

中老年人认知功能影响因素研究——基于 CHARLS 数据的实证分析



朱心雅, 封佳芸, 叶莎, 沈婕, 麦依依, 戎芬

上海中医药大学公共健康学院 (上海 201203)

【摘要】目的 基于中国健康与养老追踪调查 (China Health and Retirement Longitudinal Study, CHARLS) 数据库对中老年人认知功能的影响因素进行深入分析, 为预防中老年人认知功能损害提供一定参考依据。**方法** 选择 45 岁及以上的 9 431 名中老年人作为研究对象, 从认知功能电话问卷 (telephone interview for cognitive status, TICS)、词语回忆、画图三方面综合评估中老年人的认知状况, 使用 Lasso-Logistic 回归模型筛选认知功能的影响因素, 通过随机森林模型对影响因素进行排序, 绘制受试者工作特征 (receiver operating characteristic, ROC) 曲线, 以 ROC 曲线下面积 (area under the curve, AUC)、灵敏度、特异度评价模型的预测效能。**结果** 中老年人认知综合评分为 10.14 ± 4.44 , 低于中位数的有 4 816 人 (51.07%)。Lasso-Logistic 分析结果显示性别、年龄、居住区域、2011 年认知评分、健康自评、夜间睡眠时间、社交活动、运动习惯、养老保险、教育水平是中老年人认知功能的主要影响因素 ($P < 0.05$)。随机森林模型显示, 中老年人认知功能影响因素按重要性排序依次为教育水平、2011 年认知评分、居住区域、年龄、性别、健康自评、社交活动、运动习惯、养老保险、夜间睡眠时间。Logistic 回归模型和随机森林模型的 AUC 均为 0.82。**结论** Logistic 回归模型与随机森林模型对中老年人认知功能均有较高的预测价值, 可为临床早期识别认知功能下降风险, 及时制定有效干预措施, 预防或延缓认知功能损害的发生提供有力的工具和方法。

【关键词】 中老年人; 认知功能; 影响因素; Logistic 回归模型; 随机森林模型

【中图分类号】 C 913.6 **【文献标识码】** A

Research on influencing factors of cognitive function in the middle-aged and elderly people: an empirical analysis based on CHARLS

ZHU Xinya, FENG Jiayun, YE Sha, SHEN Jie, MAI Yiyi, RONG Fen

School of Public Health, Shanghai University of Chinese Medicine, Shanghai 201203, China

Corresponding author: RONG Fen, Email: rf97328@163.com

【Abstract】Objective Based on the China Health and Retirement Longitudinal Study (CHARLS) database, this study aimed to conduct an in-depth analysis of the factors affecting cognitive function of middle-aged and elderly people, so as to provide a reference for preventing cognitive function damage. **Methods** A total of 9,431 middle-aged and elderly people aged 45 and above were selected as the research subjects. Cognitive function was comprehensively assessed

DOI: 10.12173/j.issn.1004-4337.202511014

基金项目: 上海市教育科学研究项目 (C2024167); 上海高校市级重点课程建设 (306); 上海中医药大学第二十三期课程建设重点项目 (KECJ2024012); 上海中医药大学教学学术共同体一般项目 (2025SHUTCMTD12)

通信作者: 戎芬, 博士, 副教授, Email: rf97328@163.com

<https://slyyx.whuzhmedj.com/>

using the telephone interview for cognitive status, drawing task, and word recall. The influencing factors of cognitive function were screened by Lasso-Logistic regression model, and the influencing factors were sorted by random forest model. Receiver operating characteristic (ROC) curves were plotted, and the predictive performance of the models was evaluated using the area under the curve (AUC), sensitivity, and specificity. **Results** The mean comprehensive cognitive score of the middle-aged and elderly people, with 4,816 people (51.07%) scoring below the median. Lasso-Logistic analysis showed that gender, age, residential area, baseline cognitive test score in 2011, self-assessment of health, night sleep time, social activities, exercise habits, pension insurance and education level were the main influencing factors of cognitive function in middle-aged and elderly people ($P < 0.05$). The random forest model ranked the importance of these factors in descending order as follows: education level, baseline cognitive test score in 2011, residential area, age, gender, self-assessment of health, social activities, exercise habits, pension insurance and night sleep time. The AUC values for the Logistic regression model and the random forest model were both 0.82. **Conclusions** Logistic regression model and random forest model both have high predictive value for middle-aged and elderly people's cognitive function, which could provide a powerful tool and method for early identification of the risk of cognitive function decline in clinical practice and timely formulation of effective intervention measures to prevent or delay the occurrence of cognitive function impairment.

