

# 解释性机器学习模型对老年抑郁症患者的识别 ——基于美国国家健康和营养检测调查数据库

缪鹏程<sup>1</sup>, 陆贝尔<sup>1</sup>, 马溶基<sup>1</sup>, 钱永康<sup>1</sup>, 胡陈华<sup>1</sup>, 陈华玲<sup>1</sup>, 凡如<sup>2</sup>, 许碧云<sup>2</sup>, 陈炳为<sup>1</sup>

1. 东南大学公共卫生学院, 江苏 南京 210009; 2 南京大学医学院附属鼓楼医院统计中心

**摘要:**目的 基于 2005—2021 年美国国家健康和营养检测调查数据库, 使用可解释性机器学习方法识别 65 岁以上老年人中的抑郁症患者。方法 以 2005—2018 年及 2019—2020 年的数据分别作为训练集及测试集, 拟合 lasso logistic、随机森林、XGBoost 三种机器学习模型, 以测试集上的 AUC 最大选择较优的模型, 使用解释性机器学习模型 SHAP 进行解释。结果 XGBoost 模型 AUC 值最大, 为 0.933(0.912~0.954), 是否存在睡眠问题、是否存在健康问题和嗜酸性粒细胞计数为影响老年人抑郁症的前三重要的变量, 变量 SHAP 值的绝对值分别为 1.16、0.83、0.55; SHAP 力图呈现了每个个体的主要的影响因素, 根据 SHAP 值对每个个体进行解释。结论 机器学习在预测老年人抑郁症方面性能优于 logistic 回归模型, 解释性机器学习可以从全局和个体层面解释模型做出预测, 打开机器学习模型的黑箱, 在实际应用中可以作为机器学习模型的补充。

**关键词:** 老年人; 抑郁症; 解释性机器学习; XGBoost; SHAP

中图分类号: R749.4; B844.4 文献标志码: A 文章编号: 1003-8507(2024)05-781-07

DOI: 10.20043/j.cnki.MPM.202309307

## Identification of patients with senile depression by interpretable machine learning model—based on the US National Health and Nutrition Examination Survey

MIAO Peng-cheng\*, LU Bei-er, MA Rong-ji, QIAN Yong-kang, HU Chen-hua, CHEN Hua-ling, FAN Ru, XU Bi-yun, CHEN Bing-wei

\*School of Public Health, Southeast University, Nanjing, Jiangsu 210009, China

**Abstract: Objective** Based on the US National Health and Nutrition Survey from 2005 to 2021, an interpretable machine learning method was used to identify patients with depression in people over 65 years old. **Methods** The data of 2005–2018 and 2019–2020 were used as training set and test set, respectively, and three machine learning models of Lasso Logistic, random forest, and XG Boost were fitted. The best model of area under the curve (AUC) on the test set was selected and explained by interpretable machine learning model SHAP. **Results** The AUC value of XG Boost model was the highest, which was 0.933 (0.912–0.954). Sleep problems, health problems, and eosinophil count were the top three important variables affecting senile depression. The absolute values of SHAP were 1.16, 0.83, and 0.55, respectively, which showed the main influencing factors of each individual. **Conclusion** Machine learning is superior to logistic regression model in predicting depression in the elderly. Interpretable machine learning can explain the model from the global and individual levels to make predictions, open the black box of machine learning models, and can be used as a supplement to machine learning models in practical application.

**Keywords:** Elderly; Depression; Interpretable machine learning; XG Boost; SHAP

老年人抑郁症是一种常见的精神疾病, 会影响老年人健康相关的生活质量。2017 年的一项研究显示, 10%~15% 的老年人有临床意义的抑郁症状, 55 岁及以上人群中重度抑郁症发生率在 2% 左右, 且患病率

随着年龄的增长而上<sup>[1]</sup>。抑郁症是晚年体重减轻的主要原因<sup>[2]</sup>。抑郁症通常与心血管疾病等慢性疾病有关, 并且可能使这些疾病的过程复杂化<sup>[3]</sup>。抑郁症也与功能障碍有关, 并随着时间的推移影响残疾状态<sup>[4-5]</sup>。71% 的自杀尝试者报告以前至少有一次重度抑郁症发作<sup>[6]</sup>。在 65 岁以上老年自杀人群中, 约 83% 因抑郁症引起<sup>[6]</sup>。药物治疗和心理治疗的联合治疗对于老年抑郁症的治疗是有效的, 早期发现抑郁症并对老年抑郁

基金项目: 南京市医学重点科技发展项目(ZKX22019)

作者简介: 缪鹏程(1998—), 男, 硕士在读, 研究方向: 流行病与卫生统计专业

通信作者: 陈炳为, E-mail: drechenbw@126.com

患者的治疗非常重要<sup>[7]</sup>。因此,了解老年人抑郁的危险因素可能有助于识别高危人群,以延缓疾病进程并建立个性化干预措施<sup>[8]</sup>。

