

## · 心肌梗死专题研究 ·

## 梯度提升决策树在预测急性心肌梗死患者 PCI 后住院期间发生心力衰竭中的应用价值

扫描二维码  
查看更多

胡文标, 刘礼敬, 林晓峰, 宋清华, 陆元喜, 韦佳玲

**【摘要】** 目的 探讨梯度提升决策树 (GBDT) 在预测急性心肌梗死 (AMI) 患者 PCI 后住院期间发生心力衰竭 (HF) 中的应用价值。方法 回顾性选取 2021—2022 年于南宁市第二人民医院行 PCI 的 AMI 患者 200 例为研究对象。将患者分为训练集 (145 例) 和测试集 (55 例)。根据 PCI 后住院期间 HF 发生情况, 将训练集患者分为 HF 组 (48 例) 和非 HF 组 (97 例)。收集患者一般资料及 PCI 前实验室检查指标、心脏彩超检查指标。基于单因素分析结果, 采用 R 4.1.2 软件分别构建预测 AMI 患者 PCI 后住院期间 HF 发生风险的 GBDT 算法模型和 Logistic 回归模型; 分别采用 ROC 曲线、校准曲线分析 GBDT 算法模型、Logistic 回归模型的区分度、准确性。结果 HF 组年龄大于非 HF 组, 有糖尿病病史者占比、超敏 C 反应蛋白 (hs-CRP)、白细胞计数 (WBC)、中性粒细胞计数、肌酸激酶同工酶 (CK-MB) 高于非 HF 组 ( $P < 0.05$ )。将单因素分析中差异有统计学意义的指标纳入 GBDT 算法模型, 通过 GBDT 算法获得这 6 项指标的相对重要性, 由小到大依次为糖尿病病史 (2.220)、中性粒细胞计数 (7.713)、年龄 (14.734)、CK-MB (16.819)、WBC (24.828)、hs-CRP (33.686)。多因素 Logistic 回归分析结果显示, 年龄、hs-CRP、WBC、中性粒细胞计数、CK-MB 是训练集 AMI 患者 PCI 后住院期间发生 HF 的影响因素 ( $P < 0.05$ ), 构建 Logistic 回归模型, 其具体公式为:  $\text{logit}(P) = -18.182 + 0.147 \times \text{年龄} + 0.233 \times \text{hs-CRP} + 0.438 \times \text{WBC} + 0.242 \times \text{中性粒细胞计数} + 0.003 \times \text{CK-MB}$ 。ROC 曲线分析结果显示, GBDT 算法模型、Logistic 回归模型预测训练集 AMI 患者 PCI 后住院期间发生 HF 的 AUC 分别为 0.989 [95%CI (0.974, 1.000)]、0.864 [95%CI (0.786, 0.942)]; GBDT 算法模型、Logistic 回归模型预测测试集 AMI 患者 PCI 后住院期间发生 HF 的 AUC 分别为 0.900 [95%CI (0.817, 0.982)]、0.763 [95%CI (0.639, 0.888)]。校准曲线分析结果显示, GBDT 算法模型、Logistic 回归模型预测训练集、测试集 AMI 患者 PCI 后住院期间发生 HF 的概率分别与本组 AMI 患者 PCI 后住院期间 HF 的实际发生率一致。结论 本研究基于年龄、糖尿病病史、hs-CRP、WBC、中性粒细胞计数和 CK-MB 6 个指标构建的 GBDT 算法模型对 AMI 患者 PCI 后住院期间发生 HF 有较好的预测价值, 且优于传统 Logistic 回归模型, 这可为 AMI 患者 PCI 后预后的评估及干预治疗提供参考依据。

**【关键词】** 心肌梗死; 心力衰竭; 决策树; 梯度提升决策树; Logistic 模型; 预测

**【中图分类号】** R 542.22 R 541.62 **【文献标识码】** A DOI: 10.12114/j.issn.1008-5971.2023.00.133

**Application Value of Gradient Boosted Decision Tree in Predicting Heart Failure in Patients with Acute Myocardial Infarction during Hospitalization after PCI** HU Wenbiao, LIU Lijing, LIN Xiaofeng, SONG Qinghua, LU Yuanxi, WEI Jialing

