

·基层卫生服务·

社区原发性高血压病患者肾损伤的风险评估模型构建

邱思月¹, 郭海生¹, 郑义鹏¹, 唐品升², 王磊³

1.北京中医药大学深圳医院(龙岗),广东 深圳 518000;2.江门市五邑中医院;3.广州中医药大学东莞医院,广东 东莞 523000

摘要:目的 建立并验证社区原发性高血压病患者肾损伤风险的风险评估模型。方法 回顾性纳入 2025 年 1—12 月北京中医药大学深圳医院(龙岗)四家社区卫生服务中心共计 2 338 例原发性高血压病患者,按 7 : 3 随机分为训练集(1 637 例)和验证集(701 例)。收集相关临床资料。采用单因素分析、Lasso 及多因素回归分析进行变量筛选后,采用七种机器学习方法构建风险评估模型。使用验证集数据计算准确率、受试者工作特征曲线下面积(area under the curve, AUC)等指标评估模型效能。使用沙普利加性解释(SHapley additive exPlanations,SHAP)值评估单个变量的重要性。结果 单因素分析显示,年龄、性别等 28 个变量与肾损伤相关($P<0.05$)。Lasso 回归筛选后保留七个特征变量,多因素 logistic 回归验证显示,男性($OR=0.168, 95\%CI: 0.113 \sim 0.250$)、年龄($OR=1.155, 95\%CI: 1.133 \sim 1.178$)、饮酒史($OR=1.604, 95\%CI: 1.415 \sim 2.879$)、尿隐血($OR=1.923, 95\%CI: 1.230 \sim 3.010$)、红细胞($OR=0.702, 95\%CI: 0.508 \sim 0.968$)、尿酸($OR=1.01, 95\%CI: 1.008 \sim 1.012$)为肾损伤的独立危险因素($P<0.05$)。所有模型 AUC 值均高于 0.80,其中支持向量机(support vector machine,SVM)模型效能较好($AUC=0.85$),各性能指标均衡。SHAP 值分析显示年龄、性别、尿酸是影响模型预测的关键特征,各变量间存在非线性交互效应。结论 男性、年龄、饮酒史、尿隐血、红细胞、尿酸是社区原发性高血压病患者肾损伤的独立危险因素,以上变量构建的 SVM 模型具有较好的预测效能和临床应用价值。

关键词:原发性高血压;肾损伤;机器学习;风险评估模型

中图分类号:R544.1 文献标志码:A 文章编号:1003-8507(2026)13-2409-08

DOI:10.20043/j.cnki.MPM.202602119

Development of a risk assessment model for kidney damage in community-dwelling patients with essential hypertension

QIU Si-yue*, GUO Hai-sheng, ZHENG Yi-peng, TANG Ping-sheng, WANG Lei

*Shenzhen Hospital of Beijing University of Chinese Medicine (Longgang), Shenzhen, Guangdong 518000, China

Abstract: **Objective** To develop and validate a risk assessment model for kidney damage in community-dwelling patients with essential hypertension. **Methods** A total of 2 338 patients with essential hypertension from four community health service centers affiliated with Beijing University of Chinese Medicine Shenzhen Hospital (Longgang) were retrospectively enrolled between January and December 2025. They were randomly divided into a training set ($n=1\ 637$) and a validation set ($n=701$) in a 7:3 ratio. Relevant clinical data were collected. After variable screening using univariate analysis, LASSO regression, and multivariable regression analysis, risk assessment models were constructed using seven machine learning algorithms. Model performance was evaluated on the validation set using metrics including accuracy and the area under the receiver operating characteristic curve (AUC). The importance of individual variables was assessed using SHapley additive exPlanations (SHAP) values. **Results** Univariate analysis revealed that 28 variables, including age and sex, were associated with kidney damage ($P<0.05$). Lasso regression screened and retained seven feature variables. Multivariable logistic regression confirmed that male sex ($OR=0.168, 95\%CI: 0.113-0.250$), age ($OR=1.155, 95\%CI: 1.133-1.178$), history of alcohol consumption ($OR=1.604, 95\%CI: 1.415-2.879$), urine occult blood ($OR=1.923, 95\%CI: 1.230-3.010$), red blood cell count ($OR=0.702, 95\%CI: 0.508-0.968$), and uric acid ($OR=1.01, 95\%CI: 1.008-1.012$) were independent risk factors for kidney damage ($P<0.05$). The AUC values for all models exceeded 0.80. The support vector machine (SVM) demonstrated relatively better performance ($AUC=0.85$), with balanced performance across various metrics. SHAP value analysis indicated that age, sex, and uric acid level were key features influencing model predictions, and nonlinear interaction effects were observed among the variables. **Conclusion** Male sex, age, history of alcohol consumption, urine occult blood, red blood cell count, and uric acid are independent risk factors for kidney damage in community-dwelling patients with essential hypertension. The

基金项目:深圳市龙岗区医疗卫生科技计划项目(LGWJ2022-83);北京中医药大学深圳医院(龙岗)育龙人才计划(2023-BUCMSZYLRC15)

作者简介:邱思月(1996—),女,博士在读,主治医师,研究方向:中西医结合防治心血管疾病

通信作者:王磊,E-mail:dr.wanglei@gzucm.edu.cn

SVM model constructed using these variables exhibits good predictive performance and holds potential clinical application value.