机器学习能够较好地解决变量间的非线性复杂关系,其预测性能及泛化性能往往高于传统的统计方法,被广泛应用于医学各个领域。在精神方面,Yuan 等人<sup>[9]</sup>使用多种机器学习算法识别抑郁症患者,在这项 31 715 例的研究中,集成学习算法的受试者工作特征曲线下面积(area under curve, AUC)为 0.903 6。在一篇抑郁症中的机器学习的综述中表明,随机森林(random forest, RF)、及极限梯度提升(extreme gradient boosting, XGBoost)是应用较多的两种方法<sup>[10]</sup>。尽管这些机器学习模型提供了良好的预测性能,由于机器学习模型属于黑盒模型,在实际临床中应用中还是受到一定的限制。Lundberg 和 Lee<sup>[11]</sup>提出了 SHAP(Shapley Additive exPlanations)用以解释机器学习模型做出的决策。

本研究基于 2005—2021 年美国国家健康与营养调查数据库(National Health And Nutrition Examination Survey, NHANES),采用了 XGBoost 算法、随机森林和 lasso logistic 回归三种机器学习方法,构建识别老年人抑郁症相关的模型,并使用可解释性机器学习 SHAP 评估影响老年人抑郁症和老年人个体的主要影响因素。本研究旨在为抑郁症高危人群的早期发现和早期治疗提供科学依据。

## 1 资料与方法

**1.1 研究人群** NHANES 旨在收集美国家庭人口健康和营养的信息,评估美国个人的健康和营养状况和了解公共卫生问题,每两年调查一次,所有数据在 <https://www.cdc.gov/nchs/NHANES/> 网站可以下载。在刘悦<sup>[12]</sup>、周雯惠<sup>[13]</sup>等文章中将  $\geq 65$  岁定义为老年人,因此本研究将年龄  $\geq 65$  的个体作为本次研究对象。为了评价预测模型的外部一致性,研究将 2005—2018 年的数据作为训练集,将 2019—2020 年的数据作为测试集评价模型的预测性能。数据集中删除了 PHQ-9 量表中有任意一项缺失的个体,同时缺失比例超过 30%的变量和缺失变量比例大于 50%的个体也被删去。

**1.2 疾病定义** 在 NHANES 数据库中,采用患者健康问卷 9(patient health questionnaire-9, PHQ-9)<sup>[14]</sup>量表筛选抑郁症患者,该量表既是抑郁症严重程度的衡量标准,也是抑郁症的诊断指标。量表共有 9 个条目(在 NHANES 数据库中变量条目为 DPQ010—DPQ090),每个条目 0~3 分,PHQ-9 的总分范围为 0~27 分,以得分  $\geq 10$  分作为划分临床相关抑郁症的

界值可以得到最高的特异度和灵敏度<sup>[14]</sup>。

**1.3 影响因素** 本研究的变量分别来自 NHANES 数据库中的人口学、膳食、测量、实验室指标和问卷或量表数据集。人口学数据中包括年龄、种族、学历、收入情况、收入与贫困比例和家庭人数等变量。膳食数据集包括总营养素摄入量和补充剂中的营养素摄入量,这些数据先通过面对面访谈获取第一次数据,再通过电话跟进获得第二次数据;膳食数据集包括饮食结构,膳食营养素包括金属元素、不饱和脂肪酸、维生素和水的摄入量。测量数据集包括身体测量指标如身高、体重、腰围、血压等。实验室测量指标包括血液和(或尿液)中的生化和金属指标,如脂蛋白、胆固醇、尼古丁、可替宁、免疫细胞、维生素、汞、铁、碘、铅、镉、硒和锰等。问卷或量表数据包括吸烟、饮酒、心血管健康、消费行为、当前健康状况、残疾、吸毒、身体活动、心理健康和睡眠等数据集。

由于实验室指标中的血生化指标、血液和尿液中金属元素和膳食营养素等指标进行对数变换,当这些变量的值为 0 时,给其加上 0.01 以使其能够进行对数变换。原始数据库中舒张压和收缩压分别测量了三次,取均值收缩压和舒张压变量纳入分析。

**1.4 基线描述** 在描述人口统计学特征时,使用(均数  $\pm$  标准差)的形式描述定量变量,使用频数(百分比)描述定性变量。

**1.5 模型构建** 在训练集和测试集中采用多重填补的方式填补自变量的缺失,使用递归特征消除(recursive feature elimination, RFE)筛选变量。因为该数据集属于不平衡数据,因此对于训练集使用少数类过采样法(synthetic minority over-sampling technique, SMOTE)<sup>[15]</sup>,以使得抑郁症患者与正常人人数量一致,从而提高模型训练效果。最后对数据进行标准化、归一化处理。

选择 lasso logistic 回归、随机森林和 XGBoost 算法构建模型,机器学习模型的超参数较多,因此采用贝叶斯优化的 TPE(tree parzen estimator)<sup>[16]</sup>算法选择超参数,经过十折交叉验证保证结果的稳定性。