Department of Cardiovascular Medicine, the Second Nanning People's Hospital, Nanning 530031, China

Corresponding author: HU Wenbiao, E-mail: 281385715@qq.com

**【Abstract】 Objective** To explore the application value of gradient boosted decision trees (GBDT) in predicting heart failure (HF) in patients with acute myocardial infarction (AMI) during hospitalization after PCI. **Methods** A total of 200 patients with AMI who underwent PCI in the Second Nanning People's Hospital from 2021 to 2022 were retrospectively selected as the study subjects. The patients were divided into a training set (145 cases) and a test set (55 cases). According to the occurrence of HF during hospitalization after PCI, the patients in training set were divided into the HF group (48 cases) and the non-HF group (97 cases). The general data, laboratory examination indexes and cardiac ultrasound examination indicators before PCI of patients were collected. Based on the results of univariate analysis, GBDT algorithm model and Logistic regression model were constructed by R 4.1.2 software to predict the risk of HF in AMI patients during hospitalization after PCI. ROC curve and calibration curve were used to analyze the discrimination and accuracy of GBDT algorithm model and Logistic regression model. **Results** The age of the HF group was older than that of the non-HF group, the proportion of patients with a history of diabetes, high-sensitivity C-reactive protein (hs-CRP), white blood cell count (WBC), neutrophil

作者单位: 530031 广西壮族自治区南宁市第二人民医院心血管内科

通信作者: 胡文标, E-mail: 281385715@qq.com

count, and creatine kinase isoenzyme (CK-MB) were higher than those in the non-HF group ( $P < 0.05$ ). The indicators with statistically significant differences in univariate analysis were included in the GBDT algorithm model, and the relative importance of these six indicators was obtained through the GBDT algorithm, and the descending order was diabetes history (2.220), neutrophil count (7.713), age (14.734), CK-MB (16.819), WBC (24.828), hs-CRP (33.686). Multivariate Logistic regression analysis showed that age, hs-CRP, WBC, neutrophil count and CK-MB were the influencing factors of HF in AMI patients during hospitalization after PCI in the training set ( $P < 0.05$ ). A Logistic regression model was constructed, with the specific formula as follows:  $\text{logit}(P) = -18.182 + 0.147 \times \text{age} + 0.233 \times \text{hs-CRP} + 0.438 \times \text{WBC} + 0.242 \times \text{neutrophil count} + 0.003 \times \text{CK-MB}$ . The results of the ROC curve analysis showed that the AUC of the GBDT algorithm model and Logistic regression model for predicting HF in AMI patients during hospitalization after PCI in the training set was 0.989 [95%CI (0.974, 1.000)] and 0.864 [95%CI (0.786, 0.942)]. The AUC of the GBDT algorithm model and Logistic regression model for predicting HF in AMI patients during hospitalization after PCI in the test set was 0.900 [95%CI (0.817, 0.982)] and 0.763 [95%CI (0.639, 0.888)]. The results of the calibration curve analysis showed that the probability of HF in AMI patients during hospitalization after PCI in the training set and test set predicted by the GBDT algorithm model and Logistic regression model was consistent with the actual incidence of HF in AMI patients during hospitalization after PCI, respectively. **Conclusion** In this study, the GBDT algorithm model constructed based on age, diabetes history, hs-CRP, WBC, neutrophil count and CK-MB has an excellent predictive value for HF in AMI patients during hospitalization after PCI and is superior to the traditional Logistic regression model, which can provide a reference for the evaluation of prognosis and intervention treatment of AMI patients after PCI.

