在临床实践中的受益范围，本文对 PD 队列在不同认知障碍概率阈值下的净收益进行了量化

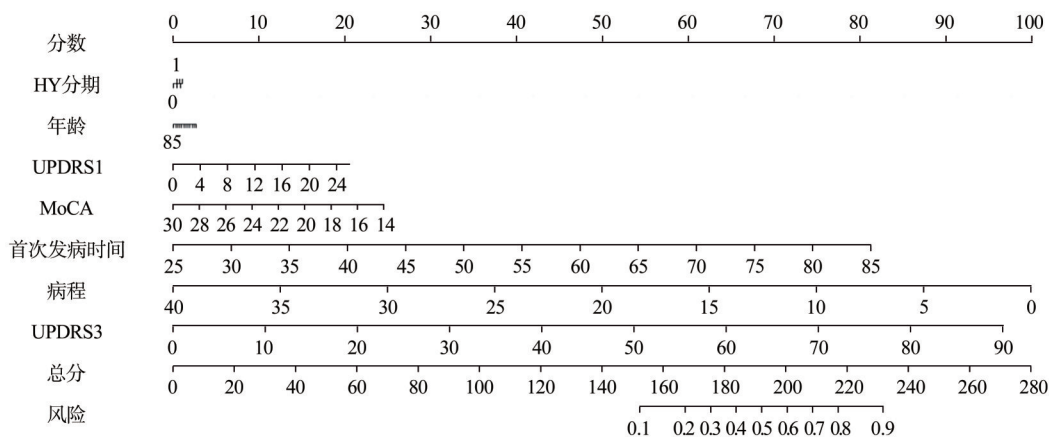


图9 预测PD患者CI风险的传统列线图模型

分析。

此外,本文依据临床决策曲线方法对模型的决策效果进行了研究。临床决策曲线是一种综合考虑诊断准确性和临床决策效果的评价方法,能够直观地展示模型在不同阈值下的决策效果。通过DCA分析,可以更全面地评估模型在临床实践中的应用价值^[27]。

图10和图11分别对训练集和验证集的预测性能进行了评估。从图中可以看出,训练集和验证集的AUC值分别为0.872和0.870,表明本文所提出的模型在训练集和验证集上均具有较高的预测准确度,能够有效地预测PD患者的认知障碍风险。

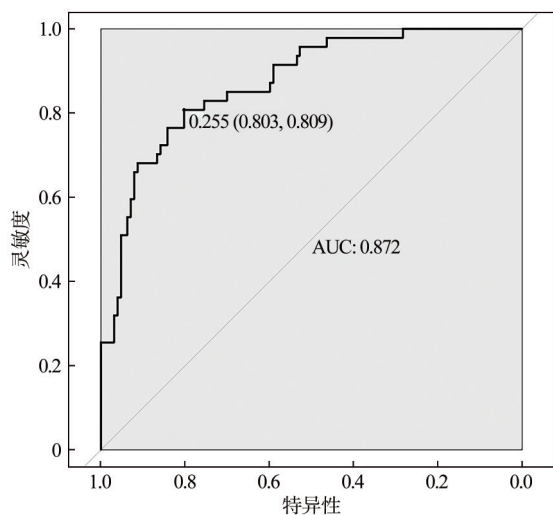


图10 训练集ROC曲线

在校正曲线上,多变量逻辑回归预测PD患者认知障碍风险列线图模型展现出高度一致性。训练集和验证集校准曲线分别如图12和图13所示,其中,实际曲线代表模型在训练过程中的表现,而偏