在测试集上评估模型性能,主要的评价指标为 AUC、准确度、灵敏度和特异度。AUC 置信区间的计算和比较使用 De Long 检验<sup>[15]</sup>, $P$  值小于 0.05 被认为有统计学差异。最后使用 SHAP 解释结果。SHAP 值起源于博弈论,Shapley 在 1950 年旨在根据玩家对游戏最终结果的贡献,在玩家之间分配收益而提出。它已发展为机器学习的一种重要解释方法,通过计算每个特征的重要性值(SHAP 值)解释变量与个体的预测,以提高模型可解释性。本研究的数据处理与统计

学检验使用 R 4.2.1 执行,模型构建、评价与解释均使用 Python 3.0 执行。检验水准  $\alpha=0.05$ 。

## 2 结果

**2.1 人口统计** 在训练集中共有 6 808 名老年人参与者接受了抑郁症评估,441 名参与者被诊断患有抑

郁症,患病率及 95%CI 为 6.48%(5.90%~7.09%),在测试集中共有 1 790 名参与者接受了抑郁症评估,131 名参与者被诊断患有抑郁症,患病率及 95%CI 为 7.32%(6.45%~9.03%)。研究人群的人口统计学特征。见表 1。

表 1 训练集和测试集中人口统计学特征 $[(\bar{x} \pm s), n(\%)]$

Table 1 Basic characteristics in the train and test datasets  $[(\bar{x} \pm s), n(\%)]$

变量	合计(n=8 598)	测试集(n=1 790)	训练集(n=6 808)
年龄(岁)	73.26 ± 5.43	72.85 ± 5.29	73.37 ± 5.47
受教育程度			
初中以下	1 215(14.13)	162(9.05)	1 053(15.47)
初中	1 172(13.63)	197(11.01)	975(14.32)
高中	2 141(24.90)	456(25.47)	1 685(24.75)
大学	2 246(26.12)	542(30.28)	1 704(25.03)
大学毕业或以上	1 811(21.06)	430(24.02)	1 381(20.28)
性别			
男性	4 352(50.62)	924(51.62)	3 428(50.35)
女性	4 246(49.38)	866(48.38)	3 380(49.65)
种族			
墨裔美国人	850(9.89)	117(6.54)	733(10.77)
拉丁裔	670(7.79)	146(8.16)	524(7.70)
非拉丁裔白人	4 896(56.94)	941(52.57)	3 955(58.09)
非拉丁裔黑人	1 625(18.90)	417(23.30)	1 208(17.74)
其他种族和混血	557(6.48)	169(9.44)	388(5.70)

**2.2 模型评价** 经过 SMOTE 算法,训练集中共有 12 734 个观测,抑郁症和非抑郁症的个体各占 50%。训练集经过 RFE 后得到特征变量 62 个。在这些变量

的基础上建立 lasso logistic 回归、随机森林和 XGBoost 模型。表 2 和图 1 总结了三种机器学习分类器在识别老年抑郁症方面的表现。

表 2 三种机器学习模型在测试集上的表现

Table 2 Performance of three machine learning models in test dataset

模型	AUC(95%CI)	准确率(%)	灵敏度(%)	特异度(%)
Lasso logistic	0.772(0.731~0.813)	68.44	74.05	67.99
随机森林	0.910(0.886~0.935)	78.72	92.37	77.64
XGBoost	0.933(0.912~0.954)	84.92	85.50	84.87

在训练集上,XGBoost 模型的 AUC、准确率和灵敏度最高,分别达到 0.933、84.92%和 84.87%。随机森林模型的灵敏度最高,为 92.37%。Lasso logistic 模型的表现最差,AUC 为 0.772。经过 De Long 检验,XGBoost 模型的 AUC 高于 lasso logistic 模型( $P<0.001$ )和 RF 模型( $P<0.001$ ),RF 模型的 AUC 高于 lasso logistic 模型( $P<0.001$ )。因此,采用 XGBoost 模型进行解释。

**2.3 模型解释** 图 2A 为训练集上 SHAP 变量重要性图,前十重要的变量为是否存在睡眠问题、是否存

在健康问题、嗜酸性粒细胞数、一个月内过度疲倦的次数、近一个月内是否有过感冒、嗜碱性粒细胞数、血清可替宁浓度、单核细胞数、受教育程度、PFA 18:4(十八碳四烯酸)。图 2B 显示不同性别之间,各个变量重要性差距接近。图 2C 为训练集上 SHAP 概要图,在重要性前十的变量中,未报告过存在睡眠问题、一个月内有感冒以及教育程度较高患抑郁症的风险会降低。与之相反,健康状况越差、一个月内感到疲倦次数越多、血清可替宁浓度越高患抑郁症的风险会升高。