**【Key words】** Myocardial infarction; Heart failure; Decision trees; Gradient boosted decision trees; Logistic models; Forecasting

急性心肌梗死 (acute myocardial infarction, AMI) 指短时间内给心脏供血的动脉发生病变, 导致冠状动脉血供急剧减少或中断, 引发心肌严重和持久性缺血的心脑血管疾病<sup>[1]</sup>。AMI具有起病急促、病死率高、致残率高等特点, 严重威胁人类健康<sup>[2]</sup>。PCI是治疗AMI及提高患者生存率的有效方式, 但是作为一种有创术式, 其术后易引起各种并发症<sup>[3-4]</sup>。AMI患者院内发生心力衰竭 (heart failure, HF) 是PCI后的常见并发症, 这不仅给患者带来更高的死亡风险, 还带来沉重的经济负担<sup>[5]</sup>。众多研究指出, 早期预测AMI患者住院期间HF发生风险并及时给予相应的干预措施是降低患者HF发生率的重要手段<sup>[6-7]</sup>。而传统Logistic回归模型在处理不平衡样本时, 可能存在过度拟合的情况<sup>[8]</sup>。梯度提升决策树 (gradient boosted decision trees, GBDT) 是机器学习 (machine learning, ML) 中的一种算法, 也是目前网络智能诊疗技术时代的产物, 由于该算法不局限于固定的建模规则, 在处理临床海量无序的数据中展现出比传统Logistic回归模型更大的优势, 利于在临床上进一步推广<sup>[9-10]</sup>。然而, 目前尚未见采用GBDT预测AMI患者住院期间HF发生风险的相关研究。因此, 本研究旨在分析GBDT在预测AMI患者PCI后住院期间发生HF中的应用价值, 以便为AMI患者PCI后住院期间HF发生风险的评估提供新思路与新方法。

## 1 对象与方法

1.1 研究对象 回顾性选取2021—2022年于南宁市第二人民医院行PCI的AMI患者200例为研究对象, 其中男122例, 女78例; 年龄47~83岁, 平均(63.4±6.6)岁; 疾病类型: ST段抬高型心肌梗死75例, 非ST段抬高型心肌梗死125例。纳入标准: (1)符合《急性心肌梗死诊断和治疗指南》<sup>[11]</sup>中AMI的诊断标准; (2)发病至入院时间≤12 h; (3)符合《中国经皮冠状动脉介入治疗指南(2016)》<sup>[12]</sup>中的PCI指征, 且成功接受PCI后血流灌注恢复正常; (4)患者及其家

属对本研究均知情同意。排除标准: (1)有HF史者; (2)非首次接受PCI者; (3)合并严重心、肝、肾功能不全者; (4)存在先天性心脏病、免疫系统疾病、造血功能障碍、恶性肿瘤者; (5)临床资料缺失者。将患者分为训练集(145例)和测试集(55例)。根据PCI后住院期间HF发生情况, 将训练集患者分为HF组(48例)和非HF组(97例)。HF的诊断标准为: (1)体格检查发现肺部啰音、颈静脉充盈、双下肢水肿、心尖冲动侧移或弥散等; (2)心电图异常; (3)X线检查显示肺淤血、肺水肿、心脏扩大等; (4)N末端脑钠肽前体(N-terminal pro-brain natriuretic peptide, NT-proBNP)≥125 ng/L或脑钠肽(brain natriuretic peptide, BNP)≥35 ng/L; (5)超声心动图检查显示心脏结构和/或功能异常<sup>[13]</sup>。

1.2 临床资料收集 从医院电子病历档案系统中收集患者一般资料(包括性别、年龄、发病至入院时间、PCI时间、疾病类型、高血压病史、糖尿病病史、心率)、PCI前实验室检查指标〔肌酐(creatinine, Cr)、总胆固醇(total cholesterol, TC)、肌钙蛋白I(troponin I, TnI)、超敏C反应蛋白(hypersensitive C-reactive protein, hs-CRP)、白细胞计数(white blood cell count, WBC)、中性粒细胞计数、肌酸激酶同工酶(creatine kinase isoenzyme, CK-MB)〕、PCI前心脏彩超检查指标〔左室内径(left ventricular internal diameter, LVID)、左心室射血分数(left ventricular ejection fraction, LVEF)〕。

