

本文引用: 王勤勇, 张芷宁, 高 俭, 等. 基于机器学习算法构建颈动脉粥样硬化风险预测模型[J]. 中国动脉硬化杂志, 2026, 34(9): 897-908. DOI: 10.20039/j.cnki.1007-3949.2026.09.010.

[文章编号] 1007-3949(2026)34-09-0897-12

· 临床研究 ·

基于机器学习算法构建颈动脉粥样硬化风险预测模型

王勤勇¹, 张芷宁², 高 俭¹, 关 婷¹, 郝淑芹³, 马 冬⁴

1. 秦皇岛市北戴河康复疗养中心综合内科, 2. 秦皇岛市北戴河康复疗养中心骨关节康复科, 河北省秦皇岛市 066100;
3. 河北医科大学第一医院健康管理中心, 河北省石家庄市 050031; 4. 河北医科大学基础医学院, 河北省石家庄市 050017

[摘要] **[目的]** 基于筛选最优机器学习算法模型, 构建并部署 shinyapps.io 网络版预测体检人群颈动脉粥样硬化(CAS)患病风险的内部验证模型。**[方法]** 回顾性收集 2024 年 1 月至 2024 年 12 月于河北医科大学第一医院体检中心接受体检的 4 096 例受试者, 采用简单随机抽样方法按 3:1 比例将样本划分为训练集(3 072 例)和验证集(1 024 例), 两组结局变量分布均衡(训练集 45.02%, 验证集 44.73%, $P=0.561$)。收集受试者的一般资料、体格检查指标、生活方式指标、血常规指标、肝肾功能指标及病史资料。采用 LASSO 回归分析(通过 10 折交叉验证筛选最优 lambda 值)筛选最佳预测因子, 将其分别纳入多种机器学习算法(Logit、GNB、DT、RF、XGBoost、SVM、KNN、Bagging), 通过网格搜索与交叉验证优化各模型超参数; 绘制多模型 ROC 曲线筛选性能最优算法, 调用 SHAP 包评估每个预测因子对模型预测结果的贡献程度, 绘制不同机器学习算法校准曲线进行性能对比并计算各模型准确率, 绘制 DCA 曲线进一步验证最优机器学习算法的性能, 最终将筛选出的最优机器学习算法 XGBoost 部署至 shinyapps.io。**[结果]** 在训练集中, LASSO 回归分析结果显示, 性别、年龄、糖尿病史、高血压史、冠心病史、天冬氨酸氨基转移酶、体重指数、腰围、大血小板比值、吸烟、收缩压、白细胞计数、中性粒细胞计数、动脉硬化指数、高密度脂蛋白胆固醇是 CAS 的影响因素, 且各变量间不存在多重共线性(VIF 值均 <10)。在训练集中, XGBoost 性能最佳, 其预测 CAS 患病风险的 AUC(95% CI)、准确率、灵敏度、特异度、F1、NLR、PLR 分别为 0.870(0.857~0.883, $P<0.001$)、78.3%、80.2%、76.7%、0.769、0.258、3.446。在验证集中, XGBoost 同样性能最佳, 其 AUC(95% CI)、准确率、灵敏度、特异度、F1、NLR、PLR 分别为 0.909(0.890~0.928, $P<0.001$)、82.5%、82.3%、82.7%、0.808、0.214、4.754, 且 XGBoost 模型 AUC 的标准偏差最小; SHAP 分析结果显示, 年龄对模型预测 CAS 风险的贡献最大, 其次为收缩压。校准曲线显示, XGBoost 模型的 Brier 分数为 0.1831, 与校准效果最优的 SVM 模型(0.1822)基本持平, 且其曲线在全概率范围内均贴近理想参考线。结合其最优的判别能力与临床净获益, 该模型综合性能最佳。临床决策曲线显示, XGBoost 模型在较宽的风险阈值范围内具有临床净获益。在训练集中, 基于 XGBoost 模型构建预测 CAS 患病风险的计算器并部署到 shinyapps.io 网络, 该模型预测概率的最佳切割率(根据约登指数最大化确定)为 44.11%。从最新的体检人群中选取 2 例, 将相关信息填入网络版计算器, 结果显示, 2 例受试者的 CAS 患病风险分别为 66.33%、11.41%, 结果均落在风险阈值概率范围内, 初步展现了该计算器的临床应用潜力。**[结论]** 本研究针对 CAS 患病风险纳入了相对全面的指标, 成功构建出准确性高、性能优良的基于 XGBoost 模型的实用型 CAS 患病风险预测网络版计算器。需要指出的是, 本研究为横断面设计, 构建的模型仅用于评估 CAS 的患病风险, 不具备时间维度上的未来预测能力, 后续仍需开展多中心、前瞻性研究验证该模型的泛化能力与预测性能。

[关键词] 体检人群; 颈动脉粥样硬化; 机器学习算法; 预测模型

[中图分类号] R5

[文献标识码] A

Construction of a risk prediction model for carotid atherosclerosis based on machine learning algorithms

Wang Qinyong¹, Zhang Zhining², Gao Jian¹, Guan Ting¹, Hao Shuqin³, Ma Dong⁴

1. Department of General Internal Medicine, Beidaihe Rehabilitation and Sanatorium Center, Qinhuangdao; 2. Department of

[收稿日期] 2026-04-04

[修回日期] 2026-08-24

[基金项目] 国家自然科学基金项目(82270508)

[作者简介] 王勤勇, 副主任医师, 研究方向为颈动脉粥样硬化与脑小血管病的相关性及其对缺血性脑卒中发生发展的影响及临床干预策略, E-mail: wangqy101524@163.com。通信作者郝淑芹, 副主任中医师, 研究方向为常见慢性疾病的预防与管理, E-mail: shuqin770308@163.com。

Bone and Joint Rehabilitation, Beidaihe Rehabilitation and Sanatorium Center, Qinhuangdao, Hebei 066100, China;

3. Health Management Center, the First Hospital of Hebei Medical University, Shijiazhuang, Hebei 050031, China;

4. School of Basic Medical Sciences, Hebei Medical University, Shijiazhuang, Hebei 050017, China

[**ABSTRACT**] **Aim** To construct and deploy an internally validated model for predicting the risk of carotid atherosclerosis (CAS) in physical examination populations on the shinyapps.io web platform, based on screening for the optimal machine learning algorithm model. **Methods** Retrospectively collected data from 4 096 subjects who underwent physical examinations at the Physical Examination Center of the First Hospital of Hebei Medical University from January to December 2024. Using simple random sampling, the sample was divided into a training set (3 072 cases) and a validation set (1 024 cases) in a 3 : 1 ratio. The distribution of the outcome variable was balanced between the two groups (45.02% in the training set, 44.73% in the validation set, $P=0.561$). General information, physical examination indicators, lifestyle indicators, blood routine indicators, liver and kidney function indicators, and medical history of the subjects were collected. LASSO regression analysis (with optimal lambda value selected via 10-fold cross-validation) was used to screen for the best predictive factors, which were then incorporated into various machine learning algorithms (Logit, GNB, DT, RF, XGBoost, SVM, KNN, Bagging). Grid search and cross-validation were employed to optimize the hyperparameters of each model. Multiple model ROC curves were plotted to select the algorithm with the best performance. The SHAP package was used to evaluate the contribution of each predictive factor to the model's prediction results. Calibration curves of different machine learning algorithms were plotted for performance comparison, and the accuracy of each model was calculated. DCA curves were plotted to further validate the performance of the optimal machine learning algorithm. Finally, the selected optimal machine learning algorithm, XGBoost, was deployed on shinyapps.io. **Results** In the training set, LASSO regression analysis identified gender, age, history of diabetes, history of hypertension, history of coronary heart disease, aspartate aminotransferase, body mass index, waist circumference, large platelet ratio, smoking, systolic blood pressure, white blood cell count, neutrophil count, atherosclerosis index, and high density lipoprotein cholesterol as factors influencing carotid artery stenosis (CAS). No multicollinearity was observed among variables (all VIF values < 10). In the training set, XGBoost demonstrated the best performance, with AUC (95% CI), accuracy, sensitivity, specificity, F1-score, negative likelihood ratio (NLR), and positive likelihood ratio (PLR) for predicting CAS risk being 0.870 (0.857 ~ 0.883, $P < 0.001$), 78.3%, 80.2%, 76.7%, 0.769, 0.258, and 3.446, respectively. In the validation set, XGBoost also performed best, with AUC (95% CI), accuracy, sensitivity, specificity, F1-score, NLR, and PLR of 0.909 (0.890 ~ 0.928, $P < 0.001$), 82.5%, 82.3%, 82.7%, 0.808, 0.214, and 4.754, respectively. Additionally, XGBoost had the smallest standard deviation in AUC . SHAP analysis revealed that age contributed the most to the model's prediction of CAS risk, followed by systolic blood pressure. Calibration curves showed that the XGBoost model had a Brier score of 0.1831, which was comparable to the SVM model with the best calibration performance (0.1822), and its curve closely aligned with the ideal reference line across the entire probability range. Considering its optimal discriminative ability and clinical net benefit, this model exhibited the best overall performance. Clinical decision curves indicated that the XGBoost model provided clinical net benefit across a wide range of risk thresholds. In the training set, a web-based calculator for predicting CAS risk was developed based on the XGBoost model and deployed on shinyapps.io. The optimal cutoff probability for the model's predicted risk (determined by maximizing Youden's index) was 44.11%. Two subjects from the latest physical examination population were selected, and their relevant information was input into the web-based calculator. The results showed CAS risk probabilities of 66.33% and 11.41% for the two subjects, both falling within the risk threshold probability range, preliminarily demonstrating the clinical application potential of this calculator. **Conclusions** This study incorporated relatively comprehensive indicators for CAS risk and successfully developed an accurate, high-performance, practical web-based CAS risk prediction calculator based on the XGBoost model. It should be noted that this study was a cross-sectional design; the constructed model was only for assessing CAS prevalence risk and lacks the ability for future prediction over time. Future multi-center, prospective studies are needed to validate the model's generalizability and predictive performance.