【Keywords】 Middle-aged and elderly people; Cognitive function; Influencing factors; Logistic regression model; Random forest model

随着我国人口老龄化程度的加深,由中老年人群多发病带来的养护问题已成为社会关注的焦点。认知功能是人脑加工、储存和提取信息的能力。作为衰老过程中的重要特征,认知功能衰退与老年人口失能状况密切相关^[1-2]。认知功能障碍者常伴随记忆力、理解力、执行力、空间定位定向等功能下降;严重者可由轻度认知障碍逐步发展为痴呆,不仅危害患者的身心健康,还增加了家庭和社会的照护与经济负担。研究表明,中老年人认知功能障碍与社会人口学特征、疾病史、生活方式等因素存在一定关联,并直接影响中老年人的生活质量与生存结局^[3-4]。因此,明确认知功能影响因素,对认知功能损害进行早预防、早识别、早治疗是提升中老年人群晚年生活质量的关键。既往我国相关研究多探讨某一特定疾病或特定因素与认知功能的关联,或局限于某一地区的横断面分析。中国健康与养老追踪调查(China Health and Retirement Longitudinal Study, CHARLS)作为一项全国性纵向队列研究,具有信息容量大、覆盖范围广的优点。Lasso回归是常用的机器学习算法之一,与传统的最小二乘法相比,在筛选变量、处理复杂模型、提高模型预测能力方面具有显著优势^[5]。本研究采用Lasso回归筛选影响中老年人认知功能的关键因素,并在

此基础上构建Logistic回归模型与随机森林预测模型,以期降低中老年人群认知功能损害风险提供一定参考。

1 资料与方法

1.1 资料来源

本研究数据来源于CHARLS数据库。CHARLS是由北京大学国家发展研究院主持的一项大型跨学科调查项目,采用入户调查方式收集具有全国代表性的45岁及以上人群及其配偶的高质量纵向调查数据,研究基线于2011年开展,每2年进行一次随访评估。在2011年CHARLS基线调查的17 708名研究对象中,完成认知测试者共14 525人。其中,年龄小于45岁404人、脑损伤/精神障碍340人、患记忆相关疾病147人,共891人因不符合纳入标准被剔除。在剩余的13 634人中,缺少2020年相关协变量者($n=4 203$)被剔除,最终纳入9 431人。

1.2 认知功能评估

本研究通过认知功能电话问卷(telephone interview for cognitive status, TICS)、词语回忆、画图三方面对研究对象的认知功能情况进行评估^[6]。TICS测试用于评估研究对象的计算能力、注意力和定向力,要求研究对象完成从100中反

复减去 7 的计算任务, 并正确回答出日期、季节和星期, 总分为 0~10 分。对 10 个中文名词的即时和延迟 (5 分钟) 回忆测试用于评估情节记忆, 每个正确词语的评分为 0.5 分, 总分为 0~10 分^[7]。画图用于评估研究对象视觉空间能力, 画出相应图形得 1 分。三项内容总分共 21 分, 得分越高者认知功能越好。

1.3 研究变量

研究变量分为 5 部分: ①人口学特征: 性别、年龄、居住区域; ②健康状况: 2011 年认知评分、健康自评、癌症、中风、高血压、糖尿病、帕金森; ③生活方式: 夜间睡眠时间、运动习惯、社交活动、吸烟、饮酒; ④社会地位: 养老保险、就业情况、教育水平; ⑤疫情影响: 焦虑或紧张感、进食量变化、疫情隔离。

1.4 统计学方法

采用 Stata/SE 16.0 软件进行数据管理, 使用 SPSS 26.0 和 R 4.4.1 软件进行统计分析。分类变量采用频数及百分比 ($n, \%$) 进行统计描述。单因素分析采用卡方检验, 用 Lasso 回归模型筛选出相关影响因素, 使用 Logistic 回归进行多因素分析。构建随机森林模型, 将研究对象按 7:3 的比例随机分为训练集和测试集, 训练集用于构建模型, 测试集用于评价模型效能。选

择最优的随机森林模型特征个数, 将 Logistic 回归中有统计学意义 ($P < 0.05$) 的变量利用随机森林算法进行重要性排序^[8]。不同变量的重要性程度用平均基尼减少量反映, 基尼指数减少的平均值越大说明变量越重要。绘制受试者工作特征 (receiver operating characteristic, ROC) 曲线, 通过曲线下面积 (area under the curve, AUC)、灵敏度、特异度评价预测模型的性能。检验水准为 $\alpha=0.05$, 采用双侧检验, $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 中老年人认知功能情况

本研究纳入的 9 431 名研究对象中, 2020 年认知测试平均分为 10.14 ± 4.44 , 中位数为 10.50 (7.00, 13.50), 其中, 高于中位数 (较高认知) 有 4 615 人 (48.93%), 低于中位数 (较低认知) 有 4 816 人 (51.07%)。单因素分析结果显示性别、年龄、居住区域、2011 年认知评分、健康自评、中风、高血压、帕金森、夜间睡眠时间、运动习惯、社交活动、吸烟、饮酒、养老保险、教育水平、焦虑或紧张感、进食量变化以及疫情隔离是中老年人认知功能的影响因素, 差异有统计学意义 ($P < 0.05$), 详见表 1。

表 1 中老年人认知功能的影响因素 ($n, \%$)

Table 1. Influencing factors for cognitive function in middle-aged and elderly adults ($n, \%$)

特征	因素	较低认知 ($n=4\ 816$)	较高认知 ($n=4\ 615$)	合计 ($n=9\ 431$)	χ^2 值	P 值
人口学特征	性别				291.85	< 0.001
	男	1 816 (41.6)	2 550 (58.4)	4 366		
	女	3 000 (59.2)	2 065 (40.8)	5 065		
	年龄 (岁)				335.26	< 0.001
	45~59	1 168 (39.5)	1 792 (60.5)	2 960		
	60~74	2 900 (53.9)	2 481 (46.1)	5 381		
	75~84	660 (66.3)	335 (33.7)	995		
	≥ 85	88 (92.6)	7 (7.4)	95		
居住区域	城市	649 (32.5)	1 347 (67.5)	1 996	398.50	< 0.001
	城镇	482 (45.9)	567 (54.1)	1 049		
	农村	3 685 (57.7)	2 701 (42.3)	6 386		
	健康状况	2011 年认知评分				1769.85
	高	3 508 (72.0)	1 363 (28.0)	4 871		
	低	1 308 (28.7)	3 252 (71.3)	4 560		
健康自评	好	1 001 (47.3)	1 117 (52.7)	2 118	122.77	< 0.001
	一般	2 280 (47.8)	2 494 (52.2)	4 774		
	差	1 535 (60.5)	1 004 (39.5)	2 539		