1.3 统计学方法 采用IBM SPSS statistics 23.0软件进行数据统计分析。符合正态分布的计量资料以( $\bar{x} \pm s$ )表示, 组间比较采用两独立样本 $t$ 检验; 偏态分布的计量资料以 $M(QR)$ 表示, 组间比较采用Wilcoxon秩和检验; 计数资料以相对数表示, 组间比较采用 $\chi^2$ 检验; 基于单因素分析结果, 采用R 4.1.2软件分别构建GBDT算法模型和Logistic回归模型以预测

AMI患者PCI后住院期间HF发生风险；分别采用ROC曲线、校准曲线分析GBDT算法模型、Logistic回归模型的区分度、准确性。以 $P < 0.05$ 为差异有统计学意义。

### 2 结果

2.1 HF组与非HF组临床资料比较 HF组与非HF组性别、发病至入院时间、PCI时间、疾病类型、有高血压病史者占比、心率、Cr、TC、TnI、LVID、LVEF比较，差异无统计学意义 ( $P > 0.05$ )；HF组年龄大于非HF组，有糖尿病病史者占比、hs-CRP、WBC、中性粒细胞计数、CK-MB高于非HF组，差异有统计学意义 ( $P < 0.05$ )，见表1。

2.2 GBDT算法模型及Logistic回归模型的构建与验证 将单因素分析中差异有统计学意义的指标（年龄、糖尿病病史、hs-CRP、WBC、中性粒细胞计数、CK-MB）纳入GBDT算法模型，通过GBDT算法获得这6项指标的相对重要性，由小到大依次为糖尿病病史（2.220）、中性粒细胞计数（7.713）、年龄（14.734）、CK-MB（16.819）、WBC（24.828）、hs-CRP（33.686）。

以年龄（实测值）、糖尿病病史（赋值：有=1，无=0）、hs-CRP（实测值）、WBC（实测值）、中性粒细胞计数（实测值）、CK-MB（实测值）为自变量，训练集AMI患者PCI后住院期间HF发生情况为因变量（赋值：发生=1，未发生=0），进行多因素Logistic回归分析，结果显示，年龄、hs-CRP、WBC、中性粒细胞计数、CK-MB是训练集AMI患者PCI后住院期间发生HF的影响因素 ( $P < 0.05$ )，基于此构建Logistic回归模型，其具体公式为： $\text{logit}(P) = -18.182 + 0.147 \times \text{年龄} + 0.233 \times \text{hs-CRP} + 0.438 \times \text{WBC} + 0.242 \times \text{中性粒细胞计数} + 0.003 \times \text{CK-MB}$ ，见表2。

2.3 GBDT算法模型及Logistic回归模型的验证与比较 ROC曲线分析结果显示，GBDT算法模型、Logistic回归模型预测训练集AMI患者PCI后住院期间发生HF的AUC分别为0.989 [95%CI (0.974, 1.000)]、0.864 [95%CI (0.786, 0.942)]，最佳截断值分别为0.398、-0.717，灵敏度分别为0.957、0.872，特异度分别为0.963、0.827，见图1；GBDT算法模型、Logistic回归模型预测测试集AMI患者PCI后住院期间发生HF的AUC分别为0.900 [95%CI (0.817, 0.982)]、0.763 [95%CI (0.639, 0.888)]，最佳截断值分别为0.196、-0.081，灵敏度分别为0.913、0.609，特异度分别为0.796、0.857，见图2。校准曲线分析结果显示，GBDT算法模型、Logistic回归模型预测训练集、测试集AMI患者PCI后住院期间发生HF的概率分别与本组AMI患者PCI后住院期间HF的实际发生率一致，见图3~6。