[**KEY WORDS**] physical examination populations; carotid atherosclerosis; machine learning algorithm; prediction model

颈动脉粥样硬化(carotid atherosclerosis, CAS)是引发脑血管事件的基础病理改变,该病变可始于

生命早期,且会在较长时间内维持无症状状态,直至进展为颈动脉斑块、颈动脉狭窄等病变阶段,具

有较高的致残率与死亡率^[1-2]。一项 Meta 分析显示,全球 CAS 患病率超过 27%,患病群体以成年人为主,且患病率随年龄增长升高,80 岁以上老年人群中颈动脉内膜中膜厚度(carotid intima-media thickness, CIMT)增加的患病率高达 92.7%^[3-4]。CAS 给社会带来沉重的疾病负担,因此在疾病发生前高效筛选精准全面的影响因素,进而制定针对性预防策略、阻断 CAS 发生,以此降低 CAS 严重并发症及心血管疾病(cardiovascular disease, CVD)的发生风险^[5],尤为重要。

多普勒超声检测 CIMT 是判断 CAS 病变程度的可靠技术^[6]。自 2009 年至今,我国基本公共卫生服务项目已扩增至 14 项^[7],但受技术难度、设备成本、专业培训等因素限制,颈动脉多普勒超声检查并未被纳入该项目,无法满足 CVD 早期防治的需求;近年来已有多个研究团队开展相关工作,构建并验证了 CAS 发生风险预测模型^[8-9],但现有研究均未通过机器学习算法性能对比筛选最优模型,导致 CAS 发生风险的预测准确率不足,基层医护人员无法及时准确识别高危人群,因此亟须探索更准确高效的

建模方法,实现早期 CAS 的精准识别。

目前,针对体检人群 CAS 的机器学习识别研究仍较为少见,本研究拟纳入 Logit、GNB、DT、RF、XG-Boost、SVM、KNN、Bagging 共 8 种机器学习模型开展性能对比,筛选出性能最优的模型,并搭建便捷的网络版 CAS 风险预测计算器,为后续临床转化应用提供工具基础。

1 资料和方法

1.1 研究对象

选取 2024 年 1 月至 2024 年 12 月在河北医科大学第一医院体检中心接受体检的 4 096 例成年受试者,所有受试者均已完成包含颈动脉多普勒超声检查的全面健康检查,相关资料通过医院电子病历系统采集。本研究对变量缺失情况进行如下处理:变量缺失占比超过 10% 的样本直接排除,缺失占比在 10% 以内的变量采用随机森林插补法补全数据,最终共纳入 4 096 例体检人群(图 1)。

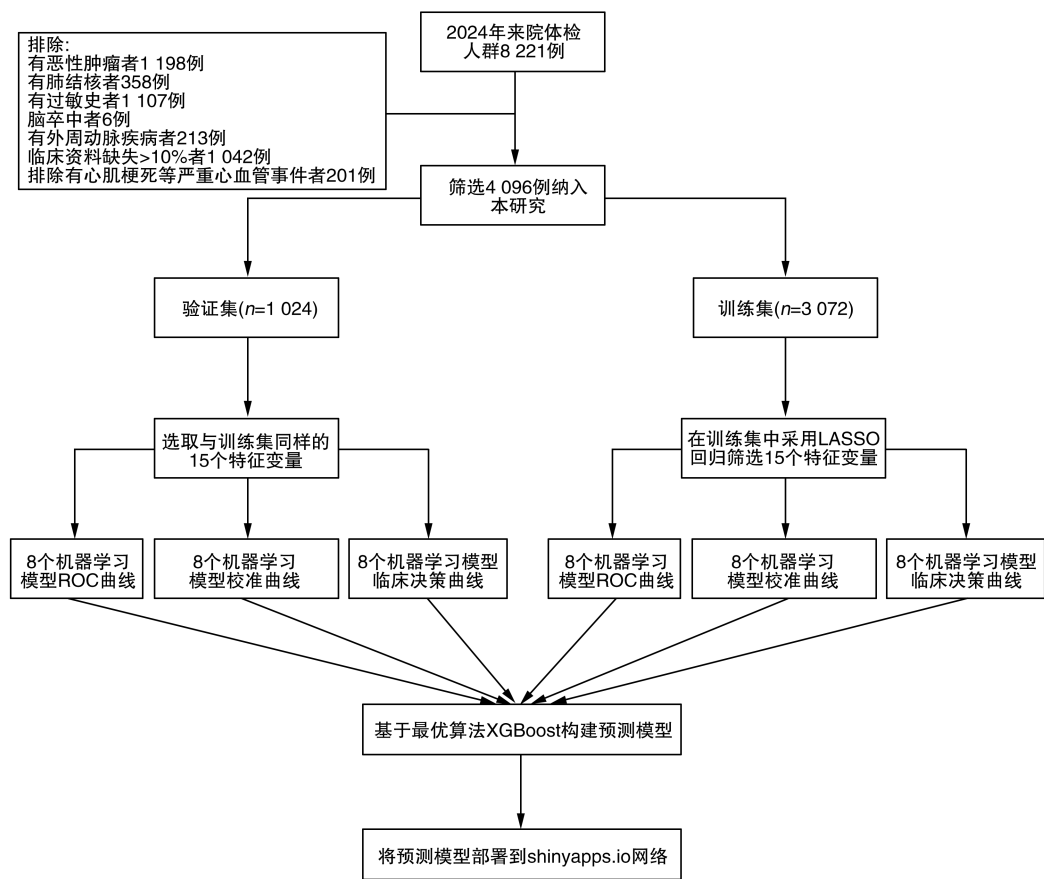


图 1. 研究对象纳入与排除流程图

Figure 1. Flowchart of participant inclusion and exclusion

纳入标准:(1)年龄>18岁;(2)接受颈动脉多普勒超声检查。排除标准:(1)存在重大疾病的个体,如罹患恶性肿瘤、严重感染性疾病,或有过敏史者;(2)既往有明确心肌梗死、经皮冠状动脉介入治疗、冠状动脉旁路移植术病史,或患有心力衰竭(NYHA 分级 $\geq$ II 级)等严重心血管疾病者;(3)已确诊卒中或外周动脉疾病(包括间歇性跛行、下肢动脉狭窄等);(4)既往已明确诊断 CAS,且正在接受阿司匹林、他汀类药物或其他调脂、抗血小板药物治疗者;(5)临床资料不全者。本研究经河北医科大学第一医院伦理委员会批准(批准号:S00290),并豁免知情同意。本研究聚焦于尚未发生明确心血管事件的社区体检人群,旨在构建适用于一级预防的 CAS 风险评估模型。