续表1

特征	因素	较低认知 (n=4 816)	较高认知 (n=4 615)	合计 (n=9 431)	χ^2 值	P值
生活习惯	癌症				1.01	0.337
	是	104 (47.7)	114 (52.3)	218		
	否	4 712 (51.1)	4 501 (48.9)	9 213		
	中风				20.68	< 0.001
	是	423 (59.2)	291 (40.8)	714		
	否	4 393 (50.4)	4 324 (49.6)	8 717		
	高血压				13.64	< 0.001
	是	2 219 (53.2)	1 952 (46.8)	4 171		
	否	2 597 (49.4)	2 663 (50.6)	5 260		
	糖尿病				0.30	0.595
	是	788 (51.7)	736 (48.3)	1 524		
	否	4 028 (50.9)	3 879 (49.1)	7 907		
	帕金森				7.64	0.007
	是	70 (64.2)	39 (35.8)	109		
	否	4 746 (50.9)	4 576 (49.1)	9 322		
	夜间睡眠时间				83.70	< 0.001
	< 6 h	2 075 (57.0)	1 565 (43.0)	3 640		
	≥ 6 h	2 741 (47.3)	3 050 (52.7)	5 791		
	运动习惯				109.93	< 0.001
	有	4 144 (49.2)	4 279 (50.8)	8 423		
无	672 (66.7)	336 (33.3)	1 008			
社交活动				98.73	< 0.001	
有	2 085 (45.8)	2 470 (54.2)	4 555			
无	2 731 (56.0)	2 145 (44.0)	4 876			
吸烟				114.25	< 0.001	
仍在吸烟	1 049 (45.3)	1 265 (54.7)	2 314			
戒烟	568 (42.4)	773 (57.6)	1 341			
从未吸烟	3 199 (55.4)	2 577 (44.6)	5 776			
饮酒				119.27	< 0.001	
每日	534 (46.1)	624 (53.9)	1 158			
每周	227 (36.4)	396 (63.6)	623			
每月	232 (39.9)	349 (60.1)	581			
从未饮酒	3 823 (54.1)	3 246 (45.9)	7 069			
社会地位	养老保险				36.48	< 0.001
有	4 044 (49.8)	4 074 (50.2)	8 118			
无	772 (58.8)	541 (41.2)	1 313			
就业情况				0.00	0.958	
是	3 029 (51.0)	2 905 (49.0)	5 934			
否	1 787 (51.1)	1 710 (48.9)	3 497			
教育水平				2 193.53	< 0.001	
文盲及未读完小学	3 221 (76.7)	981 (23.3)	4 202			
小学	857 (41.6)	1 202 (58.4)	2 059			
初中	554 (27.5)	1 459 (72.4)	2 013			
高中及以上	184 (15.9)	973 (84.1)	1 157			
疫情影响	焦虑或紧张感				67.98	< 0.001
没有或轻度	3 417 (49.9)	3 424 (50.1)	6 841			
中度	659 (47.2)	736 (52.8)	1 395			
重度	740 (61.9)	455 (29.1)	1 195			
进食量变化				8.45	0.015	
增加	93 (44.3)	117 (55.7)	210			
不变	4 146 (50.8)	4 012 (49.2)	8 158			
减少	577 (54.3)	486 (45.7)	1 063			
隔离				4.59	0.032	
是	99 (59.3)	68 (40.7)	167			
否	4 717 (50.9)	4 547 (49.1)	9 264			

2.2 认知功能影响因素筛选

2.2.1 Lasso 回归

以中老年人2020年认知功能高低为因变量(较高认知=0, 较低认知=1), 将单因素分析中有统计学意义的18个影响因素作为自变量纳入Lasso回归中进行变量筛选, 变量编码及赋值见表2。Lasso回归的结果见图1, 其中图1-A表示每个变量所对应的回归系数值, 图1-B左侧虚线代表 λ_{min} , 为均方误差最小时的 λ 值, 右侧虚线代表 λ_{1se} , 位于 λ_{min} 右侧的1个标准误

差范围内, 在 $[\lambda_{min}, \lambda_{1se}]$ 区间内模型的偏差变动幅度最小。选取 λ_{1se} 作为最佳 λ 值, 保证模型误差接近最优, 并简化模型, 提升模型对未知数据的泛化能力与稳定性。当 λ_{1se} 值为0.01时, 对应的影响因素数目为10个, 分别为: 性别($\beta=-0.11$)、年龄($\beta=-0.24$)、居住区域($\beta=-0.10$)、健康自评($\beta=-0.05$)、夜间睡眠时间($\beta=0.03$)、运动习惯($\beta=-0.19$)、社交活动($\beta=-0.04$)、养老保险($\beta=-0.09$)、教育水平($\beta=0.68$)、2011年认知评分($\beta=1.20$)。