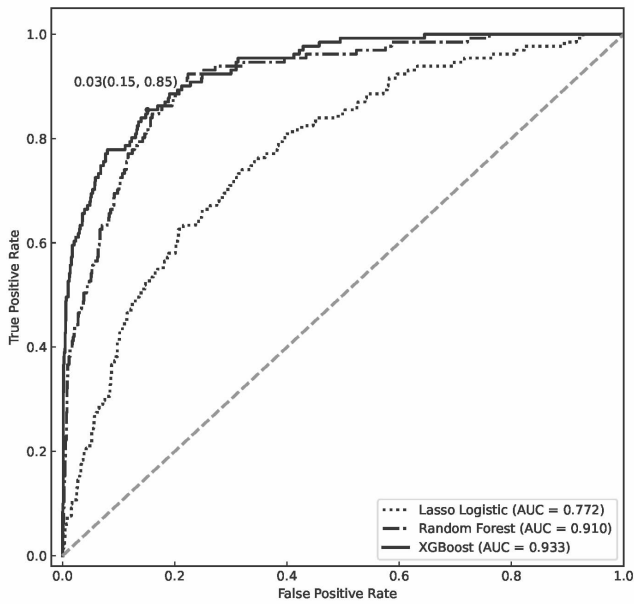
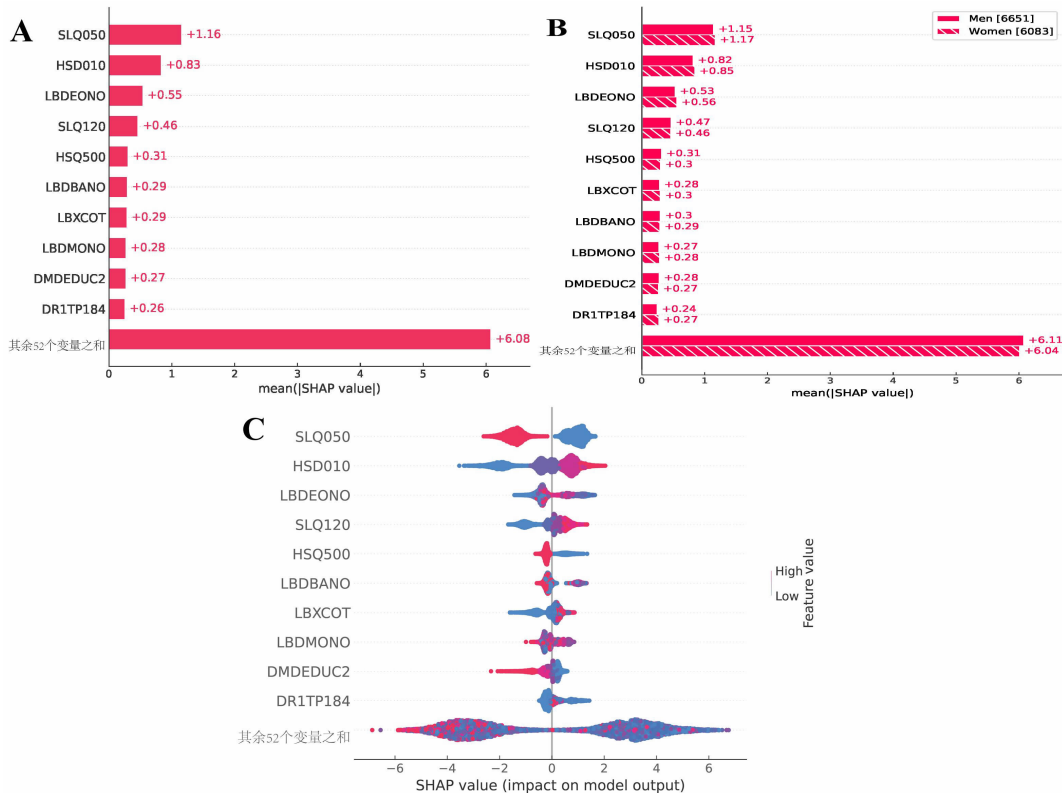


图 1 三种机器学习模型的 ROC 曲线  
Figure 1 ROC curve of three machine learning models

图 3 为 SHAP 决定图,其中图 3A 为嗜碱性粒细胞数变量的 SHAP 决定图,图 3B 为膳食中摄入的维生素 B6 的 SHAP 决定图(数值均经过对数转换)。图 3A 显示大多数患抑郁症风险较高的个体嗜碱性粒细胞数量较低;而图 3B 显示在膳食中摄入的维生素 B6 浓度较低和较高时,SHAP 值大于 0,提示患抑郁症风险较高的个体维生素 B6 摄入的浓度较高或较低,呈现“U”形的关系。

根据模型可以获得每个个体的 SHAP 力图,它可以显示影响个体的抑郁症的主要变量,整个模型的 SHAP 基础值为  $-0.022$ 。如在图 4A 中测试集里第 128 个个体,模型认为其 SHAP 值为  $0.66$  大于  $-0.022$ ,即预测其患抑郁症,主要原因为该个体近一个月有过感冒、差的健康状况、献血之外进行过艾滋病毒检测。图 4B 中展示的是第 777 个个体的力图,该个体的 SHAP 值为  $-2.33$  低于  $-0.022$  认为不存在抑郁症,主要原因为该个体无睡眠问题、膳食中较低的咖啡因摄入人等。