1.2 研究设计

本研究为回顾性横断面研究。

1.3 临床基线数据

通过医院电子病历系统收集一般资料,包括年龄、性别、体重指数(body mass index, BMI)、腰围、收缩压、舒张压、血糖、尿酸、肌酐清除率、甘油三酯、高密度脂蛋白胆固醇、白蛋白、球蛋白、丙氨酸氨基转移酶(alanine aminotransferase, ALT)、天冬氨酸氨基转移酶(aspartate aminotransferase, AST)、间接胆红素、直接胆红素、白细胞计数、大血小板比值、单核细胞计数、淋巴细胞计数、血小板计数、血小板分布宽度、中性粒细胞计数、动脉硬化指数、糖尿病史、高血压史、高脂血症史、冠心病史、吸烟史、饮酒史、CAS。CAS 被定义为 CIMT 增加 ≥ 1 mm 或存在斑块^[10]。

1.4 统计学分析

采用 Rstudio 4.3.2 和 Python 3.12 软件进行统计学分析。符合正态分布的计量资料以 $\bar{x} \pm s$ 表示,两组间比较采用独立样本 t 检验;计数资料以相对数表示,两组间比较采用 χ^2 检验。本研究以 CAS 为因变量,采用 LASSO 回归分析(采用 10 折交叉验证,根据 1 倍标准误准则选择 lambda.1se 作为最优惩罚参数,基于训练集独立筛选特征,避免信息泄露)筛选特征变量,再采用方差膨胀因子(variance inflation factor, VIF)对 LASSO 回归筛选出的变量进行多重共线性检验,以 $VIF < 10$ 作为不存在严重共线性的判断标准。本研究采用“粗筛-精调”分步优化策略对各机器学习模型进行超参数优化:粗筛阶段采用独立 5 折交叉验证(区别于 LASSO 特征筛选中的 10 折交叉验证)开展网格搜索,划定各超参数的大致最优区间,初步锁定各参数的最优区间后,为进一步提升模型性能,本研究在上述区间内以更

小步长开展精细化人工微调,通过观察验证集损失函数的收敛趋势对关键参数进行局部寻优,最终确定可使模型性能达到最优的超参数组合。随后绘制各模型在训练集与验证集中识别 CAS 的 ROC 曲线。采用基于平滑拟合的校准曲线(使用 Python sklearn.calibration.CalibrationDisplay,设置 strategy = 'sigmoid' 进行非参数平滑拟合,不采用 Hosmer-Lemeshow 分组检验)评估模型的校准度。绘制多模型临床决策曲线,比较各模型预测 CAS 风险的临床实用性,随后筛选出最优模型并确定其最佳参数,采用 SHAP 方法分析最优模型的可解释性。模型预测概率的最佳切割值根据约登指数最大化原则确定,最后基于最优模型构建 CAS 患病风险预测模型,并调用 shiny 包将其部署至 shinyapps.io 网络平台。本研究为单中心回顾性横断面设计,未设置独立的外部测试集,采用简单随机抽样的内部验证方法(训练集:验证集=3:1)评估模型性能,以 $P < 0.05$ 为差异具有统计学意义。上述所有数据预处理过程均使用训练集完成拟合和参数计算,再将得到的预处理参数直接应用于验证集,以此保证验证集的绝对独立性,避免数据泄露问题。

2 结 果

2.1 体检人群基线特征

本研究共纳入 4 096 例研究对象,采用简单随机抽样方法,按 3:1 的理论比例将研究对象分为训练集(3 072 例)与验证集(1 024 例)^[11],以 CAS 作为结局变量。基线特征分析结果显示,除吸烟史外(训练集 14.2% 比验证集 9.2%, $P < 0.001$),训练集与验证集其余变量的分布均无显著差异(均 $P > 0.05$)。训练集和验证集的 CAS 发生率分别为 45.02%、44.73%,平均年龄分别为(57.49 $\pm$ 12.12)岁和(57.47 $\pm$ 11.96)岁,BMI 均超过 24 kg/m² ($P > 0.05$;表 1)。

2.2 LASSO 回归特征筛选

以训练集是否发生 CAS 为因变量,纳入 32 项临床及实验室指标开展 LASSO 回归分析,经 10 折交叉验证确定最优惩罚参数 $\lambda = 0.009\ 215\ 17$ (lambda.1se),最终筛选出回归系数非零的 15 个预测因子,分别为年龄、性别、BMI、腰围、收缩压、HDL-C、AST、白细胞计数、大血小板比值、中性粒细胞计数、动脉硬化指数、糖尿病史、高血压史、冠心病史、吸烟史(图 2 和表 2)。VIP 检验结果显示,上述变量之间不存在多重共线性(所有 VIF 均 <10 ;表 3)。

表 1. 训练集和验证集的基线特征

Table 1. Baseline characteristics of the training and validation sets

变量	训练集 (n=3 072)	验证集 (n=1 024)	t/χ^2	P
年龄/岁	57.47±12.12	57.49±11.96	0.046	0.964
BMI/(kg/m ²)	24.26±3.24	24.15±3.12	-0.957	0.339
腰围/cm	84.42±10.09	84.02±9.72	-1.090	0.276
收缩压/mmHg	125.93±18.98	125.27±17.55	-1.024	0.325
舒张压/mmHg	80.12±10.95	79.80±10.71	-0.819	0.413
血糖/(mmol/L)	5.73±1.22	5.77±1.34	0.873	0.382
尿酸/(μmol/L)	332.19±86.01	332.41±88.43	0.072	0.943
肌酐清除率/(mL/min)	96.99±26.79	96.98±26.88	-0.009	0.993
甘油三酯/(mmol/L)	1.55±1.09	1.54±0.98	-0.266	0.79
HDLc/(mmol/L)	1.34±0.28	1.35±0.28	0.742	0.458
白蛋白/(g/L)	43.63±2.29	43.59±2.36	-0.495	0.62
球蛋白/(g/L)	29.39±3.33	29.31±3.30	-0.642	0.521
ALT/(U/L)	21.16±14.86	21.40±14.63	0.527	0.648
AST/(U/L)	23.26±8.96	23.43±8.41	0.456	0.599
间接胆红素/(mmol/L)	12.78±4.85	13.01±5.03	1.328	0.184
直接胆红素/(mmol/L)	2.35±0.91	2.37±0.99	0.739	0.746
白细胞计数/(×10 ⁹ L ⁻¹)	6.00±1.57	5.90±1.46	-1.0877	0.061
大血小板比值/%	27.03±7.05	27.33±7.14	1.181	0.238
单核细胞计数/(×10 ⁹ L ⁻¹)	0.31±0.10	0.31±0.10	-0.691	0.49
淋巴细胞计数/(×10 ⁹ L ⁻¹)	1.97±0.59	1.97±0.63	-0.135	0.893
血小板计数/(×10 ⁹ L ⁻¹)	239.83±57.64	240.98±74.00	0.514	0.608
血小板分布宽度/%	11.86±1.89	11.94±1.96	1.232	0.218
中性粒细胞计数/(×10 ⁹ L ⁻¹)	3.45±1.07	3.46±1.12	0.066	0.947
动脉硬化指数	2.91±0.90	2.92±0.93	0.126	0.9
男性/[例(%)]	1 453 (47.30)	486 (47.46)	0.003	0.957
糖尿病史/[例(%)]	215 (7.00)	86 (8.40)	2.009	0.156
高血压史/[例(%)]	713 (23.21)	228 (22.27)	0.335	0.563
高脂血症史/[例(%)]	341 (11.10)	127 (12.40)	1.161	0.281
冠心病史/[例(%)]	207 (6.74)	61 (5.96)	0.644	0.422
吸烟史/[例(%)]	436 (14.19)	94 (9.18)	16.690	<0.001
饮酒史/[例(%)]	859 (27.96)	287 (28.03)	<0.001	1
CAS/[例(%)]	1 383 (45.02)	458 (44.73)	1.096	0.561