### 3 讨论

随着精准医疗战略的逐步推进，精确预测AMI患者住院期间发生HF的风险是优化治疗方案的基石<sup>[14]</sup>。虽然PCI可降低AMI患者院内发生HF的风险，但仍有部分患者术后发生HF。本研究结果显示，训练集145例AMI患者中有48例发生HF，HF发生率为33.1%，高于李沅洋等<sup>[7]</sup>所报道的AMI患者PCI后HF发生率（30.06%），低于杨洋等<sup>[15]</sup>报道的AMI患者PCI后HF发生率（40.15%）。一项Meta分析结果显示，高龄、糖尿病病史、血糖升高、hs-CRP升高、WBC升高、中性

表1 HF组与非HF组临床资料比较

Table 1 Comparison of clinical data between HF group and non-HF group

项目	非HF组 (n=97)	HF组 (n=48)	检验统计量值	P值
性别(男/女)	61/36	27/21	0.593 <sup>a</sup>	0.441
年龄( $\bar{x} \pm s$ , 岁)	61.4 ± 6.2	66.8 ± 6.9	-4.760 <sup>b</sup>	<0.001
发病至入院时间 [M(QR), h]	4.0 (3.0)	4.0 (2.0)	-1.208 <sup>c</sup>	0.227
PCI时间 [M(QR), min]	111 (12)	113 (14)	-1.729 <sup>c</sup>	0.084
疾病类型 [n (%)]			0.284 <sup>a</sup>	0.594
ST段抬高型心肌梗死	34 (35.1)	19 (39.6)		
非ST段抬高型心肌梗死	63 (64.9)	29 (60.4)		
高血压病史 [n (%)]	67 (69.1)	36 (75.0)	0.548 <sup>a</sup>	0.459
糖尿病病史 [n (%)]	29 (29.9)	24 (50.0)	5.596 <sup>c</sup>	0.018
心率( $\bar{x} \pm s$ , 次/min)	94.4 ± 7.4	94.2 ± 8.2	-0.703 <sup>b</sup>	0.483
Cr ( $\bar{x} \pm s$ , $\mu\text{mol/L}$ )	76.4 ± 8.2	78.2 ± 8.2	-1.149 <sup>b</sup>	0.136
TC ( $\bar{x} \pm s$ , mmol/L)	4.34 ± 0.59	4.25 ± 0.73	1.009 <sup>b</sup>	0.314
TnI ( $\bar{x} \pm s$ , $\mu\text{g/L}$ )	2.9 ± 2.0	3.5 ± 2.1	-1.087 <sup>b</sup>	0.059
hs-CRP [M(QR), mg/L]	4.0 (3.9)	7.3 (6.5)	-3.918 <sup>c</sup>	<0.001
WBC [M(QR), $\times 10^9/\text{L}$ ]	8.8 (2.1)	10.1 (2.8)	-3.876 <sup>c</sup>	<0.001
中性粒细胞计数 ( $\bar{x} \pm s$ , $\times 10^9/\text{L}$ )	7.2 ± 2.4	8.8 ± 2.8	-3.608 <sup>b</sup>	<0.001
CK-MB [M(QR), U/L]	131.6 (94.4)	218.1 (106.5)	-4.152 <sup>c</sup>	<0.001
LVID [M(QR), mm]	46.0 (3.0)	45.0 (2.0)	1.609 <sup>c</sup>	0.108
LVEF [M(QR), %]	54.5 (6.0)	52.5 (8.0)	-1.583 <sup>c</sup>	0.114

注：HF=心力衰竭，Cr=肌酐，TC=总胆固醇，TnI=肌钙蛋白I，hs-CRP=超敏C反应蛋白，WBC=白细胞计数，CK-MB=肌酸激酶同工酶，LVID=左室内径，LVEF=左室射血分数；<sup>a</sup>表示 $\chi^2$ 值，<sup>b</sup>表示t值，<sup>c</sup>表示Z值