Keywords: Essential hypertension; Kidney damage; Machine learning; Risk assessment model

原发性高血压(primary hypertension,PH)是我国心血管疾病的首要危险因素,且患病率随年龄增长逐年升高,是多种心脑血管及肾脏疾病最主要、可干预的危险因素^[1]。由于高血压肾损伤起病隐匿,早期临床症状与体征常不典型,一旦进入临床蛋白尿期,肾功能损害往往已难以逆转,这给患者和社会带来了沉重的疾病负担^[2]。因此,在社区层面早期识别并干预高血压肾损伤的高危人群,对于延缓疾病进展、改善患者预后具有重要意义。

近年来,以机器学习等人工智能技术凭借其强大的数据挖掘和模式识别能力,为构建精准的疾病风险评估模型提供了新的可能^[3-4]。已有研究证实,机器学习算法在心血管事件及慢性肾病进展预测中展现出良好性能^[5],但目前尚未有研究运用机器学习对社区 PH 患者进行肾损伤风险评估模型的开发与验证。因此,本研究拟利用社区 PH 患者的多维度临床信息,采用机器学习算法构建并验证有效评估肾损伤的风险模型。期望能为基层医疗机构提供一种简便、有效的风险评估工具,从而实现对高危个体的识别和干预。

1 对象与方法

1.1 对象 收集 2025 年 1—12 月于北京中医药大学深圳医院(龙岗)四家社区卫生服务中心诊断为 PH 的患者 2 338 例,其中男性 1 232 例,女性 1 106 例,年龄(65 ± 9.27)岁,按 7:3 随机分为训练集(1 637 例)和验证集(701 例)。本研究原始数据收集集中均隐藏患者身份识别信息,并豁免患者知情同意。本研究获北京中医药大学深圳医院(龙岗)医学伦理委员会批准(KY-2025-374)。

纳入标准:(1)符合 PH 诊断标准^[6];(2)年龄为 18~80 岁。**排除标准:**(1)继发性高血压病患者;(2)既往存在肾损伤或慢性肾脏病病史;(3)合并严重基础病,如恶性肿瘤、自身免疫性疾病等影响血液指标的疾病;(4)病历资料不完整或缺失值>20%,且无法通过合理方法补全。

1.2 肾功能不全的诊断标准 依据慢性肾脏病分期标准^[7]进行分组。据患者的估算肾小球滤过率(estimated glomerular filtration rate,eGFR)水平分为二组:eGFR<90 ml/(min·1.73 m²)为肾损伤组,eGFR≥90 ml/(min·1.73 m²)为正常组。

1.3 数据收集 在数据收集过程中,参考相关的医学研究规范和模型,以确保收集的数据能够全面反映患者的疾病状态和治疗情况。数据包括:(1)人口学

信息及体格检查:性别、年龄、静息心率、身高、体重、腰围、身体质量指数(body mass index,BMI)、收缩压、舒张压;(2)既往史:冠心病史、糖尿病史、脑血管病史、吸烟史、饮酒史、锻炼史;(3)用药情况:血管紧张素转换酶抑制剂/血管紧张素 II 受体拮抗剂(angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker,ACEI/ARB)、降糖药、抗血小板药(antiplatelet drugs,ADPs)、他汀类药物(atatins)、钙通道阻滞剂(calcium channel blocker,CCB)、β 受体阻滞剂(beta-blockers,β-blockers);(4)实验室检查结果:白细胞总数(white blood cell count,WBC)、血红蛋白、血小板、红细胞总数(red blood cell count,RBC)、中性粒细胞比例(neutrophil ratio,NEUT%)、淋巴细胞比例(lymphocyte ratio,LYMPH%)、单核细胞比例(monocyte ratio,MONO%)、嗜酸性粒细胞比例(eosinophil ratio,EO%)、血糖、丙氨酸氨基转氨酶(alanine aminotransferase,ALT)、血尿酸、总胆固醇、甘油三酯、低密度脂蛋白胆固醇(low-density lipoprotein cholesterol,LDL-C)、高密度脂蛋白胆固醇(high-density lipoprotein cholesterol,HDL-C);尿液学指标包括蛋白尿、尿糖、尿酮、尿隐血、尿白细胞、尿微量白蛋白肌酐比(urinary microalbumin to creatinine ratio,UACR)。对于数据中的缺失值,采用 R 语言中的多重插补方法,确保数据集的完整性和准确性。

1.4 机器学习模型开发 采用 Lasso 回归进行变量选择,分别采用逻辑回归、决策树、随机森林、极端梯度提升树(extreme gradient boosting,XGBoost)、轻量级梯度提升机(light gradient boosting machine,LightGBM)、支持向量机(support vector machine,SVM)、人工神经网络(artificial neural network,ANN)七种机器学习算法构建 PH 患者肾损伤的风险评估模型。在数据预处理阶段对原始数据进行清洗确保数据的质量。训练集数据进行了十折交叉验证获得最优训练模型,减少过拟合风险提高模型的准确性及可靠性。考虑到数据的不平衡问题采用了合成少数类过采样技术(synthetic minority over-sampling technique,SMOTE)过采样技术帮助平衡样本分布并增强模型性能。

为评估模型的预测效能,使用验证集数据集计算各模型的受试者工作特征(receiver operating characteristic,ROC)曲线下面积(Area under the curve,AUC)、准确度、精确度、灵敏度、特异度、F1 值、决策曲线、校准曲线。使用沙普利加性解释(SHapley additive exPlanations,SHAP)值分析模型中每个变量的

重要性及影响。

1.5 统计学分析 采用 SPSS 26.0、R 4.5.2、Python 3.13.5 软件进行数据统计分析与模型构建。不符合正态分布的计量资料以 $[M(P_{25}, P_{75})]$ 表示,多组间比较采用 Kruskal-Wallis H 检验。计数资料以频数(百分比)表示,组间比较采用 χ^2 检验;当理论频数 <5 时,采用 Fisher 确切概率法。所有统计检验均为双侧检验,检验水准 $\alpha=0.05$,多重比较采用 Bonferroni 校正。采用 R 4.5.2 及 Python 3.13.5 进行模型构建绘图及效能、临床应用价值评价。