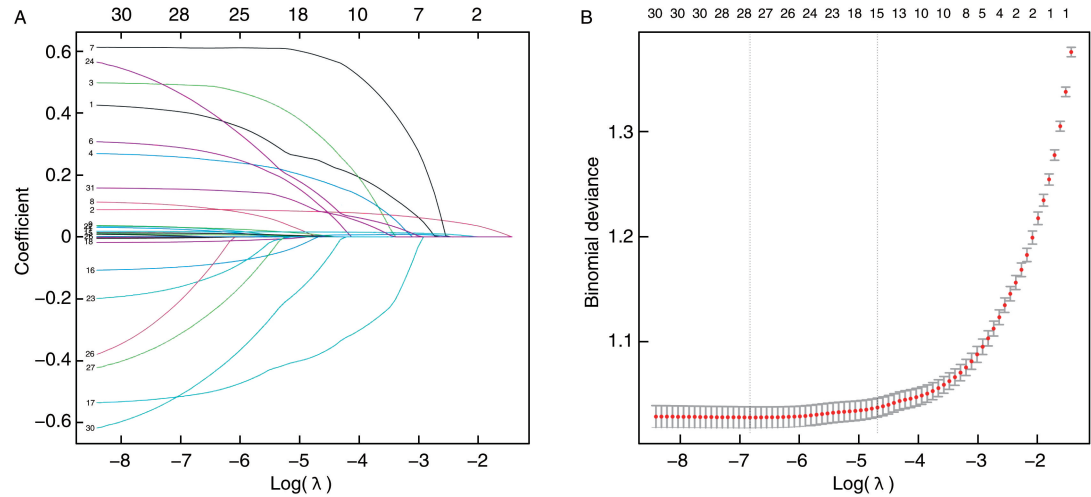


图 2. LASSO 回归系数图和交叉验证图

A 为 LASSO 回归系数图,纵轴为回归系数,横轴为 $\text{Log}(\lambda)$;B 为交叉验证图,纵轴为二项式偏差,横轴同样为 $\text{Log}(\lambda)$ 。
其中虚线分别对应左侧的 $\lambda_{\min}$ 和右侧的 λ_{1se} ,垂直虚线为经 10 折交叉验证筛选出的最优 λ 值。

Figure 2. LASSO regression coefficient plot and cross-validation plot

表 2. LASSO 回归变量系数和最佳 lambda 值
Table 2. LASSO regression variable coefficients and the optimal lambda value

变量	非零系数	lambda. 1se
年龄	0.081 800 503	0.009 215 17
BMI	0.006 859 051	
腰围	0.004 496 147	
收缩压	0.016 049 306	
HDLc	-0.380 161 630	
AST	0.001 644 575	
白细胞计数	0.137 815 023	
大血小板比值	0.002 348 829	
中性粒细胞计数	-0.107 149 606	
动脉硬化指数	0.083 475 380	
性别	0.247 867 960	
糖尿病史	0.331 010 738	
高血压史	0.183 641 261	
冠心病史	0.105 503 213	
吸烟史	0.585 030 308	

注: LASSO 回归不提供系数的标准误差及 P 值, 仅展示在最优 lambda. 1se 下回归系数非零的入选变量。

2.3 不同模型对 CAS 的判别性能对比

由于本研究未设置独立外部测试集, 因此采用随机分割内部验证的方式评估各模型的判别性能。将不同机器学习算法模型的 ROC 曲线绘制在同一坐标系中, 可直观对比各模型性能。在训练集中, XGBoost 为多模型中性能最优者, 其 $AUC(95\% CI)$ 为 0.870(0.857 ~ 0.883), 准确率为 0.783, 灵敏度为 0.802, 特异度为 0.767, F1 值为 0.769, NLR 值为 0.258, PLR 为 3.446(图 3 和表 4)。在验证集中, XGBoost 同样表现最优(图 3 和表 5)。8 个机器学习模型相比, XGBoost 模型标准偏差最小, 结果最好

(图 4)。8 种机器学习模型的最终参数分别为: Logit(penalty = “12”, solver = “liblinear”, C = 1, max_iter = 10000), GNB(priors = [0.3, 0.3, 0.4], var_smoothing = $1e-8$), DT(criterion = ‘entropy’, max_depth = 6, min_samples_leaf = 7, random_state = 0), RF(random_state = 2, max_features = 5, max_leaf_nodes = 5), XGBoost(learning_rate = 0.1, n_estimators = 28, max_depth = 4, min_child_weight = 1, seed = 0, subsample = 0.8, colsample_bytree = 0.8, gamma = 0.2, reg_alpha = 0.05, reg_lambda = 0.05), SVM(probability = True), KNN(leaf_size = 5, n_neighbors = 8), Bagging(max_features = 10, max_samples = 200)。

表 3. LASSO 回归选取的 15 个变量的多重线性检验
Table 3. Multicollinearity test of the 15 variables selected by LASSO regression

变量	R^2	VIF
性别	0.448	1.811
年龄	0.325	1.482
糖尿病史	0.085	1.093
高血压史	0.255	1.342
冠心病史	0.167	1.200
AST	0.051	1.054
BMI	0.747	3.949
腰围	0.803	5.081
大血小板比值	0.022	1.023
吸烟史	0.212	1.270
收缩压	0.374	1.598
白细胞计数	0.596	2.472
中性粒细胞计数	0.583	2.400
动脉硬化指数	0.349	1.535
HDLc	0.410	1.695

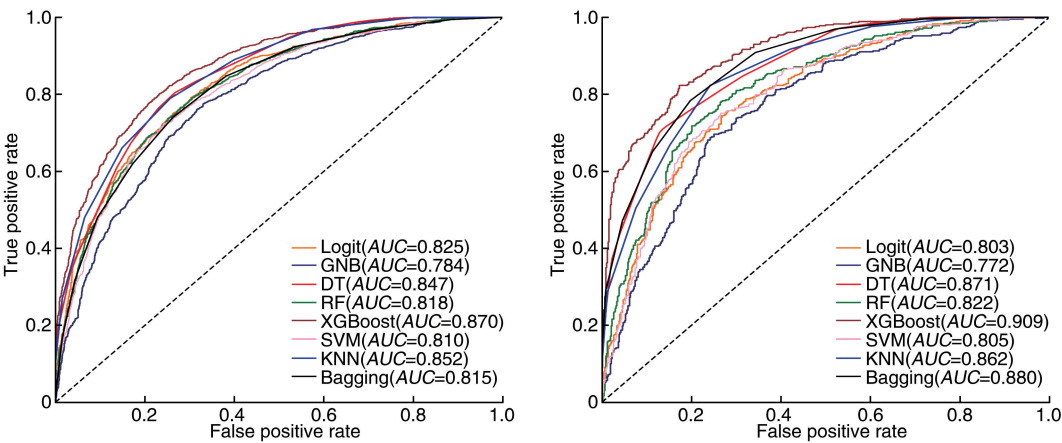


图 3. 多模型 ROC 曲线对比图

左侧是训练集多模型 ROC 曲线对比图, 右侧是验证集多模型 ROC 曲线对比图。

Figure 3. Comparison of multi-model ROC curves

表 4. 训练集中各机器学习模型的性能对比结果

Table 4. Performance comparison results of various machine learning models in the training set

模型	AUC(95% CI)	准确率	灵敏度	特异度	F1	NLR	PLR
Logit	0.825(0.810~0.840)	0.740	0.792	0.697	0.732	0.299	2.612
GNB	0.784(0.767~0.801)	0.717	0.776	0.669	0.712	0.335	2.344
DT	0.847(0.833~0.861)	0.765	0.803	0.734	0.755	0.269	3.019
RF	0.818(0.803~0.833)	0.741	0.766	0.721	0.727	0.325	2.740
XGBoost	0.870(0.857~0.883)	0.783	0.802	0.767	0.769	0.258	3.446
SVM	0.810(0.794~0.826)	0.742	0.722	0.758	0.716	0.366	2.983
KNN	0.852(0.838~0.866)	0.764	0.790	0.743	0.751	0.283	3.073
Bagging	0.815(0.799~0.831)	0.735	0.758	0.716	0.718	0.353	2.820

表 5. 验证集中各机器学习模型的性能对比结果

Table 5. Performance comparison results of various machine learning models in the validation set

模型	AUC(95% CI)	准确率	灵敏度	特异度	F1	NLR	PLR
Logit	0.803(0.775~0.831)	0.737	0.745	0.731	0.717	0.349	2.772
GNB	0.772(0.743~0.801)	0.727	0.688	0.758	0.692	0.412	2.841
DT	0.871(0.848~0.894)	0.795	0.710	0.864	0.756	0.336	4.216
RF	0.822(0.795~0.849)	0.763	0.718	0.799	0.730	0.353	3.566
XGBoost	0.909(0.890~0.928)	0.825	0.823	0.827	0.808	0.214	4.754
SVM	0.805(0.777~0.833)	0.742	0.751	0.735	0.723	0.339	2.834
KNN	0.862(0.838~0.886)	0.787	0.823	0.758	0.776	0.233	3.401
Bagging	0.880(0.858~0.902)	0.786	0.799	0.776	0.783	0.244	3.931