注:SLQ050 表示是否存在睡眠问题;HSD010 表示是否存在健康问题;LBDEONO 表示嗜酸性粒细胞数;SLQ120 表示近一个月内过度疲劳的次数;HSQ500 表示近一个月内是否有过感冒;LBDBANO 表示嗜碱性粒细胞数;LBXCOT 表示血清可替宁浓度;LBDMONO 表示单核细胞数;DMDEDUC2 表示受教育程度;DR1TP184 表示 PFA 18:4(十八碳四烯酸);图 A 为 SHAP 变量重要性图;图 B 为按照年龄分组的 SHAP 重要性;图 C 为 SHAP 概要图。

图 2 SHAP 变量重要性图  
2A SHAP variable importance plots; 2B SHAP variable importance plots by age; 2C SHAP summary plot

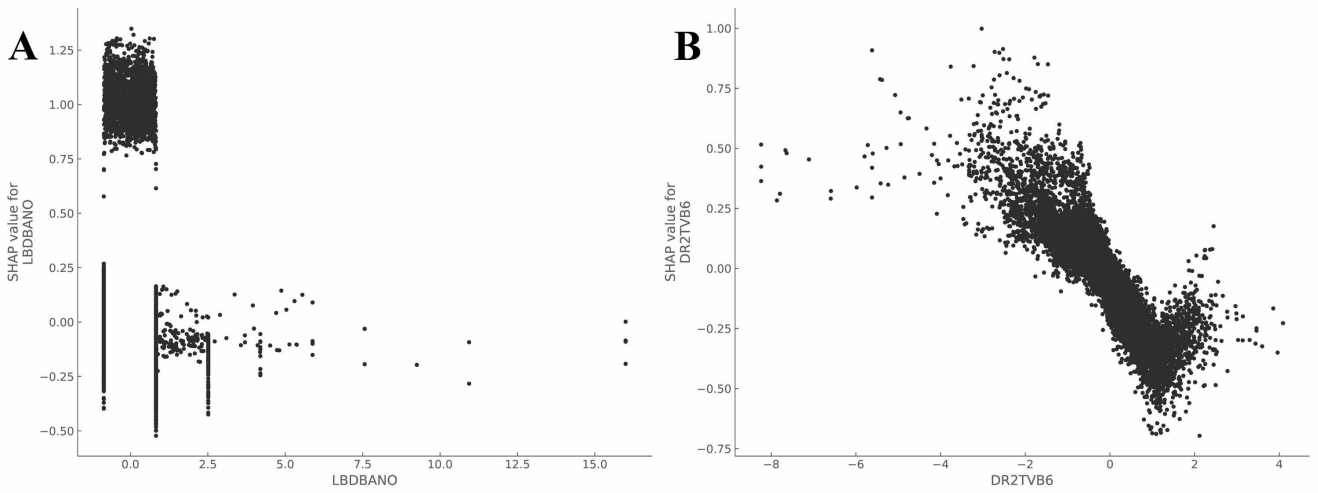
Figure 2 SHAP variable importance plots

### 3 讨论

抑郁症发病机制复杂,在人群中特别是老年人中

会造成严重后果。及时发现并尽早采取干预措施有助于提高生活质量。一方面造成抑郁症的机制复杂,

另一方面抑郁症的影响因素来自多个方面,如人口学特征、生活方式和遗传学等<sup>[17]</sup>。因此识别抑郁症的影响因素很关键。

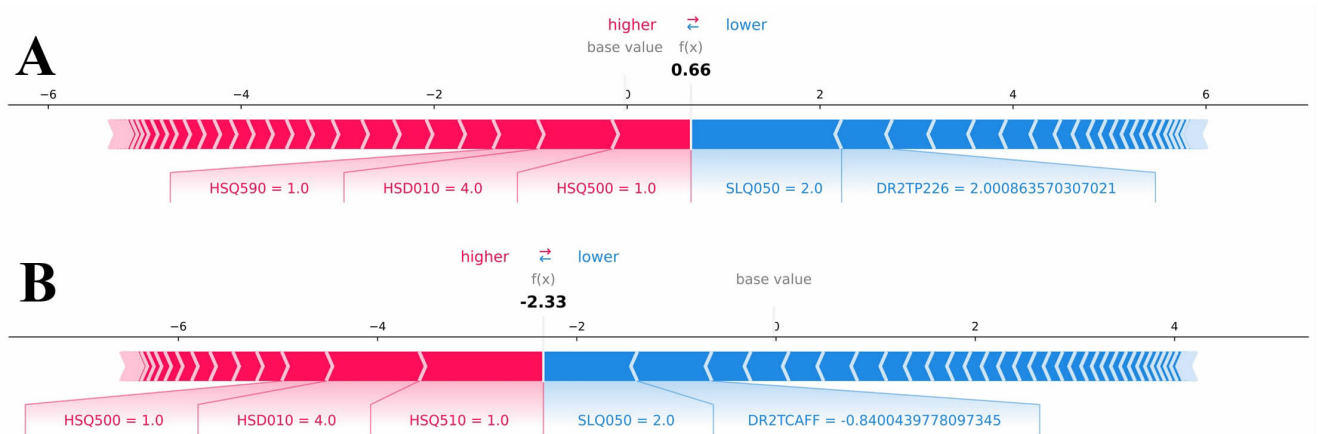
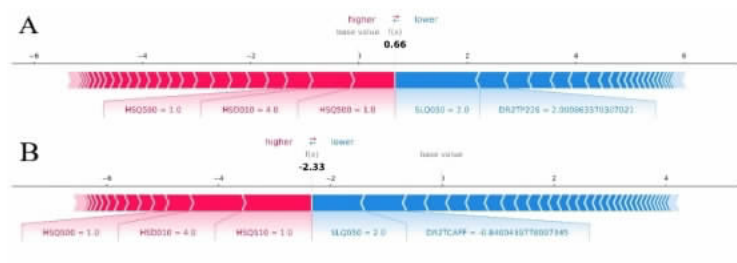


