

预测高血压合并射血分数保留的心力衰竭患者死亡风险的可解释机器学习模型

郭威¹, 田晶¹, 王雅靖¹, 张岩波², 韩清华^{1,3,4}

1. 山西医科大学第一医院心内科, 山西太原 030001; 2. 山西医科大学公共卫生学院卫生统计学教研室
重大疾病风险评估山西省重点实验室; 3. 山西省心脑血管病三高共管技术创新中心;
4. 山西医科大学细胞生理学教育部重点实验室

摘要: **目的** 构建可解释性机器学习模型, 预测高血压合并射血分数保留的心力衰竭(HFpEF)患者的全因死亡风险。**方法** 2014年4月至2019年3月期间在山西省三所三级甲等医院选取847例高血压合并HFpEF患者, 前瞻性随访至2022年4月1日。以全因死亡为结局事件。将患者按7:3的比例随机分为训练集和测试集, 训练集用于构建预测模型, 测试集用于模型性能评价。通过最小绝对收缩选择算子(LASSO)-Cox回归筛选预测因子, 分别构建极端梯度提升(XGBoost)、logistic回归、随机森林(RF)、决策树(DT)、支持向量机(SVM)和多层感知器(MLP)6种机器学习模型预测患者3年全因死亡风险。采用受试者操作特征(ROC)曲线、校准曲线和临床决策曲线评估模型性能。采用Shapley加法解释(SHAP)对简化后的最佳模型进行可解释性分析, 并运用限制性立方样条探究关键影响因素与全因死亡之间的非线性关系。**结果** 中位随访4.25年(P_{25}, P_{75} 为2.86, 6.17年), 共有224例患者(26.4%)发生全因死亡。基于LASSO-Cox回归算法, 从患者入院时的生命体征、实验室检查结果、影像学指标等临床特征中筛选出17个预测因子。ROC曲线显示, RF模型预测患者3年全因死亡曲线下面积(AUC)为0.823(95%CI: 0.693~0.950), 准确度为84.0%, 灵敏度为82.3%, 特异度为83.0%, F1评分为0.810, 预测性能优于其他5种模型。校准曲线和临床决策曲线显示, RF模型具有较好的校准度和临床适用性。SHAP特征重要性排序显示: 年龄、估算的肾小球滤过率(eGFR)、收缩压、体重指数(BMI)为高血压合并HFpEF患者全因死亡的前4位影响因素。限制性立方样条进一步分析显示: 年龄>72岁、eGFR<72.9 mL/(min·1.73 m²)、收缩压>136 mmHg、BMI>26.6 kg/m²时, 将增加患者的全因死亡风险。为提高风险预警阈值的临床适用性, 选择临床容易使用的阈值进行Cox回归分析, 结果显示, 收缩压>135 mmHg($HR=1.362, 95\%CI: 1.020\sim 1.819$)和eGFR<70 mL/(min·1.73 m²)($HR=1.519, 95\%CI: 1.135\sim 2.034$)均显著增加患者的全因死亡风险。**结论** 基于RF算法构建的高血压合并HFpEF死亡预测模型能有效地预测患者出院后3年全因死亡风险, SHAP提供的可解释性分析可为临床决策提供明确的依据。

关键词: 高血压; 射血分数保留的心力衰竭; 机器学习; Shapley加法解释(SHAP)算法; 预后

An interpretable machine learning model for predicting mortality risk in patients with hypertension and heart failure with preserved ejection fraction

GUO Wei¹, TIAN Jing¹, WANG Yajing¹, ZHANG Yanbo², HAN Qinghua^{1,3,4}

1. Department of Cardiology, The First Hospital of Shanxi Medical University, Taiyuan, Shanxi 030001, China;
2. Department of Health Statistics, School of Public Health, Shanxi Medical University, Shanxi Provincial Key Laboratory of Major Diseases Risk Assessment;
3. Shanxi Innovation Center for Integrated Management of Hypertension, Hyperlipidemia and Hyperglycemia Correlated with Cardiovascular and Cerebrovascular Diseases;
4. Key Laboratory of Cellular Physiology at Shanxi Medical University, Ministry of Education

Abstract: **Objective** To develop an interpretable machine learning model to predict all-cause mortality risk in patients with hypertension and heart failure with preserved ejection fraction (HFpEF). **Methods** A prospective cohort of 847 patients diagnosed with hypertension and HFpEF from 3 tertiary hospitals in Shanxi Province between April 2014 and March

doi: 10.16439/j.issn.1673-7245.2024-0357

基金项目: 国家自然科学基金(82103958, 81970204); 山西省重点研发计划项目(2022ZDYF089); 山西省卫生健康委员会项目(2021RC03); 山西省研究生实践创新项目(2024SJ163)

通信作者: 韩清华, E-mail: syhqh@sohu.com; 张岩波, E-mail: sxmuzyb@126.com

2019 was followed until April 1, 2022. All-cause mortality was used as the outcome event. The cohort was randomly divided into training (70%) and testing (30%) sets. The training set was used to construct prediction models, and the testing set was used for performance evaluation. Predictors were selected using the least absolute shrinkage and selection operator (LASSO)-Cox regression. Six machine learning models, including extreme gradient boosting (XGBoost), logistic regression, random forest (RF), decision tree (DT), support vector machine (SVM), and multilayer perceptron (MLP), were developed to predict the 3-year all-cause mortality risk. Model performance was evaluated using receiver operating characteristic (ROC) curves, calibration curves, and clinical decision curves. The Shapley additive explanations (SHAP) framework was applied to the simplified optimal model interpret, and restricted cubic spline models were used to explore the nonlinear relationships between key predictors and all-cause mortality. **Results** After a median follow-up of 4.25 (P_{25} , P_{75} 2.86, 6.17) years, 224 patients (26.4%) experienced all-cause mortality. Using the LASSO-Cox regression, 17 predictors were identified from patients' clinical characteristics, including vital signs, laboratory tests, and imaging results. The RF model achieved the best performance, with an area under the ROC curve (AUC) of 0.823 (95% *CI*: 0.693–0.950), accuracy of 84.0%, sensitivity of 82.3%, specificity of 83.0%, and an F1 score of 0.810. Calibration and clinical decision curves confirmed the RF model's good calibration and clinical applicability. SHAP feature importance analysis revealed that age, estimated glomerular filtration rate (eGFR), systolic blood pressure, and body mass index (BMI) were the top 4 factors influencing all-cause mortality in patients with hypertension and HFpEF. Further restricted cubic spline analysis indicated that age > 72 years, eGFR < 72.9 mL/(min \cdot 1.73 m 2), systolic blood pressure > 136 mmHg, and BMI > 26.6 kg/m 2 were associated with increased all-cause mortality risk. To enhance the clinical applicability of risk warning thresholds, clinically practical thresholds were selected for Cox regression analysis. The results showed that systolic blood pressure > 135 mmHg ($HR=1.362$, 95% *CI*: 1.020–1.819) and eGFR < 70 mL/(min \cdot 1.73 m 2) ($HR=1.519$, 95% *CI*: 1.135–2.034) were both significantly associated with an increased risk of all-cause mortality. **Conclusions** The RF-based prediction model can effectively estimate 3-year all-cause mortality risk in patients with hypertension and HFpEF after discharge. SHAP-based interpretability analysis can provide clear insights for clinical decision-making.