注:阴性似然比(negative likelihood ratio,NLR),数值越接近 0,说明模型排除疾病的能力越强;阳性似然比(positive likelihood ratio,PLR),数值越大说明模型确认疾病的能力越强。临床上通常将 NLR<0.3、PLR>3 作为模型性能可接受的标准。F1 为精确率与召回率的调和平均数,计算公式为 $F1 = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$,取值范围为 0~1,数值越高说明模型综合性能越好。

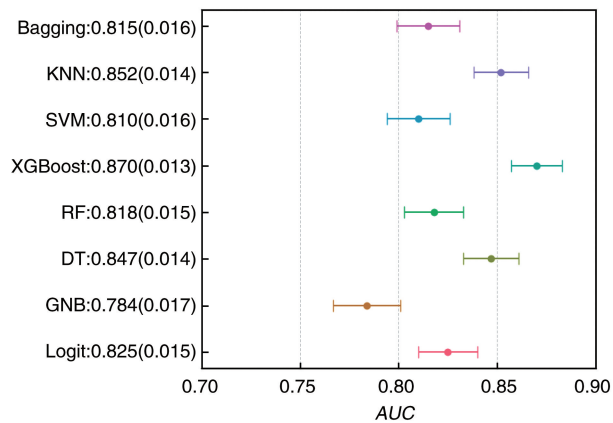


图 4. 标准偏差图

纵轴为标准差(standard deviation,SD),SD 值越小,代表模型稳定性越好。

Figure 4. Standard deviation chart

校准曲线分析显示,各模型的校准度存在一定差异:SVM 的 Brier 分数(0.182 2)略低于 XGBoost (0.183 1),提示 SVM 在概率校准方面存在微弱优

势。但校准曲线仅能反映模型拟合优度的一个维度,无法全面评价模型的临床预测效能。结合模型判别能力、临床决策曲线净获益、模型稳定性及可解释性等多维度指标来看,XGBoost 在各维度均表现最优:该模型 AUC 较高,DCA 曲线在较宽阈值范围内净获益最大,且标准差最小,提示模型泛化能力更稳定。虽然 SVM 的校准误差略小,但在灵敏度与特异度的平衡、临床实用性方面均不及 XGBoost。因此,本研究最终选择 XGBoost 作为研究模型,该模型的校准曲线贴近理想参考线,预测概率与实际观测值吻合良好,结合其优异的判别能力与临床决策价值,XGBoost 模型在识别 CAS 风险方面具备最优的综合性能(图 5)。

临床决策曲线用于评估预测模型的临床应用价值,曲线位置越高代表模型净获益越大。在本研究中,XGBoost 模型的 DCA 曲线在训练集与验证集中,均处于所有模型的上方或与最优模型持平:在训练集中,当风险阈值概率处于 5%~95% 区间时,

该模型净获益最大;在验证集中,当风险阈值概率处于 7%~75% 区间时,该模型净获益最大(图 6)。

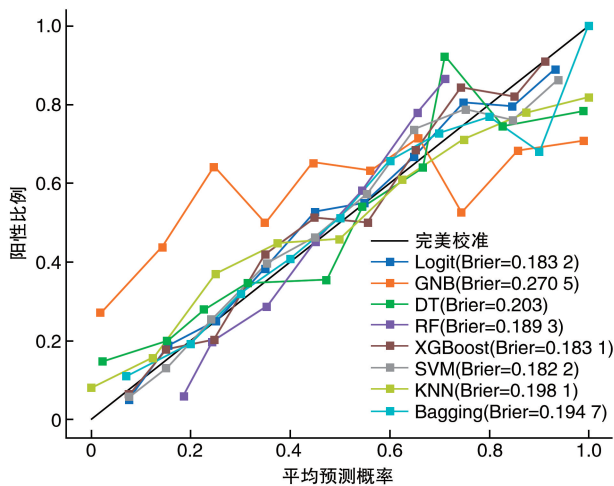


图 5. 多模型预测概率校准曲线

Figure 5. Calibration curves for predicted probabilities of multiple models

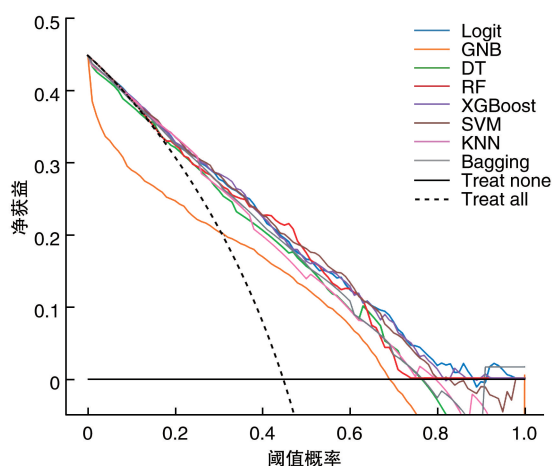
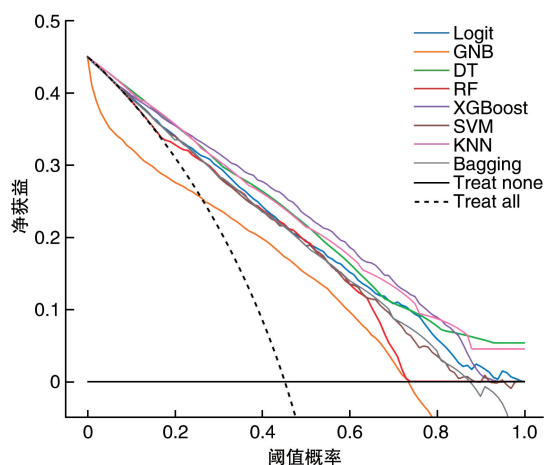


图 6. 多模型 DCA 曲线对比图

上为训练集多模型 DCA 曲线对比图,
下为验证集多模型 DCA 曲线对比图。

Figure 6. Comparison of DCA curves across multiple models

2.4 SHAP 方法分析最优模型的可解释性

采用 SHAP 方法得到 XGBoost 模型各预测因子贡献度。按照平均绝对 SHAP 值排序,对模型预测贡献最大的前 5 位变量依次为年龄、收缩压、HDLc、吸烟史和性别。此外,高血压史、白细胞计数、动脉硬化指数、腰围、BMI 等变量对模型预测也存在一定贡献;大血小板比值、中性粒细胞计数、AST、糖尿病史及冠心病史的平均 SHAP 值均较低,对模型预测的边际贡献相对有限。从影响方向来看,年龄、收缩压、吸烟史、性别(男性)、高血压史、白细胞计数及动脉硬化指数的 SHAP 值以正值为主,提示这些指标升高或阳性暴露与 CAS 患病风险增加相关;HDLc 的 SHAP 值多为负值,且其高值点(红色)集中在负半轴,表明该指标升高对 CAS 具有保护效应。其余变量(如腰围、BMI 等)的 SHAP 值分布较为分散,提示其影响方向可能受其他因素的交互调节。本研究所得结果与既往流行病学认识一致,进一步验证了该模型预测逻辑的临床合理性(图 7)。

2.5 预测模型部署 shinyapps.io

基于训练集 XGBoost 模型的预测概率,本研究以约登指数最大化为准则确定最佳切割值为 44.11%。将该切割值应用于验证集后,重新计算其判别性能可得:灵敏度为 68.79%,特异度为 77.09%,准确率为 73.14%,约登指数为 0.458 8,该值可用于区分 CAS 的高风险与低风险人群,结果提示,高于该值的人群发生 CAS 的风险较高。需要说明的是,在验证集中应用训练集确定的最佳切割值后,模型性能低于其原有表现,这是因为训练集优化得到的阈值在不同样本中通常会发生偏移,临床应用中可根据实际筛查需求(如优先保证灵敏度或特异度)灵活调整阈值。