表2 LASSO回归的变量编码与赋值

Table 2. Variable coding and assignment for LASSO regression

特征	自变量	编码	变量赋值
个体特征	性别	X1	男=0, 女=1
	年龄(岁)	X2	45~59=0, 60~74=1, 75~84=2, ≥85=3
	居住区域	X3	城市=0, 城镇=1, 农村=2
健康状况	健康自评	X4	好=0, 一般=1, 差=2
	中风	X5	是=0, 否=1
	高血压	X6	是=0, 否=1
	帕金森	X7	是=0, 否=1
	2011年认知评分	X8	较低认知=0, 较高认知=1
生活方式	夜间睡眠时间	X9	<6h=0, ≥6h=1
	运动习惯	X10	有=0, 无=1
	社交活动	X11	有=0, 无=1
	吸烟	X12	仍在吸烟=0, 戒烟=1, 从未吸烟=2
	饮酒	X13	每日=0, 每周=1, 每月=2, 从未饮酒=3
社会地位	养老保险	X14	有=0, 无=1
	教育水平	X15	文盲及未读完小学=0, 小学=1, 初中=2, 高中及以上=3
疫情影响	焦虑或紧张感	X16	没有或轻度=0, 中度=1, 重度=2
	进食量变化	X17	增加=0, 不变=1, 减少=2
	隔离	X18	是=0, 否=1

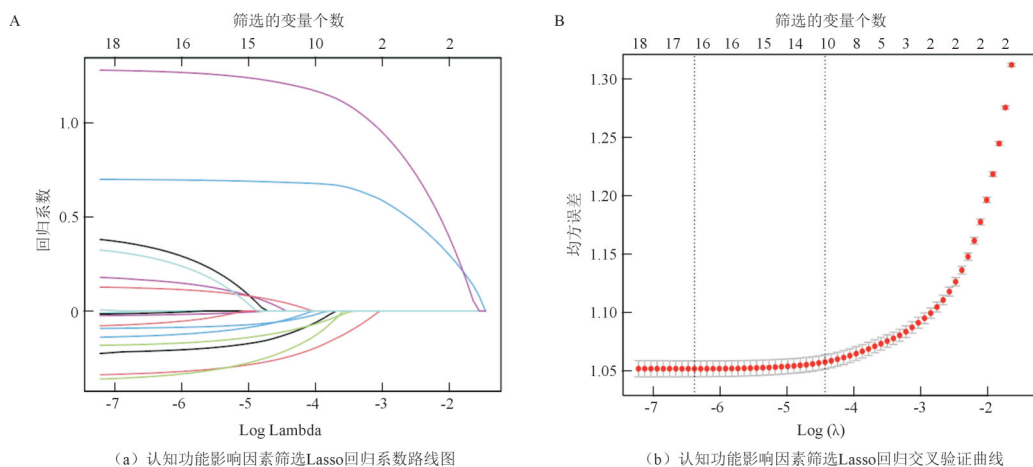


图1 Lasso回归变量筛选分析图

Figure1. Variable selection analysis plot of Lasso regression

注: A. 认知功能影响因素筛选Lasso回归系数路线图; B. 认知功能影响因素筛选Lasso回归交叉验证曲线。

2.2.2 多因素 Logistic 回归分析

将 Lasso 回归分析筛选出的 10 个特征作为自变量纳入 Logistic 回归模型中, 结果显示: 女性、高龄 (≥ 60 岁)、居住在城镇或农村、健康自评

差、睡眠时间短 (< 6 h)、无运动习惯、无社交活动、无养老保险、教育水平低、2011 年认知评分较低为中老年认知功能的危险因素 ($P < 0.05$), 见表 3。

表 3 中老年人认知测试得分影响因素的 Logistic 回归分析

Table 3. Logistic regression analysis of factors influencing cognitive test scores in middle-aged and elderly adults

因素	比较组	参照组	β	S_b	Wald χ^2	P 值	OR 值 (95%CI)
性别	女	男	0.20	0.05	13.56	< 0.001	1.22 (1.10, 1.35)
年龄	60 ~ 74	45 ~ 59	0.20	0.06	12.14	< 0.001	1.22 (1.09, 1.37)
	75 ~ 84		0.73	0.09	61.34	< 0.001	2.08 (1.73, 2.50)
	≥ 85		2.42	0.44	30.83	< 0.001	11.21 (4.78, 26.32)
居住区域	城镇	城市	0.34	0.09	12.91	< 0.001	1.40 (1.17, 1.68)
	农村		0.41	0.07	37.43	< 0.001	1.50 (1.32, 1.71)
健康自评	一般	好	-0.05	0.06	0.58	0.446	0.95 (0.84, 1.08)
	差		0.25	0.07	11.41	0.001	1.28 (1.11, 1.48)
夜间睡眠时间	< 6 h	≥ 6 h	-0.13	0.05	6.27	0.012	1.14 (1.03, 1.26)
运动习惯	无	有	0.34	0.09	16.27	< 0.001	1.41 (1.19, 1.67)
社交活动	无	有	0.15	0.05	8.98	0.003	1.16 (1.05, 1.28)
养老保险	无	有	0.30	0.07	16.53	< 0.001	1.34 (1.17, 1.55)
教育水平	小学	文盲及未读完小学	-1.18	0.06	348.14	< 0.001	0.31 (0.27, 0.35)
	初中		-1.53	0.07	481.00	< 0.001	0.22 (0.19, 0.25)
	高中及以上		-1.98	0.10	403.62	< 0.001	0.14 (0.11, 0.17)
2011 年认知评分	较低认知	较高认知	1.27	0.05	613.56	< 0.001	3.56 (3.22, 3.94)