Keywords: hypertension; heart failure with preserved ejection fraction; machine learning; Shapley additive explanations (SHAP) algorithm; prognosis

心力衰竭是各种心脏疾病的终末阶段,患者预后差^[1]。随着人口老龄化的加剧及各种心血管疾病患者生存期的延长,心力衰竭的发病率逐年攀升,给社会带来沉重的经济和医疗负担^[2]。射血分数保留的心力衰竭(heart failure with preserved ejection fraction, HFpEF)是心力衰竭的重要亚型,约占心力衰竭住院患者的50%,庞大的群体数量使之成为心力衰竭治疗的重要靶点^[3]。作为一种高度异质的临床综合征, HFpEF的病因复杂且发病机制尚不清晰,导致临床上缺乏有效的循证治疗方案,患者5年死亡率甚至超过某些恶性肿瘤^[3-4]。高血压是HFpEF的重要病因,流行病学数据显示,超过50%的HFpEF患者伴有高血压^[5],这类患者面临更高的死亡风险^[6]。因此,优化风险管理,及早发现高危患者以降低患者死亡风险至关重要。目前,少有高血压合并HFpEF患者死亡风险预测的相关研究。

随着计算机科学领域的发展,机器学习(machine learning)等人工智能技术已在心血管疾病预防、治疗及结局预测等方面展现出独特的优势^[7]。相较于传统模型,机器学习在患者死亡率预测方面具有更高的精准度^[8-10]。然而,机器学习模型的临床应用与推广仍

面临着可解释性不足的挑战。为此,本研究基于一项真实世界的前瞻性队列,在提取高血压合并HFpEF患者死亡关键影响因素的基础上,利用长期随访数据构建了6种机器学习模型预测患者的远期死亡风险并进行了验证。为提升模型的临床应用价值,采用Shapley加法解释(Shapley additive explanations, SHAP)算法对最优模型进行可解释性分析,以期助力临床早期识别高危患者,为制定个体化防治策略提供科学依据。

1 对象与方法

1.1 对象 顺序选取2014年4月至2019年3月期间在中国山西省三所三级甲等医院接受住院诊治的高血压合并HFpEF患者作为研究对象。纳入标准:①年龄 ≥ 18 岁;②高血压诊断符合《中国高血压防治指南(2024年修订版)》^[11],即在未服用抗高血压药的情况下,三次非同日收缩压 ≥ 140 mmHg(1 mmHg=0.133 kPa)和/或舒张压 ≥ 90 mmHg,或既往有高血压病史,正在使用降压药;③HFpEF诊断根据《中国心力衰竭诊断和治疗指南2018》^[12],具有典型的心力衰竭症状和/或

体征,入院时超声心动图显示左室射血分数 $\geq 50\%$,脑利钠肽升高并符合以下至少1条:左心室肥厚和/或左心房扩大;心脏舒张功能异常;纽约心脏协会(New York Heart Association, NYHA)心功能分级为II~IV级。排除标准:①近2月内发生急性心血管事件;②合并精神疾病;③合并恶性肿瘤或其他严重疾病,预计生存期 < 6 月;④拒绝参加研究。

本研究遵照赫尔辛基宣言,并经山西医科大学伦理委员会批准(审批号:2018LL128)。所有患者均同意参与本研究并签署知情同意书。技术路线图见网络版附图S1。

1.2 方法

1.2.1 数据收集 使用本研究小组自行设计的调查表收集患者住院期间的信息,包括年龄、性别、体重指数(body mass index, BMI)、入院时血压、心率、NYHA分级、个人史等一般资料;血常规、血糖、血脂、估算的肾小球滤过率(estimated glomerular filtration rate, eGFR)、尿素氮、肌酐、氨基末端脑利钠肽前体(N-terminal pro-brain natriuretic peptide, NT-proBNP)、肝功能及游离三碘甲状腺原氨酸(free triiodothyronine, FT3)、游离甲状腺素(free thyroxine, FT4)等实验室检查指标;HFpEF常见合并症如冠心病、糖尿病、高脂血症等^[1];心电图及超声心动图等影像学资料;药物与非药物[经皮冠脉介入术(percutaneous coronary intervention, PCI)或冠状动脉旁路移植术(coronary artery bypass grafting, CABG)]治疗情况等。eGFR根据肌酐通过慢性肾脏病流行病学协作组(Chronic Kidney Disease-Epidemiology Collaboration, CKD-EPI)方程计算^[13]。

1.2.2 血压及心率的测量 受试者在测量前保持安静状态至少5 min,以坐位姿势测量上臂血压,确保上臂与心脏在同一水平位置^[11]。使用上臂式医用电子血压计(OMRON, HEM-8713)测量。首诊时同时测量两侧上臂血压,选择血压读数较高的一侧作为后续测量的上臂。测量2次血压,相隔1~2 min,并取两次读数的平均值记录。若收缩压或舒张压的两次读数相差超过5 mmHg,则需再次测量,并取3次读数的平均值记录,同时记录患者心率。

1.2.3 NYHA心功能分级 由专业的医师根据NYHA心功能分级制定标准,于患者入院时评估。

1.2.4 实验室检查 被检者空腹8 h后,于住院第一日清晨采集肘静脉血样。使用全自动生化分析仪(贝克曼, AU5800)检测丙氨酸转氨酶(alanine aminotransferase, ALT)、天冬氨酸转氨酶(aspartate aminotransferase, AST)、总胆固醇、甘油三酯、高密度脂蛋白胆固醇(high-density lipoprotein cholesterol, HDL-C)、低密度脂蛋白胆固醇(low-density lipoprotein cholesterol,

LDL-C)、尿素氮、肌酐、尿酸等生化指标。使用全自动化学发光免疫分析仪(优迈, CARIS200)分析NT-proBNP。

1.2.5 心电图检查 使用12导联同步心电图仪(纳龙, RAGE-12)进行心电图检查。被检者平躺,连接肢体导联及胸导联电极后,通过心电图仪监测心电活动至少1 min,由专业的心内科医师判读心电图结果。在基线平稳且干扰最小的心电图上,选择QRS波最宽的导联,测量从QRS波起点至终点的时限,记录QRS波时限;选择QRS波清晰的导联,测量QRS波起点至T波终点的间距,记录QT间期^[14]。

1.2.6 心脏超声检查 使用超声仪(Philips, EPIQ7C)进行检查。被检者取左侧卧位,在静息状态下检测心房及心室前后径、室间隔厚度、左心室舒张末期内径、左心室后壁厚度等指标,记录左室射血分数。

1.2.7 随访与结局 患者均于出院后第1、3、6、12月,之后每隔6月通过电话随访,随访截止日期为2022年4月1日。终点事件定义为全因死亡,包括心血管疾病死亡,如致命性心肌梗死、心力衰竭死亡,意外猝死及非心血管因素导致的死亡,并记录其发生时间。

1.3 统计学方法 利用SPSS 25.0软件进行组间比较的统计分析;R 4.3.1软件进行缺失变量的处理;Python 3.7软件用于构建模型及SHAP分析。对于计量资料,本文数据均呈偏态分布,以中位数(P_{25} , P_{75}) [$M(P_{25}$, $P_{75})$]表示,两组间比较采用独立样本秩和检验;计数资料以频数(%)表示,组间比较采用卡方(χ^2)检验或Fisher确切概率法。