2 结 果

2.1 研究对象基本资料 本研究共纳入 2 338 人,训练集及验证集两组人群基本特征可比,差异无统计学意义($P>0.05$)。肾功能正常组患者 500 名,肾功能不全组患者 1 838 名,两组间年龄、性别、脑血管病史、冠心病史、糖尿病史、降糖药、锻炼史、吸烟史、饮酒史、收缩压、身高、体重、腰围、BMI、血红蛋白、RBC、LYMPH%、尿糖、ALT、血尿酸、总胆固醇、LDL-C、HDL-C、尿酮、尿隐血、尿白细胞、UACR 差异有统计学意义($P<0.05$)。见表 1。

表 1 研究对象基本特征 $[M(P_{25}, P_{75}), n(\%)]$

Table 1 Basic characteristics of study participants $[M(P_{25}, P_{75}), n(\%)]$

变量	肾功能正常患者($n=500$)	肾功能不全患者($n=1\ 838$)	Z/χ^2 值	P 值
年龄(岁)	60.00(50.00,66.00)	68.00(65.00,72.00)	19.701	<0.001
性别			231.249	<0.001
女	86(17.20)	1 020(55.50)		
男	414(82.80)	818(44.50)		
脑血管病史	5(1.03)	59(3.29)	7.115	0.008
冠心病史	7(1.40)	55(2.99)	3.861	0.049
糖尿病史	61(12.20)	503(27.37)	49.398	<0.001
用药情况				
ACEI/ARB	112(22.40)	341(18.55)	3.724	0.054
ADD	81(16.20)	219(11.92)	6.452	0.011
ADPs	31(6.20)	132(7.18)	0.584	0.445
Statins	45(9.00)	195(10.61)	1.105	0.293
CCB	128(25.60)	468(25.46)	0.004	0.950
β -blockers	43(8.60)	125(6.80)	1.908	0.167
锻炼史	349(71.81)	1416(78.84)	10.791	0.001
吸烟史	152(31.28)	319(17.77)	42.566	<0.001
饮酒史	113(23.25)	260(14.48)	21.487	<0.001
静息心率(次/分)	70.00(64.00,76.00)	69.00(63.00,77.00)	0.799	0.424
收缩压(mm Hg)	132.00(122.00,143.00)	139.50(128.00,152.00)	7.864	<0.001
舒张压(mm Hg)	87.00(80.00,92.00)	86.00(78.00,92.00)	1.702	0.089
身高(cm)	166.50(161.50,171.00)	160.00(154.00,167.00)	13.006	<0.001
体重(kg)	70.00(62.50,77.25)	63.90(57.00,71.70)	10.452	<0.001
腰围(cm)	90.00(84.00,96.00)	87.50(82.00,93.00)	5.834	<0.001
BMI(kg/m^2)	25.30(23.32,27.68)	24.90(22.92,26.77)	3.360	<0.001
HB(g/L)	151.00(141.00,159.00)	141.00(132.00,151.00)	11.968	<0.001
WBC($\times 10^9/\text{L}$)	6.36(5.30,7.44)	6.33(5.44,7.52)	0.426	0.670
PLT($\times 10^9/\text{L}$)	235.00(201.00,279.00)	231.00(196.00,272.25)	1.616	0.106
RBC($\times 10^{12}/\text{L}$)	5.04(4.75,5.33)	4.77(4.47,5.08)	10.422	<0.001
NEUT%	3.68(2.93,4.60)	3.67(2.96,4.45)	0.271	0.787
LYMPH%	1.98(1.65,2.34)	2.06(1.67,2.50)	2.192	0.028
MONO%	0.38(0.30,0.47)	0.38(0.30,0.46)	0.704	0.482
EO%	0.14(0.08,0.20)	0.14(0.08,0.21)	1.158	0.247
尿糖(mmol/L)	6.20(5.50,7.50)	5.90(5.40,7.00)	3.084	0.002
ALT(U/L)	23.00(17.00,32.00)	19.00(15.00,26.00)	7.269	<0.001
UA($\mu\text{mol}/\text{L}$)	314.00(264.00,363.00)	325.00(276.00,392.00)	3.405	<0.001
TC(mmol/L)	4.70(3.96,5.37)	4.92(4.14,5.65)	4.227	<0.001
TG(mmol/L)	1.33(0.95,2.08)	1.42(1.02,1.95)	0.776	0.438
LDL-C(mmol/L)	3.17(2.48,3.71)	3.24(2.52,3.89)	2.563	0.010
HDL-C(mmol/L)	1.18(1.01,1.38)	1.26(1.08,1.48)	5.409	<0.001
UACR(mg/g)	10.45(5.81,21.83)	14.82(8.08,31.73)	5.492	<0.001
PRO	51(10.22)	235(12.82)	2.464	0.116
尿糖	104(20.84)	198(10.80)	35.070	<0.001
尿酮	23(4.61)	38(2.07)	9.903	0.002
尿隐血	54(10.82)	447(24.39)	42.784	<0.001
尿白细胞	59(11.82)	505(27.55)	52.907	<0.001

2.2 社区 PH 患者肾功能不全风险评估模型变量选择 将上述单因素分析中 $P<0.05$ 的 28 个变量纳入 Lasso 回归分析,选择正则化参数 λ ,选择与结局相关的变量七个,包括性别为男性、年龄、饮酒史、尿糖、尿隐血、RBC、血尿酸,当 $\lambda=0.0113$ 时,所对应的非 0 系

数分别为 -1.2905 、 6.6798 、 -0.0532 、 -0.0461 、 0.1009 、 -0.2552 、 3.2072 。见图 1、2。将 Lasso 回归筛选出的七个变量进行多因素回归分析,结果表明,男性、年龄、饮酒史、尿隐血、RBC、血尿酸是社区 PH 发生肾功能不全的独立危险因素, $P<0.05$ 。见表 2。