本研究将河北医科大学第一医院最新收集的 2 名成年人体检信息输入模型进行验证,其中第 1 名情况如下:性别为男性,年龄 62 岁,无糖尿病史,无冠心病史,有高血压史,不吸烟,AST 为 19 U/L, BMI 为 24.54 kg/m²,腰围为 89 cm,大血小板比值为 26.5%,收缩压为 138 mmHg,白细胞计数为 $8.77 \times 10^9/L$,中性粒细胞计数为 $5.05 \times 10^9/L$,动脉硬化指数为 2.96, HDLc 为 1.24 mmol/L,该案例识别得到的 CAS 发生风险概率为 66.33%,高于模型确定的最佳切割值,提示该体检者当前发生 CAS 的可能性较高,建议进一步临床关注(图 8A);第 2 名情况如下:性别女性,年龄 47 岁,无糖尿病史,无冠心病史,无高血压史,不吸烟,AST 为 20.7 U/L, BMI 为 28.48 kg/m²,腰围为 91 cm,大血小板比值为 26.5%,收缩压为

106 mmHg, 白细胞计数为 $5.11 \times 10^9/L$, 中性粒细胞计数为 $2.43 \times 10^9/L$, 动脉硬化指数为 1.74, HDLC 为 1.78 mmol/L, 该案例识别得到的 CAS 发生风险概率为 11.41%, 低于模型确定的最佳切割值, 提示

该体检者 CAS 发生风险很低(图 8B)。本研究开发了基于 XGBoost 模型识别 CAS 发生风险的网络版计算器(URL <https://suabvcs.shinyapps.io/JDMZYH/>), 目前为内部验证版本, 可供临床医生试用并反馈意见。

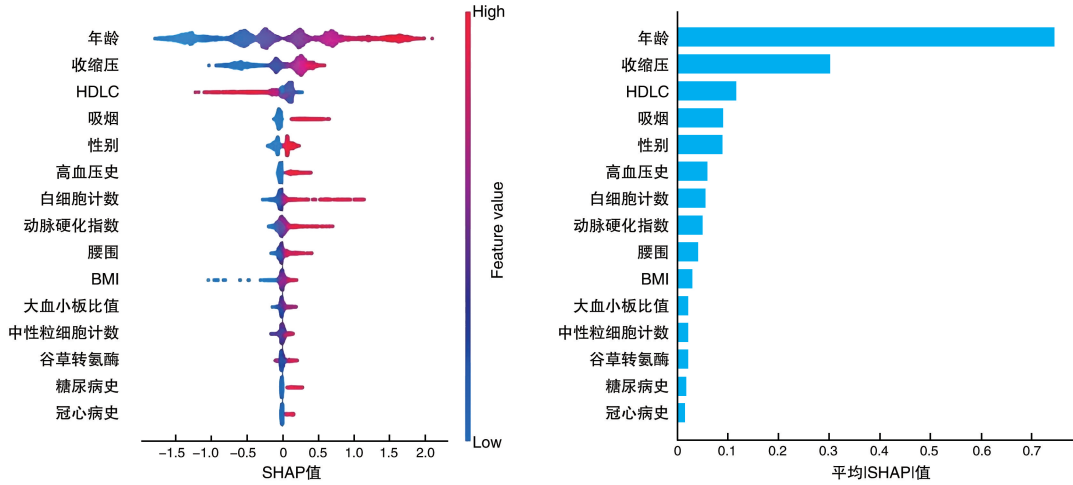


图 7. XGBoost SHAP 值图及 XGBoost 特征重要性平均|SHAP|值图

Figure 7. XGBoost SHAP value plot and mean |SHAP| value plot for feature importance

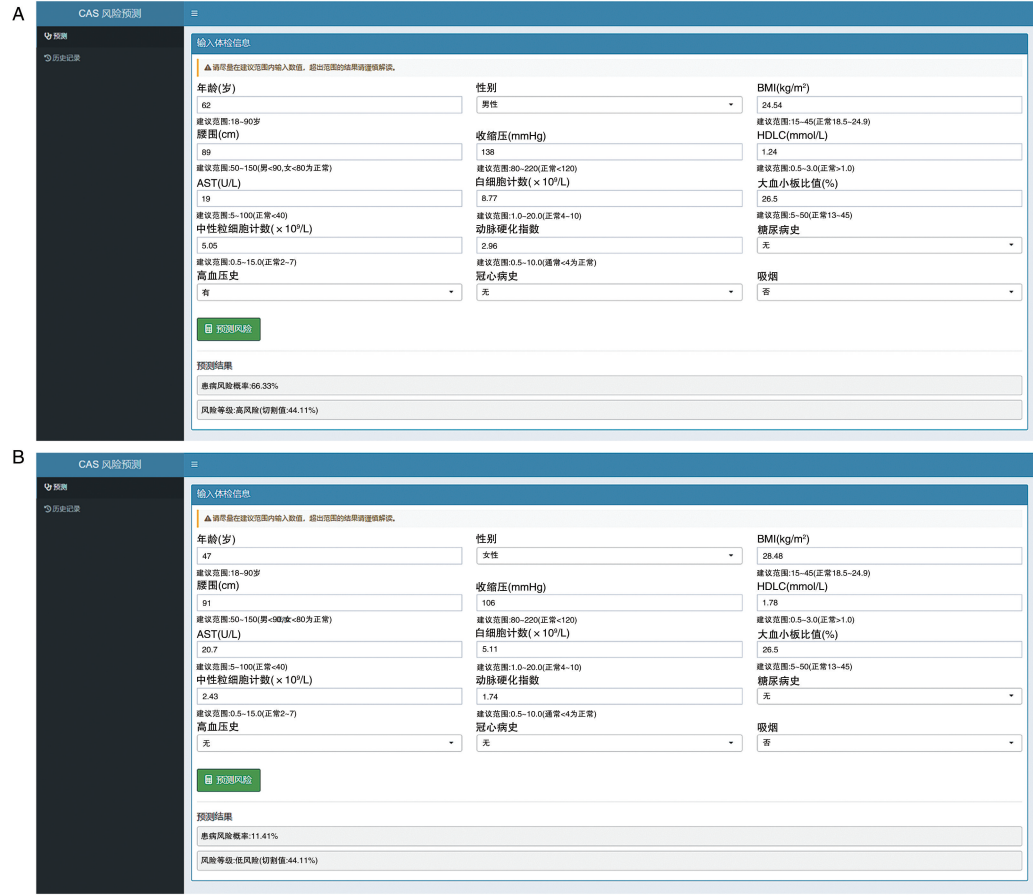


图 8. 高风险及低风险患者预测示例

模型最佳切割值为 44.11%, 当预测概率 $\geq 44.11\%$ 时提示为高风险(上图 66.33%), 反之则为低风险(下图 11.41%)。

Figure 8. Prediction examples for high-risk and low-risk patients

3 讨 论

基于 XGBoost 算法构建的预测模型,能够较好预测体检人群发生 CAS 的风险。本研究显示,纳入的体检人群中 CAS 发生率约为 44.95%,高于已有国外研究报道^[12-14];年龄、BMI、腰围、收缩压、HDLc、AST、白细胞计数、大血小板比值、中性粒细胞计数、动脉硬化指数、性别、糖尿病史、高血压史、冠心病史、吸烟史均与 CAS 的发生风险存在相关性。SHAP 分析显示,年龄、收缩压、HDLc、吸烟史、性别对模型的贡献最大,结果与既往研究一致。

与经典的 Framingham 评分(该评分主要用于冠心病预测^[15],也被应用于 CAS 研究)相比,本研究纳入的预测因子更为全面,包括血常规、肝肾功能相关指标,并且能够通过机器学习算法捕捉变量间的非线性关系。在本研究内部验证中,XGBoost 模型的 AUC(0.870~0.909)显著优于传统 Logistic 回归模型,提示 XGBoost 在 CAS 风险识别中具备一定优势。需要指出的是,Framingham 评分是基于长期随访的大样本队列开发而来^[15],而本研究仅为横断面内部验证,模型的外推性仍需进一步验证,且未直接比较 Framingham 评分在本研究队列中的实际表现。

CAS 在普通人群中十分普遍,是中风和冠心病的独立危险因素,且 CAS 的发生与多种可调控因素相关,这类因素占比约为 60%~70%,因此不同个体罹患 CAS 的概率存在一定差异,对 CAS 进行早期干预具有重要临床价值^[16-18]。既往研究表明,以每日 5 000 步为活动量分界线,与日步数低于 5 000 步的人群相比,活动量达标人群脂代谢指标下降幅度更大,说明适当运动可降低 CAS 发生风险^[19];另有研究显示,与低脂饮食模式相比,坚持地中海饮食模式可显著降低 CIMT^[20];阻塞性睡眠呼吸暂停患者炎症指标水平升高,可能会推动动脉粥样硬化进展^[21]。