随机将70%的患者划分为一个训练集,余下的30%的患者作为独立的测试集。在训练集上采用10折交叉验证法进行特征筛选和模型训练,具体方法为将训练集数据随机划分为10个小组,其中9个小组用于模型的训练,1个小组用于模型的性能评估。将训练集中所有可能的训练小组和测试小组进行折叠组合,重复该过程10次,用于确定最优的模型和参数配置。最后,在独立的测试集上对模型进行评估并计算评价指标。在训练集中,利用最小绝对收缩选择算子(least absolute shrinkage and selection operator, LASSO)-Cox回归算法,提取全因死亡的预测因子纳入模型。分别采用极端梯度提升(extreme gradient boosting, XGBoost)、logistic回归、随机森林(random forest, RF)、决策树(decision tree, DT)、支持向量机(support vector machine, SVM)和多层感知器(multilayer perceptron, MLP)6种机器学习算法构建高血压合并HFpEF患者死亡风险模型,并在独立测试集上利用受试者操作特征(receiver operator characteristic, ROC)曲线的曲线下面积(area under the curve, AUC)、准确度、灵敏度、特

异度和 F1 评分对模型的预测性能进行评价,采用校准曲线和临床决策曲线分析(decision curve analysis, DCA)对模型的校准度和临床适用性进行评估。SHAP图是基于SHAP值生成的图形,用于直观地展示模型预测结果中各个变量的重要性排序及其影响,提升模型的可解释性和透明度。本文利用SHAP工具包对最优模型进行可解释性分析。对于缺失变量,删除缺失率超过30%的变量,其余使用缺失森林填补和多重插补填补两种策略填补。对绝大多数变量而言,两种填补方法差异无统计学意义,见网络版附表S1。进一步,对两种方法填补后的数据与填补前的基线资料进行比较,发现缺失森林方法填补后的数据较多重插补与原始数据更接近,组间平衡性更优,因此最终选择R 4.3.1中的MissForest包^[15]对缺失变量进行填补。采用限制性立方样条分析影响因素与全因死亡之间的非线性关系,节点数选择3个。采用Cox比例风险回归模型分析收缩压、eGFR与全因死亡的关系。

以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 患者基线特征 本研究纳入902例患者,排除55例(6.10%)失访后,最终纳入847例患者。患者年龄中位数为70(P_{25}, P_{75} 为59, 79)岁,女性399例(47.1%)。在中位随访4.25(P_{25}, P_{75} 为2.86, 6.17)年期间,224例患者(26.4%)发生全因死亡。存活组和死亡组患者在年龄、BMI、入院时血压、NYHA分级、合并症(糖尿病、高脂血症、冠心病等)、红细胞分布宽度、血红蛋白、ALT、白蛋白、糖化血红蛋白、血脂(除了HDL-C)、eGFR、尿酸、尿素氮、钾、钠、氯、肌钙蛋白I、FT3、FT4、IgNT-proBNP、QT间期、左心房前后径、右心房前后径、室间隔厚度、左室射血分数、抗血小板药物、他汀类药物治疗以及PCI或CABG等方面差异具有统计学意义(均 $P < 0.05$),见表1。

表1 两组患者基线主要临床资料比较

组别	例数	年龄 (岁)	女性 [例(%)]	吸烟史 [例(%)]	饮酒史 [例(%)]	收缩压 (mmHg)	舒张压 (mmHg)	心率 (次/min)
存活	623	66 (55, 76)	284 (45.6)	252 (40.4)	153 (24.6)	134 (124, 150)	80 (70, 88)	70 (63, 78)
死亡	224	80 (71, 85)	115 (51.3)	86 (38.4)	36 (16.1)	140 (130, 156)	82 (70, 92)	71 (63, 80)
Z χ^2 值		-12.777	2.189	0.707	6.846	-3.675	-2.350	-1.895
P值		<0.001	0.139	0.702	0.009	<0.001	0.018	0.058
组别	NYHA分级 [例(%)]			BMI (kg/m ²)	糖尿病 [例(%)]	高脂血症 [例(%)]	瓣膜病 [例(%)]	
	II级	III级	IV级					
存活	304 (48.8)	257 (41.3)	62 (10.0)	25.5 (23.4, 27.6)	205 (32.9)	165 (26.5)	49 (7.9)	
死亡	62 (27.7)	104 (46.4)	58 (25.9)	23.6 (21.2, 26.4)	92 (41.1)	31 (13.8)	23 (10.3)	
Z χ^2 值		47.592		-5.972	4.825	14.813	1.223	
P值		<0.001		<0.001	0.028	<0.001	0.269	
组别	冠心病 [例(%)]	心房颤动 [例(%)]	脑卒中 [例(%)]	陈旧性心肌梗死 [例(%)]	慢性阻塞性肺疾病 [例(%)]	慢性肾脏病 [例(%)]	贫血 [例(%)]	
存活	603 (96.8)	123 (19.7)	155 (24.9)	165 (26.5)	68 (10.9)	61 (9.8)	14 (2.2)	
死亡	206 (92.0)	89 (39.7)	73 (32.6)	74 (33.0)	54 (24.1)	59 (26.3)	21 (9.4)	
Z χ^2 值	8.953	35.082	4.978	3.491	23.257	37.102	21.130	
P值	0.003	<0.001	0.026	0.062	<0.001	<0.001	<0.001	
组别	红细胞分布宽度 (%)	血红蛋白 (g/L)	血小板 (10 ⁹ /L)	ALT (U/L)	AST (U/L)	白蛋白 (g/L)	血清总胆红素 (μ mol/L)	
存活	13.6 (13.1, 14.3)	135 (125, 148)	183 (153, 222)	19.0 (14.0, 29.0)	23.0 (19.2, 30.0)	42 (40, 45)	13.5 (10.2, 16.9)	
死亡	14.2 (13.6, 15.3)	126 (111, 140)	176 (138, 227)	15.9 (11.0, 23.0)	24.3 (19.0, 32.0)	41 (38, 44)	12.9 (10.3, 17.2)	
Z χ^2 值	-7.373	-6.092	-1.071	-4.390	-0.913	-4.079	-0.330	
P值	<0.001	<0.001	0.284	<0.001	0.361	<0.001	0.741	
组别	血糖 (mmol/L)	糖化血红蛋白 (%)	HDL-C (mmol/L)	LDL-C (mmol/L)	总胆固醇 (mmol/L)	甘油三酯 (mmol/L)	eGFR [mL/(min·1.73 m ²)]	
存活	5.2 (4.6, 6.2)	5.00 (4.80, 6.78)	1.0 (0.9, 1.2)	2.4 (1.9, 3.0)	4.0 (3.4, 4.8)	1.5 (1.1, 2.0)	80.7 (65.9, 95.3)	
死亡	5.3 (4.6, 6.6)	5.98 (5.00, 6.21)	1.0 (0.9, 1.2)	2.3 (1.8, 2.8)	3.8 (3.2, 4.6)	1.2 (0.9, 1.7)	59.1 (40.9, 80.0)	
Z χ^2 值	-1.021	-9.931	-1.131	-2.025	-2.682	-5.047	-9.282	
P值	0.307	<0.001	0.258	0.043	0.007	<0.001	<0.001	