2.2.3 随机森林模型

将 Logistic 回归中有统计学意义 ($P < 0.05$) 的 10 个影响因素作为自变量, 以认知功能高低为因变量, 构建中老年人认知功能影响因素的随机森林模型, 使用 6 601 个训练集用于建立随机森林模型。设置随机种子数为 123, 固定随机树数量 (n_{tree}), 确定最佳特征选择量 (m_{try}), 当 $m_{try}=2$ 时, 袋外误差率最小, 调整决策树数量, 训练多个模型并记录 OOB 误差率, 得到 $m_{try}=2$, $n_{tree}=400$ 时为最优模型, 随机森林预测误差与随机树的动态变化关系见图 2。根据各影响因素在随机森林模型中的基尼指数减少平均值 (MeanDecreaseGini), 对 10 个变量进行重要性评价, 结果显示, 变量按重要性排序依次为: 教育水平、2011 年认知评分、居住区域、年龄、性别、健康自评、社交活动、运动习惯、养老保险、夜间睡眠时间, 详见图 3。

2.3 预测效能评价

使用 ROC 曲线对 Logistic 回归模型和随机森林模型的 AUC、敏感度和特异度进行评价, 根据约登指数确定敏感度和特异度, 见图 4。Logistic 回归模型的 AUC 为 0.82 (95%CI: 0.81 ~ 0.83), 随机森林模型的 AUC 为 0.82 (95%CI: 0.81 ~

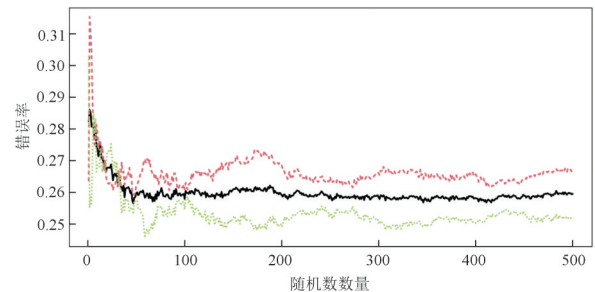


图 2 随机森林错误率评估图

Figure 2. Random forest error rate evaluation plot

注: 绿色, 表示较高认知功能; 黑色, 表示袋外数据; 红色, 表示较低认知功能。

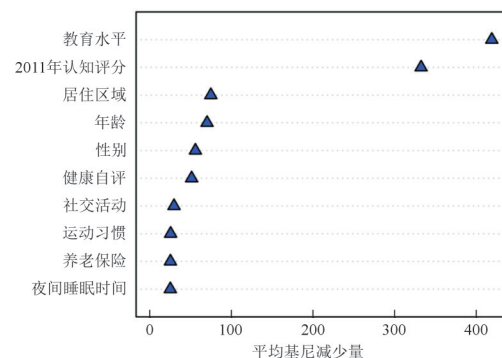


图 3 随机森林模型变量重要性排序

Figure 3. Importance ranking of variables in random forest model

0.84)。Logistic 回归模型和随机森林模型对中老年人认知功能均有较好的预测效能，见表 4。

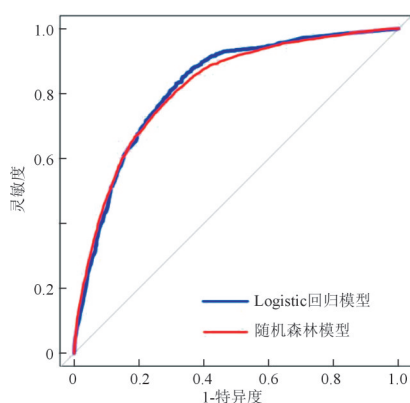


图 4 Logistic 回归模型和随机森林模型预测受试者工作特征曲线

Figure 4. Receiver operating characteristic curves of Logistic regression model and random forest model

表 4 Logistic 回归模型和随机森林模型预测受试者工作特征曲线分析

Table 4. Receiver operating characteristic curve analysis of Logistic regression model and random forest model

模型	AUC (95%CI)	标准误	P 值	灵敏度 (%)	特异度 (%)
Logistic 回归模型	0.82 (0.81, 0.83)	0.0043	< 0.001	80.4	69.0
随机森林 模型	0.82 (0.81, 0.84)	0.0002	< 0.001	84.8	66.8

注：AUC，area under the curve，曲线下面积。

3 讨论

人口老龄化已成为全社会普遍关注的重要问题^[9]。中老年人群由于生理、心理、社会等多方面因素的综合作用，认知功能逐渐减退，随着疾病进展，可从轻度认知障碍（mild cognitive impairment, MCI）发展为痴呆，不仅影响其日常生活，还会给家庭及社会带来沉重的疾病负担^[10]。因此，深入分析中老年人认知功能的影响因素，对中老年人认知功能障碍进行早期干预具有积极作用。

本研究利用 Lasso 回归筛选出的 10 个变量建立 Logistic 回归模型。结果显示，性别、年龄、居住区域、健康自评、2011 年认知评分，夜间睡眠时间、运动习惯、社交活动、养老保险及教育水平是中老年人认知功能的影响因素（ $P < 0.05$ ）。Xiong 等^[11]的研究指出，中老年女性有较低的认