表2 训练集AMI患者PCI后住院期间发生HF影响因素的多因素Logistic回归分析

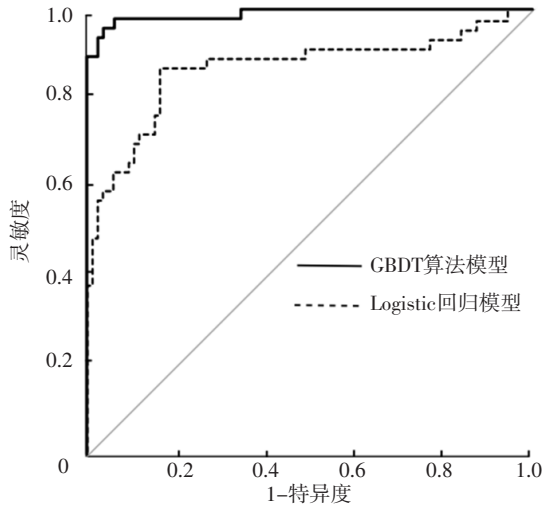
Table 2 Multivariate Logistic regression analysis of influencing factors of HF in AMI patients during hospitalization after PCI in training set

变量	$\beta$	SE	Wald $\chi^2$ 值	P值	OR值	95%CI
年龄	0.147	0.041	12.733	<0.001	1.158	(1.058, 1.256)
糖尿病病史	-0.316	0.512	0.382	0.537	0.729	(0.267, 1.987)
hs-CRP	0.233	0.075	9.624	0.002	1.263	(1.090, 1.463)
WBC	0.438	0.152	8.297	0.004	1.550	(1.150, 2.088)
中性粒细胞计数	0.242	0.093	6.818	0.009	1.274	(1.062, 1.528)
CK-MB	0.003	0.001	6.179	0.013	1.003	(1.001, 1.006)

粒细胞计数升高、CK-MB升高是AMI患者PCI后住院期间发生HF的危险因素<sup>[16]</sup>。鉴于HF的发生发展受多种影响因素影响，仅依靠单一因素预测AMI患者PCI后住院期间HF发生风险可能存在较大偏差，因此需要对可能导致HF的高危因素进行综合考量，进而建立一个模型以对AMI患者PCI后住院期间HF发生风险进行精准预测。本研究利用GBDT算法模型预测AMI患者PCI后住院期间发生HF的风险。

笔者回顾既往研究发现，利用分子生物学技术探索具有更高疾病预测价值的分子标志物虽然是当前研究的热点，但预测标准不同和医疗费用高昂等多种因素的限制导致短时间内无法将其应用于临床<sup>[17]</sup>。因此目前的大多数研究仍立足于现有患者的临床资料，通过构建预测模型来预测AMI患者





注: GBDT=梯度提升决策树

图1 GBDT算法模型、Logistic回归模型预测训练集AMI患者PCI后住院期间发生HF的ROC曲线

Figure 1 ROC curve of GBDT algorithm model and Logistic regression model in predicting HF in AMI patients during hospitalization after PCI in training set

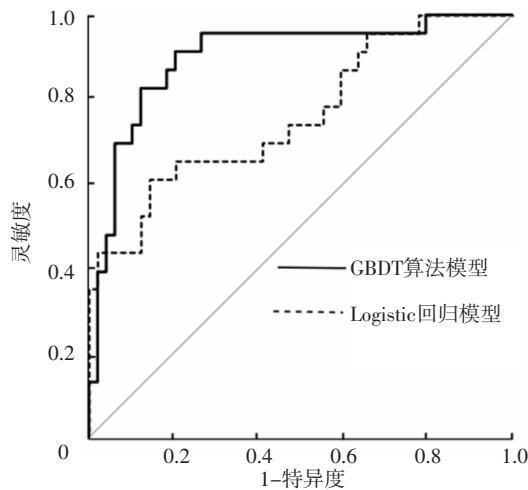


图2 GBDT算法模型、Logistic回归模型预测测试集AMI患者PCI后住院期间发生HF的ROC曲线

Figure 2 ROC curve of GBDT algorithm model and Logistic regression model in predicting HF in AMI patients during hospitalization after PCI in test set