续表 1

组别	尿酸 ($\mu\text{mol/L}$)	尿素氮 (mmol/L)	钾 (mmol/L)	钠 (mmol/L)	氯 (mmol/L)	肌钙蛋白I ($\mu\text{g/L}$)	FT3 (pmol/L)
存活	347 (285, 415)	5.6 (4.5, 6.9)	4.0 (3.8, 4.2)	141.0 (139.0, 142.3)	104.0 (102.0, 106.3)	0 (0, 0.2)	4.36 (4.29, 4.43)
死亡	378 (291, 473)	7.0 (5.4, 10.4)	4.1 (3.8, 4.5)	139.3 (137.0, 141.0)	103.8 (100.0, 106.0)	0 (0, 0.4)	4.36 (3.98, 4.75)
Z/ χ^2 值	-2.973	-7.577	-3.777	-5.533	-2.996	-2.112	-2.954
P值	0.003	<0.001	<0.001	<0.001	0.003	0.031	0.003
组别	FT4 (pmol/L)	TSH (mU/L)	lg[NT-proBNP (ng/L)]	QRS波时限 (ms)	QT间期 (ms)	左心房前后径 (mm)	右心房前后径 (mm)
存活	13.04 (12.11, 13.92)	3.01 (2.14, 3.76)	2.5 (2.3, 2.9)	93 (85, 102)	422 (404, 446)	38 (35, 41)	34 (31, 39)
死亡	13.24 (12.76, 14.52)	3.01 (2.31, 3.89)	3.1 (2.7, 3.5)	94 (86, 106)	436 (412, 470)	39 (37, 43)	36 (32, 45)
Z/ χ^2 值	-2.343	-0.313	-10.136	-1.716	-4.739	-4.895	-3.221
P值	0.019	0.754	<0.001	0.086	<0.001	<0.001	<0.001
组别	室间隔厚度 (mm)	左心室舒张末期 内径(mm)	左心室后壁厚度 (mm)	右心室前后径 (mm)	左室射血分数 (%)	抗血小板类药物 [例(%)]	他汀类药物 [例(%)]
存活	10 (9, 11)	49 (46, 52)	10.0 (9.0, 10.1)	21 (20, 24)	60.0 (55.0, 66.0)	581 (93.3)	566 (90.9)
死亡	11 (9, 12)	48 (45, 52)	10.0 (9.0, 10.6)	21 (19, 24)	58.0 (54.0, 65.0)	190 (84.8)	189 (84.4)
Z/ χ^2 值	-2.163	-1.212	-0.056	-0.053	-2.032	14.359	7.136
P值	0.028	0.225	0.955	0.958	0.042	<0.001	0.008
组别	舒张冠状动脉类药物 [例(%)]	ACEI/ARB [例(%)]	β 受体阻滞剂 [例(%)]	利尿剂 [例(%)]	钙通道阻滞药 [例(%)]	强心类药物 [例(%)]	PCI/CABG [例(%)]
存活	505 (81.1)	349 (56.0)	160 (25.7)	319 (51.2)	159 (25.5)	34 (5.5)	189 (30.3)
死亡	174 (77.7)	113 (50.4)	37 (16.5)	154 (68.8)	109 (48.7)	21 (9.4)	39 (17.4)
Z/ χ^2 值	1.184	2.064	7.752	20.571	40.785	4.164	-12.777
P值	0.276	0.151	0.005	<0.001	<0.001	0.041	<0.001

注:非正态分布的计量资料用中位数(P_{25}, P_{75})表示。NYHA为纽约心脏协会;BMI为体重指数;ALT为丙氨酸转氨酶;AST为天冬氨酸转氨酶;HDL-C为高密度脂蛋白胆固醇;LDL-C为低密度脂蛋白胆固醇;eGFR为估算的肾小球滤过率;FT3为游离三碘甲状腺原氨酸;FT4为游离甲状腺素;TSH为促甲状腺激素;lgNT-proBNP为氨基末端脑利钠肽前体的对数值;ACEI为血管紧张素转换酶抑制剂;ARB为血管紧张素受体阻滞药;PCI为经皮冠脉介入术;CABG为冠状动脉旁路移植术。

2.2 特征筛选 为避免模型过度拟合,降低指标间的共线性,利用 LASSO-Cox 回归算法对 62 个候选变量进行特征选择。通过 10 折交叉验证方法对正则化参数进行调优,如图 1 所示。当惩罚系统 $\lambda=0.024$ 时,模型均方误差最小,此时筛选出 17 个全因死亡的预测因子,分别为:年龄、收缩压、NYHA 分级、BMI、吸烟史、红细胞分布宽度、尿素氮、eGFR、lgNT-proBNP、QRS 波时限、QT 间期、糖尿病、慢性肾脏病、慢性阻塞性肺疾病、贫血、 β 受体阻滞剂和 PCI/CABG。

2.3 死亡预测模型的构建与评估 将 LASSO-Cox 提取出的 17 个关键影响因子纳入模型,在训练集上分别基于机器学习方法构建了 6 种不同的模型用于预测患者 3 年全因死亡风险。在测试集上进行性能评估,结果见表 2。ROC 曲线(图 2)显示,与其他 5 种模型相比,RF 模型的 AUC 值(0.823, 95%CI: 0.693~0.950)最高。除特异度指标(0.830, 95%CI: 0.800~0.860)略低于 XGBoost(0.834, 95%CI: 0.810~0.864)外,其他指标均表现最佳。进一步对 6 种机器学习模型的校准曲线

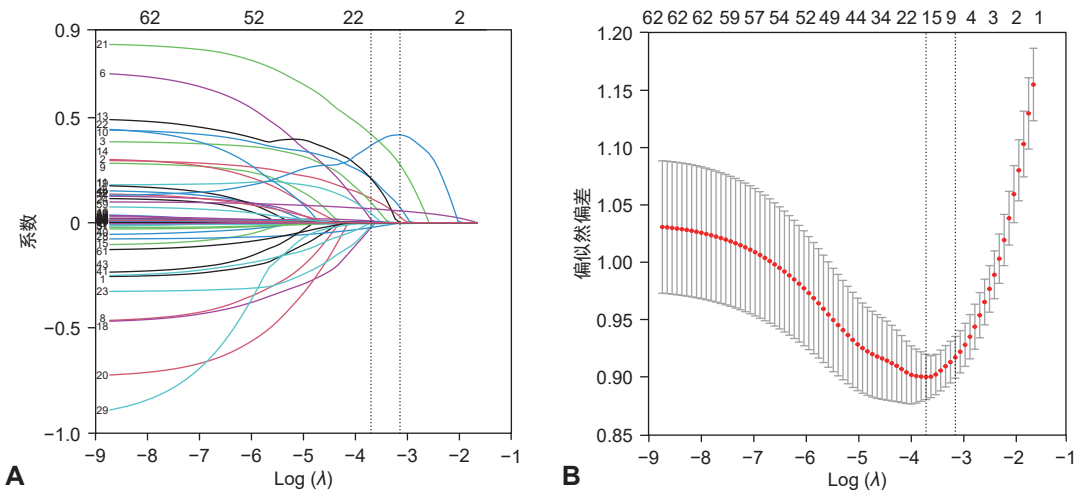
和 DCA 曲线进行比较,结果如图 3 所示。与其他 5 种模型相比,RF 模型的校准曲线显示预测结果与实际情况具有较好的吻合度,校准度(0.135)低于其他模型。此外,DCA 曲线显示,在 0~85% 的阈值范围内,RF 模型预测患者全因死亡风险的临床净收益均高于其他 5 种模型及“全干预”和“不干预”方案,表明该模型具有良好的临床适用性。