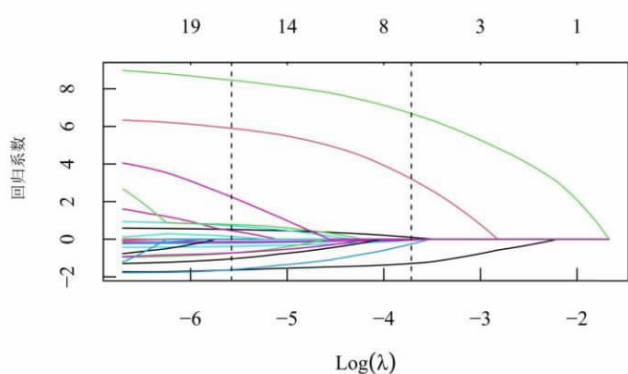
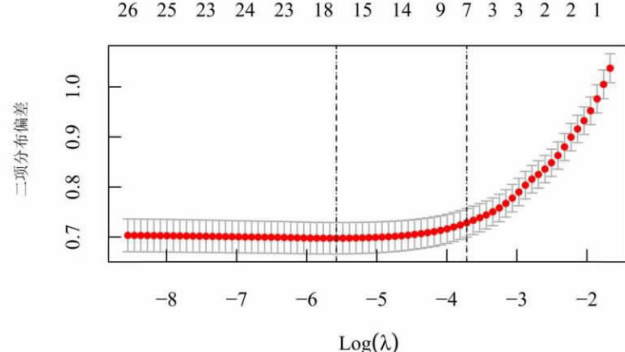


图 1 基于 Lasso 回归分析的社区 PH 患者肾功能不全发生风险变量选择路径图

Figure 1 Path diagram of variable selection for the risk of renal insufficiency in community patients with PH based on Lasso regression analysis



注:左侧虚线对应最小偏差的 λ 值($\lambda_{\min}$);右侧虚线代表一个标准误差范围内的最大 λ 值(λ_{1se});本研究选择左侧虚线的 λ 值($\lambda=0.0113$)。

图 2 Lasso 回归分析筛选影响肾功能不全发生的关键特征的交叉验证结果

Figure 2 Cross-validation results of Lasso regression for identifying key features associated with renal insufficiency

表 2 社区 PH 患者合并肾功能不全的多因素 logistic 回归分析

Table 2 Multivariable logistic regression analysis of renal insufficiency in community patients with PH

变量	β	$\bar{x}$	Wald	OR(95%CI)	P 值
男性	-1.784	0.202	-8.818	0.168(0.113 ~ 0.250)	<0.001
年龄	0.144	0.010	14.514	1.155(1.133 ~ 1.178)	<0.001
饮酒史	0.504	0.191	2.636	1.604(1.415,2.879)	0.008
尿糖	-0.224	0.206	-1.087	0.799(0.543,1.197)	0.277
尿隐血	0.654	0.228	2.868	1.923(1.230,3.010)	0.004
RBC	-0.354	0.164	-2.155	0.702(0.508,0.968)	0.031
血尿酸	0.010	0.001	9.103	1.010(1.008,1.012)	<0.001

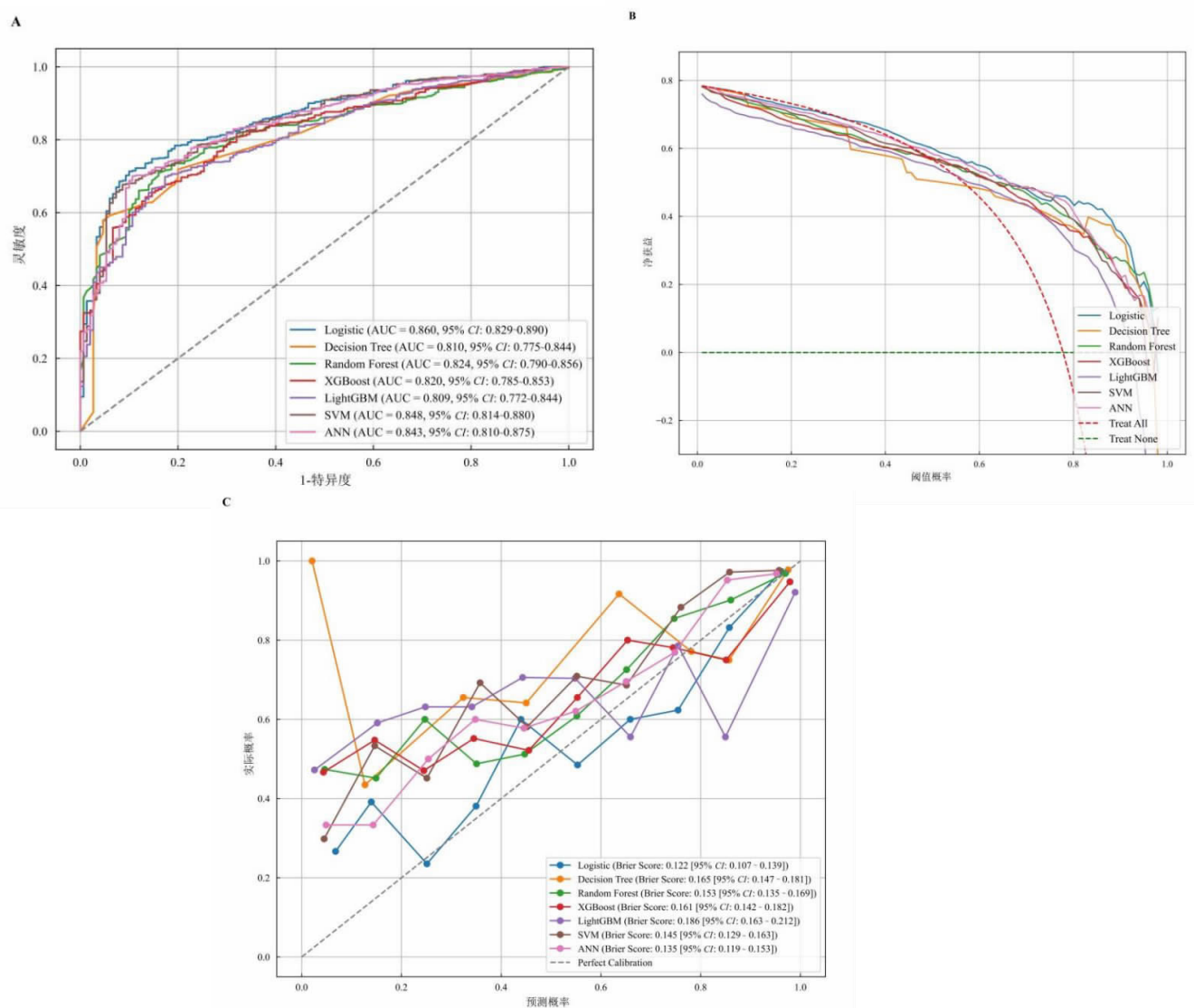
2.3 不同机器学习模型的预测效能评估 基于训练集构建的所有模型在验证集上评估预测效能,七种机器学习模型在验证集上均表现出良好的预测能力,整体区分度良好,见图 3A。SVM 模型在各项指标中表现均衡,且 F1 值处于较高水平。校准曲线分析显示