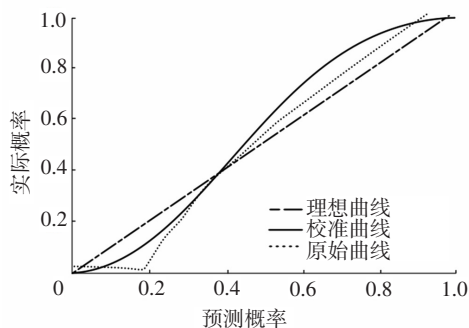


图3 GBDT算法模型预测训练集AMI患者PCI后住院期间发生HF的校准曲线

Figure 3 Calibration curve of GBDT algorithm model in predicting HF in AMI patients during hospitalization after PCI in training set

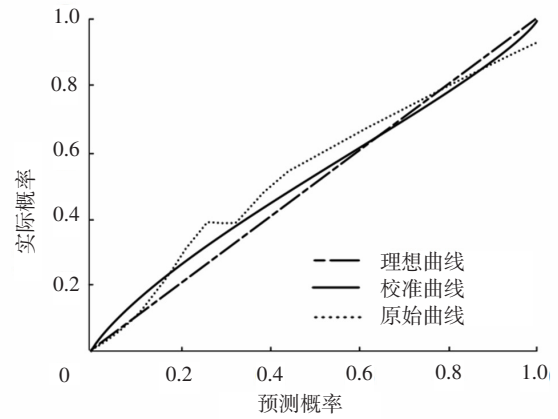


图4 GBDT算法模型预测测试集AMI患者PCI后住院期间发生HF的校准曲线

Figure 4 Calibration curve of GBDT algorithm model in predicting HF in AMI patients during hospitalization after PCI in test set

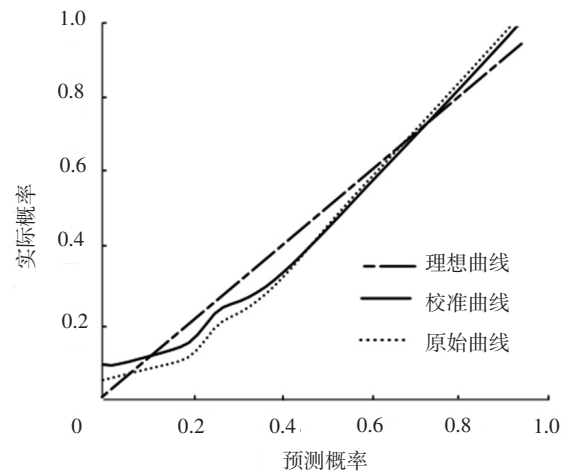


图5 Logistic回归模型预测训练集AMI患者PCI后住院期间发生HF的校准曲线

Figure 5 Calibration curve of Logistic regression model in predicting HF in AMI patients during hospitalization after PCI in training set

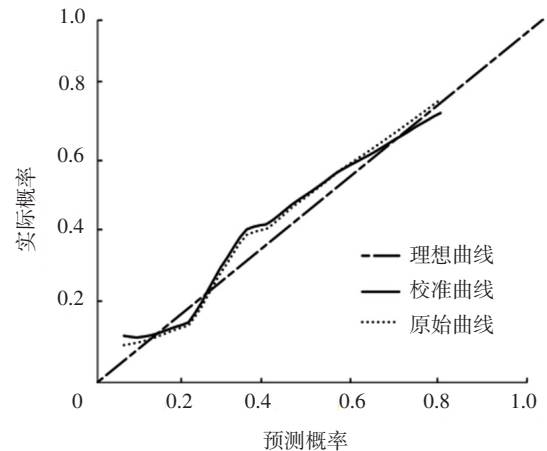


图6 Logistic回归模型预测测试集AMI患者PCI后住院期间发生HF的校准曲线

Figure 6 Calibration curve of Logistic regression model in predicting HF in AMI patients during hospitalization after PCI in test set