2.4 RF 模型的可解释性 SHAP 分析 基于 SHAP 算法,对 RF 模型进行分析,得到各临床变量的重要性排序依次为:年龄、eGFR、收缩压、BMI、lgNT-proBNP、尿素氮、QT 间期、QRS 波时限、红细胞分布宽度、吸烟史、NYHA 分级、慢性肾脏病、糖尿病、慢性阻塞性肺疾病、PCI/CABG、 β 受体阻滞剂和贫血。

为方便模型的临床应用,根据上述 SHAP 分析结果,对纳入 RF 模型的变量进行逐步精简,并记录该过程模型预测性能的变化,如图 4 所示。当特征数量由 17 个逐步缩减至 10 个时,模型预测性能变化幅度较小,包含 10 个特征的模型与 17 个特征的模型性能

接近($\Delta AUC_{\text{训练集}} = 0.018, \Delta AUC_{\text{测试集}} = 0.006$)。然而,当纳入变量数减至5~9个时,模型性能较17个特征的模型降幅较大。因此,选择10个特征变量的模型作

为临床预测模型。此时,RF模型在测试集上预测患者3年全因死亡风险的AUC为0.817,灵敏度为0.826,特异度为0.813,准确度为0.832,F1分数为0.792。



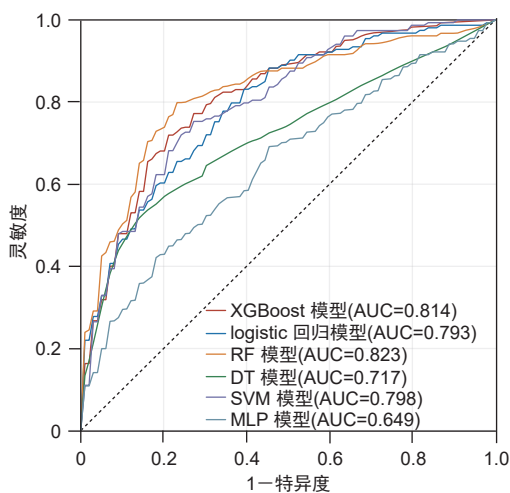
注: A, 全因死亡预测变量的最小绝对收缩选择算子(LASSO)回归模型系数轨迹。B, 调优参数(Lambda, λ)选择交叉验证的误差曲线。在最小值准则和1-SE准则给出的最佳值处绘制垂直线。左线以模型偏差最小($\lambda = 0.024$)识别。即当参数 $\lambda = 0.024$ 时,模型拟合效果最好。在10倍交叉验证选择的值处画垂直线,包括优化的17个非零系数。

图1 LASSO回归在训练集筛选变量($n=593$)

表2 各模型在测试集的性能评价($n=254$)

分类模型	AUC (95%CI)	准确度 (95%CI)	灵敏度 (95%CI)	特异度 (95%CI)	F1分数 (95%CI)
XGBoost	0.814 (0.691~0.937)	0.810 (0.790~0.830)	0.780 (0.740~0.820)	0.834 (0.810~0.864)	0.785 (0.750~0.820)
logistic回归	0.793 (0.661~0.925)	0.715 (0.675~0.754)	0.732 (0.664~0.801)	0.708 (0.648~0.768)	0.574 (0.531~0.616)
RF	0.823 (0.693~0.950)	0.840 (0.820~0.860)	0.823 (0.780~0.860)	0.830 (0.800~0.860)	0.810 (0.780~0.840)
DT	0.717 (0.572~0.863)	0.784 (0.751~0.817)	0.620 (0.580~0.660)	0.850 (0.810~0.890)	0.650 (0.610~0.690)
SVM	0.798 (0.671~0.925)	0.740 (0.710~0.770)	0.720 (0.680~0.760)	0.750 (0.720~0.780)	0.710 (0.670~0.750)
MLP	0.649 (0.490~0.809)	0.684 (0.636~0.732)	0.473 (0.310~0.637)	0.758 (0.659~0.857)	0.423 (0.337~0.510)

注: AUC为曲线下面积; XGBoost为极端梯度提升; RF为随机森林; DT为决策树; SVM为支持向量机; MLP为多层感知器。



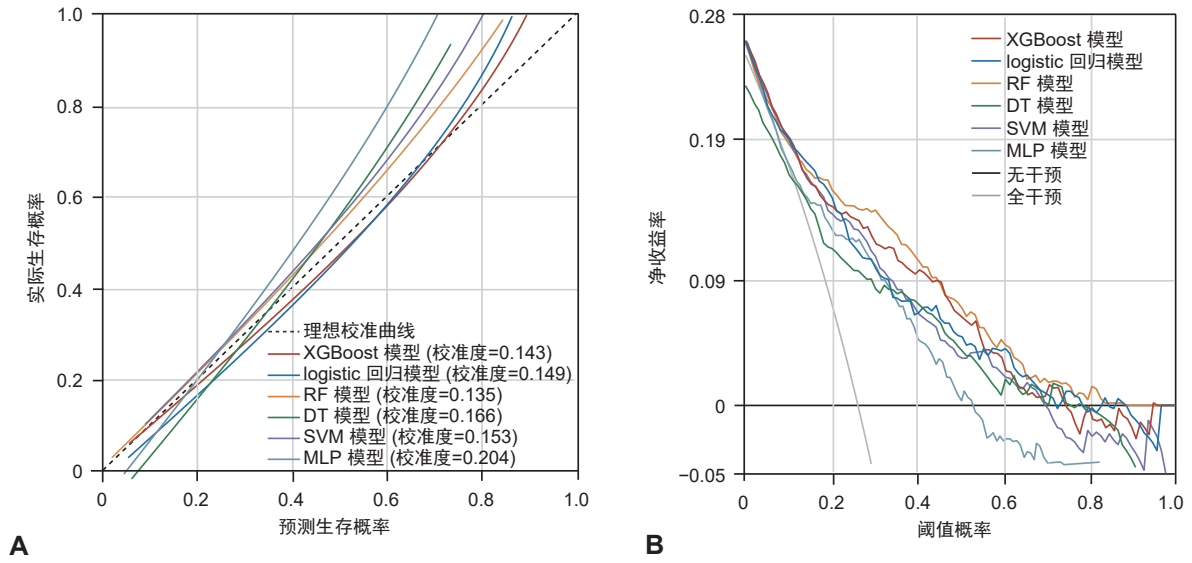
注: XGBoost为极端梯度提升; RF为随机森林; DT为决策树; SVM为支持向量机; MLP为多层感知器; AUC为曲线下面积。

图2 6种机器学习模型在测试集预测3年全因死亡的受试者操作特征(ROC)曲线($n=254$)

为了更直观地显示模型和数据间的内在联系,采用SHAP算法对模型进行可解释性分析。图5展示了各临床变量对RF模型预测结果的重要性排序以及SHAP值的分布情况。每个点表示一个变量,位置显示其SHAP值,即该变量对模型预测结果的贡献度。SHAP值为正表示该变量对预测结果产生正面影响,为负则表明产生负面影响。红色和蓝色分别标记高风险与低风险,颜色越深表示该特征对目标变量的影响越强。