移校正曲线代表经过反复自助抽样训练得到的模型的表现。

决策曲线是在特定条件下,不同解决方案所对应的风险和收益之间关系的曲线。训练集(training set)和验证集(validation set)决策曲线分别如

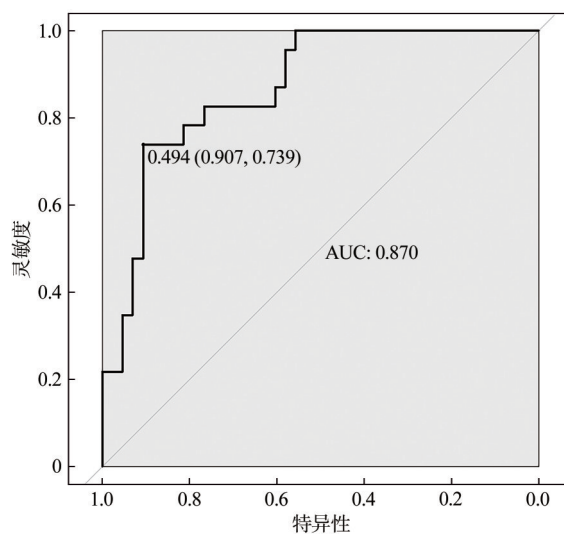


图11 验证集ROC曲线

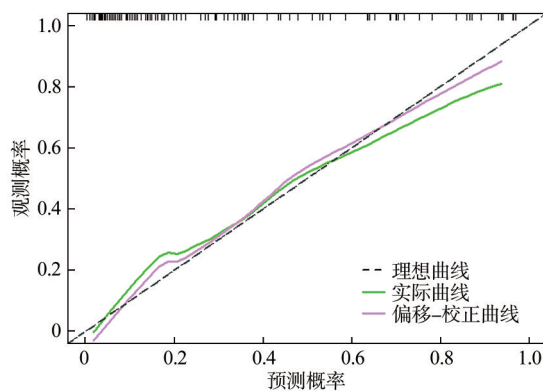


图12 训练集校准曲线

图14和图15所示,其中All代表所有患者接受了临床干预(如用药、手术等)后的净收益曲线, None为所有患者未接受干预的净收益曲线。结果显示,无论是训练集还是测试集,除了0.03的极小范围外,患者的临床净收益最高。这表明,所提出的列线图模型在大多数情况下可以为临床医生提供更准确的决策支持,使他们能够根据患者的特定风险因素评估其认知障碍发生的可能性,并制定相应的干预措施。

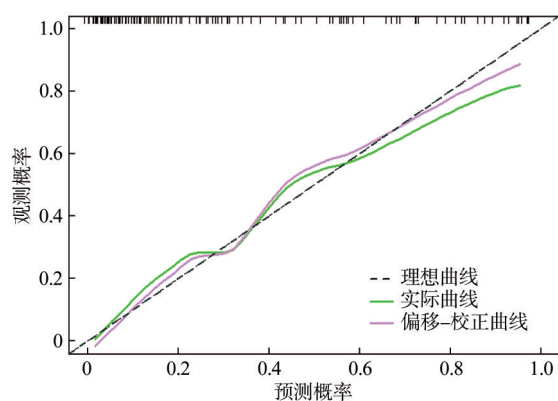


图13 验证集校准曲线

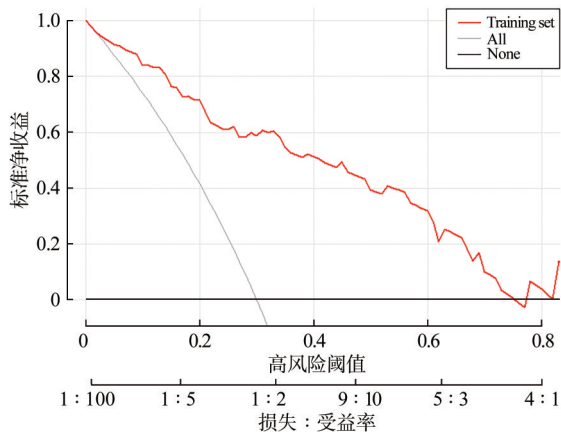


图14 训练集决策曲线

最后,为了验证模型的稳定性和可靠性,使用了 K 折交叉验证对模型进行了验证,从图16中可以看出在十折交叉验证中模型的准确率均在85%以上,最高达94%,说明模型具有较好的稳定性和可靠性。