注:图 A 为嗜碱性粒细胞数变量的 SHAP 决定图;图 B 为摄入维生素 B6 的 SHAP 决定图。

图 3 SHAP 决定图

3A SHAP dependence plot of count of basophils; 3B SHAP dependence plot of vitamin B6 taken

Figure 3 SHAP dependence plot



注:HSQ590 表示除了献血外,是否进行过艾滋病病毒检测;HSD010 表示是否存在健康问题;HSQ500 表示近一个月内是否有过感冒;SLQ050 表示是否存在睡眠问题;DR2TP226 表示PFA 22:6 (二十二碳六烯酸, DHA)(标化值);HSQ510 表示30 天内是否患有胃病或肠道疾病,伴有呕吐或腹泻;DR2TCAFF 表示膳食中摄入的咖啡因浓度(mg)(标化值);图 A 为第 128 个个体的力图;图 B 为第 777 个个体的 SHAP 力图。

图 4 测试集上的 SHAP 力图

4A: SHAP force plot of 128th individual; 4B: SHAP force plot of 777th individual

Figure 4 SHAP force plots in test dataset

近年来数据挖掘技术在探寻疾病的影响因素方面应用广泛,Su 等人<sup>[18]</sup>基于中国健康调查数据库,使用长短期记忆模型(LSTM)和其他 6 个机器学习模型预测老年人抑郁情况和寻找抑郁的影响因素,最终在

机器学习模型认为,日常生活活动(ADL)/工具性 ADL(IADL)、自我评估的健康状况、婚姻状况、关节炎和同居次数是抑郁症老年人最重要的预测因素。Kessler 等人<sup>[19]</sup>使用机器学习模型应用于 1 056 名参

与者,预测重度抑郁症,最终发现机器学习模型的性能优于传统 logistic 回归模型。Zhang 等人<sup>[20]</sup>使用 Catboost 模型识别 2011—2018 年 NHANES 数据库中中老年人抑郁症,结果显示,在中老年人抑郁症识别方面,Catboost 模型性能最优;在中老年人中抑郁症的最重要的影响因素为家庭收入与贫困的比率、一般健康状况和存在睡眠困难;在老年人中,抑郁症最主要的影响因素为一般健康状况、存在睡眠困难和经历困惑或记忆问题。

虽然机器学习模型在医学领域的研究很多,但实际在临床上的应用受到了一定限制<sup>[21]</sup>。通常机器学习模型被视为黑盒模型,因此想要获得医患的信任,就必须具有解释性。SHAP 作为一种与模型无关的解释性机器学习方法,可以解释模型是如何做出决策的,给出各个变量的贡献大小和方向。同时,SHAP 还可以解释影响个体的主要因素,为精准化预防和控制疾病提供帮助。

本研究数据集来源于 2005—2021 年的 NHANES 数据库,使用机器学习模型识别 65 岁以上老年抑郁症患者。因为数据属于不平衡数据,因此采用 SMOTE 对数据集进行重抽样,以增加模型预测的准确性,使用 lasso logistic、随机森林和 XGBoost 三种机器学习模型,最终 XGBoost 模型的 AUC、准确率和特异度达到最高,随机森林模型的灵敏度最高。利用可视化的方式展示 SHAP 是如何做出决策的,其表现在变量重要性与个体解释两方面。第一方面,通过计算 SHAP 绝对值来比较变量重要性并通过柱状图展示,SHAP 认为最主要的影响因素是是否存在睡眠问题、是否存在健康问题、嗜酸性粒细胞数,这和 Zhang 等人<sup>[20]</sup>的研究结果基本一致。SHAP 概要图在此基础上还可以说明变量作用的方向以及该变量中个体值的分布情况,SHAP 概要图显示嗜酸性粒细胞浓度较高时,会增大患抑郁症的风险,在更高浓度时和较低浓度时会降低抑郁的风险,提示炎症水平可能和抑郁症相关这与 Lin 等人<sup>[22]</sup>的研究一致。同时模型还提示在生活方式和生生化指标以外,膳食影响因素如不饱和脂肪酸中的十八碳四烯酸、DHA 等还有维生素 B6 等可能也与抑郁症相关。第二方面体现在个体的预测,SHAP 力图展示了单个个体的影响因素及其贡献(因素的 SHAP 值大小),并求和得到个体的 SHAP 值,并通过比较与模型的 SHAP 基础值(本数据为 -0.022)大小判断是否存在抑郁症,有助于为个体精准化防治提供依据。