各模型预测概率与实际发生概率具有良好的一致性,见图 3C;校准曲线显示模型在较宽的阈值概率范围内具有正向的临床净获益,见图 3B,具备临床应用价值。各模型的性能指标见表 3。

表 3 不同机器学习模型对社区 PH 患者发生肾功能不全的预测效能

Table 3 Predictive performance of different machine learning models for renal insufficiency in community patients with PH

模型	AUC	准确率	精确度	灵敏度	特异度	F1 值
逻辑回归	0.86	0.82	0.85	0.93	0.39	0.89
决策树	0.81	0.72	0.93	0.70	0.80	0.80
随机森林	0.82	0.78	0.89	0.82	0.63	0.86
XGBoost	0.82	0.78	0.89	0.83	0.62	0.86
LightGBM	0.81	0.76	0.87	0.82	0.57	0.85
SVM	0.85	0.79	0.90	0.82	0.67	0.86
ANN	0.84	0.80	0.88	0.87	0.57	0.87



注:A 为基于机器学习模型的验证集 ROC;B 为决策曲线分析;C 为校准曲线,虚线为理想校准线。

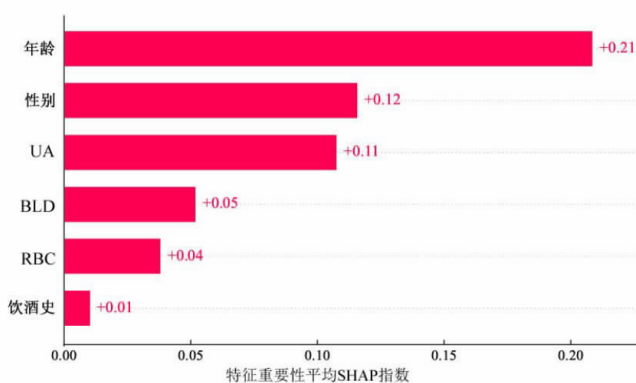
图 3 不同机器学习模型对社区 PH 患者发生肾功能不全预测效能的评估曲线

Figure 3 Evaluation curves of predictive performance of different machine learning models for renal insufficiency in community patients with PH

SHAP 分析结果显示,在缺少肌酐指标的条件下,年龄、性别、血尿酸对模型预测的重要性高于尿糖、尿隐血及 RBC,见图 4。蜂群图与散点图显示,女性、饮酒史、尿隐血阳性、年龄增大、RBC 降低、血尿酸升高对模型输出具有正向影响,见图 5。

单样本特征贡献解释表明,模型中各特征呈现明显的非线性交互效应,同一特征在不同个体中的影响方向与强度存在显著差异。在阴性预测样本中,血尿酸 =225 表现为强正向贡献($\text{SHAP}=+0.16$),而女性 ($\text{male}=0, \text{SHAP}=-0.14$)与高龄 ($\text{age}=70, \text{SHAP}=-0.14$)的负向贡献占主导,共同将预测概率由基准值 0.339 降至 0.236,提示年龄与性别是该样本实现阴性预测的关键因素,见图 6A。在阳性预测样本中,高龄、女性、血尿(urinary_blood=1)为核心正向驱动特征,见图 6B ~ D;在强阳性预测样本中,三者协同作用可将预