3 结论

帕金森病认知障碍对患者的生活造成严重影响,为了辅助临床医生进行认知障碍的诊断,并在PD早期对患者进行干预和治疗,本文首先采用

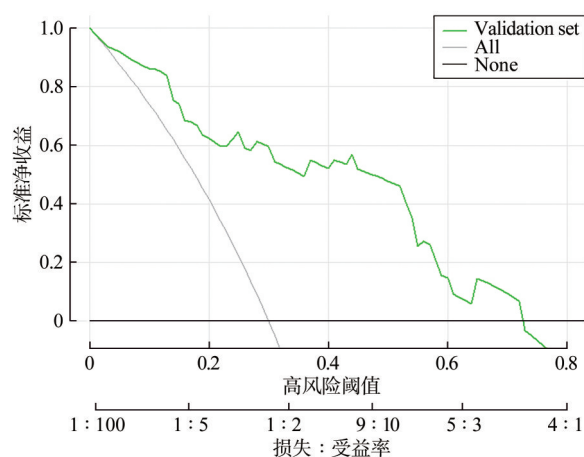


图15 验证集决策曲线

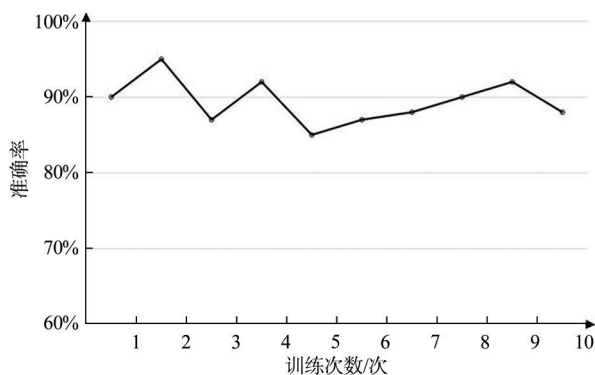


图16 十折交叉验证准确率

LASSO 算法^[28]设计了筛选策略,成功筛选出7个高风险的临床影响变量。这些变量在临床实践中易于获取且具有高准确度,为模型的建立提供了可靠的基础。其次,本文提出了一种基于多变量逻辑回归的新型列线图模型。相较于传统列线图,该模型具有更高的清晰度和可视化程度,能够对同一影响因子不同值的患者分布情况进行更直观的对比分析。这一改进有助于提高临床医生的诊断效率,并为每位患者量身定制个性化的疾病进展管理策略和咨询服务,有助于尽早发现认知障碍并进行干预。

实验验证和分析显示,该模型在训练集和验证集上的ROC曲线下面积分别为0.872和0.870,表现出优良的预测性能。预测概率与实际观测概率之间具有良好的一致性,验证了模型准确地预测患者发生认知障碍的风险。在临床获益方面,按照本文提出的列线图进行认知障碍患者的早期干预,可以为患者提供净收益。因此,该模型具有较高的准确性、一致性和临床实用性,为临床医生提供了有力的决策支持工具。通过该模型,识别高危帕金森病

患者发生认知障碍的人群,并对其进行治疗和干预,可以有效地控制症状,延缓病情进展。

综上所述,该模型在帕金森病认知障碍的早期预测和干预方面具有一定的临床诊断价值,并且在未来的研究中有望进一步应用于不同地区 and 人群,为更多患者提供更为准确的诊断和治疗策略。

参考文献:

- [1] VOSSIUS C, LARSEN J P, JANVIN C, et al. The economic impact of cognitive impairment in Parkinson's disease[J]. *Movement Disorders*, 2011, 26(8): 1541-1544.
- [2] LEROI I, MCDONALD K, PANTULA H, et al. Cognitive impairment in Parkinson disease: impact on quality of life, disability, and caregiver burden[J]. *Journal of Geriatric Psychiatry and Neurology*, 2012, 25(4): 208-214.
- [3] KRAMER A A, ZIMMERMAN J E. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited[J]. *Critical Care Medicine*, 2007, 35(9): 2052-2056.
- [4] 李伟凯, 高欣, 纪同俭. 用于轻度认知障碍诊断的群体相似约束功能脑网络建模方法[J]. *智能科学与技术学报*, 2019, 1(2): 145-153.
LI W K, GAO X, JI T J. Method of functional brain network modeling with group similarity constraint for mild cognitive impairment classification[J]. *Journal of Intelligent Science and Technology*, 2019, 1(2): 145-153.
- [5] SAUERBREI W, ROYSTON P, BINDER H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building[J]. *Stat Med*, 2007, 26(30): 5512-5528.
- [6] HU M T, SZEWCZYK-KRÓLIKOWSKI K, TOMLINSON P, et al. Predictors of cognitive impairment in an early stage Parkinson's disease cohort[J]. *Movement Disorders*, 2014, 29(3): 351-359.
- [7] 钱忠立, 缪建良, 陈旭. 应用蒙特利尔认知评估量表评价睡眠剥夺对认知功能影响的研究[J]. *中国疗养医学*, 2023, 32(7): 721-724.
QIAN Z L, MIAO J L, CHEN X. Study on the effect of sleep deprivation on cognitive function using the montreal cognitive assessment scale[J]. *Chinese Convalescent Medicine*, 2023, 32(7): 721-724.
- [8] HOBSON P, MEARA J. Mild cognitive impairment in Parkinson's disease and its progression onto dementia: a 16-year outcome evaluation of the Denbighshire cohort[J]. *International Journal of Geriatric Psychiatry*, 2015, 30(10): 1048-1055.
- [9] NASKAR A, STEZIN A, DHARMAPPA A, et al. Fibrinogen and complement factor H are promising CSF protein biomarkers for Parkinson's disease with cognitive impairment — a proteomics-ELISA-based study[J]. *ACS Chemical Neuroscience*, 2022, 13(7): 1030-1045.
- [10] CHEN B, SONG L, YANG J, et al. P-roteomics of serum exosomes identified fibulin-1 as a novel biomarker for mild cognitive impairment[J]. *Neural Regeneration Research*, 2023, 18(3): 587-593.
- [11] LIU W L, LIN H W, HE X J, et al. Neurogranin as a cognitive biomarker in cerebrospinal fluid and blood exosomes for Alzheimer's disease and mild cognitive impairment[J]. *Transl Psychiatry*, 2020, 10(125): 1-9.
- [12] COSGROVE J, ALTY J E, JAMIESON S. Cognitive impairment in Parkinson's disease[J]. *Postgraduate Medical Journal*, 2015, 91(1074): 212-220.
- [13] MAREK K, JENNINGS D, LASCH S, et al. The Parkinson progression marker initiative (PPMI)[J]. *Progress in Neurobiology*, 2011, 95(4): 629-635.
- [14] 杜淑慧, 何小海, 赵晓玲, 等. 基于时空和频域特征的 EEG 帕金森疾病识别[J]. *电子测量技术*, 2023, 46(3): 121-127.
DU S H, HE X H, ZHAO X L, et al. Parkinson's disease recognition by EEG based on spatiotemporal and frequency domain features[J]. *Electronic Measure-Ment Technique*, 2023, 46(3): 121-127.
- [15] TAJIRI Y, WADA-ISOE K, HAMADA M, et al. Comparison of MoCA and M-MSE for evaluating cognitive function in Japanese patients with Parkinson's disease[J]. *Movement Disorders*, 2014: S367-S367.