近年来,已有多项临床研究探索通过生物标志物评估 CAS 发生风险^[22],例如 microRNA^[23]、低 CD4 细胞计数^[24]、血清维生素 D 水平^[25]、血浆 microRNA-29C 水平^[26]等,均可辅助识别 CAS;同时研究者也在不断探寻降低 CAS 发生风险的干预方案。尽管这类生物标志物对 CAS 的识别效果较好,但复杂的采样流程与高昂的检测成本,导致其难以在临床场景中广泛推广应用。与之相比,基于常规体检指标构建的预测模型具备成本低廉、指标易获取的突出优势。

本研究纳入的指标较为全面,包括一般资料、体格检查指标、生活方式指标、血常规指标、肝肾功能指标与病史资料,可显著提升模型的区分能力与

预测准确性,对识别 CAS 高风险人群具备较高的临床实用价值。本研究通过 LASSO 回归共筛选出 15 个最佳预测因子,筛选结果与既往研究结论基本一致。已有荟萃分析显示,颈动脉斑块与颈动脉狭窄的患病率会随年龄增长不断升高,且男性的升高幅度高于女性^[3-4];一项前瞻性队列研究也表明,体重增加会提升动脉粥样硬化的发病风险^[27]。腰围是反映腹部肥胖的直接指标,研究证实腰围与临床前动脉粥样硬化存在显著关联,且该关联存在性别差异^[28];收缩压与 CIMT 呈正相关,而 CIMT 增厚正是 CAS 的早期标志^[29]。HDLc 是 CVD 的保护因素,HDLc 水平升高可有效降低 CAS 发生风险^[30]。一项观察性研究显示,CIMT 增厚与非酒精性脂肪性肝病(non-alcoholic fatty liver disease, NAFLD)分级升高显著相关,而 NAFLD 会进一步导致 AST 水平升高^[31]。白细胞计数与 CAS 斑块显著相关^[32];本研究中,中性粒细胞计数在模型中呈现负向贡献,但这一关联并不代表生物学层面的保护作用,该结果更多是由于白细胞计数与中性粒细胞计数在临床检测中存在高度共线性,二者同时进入 XGBoost 模型进行非线性计算后,模型为维持整体预测精度,在权重分配上产生的统计学伪关联,二者各自的独立效应目前尚不明确^[33]。大血小板比值升高反映血小板活化水平增强,活化的血小板可释放促炎介质,因此与动脉粥样硬化的严重程度相关^[34]。动脉硬化指数是用于评估动脉粥样硬化的指标,可通过脉搏波传导速度或踝臂指数等方法测量,其数值与病变严重程度呈正相关^[35]。高血压、糖尿病、高脂血症、冠状动脉疾病均为 CAS 的显著预测因子^[36]。此外,既往研究显示,戒烟可延缓 CIMT 增厚进展,降低斑块发生率^[37]。

本研究基于 XGBoost 算法构建了体检人群 CAS 发生风险预测模型,可早期识别 CAS 高风险人群,便于及时采取预防措施。与同类研究相比,本研究通过对比 8 种机器学习模型后选取性能最优的 XGBoost 建模,该算法通常能获得较高预测精度,可捕捉特征间的复杂非线性关系,与 CAS 风险预测的需求相适配;同时通过正则化和早停策略能有效防止过拟合,适合处理大规模数据^[38],这也是本研究的亮点之一。其次,本研究在 shinyapps.io 平台部署了基于该算法的预测模型,临床医师可通过浏览器直接访问,工具兼容多平台,界面直观,可实时计算并显示预测结果,这是本研究的另一大亮点。

本研究存在一定局限性:首先,本研究为单中心回顾性研究,样本来源存在局限;此外本研究中

XGBoost 验证集 AUC 略高于训练集,尽管本研究严格执行了训练集调参、验证集独立评估的隔离策略,仍难以在内部验证中完全排除交叉验证与早停策略带来的潜在信息渗漏风险。其次,吸烟变量虽在组间分布存在一定差异,但其已作为特征纳入模型,模型可据此实现对基线偏差的统计校准,该模型真实的泛化性能仍有待前瞻性多中心外部研究数据证实。另外,本研究采用基于线性假设的 LASSO 回归进行降维筛选后,将预测因子直接输入非线性的 XGBoost 模型,该方式虽有效控制了多重共线性(VIF 均 <10),但二者在特征贡献逻辑层面的理论契合度仍存在不足,属于现阶段单中心探索性研究的折中处理方案。后续研究拟引入嵌套交叉验证或 SHAP 递归消除等逻辑一致性更强的策略,进一步优化特征选择流程。本研究明确排除了既往存在心肌梗死、冠状动脉血运重建史或心力衰竭等重大心血管事件的患者,同时也排除了已确诊脑卒中或外周动脉疾病的受试者。这一纳入排除标准可能导致本模型不适用于已发生严重 ASCVD 事件的人群;本模型面向尚未出现明确心血管事件的体检人群开发,用于早期识别首次 CAS 发生风险,若将其外推至已接受用药或既往已确诊的人群需格外谨慎。同时,本研究未对斑块性质进行区分,而不同性质斑块的危险因素与预后可能存在差异;本研究将 CAS 定义为(IMT 增厚 ≥ 1 mm 或存在斑块,也未进一步对软斑块、硬斑块及混合斑块分层。未来开展前瞻性研究时,应统一超声报告的评估标准,进一步构建针对易损斑块的亚型预测模型。既往研究表明,软斑块的脑血管事件风险显著高于硬斑块,不同性质斑块的危险因素谱可能存在差异。受回顾性研究数据源中超声报告描述标准不一致的限制,本研究未能开展斑块性质分层分析。尽管本研究已经纳入了较为全面的指标,但仍可能遗漏氧化应激指标、饮食运动等生活方式相关的重要因素;最后,本研究得出的最佳切割值 44.11% 基于训练集确定,其稳定性仍需独立人群中加以验证。后续研究将开展多中心前瞻性队列研究,纳入更多潜在预测因子,并进一步探索该模型在真实世界中的临床效用。

综上,本研究对 8 种机器学习模型进行对比,通过 ROC 曲线、校准曲线、DCA 曲线完成模型性能验证,最终选择最优的 XGBoost 模型识别 CAS 发生风险,并将其部署于 shinyapps.io 平台,构建了经内部验证的预测工具。该模型在内部验证中表现良好,适用于尚未发生心肌梗死、脑卒中等严重心血管事

件的体检人群,具备早期风险筛查的临床潜力,但仍需进一步开展多中心前瞻性研究验证其泛化能力。本研究所有分析代码与匿名化数据示例已公开于 GitHub: <https://github.com/asaaasasd/jubilant-guacamole.git>。

[参考文献]

- [1] 陈佳琳,唐志晗,刘录山,等. PCSK9 通过诱导血管平滑肌细胞表型转换促进动脉粥样硬化[J]. 中国动脉硬化杂志, 2026, 34(4): 287-296.
Chen J L, Tang Z H, Liu L S, et al. PCSK9 promotes atherosclerosis by inducing phenotypic transformation of vascular smooth muscle cells[J]. Chin J Arterioscler, 2026, 34(4): 287-296.
- [2] GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013[J]. Lancet, 2015, 385(9963): 117-171.
- [3] Song P, Fang Z, Wang H. Global and regional prevalence, burden, and risk factors for carotid atherosclerosis: a systematic review, meta-analysis, and modelling study[J]. Lancet Glob Health, 2020, 8(5): e721-e729.
- [4] Fu J Z, Deng Y H, Ma Y, et al. National and provincial-level prevalence and risk factors of carotid atherosclerosis in Chinese adults [J]. JAMA Netw Open, 2024, 7(1): e2351225.
- [5] 杨兴丹,王秋丽,李国锋,等. 缺血性脑卒中患者颈动脉钙化影响因素的列线图模型构建及验证[J]. 中国动脉硬化杂志, 2024, 32(1): 40-48.
Yang X D, Wang Q L, Li G F, et al. Construction and validation of a columnar line graph model based on factors influencing carotid artery calcification in patients with ischaemic stroke[J]. Chin J Arterioscler, 2024, 32(1): 40-48.
- [6] Johri A M, Nambi V, Naqvi T Z, et al. Recommendations for the assessment of carotid arterial plaque by ultrasound for the characterization of atherosclerosis and evaluation of cardiovascular risk: from the American Society of Echocardiography[J]. J Am Soc Echocardiogr, 2020, 33(8): 917-933.
- [7] 田昌伟,郑艳敏,孙乃玲,等. 我国基本公共卫生服务标准化现状分析暨均等化的标准化策略初探[J]. 中华流行病学杂志, 2020, 41(10): 1723-1730.
Tian C W, Zheng Y M, Sun N L, et al. Current status of standardization of basic public health services and standardization strategy of service equalization in China[J]. Chin J Epidemiol, 2020, 41(10): 1723-1730.
- [8] Huang G Q, Jin Q K, Tian X Q, et al. Development and validation of a carotid atherosclerosis risk prediction model based on a Chinese population[J]. Front Cardiovasc Med, 2022, 9: 946063.
- [9] Tusongtuoheti X, Huang G Q, Mao Y S. Development and internal validation of a risk prediction model for carotid atherosclerosis in the hyperuricemia population[J]. Vasc Health Risk Manag, 2024, 20: 195-205.
- [10] Wang X J, Li W Z, Song F J, et al. Carotid atherosclerosis detec-