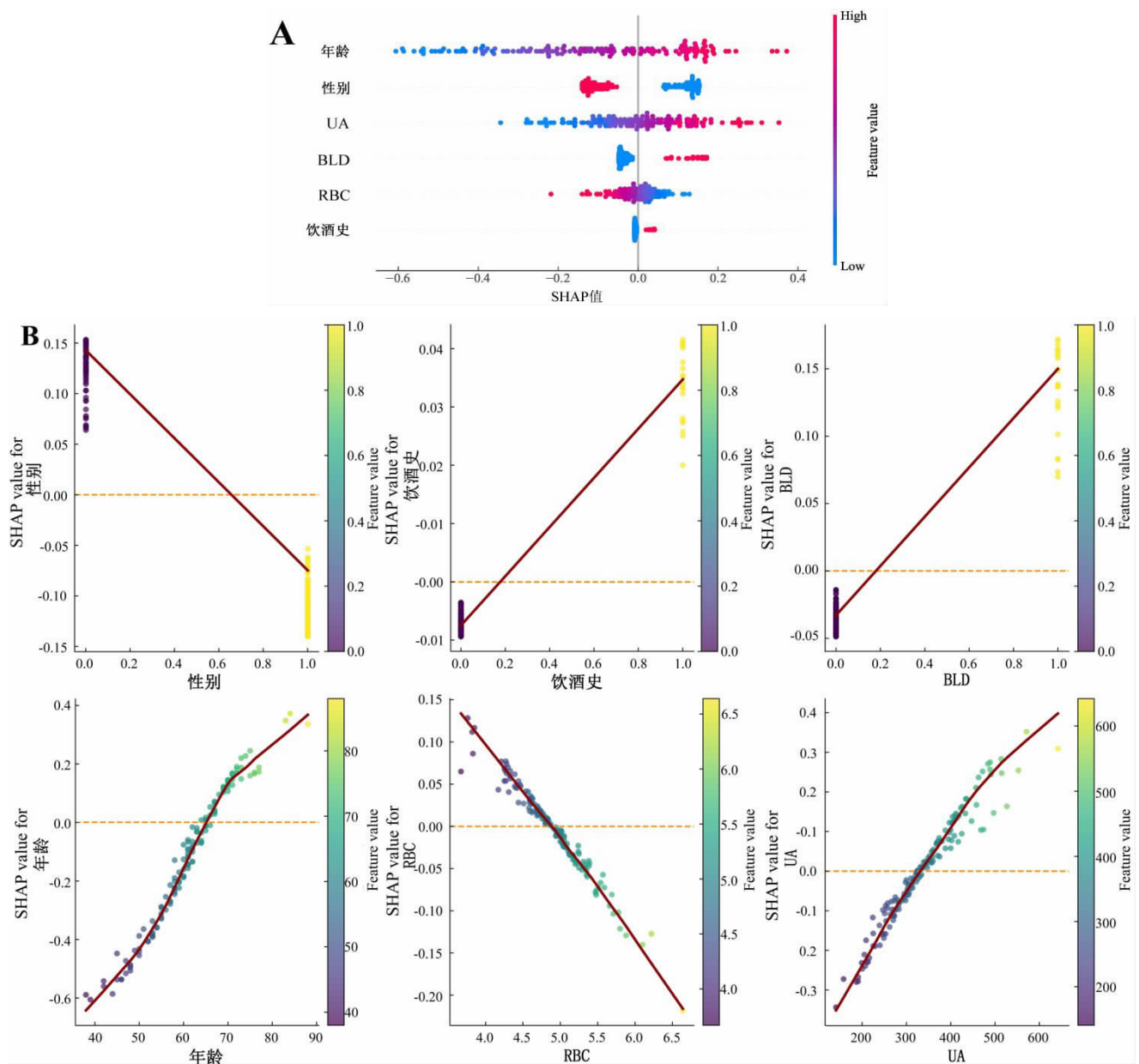
测概率由 0.660 9 提升至 0.97,显示该特征组合对阳性结局具有强判别能力,见图 6D。