知功能可能与绝经后卵泡刺激素增高有关。高龄是认知障碍的不可控危险因素，年龄越大，罹患该疾病的风险越高，且 85 岁以后出现认知损害的概率更大^[12]。在本研究中，2011 年基线认知功能评分对 2020 年的认知功能评分有一定影响，且认知测试评分随年龄增长呈现降低趋势，提示认知功能会随着年龄的增长而衰退。居住在城市的中老年人享有发达的基础设施、便利的交通、完善的医疗设施建设，可能拥有较好的认知功能，这与 Wang 等^[13]的研究结果一致。生活方式和社会地位方面，夜间睡眠时间较短、无社交活动、无运动习惯、教育水平较低以及无养老保险与较低认知功能相关。余成等^[14]的研究表明，睡眠时长的减少会导致认知功能下降。衰老过程中，老年人的睡眠结构发生改变，总睡眠时间缩短以及入睡困难程度增加，可导致认知功能障碍相关疾病加速进展^[15]。中老年人在参与社会活动时获得的认知刺激能够加强各种神经生物学通路，其获得的情感、社会以及资源支持能够缓冲负面情绪对健康带来的不良影响，从而降低中老年人认知障碍的发生风险，这与朱正杰等^[16]的研究结果一致。刘东祺等^[17]认为运动不仅可以锻炼中老年人的身体协调性、注意力、记忆力以及行动力，还能够减少大脑萎缩体积、增加脑血流量、促进神经网络的建立和改善脑组织代谢，具有运动习惯的中老年人会拥有较高的认知功能。缴纳养老保险者的认知功能较好，意味着在退休后享有基本养老金以及就医、丧葬补助等福利待遇，对其健康产生正向促进作用，这与既往研究结果一致^[18-19]。教育水平对认知功能呈现出显著保护效应，学历水平越高，发生认知障碍的风险越低，即教育训练在提高脑神经突触密度和效率的同时还能改善中老年人的神经网络，从而提供抵抗认知受损的储备和缓冲^[20]。

随机森林模型是近年来常用的集成学习方法，能通过构建多个决策树并结合其预测结果来提高模型的准确性，具有准确性高、抗过拟合能力强、可处理高维数据并进行特征重要性评估等优势^[21]。本研究对 Logistic 回归中有统计学意义（ $P < 0.05$ ）的 10 个变量建立了中老年人认知功能影响因素的随机森林预测模型，对各影响因素按其重要性进行排序。结果表明，影响中老年人认知功能的因素按重要性排序依次为教育水平、

2011 年认知评分、居住区域、年龄、性别、健康自评、社交活动、运动习惯、养老保险和夜间睡眠时间。Logistic 回归模型和随机森林模型预测中老年认知功能影响因素的 AUC 均为 0.82 (> 0.8), 呈现了较好的预测准确性和稳定性, 具有较高的预测效能。Logistic 回归作为线性分类模型, 可输出单个因素的量化效应, 以其简洁、直观、结果可解释性强的特点, 为临床实践提供量化支持^[22]。随机森林模型是非参数化集成模型, 通过捕捉变量非线性交互, 预测稳定性较强。当随机森林模型和 Logistic 回归模型在预测性能无显著差异时, Logistic 回归的简单性和透明度通常是精准干预更实用的选择。

综上所述, 教育水平、2011 年认知评分、居住区域、年龄、性别、健康自评、社交活动、运动习惯、养老保险和夜间睡眠时间与中老年人认知功能相关, 且重要性依次降低。采用 Lasso 回归筛选变量后建立的 Logistic 回归模型和随机森林模型有较好的临床使用价值, 提示临床实践中可以通过上述模型预测中老年人罹患认知功能损害的风险, 有助于积极预防认知功能损害的发生, 尽早制定有效的认知功能康复策略。本研究尚存在以下局限性: 认知功能主要依赖问卷和行为测试评估, 缺乏脑影像、生物标志物等客观数据, 可能遗漏潜在的生物学关联; 尽管已控制多种协变量, 但仍有一些潜在混杂因素, 如家庭收入、痴呆家族史、抑郁病史等, 尚未纳入, 有待在今后的研究中进一步完善。