- ted by ultrasonography: a national cross-sectional study[J]. *J Am Heart Assoc*, 2018, 7(8): e008701.
- [11] Li W, Xie B, Qiu S H, et al. Non-lab and semi-lab algorithms for screening undiagnosed diabetes: a cross-sectional study[J]. *EBio-Medicine*, 2018, 35: 307-316.
- [12] Huang W, Wang H, Shen Z M, et al. Association between TyG index and risk of carotid atherosclerosis in NAFLD patients: a retrospective cohort study[J]. *Front Endocrinol (Lausanne)*, 2024, 15: 1448359.
- [13] Dai G K, Cai X S, Ye C J, et al. A cross-sectional study of factors associated with carotid atherosclerosis[J]. *Front Physiol*, 2024, 15: 1434173.
- [14] Lin Q X, Liu J, Hu P, et al. Incidence and predictors of progression of carotid atherosclerosis in a low-income Chinese population: a prospective cohort study[J]. *Postgrad Med*, 2022, 134(1): 85-95.
- [15] Damen J A, Pajouheshnia R, Heus P, et al. Performance of the framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis[J]. *BMC Med*, 2019, 17(1): 109.
- [16] Saba L, Cau R, Murgia A, et al. Carotid plaque-RADS: a novel stroke risk classification system[J]. *JACC: Cardiovasc Imaging*, 2024, 17(1): 62-75.
- [17] Bos D, Arshi B, van den Bouwhuijsen Q J A, et al. Atherosclerotic carotid plaque composition and incident stroke and coronary events [J]. *J Am Coll Cardiol*, 2021, 77(11): 1426-1435.
- [18] Fu J Z, Deng Y H, Ma Y, et al. Adherence to a healthy diet and risk of multiple carotid atherosclerosis subtypes: insights from the China MJ health check-up cohort [J]. *Nutrients*, 2024, 16(14): 2338.
- [19] Chen H J, Chen J P, Zhang F P, et al. Effective management of atherosclerosis progress and hyperlipidemia with nattokinase: a clinical study with 1 062 participants[J]. *Front Cardiovasc Med*, 2022, 9: 964977.
- [20] Jimenez-Torres J, Alcalá-Díaz J F, Torres-Peña J D, et al. Mediterranean diet reduces atherosclerosis progression in coronary heart disease: an analysis of the CORDIOPREV randomized controlled trial[J]. *Stroke*, 2021, 52(11): 3440-3449.
- [21] Ji P, Kou Q X, Zhang J W. Study on relationship between carotid intima-media thickness and inflammatory factors in obstructive sleep apnea[J]. *Nat Sci Sleep*, 2022, 14: 2179-2187.
- [22] Fu Q, Wang X J, Wu T L, et al. Carotid atherosclerosis biomarkers in cardiovascular diseases prevention: a systematic review and bibliometric analysis[J]. *Eur J Radiol*, 2020, 129: 109133.
- [23] Witarto B S, Visuddho V, Aldian F M, et al. Blood-based circulating microRNAs as diagnostic biomarkers for subclinical carotid atherosclerosis: a systematic review and meta-analysis with bioinformatics analysis[J]. *Diabetes Metab Syndr*, 2023, 17(10): 102860.
- [24] Siwamogsatham S, Chutinet A, Vongsayan P, et al. Low CD4 cell counts are associated with carotid plaque and intima-media thickness in virologically suppressed HIV-infected Asians older than 50 years [J]. *AIDS Res Hum Retroviruses*, 2019, 35(11/12): 1160-1169.
- [25] Chen F H, Liu T, Xu Li, et al. Association of serum vitamin D level and carotid atherosclerosis: a systematic review and meta-analysis[J]. *J Ultrasound Med*, 2018, 37(6): 1293-1303.
- [26] Huang Y Q, Li J, Huang C, et al. Plasma microRNA-29c levels are associated with carotid intima-media thickness and is a potential biomarker for the early detection of atherosclerosis [J]. *Cell Physiol Biochem*, 2018, 50(2): 452-459.
- [27] Liu J, Sze N S K, Narushima M, et al. Weight change and risk of atherosclerosis measured by carotid intima-media thickness (cIMT) from a prospective cohort-analysis of the first-wave follow-up data of the Canadian longitudinal study on aging (CLSA)[J]. *J Cardiovasc Dev Dis*, 2023, 10(10): 435.
- [28] Kamon T, Kaneko H, Itoh H, et al. Association between waist circumference and carotid intima-media thickness in the general population[J]. *Int Heart J*, 2020, 61(1): 103-108.
- [29] Takase M, Nakaya N, Nakamura T, et al. Carotid intima media thickness and risk factor for atherosclerosis: tohoku medical megabank community-based cohort study [J]. *J Atheroscler Thromb*, 2023, 30(10): 1471-1482.
- [30] Caimi G, Lo Presti R, Urso C, et al. Neutrophil/HDL-C, lymphocyte/HDL-C and monocyte/HDL-C in subjects with asymptomatic carotid atherosclerosis[J]. *Clin Hemorheol Microcirc*, 2024, 88(1): 1-11.
- [31] Vu H, Khanh Tuong T T, Hoang Lan N, et al. Association between nonalcoholic fatty liver disease and carotid intima-media thickness[J]. *Clin Ter*, 2023, 174(1): 42-47.
- [32] Liu Y H, Zhu Y J, Jia W R, et al. Association of the total white blood cell, neutrophils, and monocytes count with the presence, severity, and types of carotid atherosclerotic plaque [J]. *Front Med*, 2020, 7: 313.
- [33] Zhao J J, Zhang W X, Wei H Y, et al. Neutrophil count and reduced risk of venous thromboembolism: a Mendelian randomization study[J]. *Hematology*, 2024, 29(1): 2428481.
- [34] Totoń-Żurańska J, Mikolajczyk T P, Saju B, et al. Vascular remodelling in cardiovascular diseases: hypertension, oxidation, and inflammation[J]. *Clin Sci (Lond)*, 2024, 138(13): 817-850.
- [35] Suárez-Cuenca J A, Montoya-Ramírez J, Gutiérrez-Buendía J A, et al. Pulse-wave velocity, flow-mediated dilation, and carotid intima-media thickness to assess cardiovascular risk in population with metabolic syndrome[J]. *J Vis Exp*, 2024. DOI: 10.3791/65842.
- [36] Hsiao C L, Chen P Y, Hsu P J, et al. Nomogram and carotid risk score for predicting moderate or high carotid atherosclerosis among asymptomatic elderly recycling volunteers [J]. *Diagnostics (Basel)*, 2022, 12(6): 1407.
- [37] McEvoy J W, Nasir K, DeFilippis A P, et al. Relationship of cigarette smoking with inflammation and subclinical vascular disease: the multi-ethnic study of atherosclerosis[J]. *Arterioscler Thromb Vasc Biol*, 2015, 35(4): 1002-1010.
- [38] Karbassiayzdi E, Fattahi F, Yousefi N, et al. XGBoost model as an efficient machine learning approach for PFAS removal: effects of material characteristics and operation conditions[J]. *Environ Res*, 2022, 215(Part 1): 114286.