2.5 年龄、eGFR、收缩压及BMI与全因死亡的关联

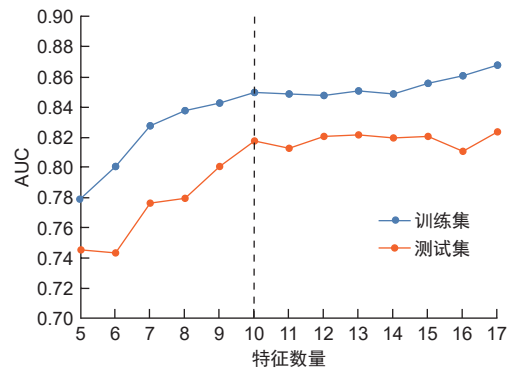
通过绘制限制性立方样条曲线进一步探索SHAP图中全因死亡预测贡献前4位的临床特征(年龄、eGFR、收缩压及BMI)与全因死亡间的非线性关联。结果显示,年龄($P_{\text{非线性}}=0.011$)、eGFR($P_{\text{非线性}} < 0.001$)、收缩压($P_{\text{非线性}}=0.048$)及BMI($P_{\text{非线性}}=0.002$)均与全因死亡具有非线性关联。



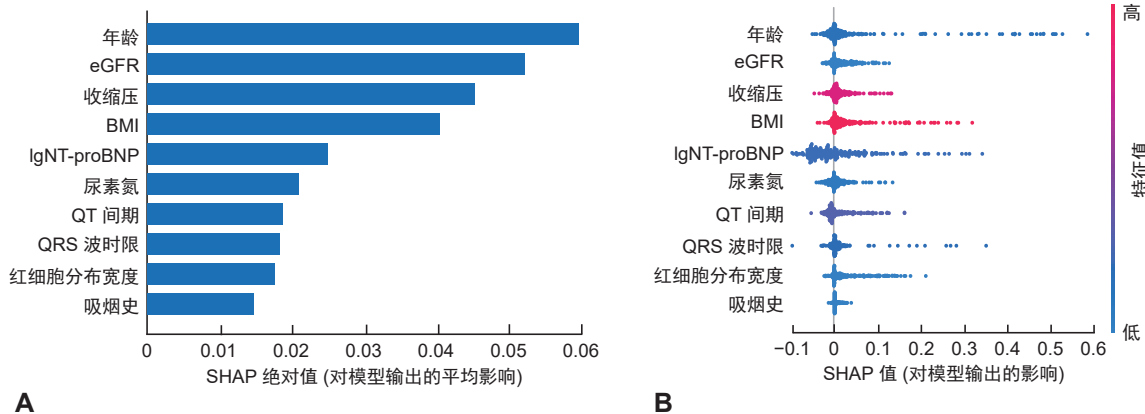
注: A为校准曲线; B为DCA。XGBoost为极端梯度提升; RF为随机森林; DT为决策树; SVM为支持向量机; MLP为多层感知器。
图3 6种机器学习模型在测试集的校准曲线与临床决策曲线分析(DCA)($n=254$)

在校正其他混杂因素后, 研究发现: 年龄 > 72 岁, $eGFR < 72.9 \text{ mL}/(\text{min} \cdot 1.73 \text{ m}^2)$, 收缩压 $> 136 \text{ mmHg}$ 时, 患者全因死亡风险增加。对于 BMI, 当 $23.4 \leq \text{BMI} \leq 26.6 \text{ kg}/\text{m}^2$ 时, BMI 可降低患者全因死亡风险, 而当 $\text{BMI} > 26.6 \text{ kg}/\text{m}^2$ 时, 患者全因死亡风险随着 BMI 的升高而增加(图 6)。

为提高风险预警阈值的临床适用性, 选择临床容易使用的阈值, 即: “收缩压 $> 135 \text{ mmHg}$, $eGFR < 70 \text{ mL}/(\text{min} \cdot 1.73 \text{ m}^2)$ ”, 采用 Cox 回归分析其与全因死亡的关联。结果如表 3 所示, 收缩压 $> 135 \text{ mmHg}$ ($HR=1.362$, $95\%CI: 1.020 \sim 1.819$) 和 $eGFR < 70 \text{ mL}/(\text{min} \cdot 1.73 \text{ m}^2)$ ($HR=1.519$, $95\%CI: 1.135 \sim 2.034$) 均显著增加了患者全因死亡风险。

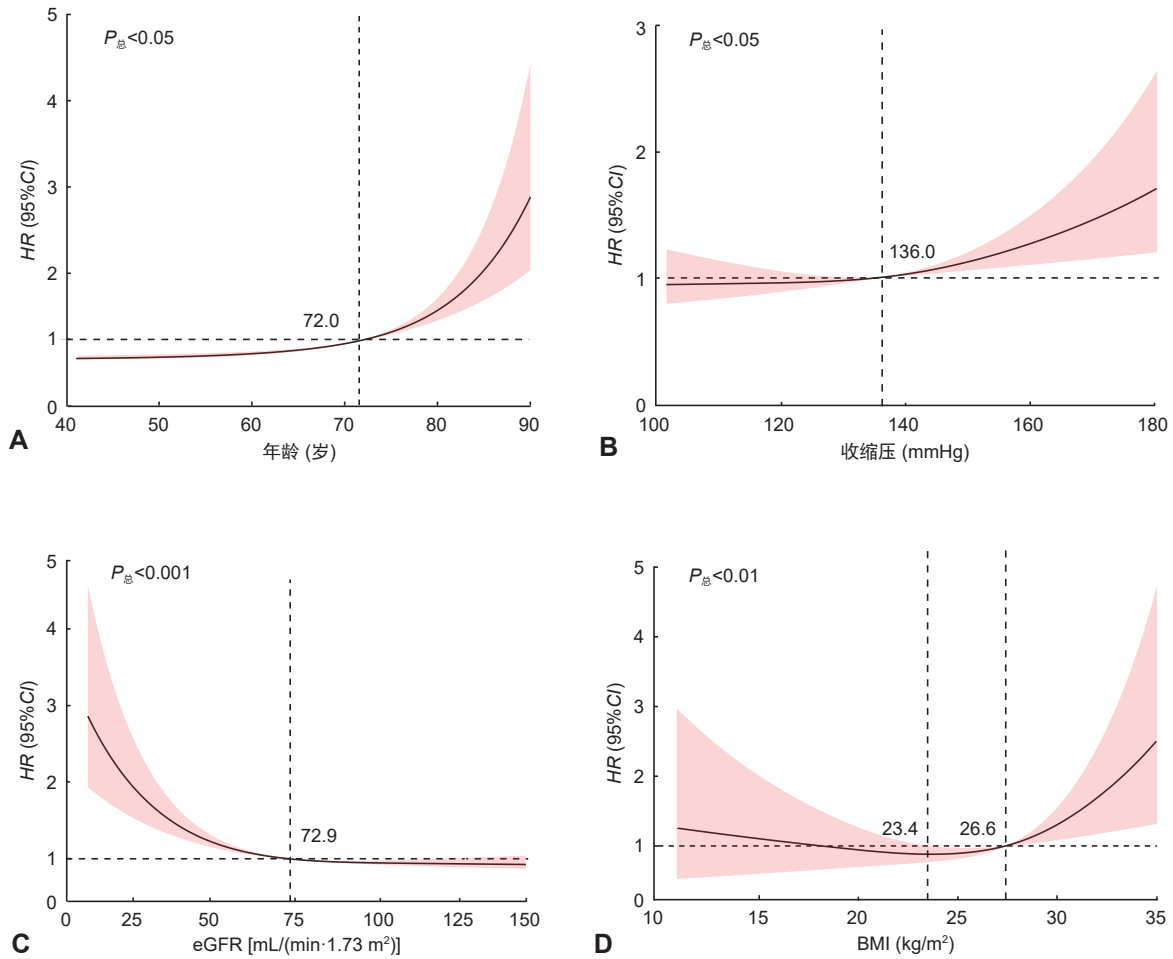


注: RF为随机森林; AUC为受试者操作特征(ROC)曲线下面积。
图4 纳入不同数量特征 RF 模型训练集($n=593$)和测试集($n=254$)的预测性能