同时本研究还存在一些不足,本研究的数据为横断面数据,获得因素仅能进行抑郁症患者的识别,而不能进行因果推断。

**利益冲突声明** 本研究不存在任何利益冲突

#### 参考文献

- [1] Kok RM, Reynolds CF3. Management of depression in older adults: a review [J]. JAMA : the Journal of the American Medical Association, 2017, 317(20): 2114–2122.
- [2] 程小伟,黄俭,朱向阳,等. 抑郁症患者心理韧性在述情障碍与情绪自我效能感间的中介效应 [J]. 现代预防医学, 2023, 50(11): 2062–2066.  
Cheng XW, Huang J, Zhu XY, et al. Mediating effect of psychological resilience between alexithymia and emotional self-efficacy in patients with depression [J]. Modern Preventive Medicine, 2023, 50(11): 2062–2066.
- [3] 谢静静,李丽霞,柳学华,等. 正念减压疗法和正念认知疗法安全性的 meta 分析[J]. 中国心理卫生杂志, 2024, (1): 73–83.  
Xie JJ, Li LX, Liu XH, et al. A meta-analysis of safety of mindfulness-based stress reduction therapy and mindfulness-based cognitive therapy [J]. Chinese Mental Health Journal, 2024, (1): 73–83.
- [4] Bruce ML. Depression and disability in late Life: directions for future research [J]. The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry, 2001, 9(2): 102–112.
- [5] Beghi M, Butera E, Cerri CG, et al. Suicidal behaviour in older age: A systematic review of risk factors associated to suicide attempts and completed suicides [J]. Neuroscience and Biobehavioral Reviews, 2021, 127: 193–211.
- [6] 蔡利强,游亚凤,魏丽丽,等. 老年抑郁症患者自杀观念与多导睡眠图参数的相关性分析[J]. 浙江大学学报:医学版, 2020, 49(4): 462–467.  
Cai LQ, You YF, Wei LL, et al. Correlation between suicidal ideation and polysomnography parameters in late-life depression patients[J]. Journal of Zhejiang University. Medical Sciences, 2020, 49(4): 462–467.
- [7] 王威,第五永长,杨谦,等. 轻度认知障碍与老年期痴呆患者中医证候要素及影响因素的横断面调查[J]. 中医杂志, 2022, 63(11): 1064–1072.  
Wang W, Di Wu YC, Yang Q, et al. Traditional Chinese medicine syndrome elements and influencing factors of patients with mild cognitive impairment and senile dementia: a cross-sectional study[J]. Journal of Traditional Chinese Medicine, 2022, 63(11): 1064–1072.
- [8] 石萌,邹宇量. 中国中老年人午睡时长与抑郁症状的关联——基于 CHARLS 数据分析 [J]. 现代预防医学, 2023, 50(8): 1461–1467.  
Shi M, Zou YL. Relationship between nap duration and depressive symptoms among middle-aged and elderly Chinese——based on CHARLS data analysis[J]. Modern Preventive Medicine, 2023, 50(8): 1461–1467.
- [9] Sun YH, Liu QJ, Lee NY, et al. A novel machine learning approach to shorten depression risk assessment for convenient uses[J]. Journal of Affective Disorders, 2022, 312: 275–291.
- [10] Aleem S, Huda NU, Amin R, et al. Machine learning algorithms for depression: diagnosis, insights, and research directions [J]. Electronics, 2022, 11(7): 1111.
- [11] Lundberg S, Lee S L A unified approach to interpreting model predictions