注:其值表示每个特征对最终预测结果的影响程度,高 SHAP 值表明该特征对模型的预测影响较大。

图 4 SHAP 特征重要性条形图

Figure 4 Bar plot of SHAP feature importance



注:A 蜂群图显示每个变量的 SHAP 值分布,颜色编码为蓝色表示低特征值、红色为高特征值;B 散点图显示每个特征如何影响模型在不同值范围的预测输出。

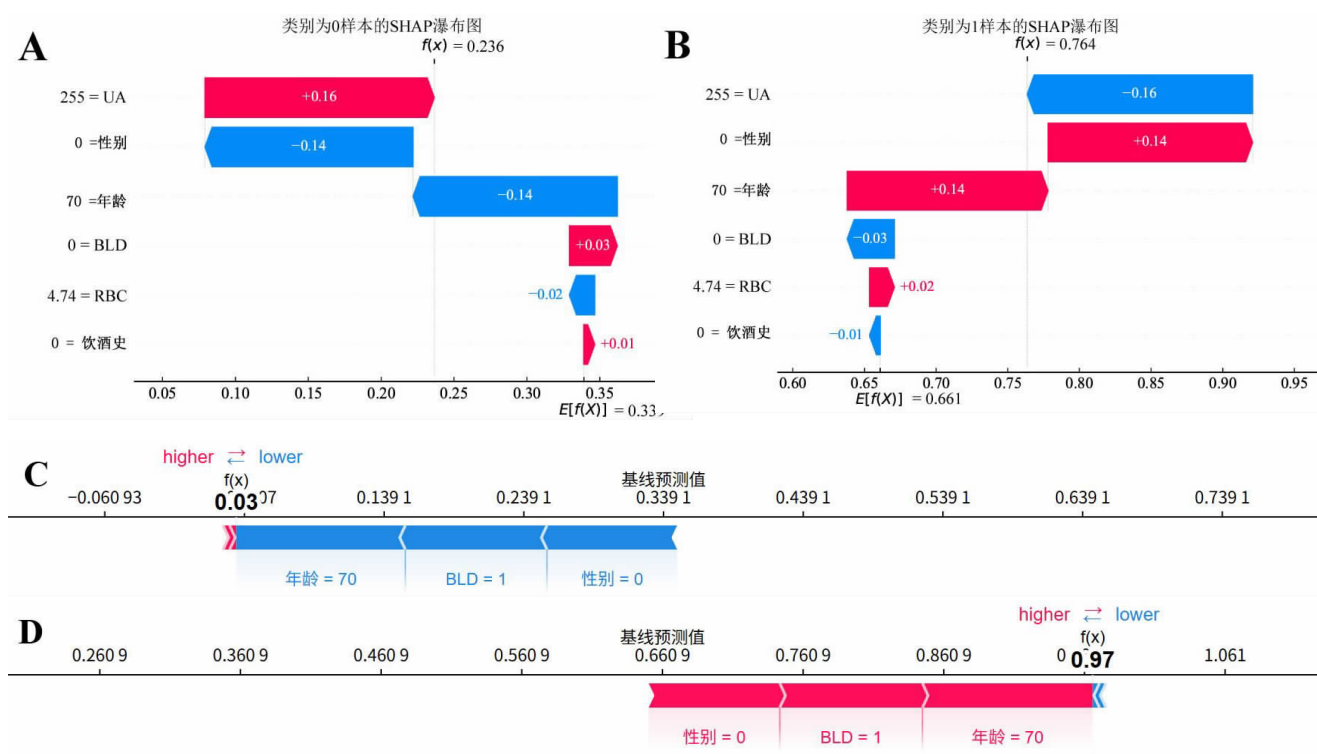
图 5 特征贡献度 SHAP 蜂群图与散点图

Figure 5 SHAP beeswarm plot and scatter plot of feature contribution

3 讨论

高血压肾损伤是 PH 最常见的靶器官损害之一,其起病、进展隐匿,社区层面的早期精准筛查尤为关键。传统风险评估模型多依赖单一逻辑回归算法,受限于线性假设,难以处理临床数据中高维度、非线性及变量间复杂交互关系的特点,预测精准度有限。本研究回顾性纳入 2 338 例社区 PH 患者,采用七种机器学习算法建模,所有模型 AUC 值均高于 0.80, SHAP 值明确了各变量的预测重要性及交互效应,为社区基层开展高血压肾损伤早期筛查与干预提供了可靠的工具与理论依据,提示机器学习算法在社区高血压肾损伤预测中具有显著优势^[8-9]。

本研究筛选出男性、年龄、饮酒史、尿隐血、RBC、血尿酸为社区 PH 患者肾损伤的独立危险因素,这一结果与现有临床研究规律基本契合^[10-11]。年龄在模型中表现出稳定的预测作用 ($OR=1.155, P<0.001$),其与肾损伤的正向关联源于年龄增长导致的动脉硬化加重、肾小球滤过功能自然减退,而高血压进一步加速此进程。血尿酸 ($OR=1.01, P<0.001$) 是模型中重要的生化预测变量,其通过尿酸结晶沉积、诱发炎症反应及氧化应激等机制损伤肾组织,与临床中高尿酸血症作为高血压肾损伤危险因素的认知一致,证实了血尿酸作为肾损伤预测指标的可靠性^[12]。男性作为保护性因素 ($OR=0.168, P<0.001$),其肾损伤风险



注:图 A、B 为瀑布图,瀑布图中蓝色表示对本样本预测结果的负贡献、红色表示正贡献;图 C、D 为力图,力图说明每个特征对最终预测值 $f(x)$ 做出的贡献,从模型的基值开始,蓝色箭头表示降低预测值的特征,红色箭头表示提高预测值的特征。

图 6 SHAP 对典型单个样本的解释

Figure 6 SHAP explanation for typical individual samples

显著低于女性,推测与女性绝经后雌激素保护作用减退相关,该结果虽与部分既往研究存在差异,但可能与本研究纳入的社区女性患者年龄偏高、合并症较多的样本特征相关,其具体机制仍需更大样本、多中心研究进一步验证^[13]。尿隐血患者肾损伤发生风险是阴性患者的 1.923 倍,是模型中灵敏度较高的预测变量,长期微量出血可诱发肾间质炎症,破坏肾组织结构。RBC ($OR=0.702, 95\%CI:0.508 \sim 0.968$) 是模型中重要的血液学预测变量,其变化通过影响肾组织氧供间接调控肾损伤风险,丰富了模型的风险评估维度^[14]。饮酒史 ($OR=1.604, 95\%CI:1.415 \sim 2.879$) 是模型中唯一可干预的行为学预测变量,通过酒精代谢产物的直接肾毒性及血压、血脂代谢异常等间接作用加速肾损伤进展,这一发现为社区高血压患者的健康管理提供了明确的干预方向。

在模型效能方面,逻辑回归模型准确率较高 (0.82) 但特异度偏低 (0.39),提示其在识别阴性样本时存在局限;决策树模型精确度最高 (0.93) 但准确率最低 (0.72),易出现过拟合现象;SVM 模型 AUC 值达 0.85,准确率、精确度、灵敏度、特异度及 F1 值均处于较高水平且更为均衡,是本研究中最优的风险评估模型,其中 SVM 模型因泛化能力强、抗噪声优,在小样本高维度的社区场景中表现优异,各项性能指标更均

衡。此外,本研究采用 SMOTE 过采样技术解决样本不平衡问题,通过十折交叉验证优化模型参数,有效降低了过拟合风险,提高了模型的稳定性与泛化能力,确保模型在社区基层推广应用的可靠性^[15]。

本研究采用 SHAP 可解释性分析对机器学习模型进行解析,旨在突破传统模型黑箱局限,明确各特征对预测结果的贡献方向、作用强度及临床意义。SHAP 分析显示,年龄、性别、尿酸为模型最核心预测特征,其贡献度显著高于其他变量,提示在缺少血肌酐等传统肾功能指标时,上述三项特征仍可实现稳定可靠的风险判断,提升了模型在基层社区筛查场景下的实用性与可及性。同时,SHAP 分析揭示各变量对肾损伤风险的影响并非简单线性关系,而是存在复杂的非线性交互效应,其中高龄、女性、尿隐血阳性三者协同作用可显著升高肾损伤发生概率,这一发现提示基层医务人员在临床工作中,不应仅关注单一指标异常,而应重视多因素联合评估,这与既往多中心研究结论一致^[16]。该结果为基层医务工作者提供了清晰可操作的风险识别策略,对具备上述特征的高血压人群强化监测与早期干预,有助于实现肾损伤的精准预防与前移管理。