注: A为 SHAP 特征重要性排序; B为特征对模型输出影响的 SHAP 分布情况。eGFR 为估算的肾小球滤过率; BMI 为体重指数; lgNT-proBNP 为氨基末端脑利钠肽前体的对数值。

图5 Shapley 加法解释(SHAP)汇总图



注: eGFR 为估算的肾小球滤过率; BMI 为体重指数。

图 6 年龄(A)、收缩压(B)、eGFR(C)、BMI(D)与全因死亡关联的限制性立方条图($n=847$)

表 3 高血压合并射血分数保留的心力衰竭患者收缩压、eGFR 与全因死亡的 Cox 回归分析($n=847$)

自变量	<i>B</i>	<i>SE</i>	Wald χ^2	<i>P</i> 值	<i>HR</i>	95% <i>CI</i>
收缩压 ≤ 135 mmHg					1.000	
收缩压 > 135 mmHg	0.309	0.148	4.378	0.036	1.362	1.020~1.819
eGFR ≥ 70 mL/(min $\cdot 1.73$ m 2)					1.000	
eGFR < 70 mL/(min $\cdot 1.73$ m 2)	0.418	0.149	7.891	0.005	1.519	1.135~2.034

注:校正了年龄、性别、吸烟史、饮酒史、NYHA 分级、收缩压(分析收缩压时不校正)、eGFR(分析eGFR时不校正)、BMI、lgNT-proBNP、尿素氮、QRS波时限、QT间期、红细胞分布宽度、脑卒中、慢性阻塞性肺疾病、慢性肾脏病、糖尿病、贫血、 β 受体阻滞剂、ACEI/ARB、利尿剂、PCI/CABG。NYHA为纽约心脏协会; eGFR为估算的肾小球滤过率; BMI为体重指数; lgNT-proBNP为氨基末端脑利钠肽前体的对数值; ACEI为血管紧张素转换酶抑制剂; ARB为血管紧张素受体阻滞剂; PCI为经皮冠脉介入术; CABG为冠状动脉旁路移植术。

3 讨论

本研究基于一项真实世界的前瞻性队列研究,利用机器学习算法构建了6种死亡风险预测模型,评估并比较了其预测高血压合并HFpEF患者出院后3年全因死亡的价值。进一步对预测性能最好的模型利用SHAP算法进行了可解释性分析,以促进模型在临床的应用。结果显示,与其他模型相比,RF模型预测性能最好。SHAP分析显示,对全因死亡影响排在前三

四的临床特征分别为:年龄、eGFR、收缩压和BMI。

为简化模型,降低变量的共线性,基于62个候选临床特征,采用LASSO-Cox回归算法进行特征筛选。基于提取出的17项关键特征,分别构建了XGBoost、logistic回归、RF、DT、SVM和MLP 6种死亡风险评估模型,并在测试集上通过ROC曲线、临床校准曲线和DCA等对模型预测患者远期死亡率的性能进行评估。其中,RF模型的预测性能最高(AUC=0.823, 95%CI: 0.693~0.950)。校准曲线和DCA显示,RF模

型的预测结果与实际具有较好的吻合度(校准度 0.135), 在广泛的阈值范围内(0~85%), 其临床净收益均高于其他几种模型及“全干预”和“不干预”方案, 具有良好的临床适用性。这一研究结果与近期一项预测 HFpEF 门诊患者 3 年死亡率的研究相吻合, 在这项研究中, RF 同样是表现最佳的模型, 其预测死亡率的 AUC 为 0.72(95%CI: 0.69~0.75), 高于 logistic 回归模型(AUC=0.65)及 SVM 模型(AUC=0.66)^[10]。RF 模型是一种高度先进的集成学习方法, 通过结合多个决策树的预测结果来显著提升整体的预测性能和稳定性, 具有高预测精度及高效处理复杂数据的能力^[16]。目前, RF 已在多个心力衰竭患者死亡率的预测研究中被证实具有稳定出色的预测性能与准确性^[17-20]。因此, 本研究基于 RF 模型构建的高血压合并 HFpEF 死亡风险模型具有广阔的临床应用前景。需要注意的是, 尽管 RF 模型在本研究中表现最佳, 但这并不意味着它是所有临床预测模型的最优选择。在实际应用中, 模型选择仍需紧密结合临床实际需求, 并综合考量数据集特性、预测效能及模型的可解释性。

然而, 由于机器学习模型的高度复杂性, 在临床应用中常导致其缺乏透明度, 使用者难以直观理解模型的决策过程, 这一问题通常被称作“黑匣子”问题。在本研究中, 为增强临床可理解性, 采用 SHAP 算法对构建的 RF 模型进行了可解释分析。通过量化各个特征对模型输出的贡献, 揭示了模型的内部决策机制。SHAP 分析显示, 对全因死亡预测贡献排在前三位的影响因子分别为: 年龄、eGFR、收缩压和 BMI。年龄是患者全因死亡最重要的独立预测因子, 与既往研究结果一致^[21-23]。进一步通过限制性立方样条分析, 发现当年龄 > 72 岁时, 患者的死亡风险升高。该风险阈值低于 Lopuszynski 等^[24]的研究, 他们发现年龄 > 80 岁是 HFpEF 全因死亡的预测因素。这可能与该队列患者平均年龄 [(80.5±10.9) 岁] 较本研究 [70(59, 79) 岁] 更高有关。肾功能损害可通过水钠潴留、电解质紊乱、交感神经激活等机制加重心力衰竭, 导致心力衰竭患者死亡风险增加^[25]。作为反映肾功能的重要指标, eGFR 水平降低已被证实是心力衰竭患者死亡的危险因素^[26-27]。本研究在绘制限制性立方样条图的基础上, 结合 Cox 比例风险回归模型进一步分析了这种关联, 结果显示, eGFR < 70 mL/(min·1.73 m²) 时, 患者的全因死亡风险增加。血压管理有助于降低心血管事件的发生率。本研究结果表明, 收缩压控制在 135 mmHg 以下可降低患者的全因死亡风险。虽然目前尚缺乏大型随机对照试验指导高血压合并 HFpEF 患者的血压控制目标, 但本研究的结果为《射血分数保

留的心力衰竭伴高血压患者管理中国专家共识》^[28] 推荐的该部分患者收缩压的降压目标 (< 140 mmHg) 提供了支持。此外, BMI 作为国际通用的评估人体肥胖程度及体重健康状态的标准, 也是心力衰竭预后评估的常用指标。相关研究表明, 高 BMI 通常预示着更高的心血管疾病风险^[29]。然而, 一些荟萃分析报道了较高的 BMI(特别是超重到轻度肥胖范围内)与更好的预后相关, 即“肥胖悖论”^[30]。本研究同样证实了这一点, 当 BMI 在 23.4 ~ 26.6 kg/m² 时, 可降低患者的全因死亡风险, 而当 BMI 超过 26.6 kg/m² 后, 患者的死亡风险则随着 BMI 上升而增加, 即 BMI 与患者全因死亡风险存在 U 形关联。这也提示在 HFpEF 患者的管理中, 需权衡 BMI 的影响。对 BMI 过低的患者, 应关注营养与体重管理, 而 BMI 过高者, 则需通过生活方式干预、药物乃至手术等治疗方式避免过度肥胖。