- [M]. Neural Information Processing Systems, 2017, 4: 4768–4777.
- [ 12 ] 刘悦. 基于机器学习的老年人抑郁症状的预测[D]. 济南:山东大学, 2023.  
Liu Y. Prediction of depression in the elderly based on machine learning[D]. Jinan: Shandong University, 2023.
- [ 13 ] 周雯惠. 中国老年人抑郁症状影响因素研究[D]. 南京:南京邮电大学, 2022.  
Zhou WH. Study on influencing factors of depressive symptoms in Chinese Elderly—Based on 2018 CLHLS data [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2022.
- [ 14 ] Zeng Z, Li Q, Caine ED, et al. Prevalence of and optimal screening tool for postpartum depression in a community-based population in China[J]. Journal of Affective Disorders, 2024, 348: 191–199.
- [ 15 ] Dablain D, Krawczyk B, Chawla NV. DeepSMOTE: fusing deep learning and SMOTE for imbalanced data [J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(9): 6390–6404.
- [ 16 ] Bergstra J, Bardenet R, Bengio Y, et al. Algorithms for hyper-parameter optimization[C]//Proceedings of the 24th International Conference on Neural Information Processing Systems. Granada, Spain, Red Hook, NY, USA: Curran Associates Inc., 2011.
- [ 17 ] Monroe SM, Harkness KL. Major depression and its recurrences: Life course matters [J]. Annual Review of Clinical Psychology, 2022, 18: 329–357.
- [ 18 ] Jiang YW, Xu XJ, Wang R, et al. Radiomics analysis based on lumbar spine CT to detect osteoporosis [J]. European Radiology, 2022, 32: 8019–8026.
- [ 19 ] Kessler RC, Van Loo HM, Wardenaar KJ, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports [J]. Molecular Psychiatry, 2016, 21(10): 1366–1371.
- [ 20 ] Zhang CY, Chen XF, Wang S, et al. Using CatBoost algorithm to identify middle-aged and elderly depression, National health and nutrition examination survey 2011–2018 [J]. Psychiatry Research, 2021, 306: 114261.
- [ 21 ] Du TC, Tran TQB, Deo N, et al. Survey and evaluation of hypertension machine learning research [J]. Journal of the American Heart Association, 2023, 12(9): e027896.
- [ 22 ] Lin ZQ, Lawrence WR, Huang YH, et al. Classifying depression using blood biomarkers: A large population study [J]. Journal of Psychiatric Research, 2021, 140: 364–372.

收稿日期: 2023–09–17

(上接第 780 页)

- Journal of Internal Medicine Concepts & Practice, 2022, 17 (4): 273–277.
- [ 17 ] 国际高原医学会慢性高原病专家组. 第六届国际高原医学和低氧生理学术大会颁布慢性高原病青海诊断标准[J]. 青海医学院学报, 2005, 26(1): 3–5.  
International Society of Plateau Medicine Chronic Plateau Disease Expert Group. Qinghai diagnostic criteria for Chronic Mountain Sickness (CMS)[J]. Journal of Qinghai Medical College, 2005, 26(1): 3–5.
- [ 18 ] Zeng Y, Li FF, Yuan SQ, et al. Prevalence of hyperhomocysteinemia in China: an updated Meta-Analysis [J]. BIOLOGY-BASEL, 2021, 10(10): 959.
- [ 19 ] 李军, 白玛旺扎. 西藏山南地区藏族人群血清同型半胱氨酸和瘦素水平[J]. 基础医学与临床, 2023, 43(9): 1408–1411.  
Li J, Bai MWZ. Serum levels of homocysteine and leptin in Tibet Tibetans at Shannan area[J]. Basic & Clinical Medicine, 2023, 43(9): 1408–1411.
- [ 20 ] 李鹏昌, 邹雨桐, 刘治娟, 等. 西藏不同海拔地区藏族人群同型半胱氨酸水平 [J]. 基础医学与临床, 2023, 43 (12): 1841–1846.  
Li PC, Zou YT, Liu ZJ, et al. Homocysteine level of Tibetan population settled down at different altitudes [J]. Basic & Clinical Medicine, 2023, 43(12): 1841–1846.
- [ 21 ] Haj mouhamed D, Ezzaher A, Neffati F, et al. Effect of cigarette smoking on plasma homocysteine concentrations [J]. Clinical Chemistry and Laboratory Medicine : CCLM / FESCC, 2011, 49(3): 479–483.
- [ 22 ] 高官莉, 申东北, 魏晓龙, 等. 同型半胱氨酸与肠道疾病关系的研究进展[J]. 广西医学, 2023, 45(17): 2134–2137.  
Gao GL, Shen DB, Wei XL, et al. Research progress on the relationship between homocysteine and intestinal diseases [J]. Guangxi Medical Journal, 2023, 45(17): 2134–2137.
- [ 23 ] 冯胜强, 叶平, 骆雷鸣, 等. 北京市 1680 名社区居民血清同型半胱氨酸与代谢综合征关系的横断面研究[J]. 中华流行病学杂志, 2012, 33(3): 256–259.  
Feng SQ, Ye P, Luo LM, et al. Relationship between serum homocysteine and metabolic syndrome: a cross-sectional study [J]. Chinese Journal of Epidemiology, 2012, 33(3): 256–259.
- [ 24 ] 王雪颖, 袁维运, 周详, 等. 高同型半胱氨酸血症患者血液流变学和血脂水平的相关性研究 [J]. 心脏杂志, 2022, 34(6): 659–662.  
Wang XY, Yuan WY, Zhou X, et al. Correlation between hemorheology, blood lipid level and hyperhomocysteinemia [J]. Chinese Heart Journal, 2022, 34(6): 659–662.
- [ 25 ] 伍伟明, 李晓娜, 于梅花, 等. 高原红细胞增多症患者血清同型半胱氨酸、维生素 B12 和叶酸的检测及意义[J]. 高原医学杂志, 2017, 27(1): 53–54.  
Wu WM, Li XN, Yu MH, et al. Detection and significance of serum homocysteine, vitamin B12 and folate in patients with plateau erythrocytosis [J]. Journal of High Altitude Medicine, 2017, 27(1): 53–54.

收稿日期: 2023–07–07