参考文献

- 刘凌晨, 程宏宇, 彭希哲. 中国老年人社会参与模式对认知衰退的影响[J]. 中国人口科学, 2022, (04): 103-114. [Liu LC, Cheng HY, Peng XZ. The influence of social participation patterns on cognitive decline in older adults in China[J]. Chinese Journal of Population Science, 2022, (4): 103-114, 128. <https://d.wanfangdata.com.cn/periodical/ChdQZXJpb2RpY2FsQ0hJU29scjlrRdWljaxIPemdyA2t4MjAyMjA0MDA5GghieDE5cDcyQ%3D%3D>
- 黄晓彤, 王琼, 刘晨, 等. 童年社会经济地位对中老年认知功能影响的中介效应路径分析[J]. 中国全科医学, 2024, 27(3): 322-327, 334. [Huang XT, Wang Q, Liu C, et al. A path analysis of the effect of childhood socioeconomic status on cognitive function in middle-aged and older adults [J]. Chinese General Practice, 2024, 27(3): 322-327, 334.] DOI: [10.12114/j.issn.1007-9572.2023.0359](https://doi.org/10.12114/j.issn.1007-9572.2023.0359).
- 班春霞, 石文磊, 高红艳, 等. 社区老年人认知功能障碍现状及影响因素调查[J]. 国际精神病学杂志, 2025, 52(3): 852-856. [Ban CX, Shi WL, Gao HY, et al. Investigation on the status quo and related influencing factors of cognitive dysfunction in elderly people in community[J]. Journal of International Psychiatry, 2025, 52(3): 852-856.] DOI: [10.13479/j.cnki.jip.2025.03.051](https://doi.org/10.13479/j.cnki.jip.2025.03.051).
- 高娇娇, 王小楠, 吴抗, 等. 上海浦东新区中老年 2 型糖尿病患者轻度认知功能障碍的现状及其影响因素分析[J]. 公共卫生与预防医学, 2025, 36(3): 32-36. [Gao JJ, Wang XN, Wu K, et al. Current status and influencing factors of mild cognitive impairment in middle-aged and elderly patients with type 2 diabetes in Pudong New Area of Shanghai[J]. Journal of Public Health and Preventive Medicine, 2025, 36(3): 32-36.] DOI: [10.3969/j.issn.1006-2483.2025.03.007](https://doi.org/10.3969/j.issn.1006-2483.2025.03.007).
- 张沥今, 魏夏琰, 陆嘉琦, 等. Lasso 回归: 从解释到预测[J]. 心理科学进展, 2020, 28(10): 1777-1791. [Zhang LJ, Wei XY, Lu JQ, et al. Lasso regression: from explanation to prediction[J]. Advances in Psychological Science, 2020, 28(10): 1777-1791. <https://d.wanfangdata.com.cn/periodical/ChdQZXJpb2RpY2FsQ0hJU29scjlrRdWljaxIOeGx4ZHQyMDIwMTAwMTUaCDRhbXJiaDg4>.
- 宋雅云. 中老年人认知功能与握力的相关性研究——基于 CHARLS 数据[J]. 中华养生保健, 2022, 40(19): 1-5. [Song YY. Association between cognitive functions and grip strength among the older adults: based on CHARLS data[J]. ZHONGHUA YANGSHENG BAOJIAN, 2022, 40(19): 1-5. <https://d.wanfangdata.com.cn/periodical/QKBJBD2022022100100007910>
- Hua J, Dong J, Chen GC, et al. Trends in cognitive function before and after stroke in China[J]. BMC Med, 2023, 21(1): 204. DOI: [10.1186/s12916-023-02908-5](https://doi.org/10.1186/s12916-023-02908-5).
- 盛丹, 高宁, 唐绍琰, 等. 基于 LASSO-logistic 回归模型的糖尿病周围神经病变预测模型[J]. 中国慢性病预防与控制, 2024, 32(10): 733-740. [Sheng D, Gao N, Tang SY, et al. Prediction model of diabetic peripheral neuropathy based on LASSO-logistic regression model[J]. Chinese Journal of Prevention and Control of Chronic Diseases, 2024, 32(10): 733-740.] DOI: [10.16386/j.cjpcd.issn.1004-6194.2024.10.003](https://doi.org/10.16386/j.cjpcd.issn.1004-6194.2024.10.003).
- 王洪川, 陈怡莹, 王聪. 人口老龄化背景下体育消费的健康效应——基于全国消费大数据的实证证据[J]. 上海体育大学学报, 2024, 48(3): 75-84. [Wang HC, Chen YY, Wang C. Health effect of sports consumption in aging society: empirical evidence from nationwide consumption big data[J]. Journal of Shanghai University of Sport, 2024, 48(3): 75-84.] DOI: [10.16099/j.sus.2023.04.04.0002](https://doi.org/10.16099/j.sus.2023.04.04.0002).
- 冯薇. 老年人认知功能障碍与阿尔茨海默病的研究进展[J]. 公共卫生与预防医学, 2024, 35(6): 123-126. [Feng W. Research progress of elderly cognitive impairment and Alzheimer's disease[J]. Journal of Public Health and Preventive Medicine, 2024, 35(6): 123-126.] DOI: [10.3969/j.issn.1006-2483.2024.06.028](https://doi.org/10.3969/j.issn.1006-2483.2024.06.028).
- Xiong J, Kang SS, Wang Z, et al. FSH blockade improves cognition in mice with Alzheimer's disease[J]. Nature, 2022, 603(7901): 470-476. DOI: [10.1038/s41586-022-04463-0](https://doi.org/10.1038/s41586-022-04463-0).
- 景璟, 王丹阳. 社区老年人群轻度认知障碍发生情况及其相关危险因素调查分析[J]. 安徽医药, 2023, 27(3): 542-545. [Jing J, Wang DY. Investigation and analysis of the occurrence of mild