本研究存在一定的局限性,单中心回顾性研究使得样本代表性有限,可能存在选择偏倚,模型泛化能

力需多中心前瞻性研究验证,且模型尚未临床转化,缺乏便捷筛查工具,需进一步推进转化研究。

综上,本研究采用多种机器学习算法构建的社区 PH 患者肾损伤风险评估模型,以 SVM 模型效能最优,筛选出的六项常规可及指标,操作便捷、稳定性强的独立危险因素可有效识别肾损伤高危人群,结合 SHAP 值的可解释性分析的明确了各变量的作用机制与交互效应,解决了传统模型精度不足、可解释性差的问题。该模型可作为社区基层高血压肾损伤早期筛查的有效工具,有助于实现高危人群的精准识别与前瞻性干预,延缓疾病进展,改善患者预后,为社区高血压肾损伤的防治提供新的思路与方法。

利益冲突声明 本研究不存在任何利益冲突

参考文献

- [1] 佚名. 中国心血管健康与疾病报告 2024 概要[J]. 中国循环杂志, 2025, 40(6): 521-559.
Anonym. Summary of Report on Cardiovascular Health and Diseases in China 2024 [J]. Chinese Circulation Journal, 2025, 40 (6): 521-559.(In Chinese)
- [2] 焦艳红. 原发性高血压肾损伤患者血清 CysC、Hey 水平的变化及临床意义研究 [J]. 现代诊断与治疗, 2024, 35 (10): 1518-1520.
Jiao YH. Study on changes and clinical significance of serum CysC and Hey levels in patients with renal injury due to essential hypertension [J]. Modern Diagnosis & Treatment, 2024, 35 (10): 1518-1520.(In Chinese)
- [3] Wu Y, Ma G, Sun H, et al. Prediction of renal function damage in patients with essential hypertension based on stepwise regression equation scanning by AASI[J]. Scanning, 2022, 28: 4728921.
- [4] Sanmarchi F, Fanconi C, Golinelli D, et al. Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review[J]. Journal of Nephrology, 2023, 36(4): 1101-1117.
- [5] Bakris G, Lin PP, Xu C, et al. Prediction of cardiovascular and renal risk among patients with apparent treatment-resistant hypertension in the United States using machine learning methods [J]. Journal of Clinical Hypertension (Greenwich, Conn.), 2024, 26(5): 500-513.
- [6] 《中国高血压防治指南》修订委员会. 中国高血压防治指南 2018 年修订版[J]. 心脑血管病防治, 2019, 19(1): 1-44.
Revision Committee of the Chinese Hypertension Prevention and Treatment Guidelines. Chinese guidelines for the prevention and treatment of hypertension (2018 Revised Edition)[J]. Prevention and Treatment of Cardio-Cerebral-Vascular Disease, 2019, 19(1): 1-44. (In Chinese)
- [7] National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification [J]. American Journal of Kidney Diseases, 2002, 39 (2 Suppl 1): S1-266.
- [8] Ansari GA, Shafi S, Ansari MD, et al. Advanced supervised machine learning methods for precise diabetes mellitus prediction using feature selection[J]. Front Med(Lausanne), 2025, 12: 1620268.
- [9] Wang X, Xu X, Wang Y, et al. Evaluation of the clinical value of 10 estimating glomerular filtration rate equations and construction of a prediction model for kidney damage in adults from central China[J]. Front Mol Biosci, 2024, 11: 1408503.
- [10] 刘佳仪, 蔡广研. 高血压肾损害的诊治[J]. 中国实用内科杂志, 2025, 45(3): 194-198.
Liu JY, Cai GY. Diagnosis and treatment of hypertensive nephropathy [J]. Chinese Journal of Practical Internal Medicine, 2025, 45(3): 194-198.(In Chinese)
- [11] 高血压肾病诊治中国专家共识组成员. 高血压肾病诊断和治疗中国专家共识 (2022)[J]. 中华高血压杂志, 2022, 30(4): 307-317.
Members of the Expert Consensus Group on Diagnosis and Treatment of Hypertensive Nephropathy in China. Chinese expert consensus on diagnosis and treatment of hypertensive nephropathy (2022) [J]. Chinese Journal of Hypertension, 2022, 30(4): 307-317.(In Chinese)
- [12] 王娟, 胡淑霞, 罗代平. 血清 Hcy、UA、BUN 水平与难治性高血压患者早期肾损害的关系[J]. 中国卫生工程学, 2025, 24(6): 794-796, 800.
Wang J, Hu SX, Luo DP. Relationship between serum Hcy, UA, BUN levels and early renal damage in patients with refractory hypertension[J]. Chinese Journal of Health Engineering, 2025, 24(6): 794-796, 800.(In Chinese)
- [13] 庄贤文, 郑颖琨, 陈玲, 等. 绝经后高血压女性患者性激素和肾功能紊乱的相关研究 [J]. 中国实验诊断学, 2026, 30(1): 87-91.
Zhuang XW, Zheng YK, Chen L, et al. A study on the correlation between sexual hormones and renal dysfunction in hypertensive women of different ages [J]. Chinese Journal of Laboratory Diagnosis, 2026, 30(1): 87-91.(In Chinese)
- [14] 胡乃青, 吕晋琳, 尹雪艳, 等. 高海拔地区红细胞计数与高血压的相关性研究[J]. 大理大学学报, 2021, 6(8): 54-57.
Hu NQ, Lv JL, Yin XY, et al. Study on the correlation between red blood cell count and hypertension at high altitude [J]. Journal of Dali University, 2021, 6(8): 54-57.(In Chinese)
- [15] Wang M. Explainable machine-learning-based cardiovascular disease prediction in patients with hypertension: Algorithm comparison and SHapley Additive exPlanations (SHAP) analysis[J]. Archives of Cardiovascular Diseases, 2026, 119(4): 273-282.
- [16] He JY, Wang X, Zhu PQ, et al. Identification and validation of an explainable early-stage chronic kidney disease prediction model: a multicenter retrospective study [J]. eClinicalMedicine, 2025, 84: 103286.

收稿日期: 2026-02-11