- [16] ZORZON M, CAPUS L, PELLEGRINO A, et al. Familial and environmental risk factors in Parkinson's disease: a case-control study in north-east Italy[J]. *Acta Neurologica Scandinavica*, 2002, 105(2): 77-82.
- [17] TIBSHIRANI R. Regression shrinkage and selection via the Lasso[J]. *Journal of the Royal Statistical Society*, 1996, 58(1): 267-288.
- [18] 段俊伟, 许林灿, 全渝娟. 基于图正则化的贝叶斯宽度学习系统[J]. *智能科学与技术学报*, 2022, 4(1): 109-117.
DUAN J W, XU L C, QUAN Y J. Bayesian width learning system based on graph regularization[J]. *Journal of Intelligent Science and Technology*, 2022, 4(1): 109-117.
- [19] 龚永丽, 刘明, 刘蓉, 等. 基于 Logistic 回归的城市资源环境承载力预测研究——以武汉市为例[J]. *电子测量技术*, 2018, 41(17): 42-46.
GONG Y L, LIU M, LIU R, et al. Prediction of urban resource and environment carrying capacity based on Logistic regression: a case study of Wuhan[J]. *Electronic Measurement Technology*, 2018, 41(17): 42-46.
- [20] 曹汉平, 张晓晶, 祝容杰, 等. 数字金融时代机器学习模型在实时反欺诈中的应用与实践[J]. *智能科学与技术学报*, 2019, 1(4): 342-351.
CAO H P, ZHANG X J, ZHU R J, et al. Application and practice of machine learning model in real-time anti-fraud in digital finance era[J]. *Journal of Intelligent Science and Technology*, 2019, 1(4): 342-351.
- [21] 陈乐陶, 杨士保, 陈橙, 等. 决策曲线分析在 R 语言中的实现[J]. *中国卫生统计*, 2018, 35(6): 955-957.
CHEN L T, YANG T B, CHEN C, et al. Implementation of decision curve analysis in R language[J]. *China Health Statistics*, 2018, 35(6): 955-957.
- [22] 梁子超, 李智炜, 赖铿, 等. 10 折交叉验证用于预测模型泛化能力评价及其 R 软件实现[J]. *中国医院统计*, 2020, 18(4): 289-292.
LIANG Z C, LI Z W, LAI K, et al. Application of 10-fold cross-validation in the evaluation of generalization ability of prediction models and the realization in R[J]. *Chinese Hospital Statistics*, 2020, 18(4): 289-292.
- [23] LITVAN I, GOLDMAN J G, TRÖSTER A I, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: movement disorder society task force guidelines[J]. *Movement disorders*, 2012, 27(3): 349-356.
- [24] WANG G, LAURI F, EL HASSANI A H. Feature selection by mRMR method for heart disease diagnosis[J]. *IEEE Access*, 2022, 10: 100786-100796.
- [25] GUAN H, YAN R, TANG H, et al. Intelligent fault diagnosis of hydraulic multi-way valve using the improved SECNN-GRU method with mRMR feature selection[J]. *Sensors*, 2023, 23(23): 9371.
- [26] LUSTED L B. Signal detectability and medical decision-making: signal detectability studies help radiologists evaluate equipment systems and performance of assistants[J]. *Science*, 1971, 171(3977): 1217-1219.
- [27] PENCINA M J, D'AGOSTINO R B. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation[J]. *Statistics in Medicine*, 2004, 23(13): 2109-2123.

- [28] 郭陈凤, 伍冬睿. 基于典型相关分析的多视图学习方法综述[J]. 智能科学与技术学报, 2022, 4(1): 14-26.

GUO C F, WU D R. A survey on canonical correlation analysis based multi-view learning[J]. Journal of Intelligence Science and Technology, 2022, 4(1): 14-26.

[作者简介]



巴梦茹 (1997-), 女, 青岛科技大学自动化与电子工程学院硕士生, 主要研究方向为机器学习、智能优化、智慧医疗。



尹晓红 (1985-), 女, 博士, 青岛科技大学自动化与电子工程学院副教授, 主要研究方向为复杂系统建模与优化、预测控制理论及其应用、绿色建筑节能优化。



李少远 (1965-), 男, 博士, 青岛科技大学自动化与电子工程学院教授, 主要研究方向为网络化分布式系统的自适应预测控制、满意优化控制、生产全过程系统的优化控制、智能控制及工业应用。