- cognitive impairment and related risk factors in the elderly in the community[J]. *Anhui Medical and Pharmaceutical Journal*, 2023, 27(3): 542–545. DOI: 10.3969/j.issn.1009-6469.2023.03.026..
- 13 Wang Y, Jiang Y, Wu W, et al. Education, neighborhood environment, and cognitive decline: findings from two prospective cohort studies of older adults in China[J]. *Alzheimer's Dement.* 2023, 19(2): 560–568. DOI: 10.1002/alz.12679.
- 14 余成, 孙景贤. 农村地区轻度认知障碍老年人睡眠时长与睡眠效率对认知功能的影响[J]. *实用老年医学*, 2022, 36(3): 295–298. [Yu C, Sun JX. Effects of sleep duration and sleep efficiency on cognitive function of the elderly with mild cognitive impairment in rural areas[J]. *Practical Geriatrics*, 2022, 36(3): 295–298.] DOI: 10.3969/j.issn1003-9198.2022.03.021.
- 15 曹璐, 王迎斌, 张伟, 等. 睡眠障碍对老年认知功能的影响机制及防治研究进展[J]. *世界睡眠医学杂志*, 2024, 11(4): 778–783. [Cao L, Wang YB, Zhang W, et al. Research progress on the mechanism of sleep disorders affecting cognitive function in elderly patients and its prevention and treatment strategies [J]. *World Journal of Sleep Medicine*, 2024, 11(4): 778–783.] DOI: 10.3969/j.issn.2095-7130.2024.04.019.
- 16 朱正杰, 施正丽, 张浩, 等. 我国老年人社会参与情况及其与认知障碍的关系[J]. *医学与社会*, 2023, 36(12): 53–58. [Zhu ZJ, Shi ZL, Zhang H, et al. Status of social participation and its relationship with cognitive impairment among the elderly in China [J]. *Medicine and society*, 2023, 36(12): 53–58.] DOI: 10.13723/j.xysh.2023.12.009.
- 17 刘东祺, 李荣梅, 张美琪, 等. 有氧运动干预老年轻度认知功能障碍的Meta分析[J]. *中国组织工程研究*, 2019, 23(35): 5727–5731. [Liu DQ, Li RM, Zhang MQ, et al. Meta-analysis of the effect of aerobic exercise on mild cognitive impairment in the elderly[J]. *Chinese Journal of Tissue Engineering Research*, 2019, 23(35): 5727–5731.] DOI: 10.3969/j.issn.2095-4344.1943.
- 18 朱雅丽, 李慧, 赵豪. 养老保险提高了老年人的健康水平吗?——基于 CLHLS 数据的实证分析[J]. *中国经贸导刊*, 2020, (05): 136–138. [Zhu YL, Li H, Zhao H. Does endowment insurance improve the health status of the elderly?—An empirical analysis based on CLHLS data[J]. *China Economic & Trade Herald*, 2020, (05): 136–138.] DOI: 10.3969/j.issn.1007-9777.2020.14.041.
- 19 毛晓雨, 黄雅莲, 蔡清艳, 等. 中国失能老年人认知功能现状及影响因素[J]. *中华老年多器官疾病杂志*, 2022, 21(2): 91–94. [Mao XY, Huang YL, Cai QY, et al. Cognitive function and its influencing factors in the disabled elderly in China[J]. *Chinese Journal of Multiple Organ Diseases in the Elderly*, 2022, 21(2): 91–94.] DOI: 10.11915/j.issn.1671-5403.2022.02.020.
- 20 袁菲, 刘溯, 梁小姐, 等. 社区老年人群高血压情况与痴呆症发病风险的关联研究[J]. *中国临床神经科学*, 2021, 29(5): 519–526. [[Yuan F, Liu S, Liang XN, et al. Relationship between hypertension and incidence risk of dementia in community elderly[J]. *Chinese Journal of Clinical Neuroscience*, 2021, 29(5): 519–526. <https://d.wanfangdata.com.cn/periodical/ChdQZXJpb2RpY2FsQ0hJU29scjIRdWljaxlRemdsY3Nqa3gyMDIxMDUwMDYaCDZjdW90YXBt>
- 21 张华, 高少芳, 赵欣, 等. 基于随机森林、Logistic 回归的角膜塑形镜近视控制效果及预测模型构建和评价[J]. *眼科新进展*, 2024, 44(8): 637–642. [Zhang H, Gao SF, Zhao X, et al. Myopia control effect of orthokeratology lens and construction and evaluation of prediction model based on random forest and Logistic regression[J]. *Recent Advances in Ophthalmology*, 2024, 44(8): 637–642.] DOI: 10.13389/j.cnki.rao.2024.0122.
- 22 胡小琴, 黄原义, 王勤, 等. 基于 LASSO-Logistic 回归的肺亚实性结节浸润性预测模型的建立与验证[J]. *放射学实践*, 2025, 40(11): 1410–1415. [Hu XQ, Huang YY, Wang Q, et al. Establishment and validation of the LASSO-logistic regression-based model for invasiveness prediction in pulmonary subsolid nodules[J]. *Radiologic Practice*, 2025, 40(11): 1410–1415.] DOI: 10.13609/j.cnki.1000-0313.2025.11.011.

收稿日期: 2025 年 11 月 06 日 修回日期: 2026 年 05 月 26 日
 本文编辑: 梁亮艺 黄 笛

引用本文: 朱心雅, 封佳芸, 叶莎, 等. 中老年人认知功能影响因素研究——基于 CHARLS 数据的实证分析[J]. *数理医药学杂志*, 2026, 39(5): 319–327. DOI: 10.12173/j.issn.1004-4337.202511014.
 Zhu XY, Feng JY, Ye S, et al. Research on influencing factors of cognitive function in the middle-aged and elderly people: an empirical analysis based on CHARLS[J]. *Journal of Mathematical Medicine*, 2026, 39(5): 319–327. DOI: 10.12173/j.issn.1004-4337.202511014.