尽管本研究取得了一些成果, 但仍存在一些局限性。首先, 本研究的数据集来自山西省三所三级甲等医院, 虽然患者来源于山西各地区, 仍缺乏更广泛的代表性, 可能存在选择偏倚。其次, 仍有不可避免的临床混杂因素影响结果, 期待更大规模的国内外多中心、大样本量的前瞻性研究进一步验证模型的外延性。再次, 仍有其他潜在的预测因子未被纳入模型, 如血气分析, 主动脉球囊反搏术等, 这些变量由于缺失率超过 30% 而被排除。最后, 模型缺乏外部验证, 在未来的研究中, 有必要通过更大规模和更具代表性的外部队列来验证和优化模型。

本主题国内外已有的结论

• 高血压是 HFpEF 的重要病因, 超过 50% 的患者合并高血压且远期预后差, 优化这部分患者的风险管理已成为心力衰竭治疗的重要目标之一。

本文特色与见解

• 基于临床电子数据库, 以及可解释性机器学习算法构建的 RF 预测模型显示出良好的预测能力, 可以明确地指导高血压合并 HFpEF 患者的风险预测。

参考文献

- [1] McDonagh TA, Metra M, Adamo M, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure[J]. Eur Heart J, 2021, 42(36): 3599-3726.
- [2] Hao G, Wang X, Chen Z, et al. Prevalence of heart failure and left ventricular dysfunction in China: the China Hypertension Survey, 2012-2015[J]. Eur J Heart Fail, 2019, 21(11): 1329-1337.
- [3] Dunlay SM, Roger VL, Redfield MM. Epidemiology of heart failure with preserved ejection fraction[J]. Nat Rev Cardiol, 2017, 14(10):

- 591-602.
- [4] Shah KS, Xu H, Matsouaka RA, et al. Heart failure with preserved, borderline, and reduced ejection fraction: 5-year outcomes[J]. *J Am Coll Cardiol*, 2017, 70(20): 2476-2486.
- [5] 张健, 张宇辉. 多中心、前瞻性中国心力衰竭注册登记研究——病因、临床特点和治疗情况初步分析[J]. *中国循环杂志*, 2015, 30(5): 413-416.
- [6] Myhre PL, Selvaraj S, Solomon SD. Management of hypertension in heart failure with preserved ejection fraction: is there a blood pressure goal?[J]. *Curr Opin Cardiol*, 2021, 36(4): 413-419.
- [7] Mathur P, Srivastava S, Xu X, et al. Artificial intelligence, machine learning, and cardiovascular disease[J]. *Clin Med Insights Cardiol*, 2020, 14: 1179546820927404.
- [8] Shin S, Austin PC, Ross HJ, et al. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality[J]. *ESC Heart Fail*, 2021, 8(1): 106-115.
- [9] Jing L, Ulloa Cerna AE, Good CW, et al. A machine learning approach to management of heart failure populations[J]. *JACC Heart Fail*, 2020, 8(7): 578-587.
- [10] Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction[J]. *JACC Heart Fail*, 2020, 8(1): 12-21.
- [11] 中国高血压防治指南修订委员会, 高血压联盟(中国), 中国医疗保健国际交流促进会高血压病学分会, 等. 中国高血压防治指南(2024年修订版)[J]. *中华高血压杂志(中英文)*, 2024, 32(7): 603-700.
- [12] 中华医学会心血管病学分会心力衰竭学组, 中国医师协会心力衰竭专业委员会, 中华心血管病杂志编辑委员会. 中国心力衰竭诊断和治疗指南 2018[J]. *中华心血管病杂志*, 2018, 46(10): 760-789.
- [13] Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate[J]. *Ann Intern Med*, 2009, 150(9): 604-612.
- [14] 《心电图测量技术指南》编写专家组. 心电图测量技术指南[J]. *实用心电学杂志*, 2019, 28(2): 77-86.
- [15] Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data[J]. *Bioinformatics*, 2012, 28(1): 112-118.
- [16] Dietrich S, Floegel A, Troll M, et al. Random survival forest in practice: a method for modelling complex metabolomics data in time to event analysis[J]. *Int J Epidemiol*, 2016, 45(5): 1406-1420.
- [17] Segar MW, Hall JL, Jhund PS, et al. Machine learning-based models incorporating social determinants of health vs traditional models for predicting in-hospital mortality in patients with heart failure[J]. *JAMA Cardiol*, 2022, 7(8): 844-854.
- [18] Xu C, Li H, Yang J, et al. Interpretable prediction of 3-year all-cause mortality in patients with chronic heart failure based on machine learning[J]. *BMC Med Inform Decis Mak*, 2023, 23(1): 267.
- [19] Cai A, Chen R, Pang C, et al. Machine learning model for predicting 1-year and 3-year all-cause mortality in ischemic heart failure patients[J]. *Postgrad Med*, 2022, 134(8): 810-819.
- [20] Dai Q, Sherif AA, Jin C, et al. Machine learning predicting mortality in sarcoidosis patients admitted for acute heart failure[J]. *Cardiovasc Digit Health J*, 2022, 3(6): 297-304.
- [21] Sciomer S, Moscucci F, Salvioni E, et al. Role of gender, age and BMI in prognosis of heart failure[J]. *Eur J Prev Cardiol*, 2020, 27(2_suppl): 46-51.
- [22] Hamo CE, DeJong C, Hartshorne-Evans N, et al. Heart failure with preserved ejection fraction[J]. *Nat Rev Dis Primers*, 2024, 10(1): 55.
- [23] 郭威, 田晶, 张雅婧, 等. 射血分数轻度降低的心力衰竭死亡影响因素分析及预测模型的构建[J]. *临床心血管病杂志*, 2024, 40(6): 467-474.
- [24] Lopuszynski JB, Downing AJ, Finley CM, et al. Prognosticators of all-cause mortality in patients with heart failure with preserved ejection fraction[J]. *Am J Cardiol*, 2021, 158: 66-73.
- [25] Ostrominski JW, Thierer J, Claggett BL, et al. Cardio-renal-metabolic overlap, outcomes, and dapagliflozin in heart failure with mildly reduced or preserved ejection fraction[J]. *JACC Heart Fail*, 2023, 11(11): 1491-1503.
- [26] Löfman I, Szummer K, Evans M, et al. Incidence of, associations with and prognostic impact of worsening renal function in heart failure with different ejection fraction categories[J]. *Am J Cardiol*, 2019, 124(10): 1575-1583.
- [27] Su J, Su K, Song Y, et al. Clinical characteristics and prognosis of heart failure with preserved ejection fraction across diverse ejection fraction ranges[J]. *Rev Cardiovasc Med*, 2024, 25(5): 177.
- [28] 射血分数保留的心力衰竭伴高血压患者管理中国专家共识编写委员会. 射血分数保留的心力衰竭伴高血压患者管理中国专家共识[J]. *中华高血压杂志*, 2021, 29(7): 612-617.
- [29] Kenchaiah S, Evans JC, Levy D, et al. Obesity and the risk of heart failure[J]. *N Engl J Med*, 2002, 347(5): 305-313.
- [30] Horwich TB, Fonarow GC, Hamilton MA, et al. The relationship between obesity and mortality in patients with heart failure[J]. *J Am Coll Cardiol*, 2001, 38(3): 789-795.

收稿日期: 2024-10-15 责任编辑: 张刘锋