PCI后住院期间发生HF的风险<sup>[18-20]</sup>。与此同时,不少研究也相继指出ML模型可以提高疾病预测效能,如范焯等<sup>[19]</sup>研究发现,梯度提升机模型对肝内胆管癌患者手术预后有良好的预测效果,这在大数据时代具有重要的临床价值;ASKARI等<sup>[20]</sup>研究显示,梯度提升随机森林技术在预测COVID-19患者住院时间方面优于其他技术。从GBDT算法模型处理数据的原理上来看,GBDT算法基于多个独立的分类与回归树,可将这些决策树整合成一个强分类器以加强预测的精确度和稳定性,这是一种综合决策的设计思想,不仅可高速处理海量数据,还能减少缺失数值的分析偏倚,更重要的是,GBDT算法模型的结构可拆解,结果也便于临床医师解读<sup>[21]</sup>。

本研究基于单因素分析结果,将年龄、糖尿病病史、hs-CRP、WBC、中性粒细胞计数和CK-MB这6项指标纳入了GBDT算法模型,并确定了各指标的相对重要性。有研究发现,年龄、糖尿病病史均是AMI患者PCI后发生主要不良心血管事件(包含HF)的独立危险因素,其可作为预测模型的相关变量<sup>[22]</sup>。据有关数据统计,年龄>70岁的老年人HF发生率为10%,同时在糖尿病人群中有将近22%的患者发生了HF,是健康人群的4倍<sup>[23-24]</sup>。分析原因为,高龄会降低心脏功能,改变心脏结构,极易影响患者血管顺应性,提高内皮功能障碍和心室重塑等不良事件的发生率,从而导致不良预后(HF)的发生<sup>[25]</sup>。其次,有研究表明,糖尿病与AMI患者心肌功能密切相关,合并糖尿病的AMI患者PCI后心脏舒张功能障碍发生风险明显升高<sup>[26]</sup>。但本研究中糖尿病病史没有进入Logistic回归方程,可能是选择偏倚导致的。此外,一些实验室检查指标如hs-CRP、WBC、中性粒细胞计数、CK-MB等也可影响AMI患者PCI后HF的发生发展<sup>[16]</sup>,本研究结果与之相似。hs-CRP属于炎症因子之一,对心血管疾病(如冠状动脉粥样硬化性心脏病)及患者PCI后的远期预后具有一定预测价值<sup>[27]</sup>。WBC水平升高是合并冠心病的2型糖尿病患者发生HF的预测因子,可以反映患者左心室功能和心肌损伤程度<sup>[28]</sup>,中性粒细胞计数升高则提高了AMI患者早期充血性HF的发生率<sup>[29]</sup>,CK-MB作为心肌损伤标志物,是衡量心肌受损程度的敏感指标,既往研究结果也表明,CK-MB升高要预防AMI患者PCI后院内发生HF<sup>[30]</sup>。

本研究ROC曲线分析结果显示,GBDT算法模型、Logistic回归模型预测训练集AMI患者PCI后住院期间发生HF的AUC分别为0.989、0.864,预测测试集AMI患者PCI后住院期间发生HF的AUC分别为0.900、0.763,提示GBDT算法模型对AMI患者PCI后住院期间发生HF的预测效能优于Logistic回归模型,与国外一项应用GBDT算法模型预测重症监护病房脓毒症患者住院死亡率的研究结果<sup>[31]</sup>相似。综上,GBDT算法模型不同于Logistic回归模型的线性处理思想,强大的分类决策功能使其处理数据的效率不断提高,这在某种程度上可弥补Logistic回归模型的不足,从而进一步提高模型的预测准确度,也避免了线性处理的过度拟合情况<sup>[32-33]</sup>。

综上所述,本研究基于年龄、糖尿病病史、hs-CRP、WBC、中性粒细胞计数和CK-MB 6个指标构建的GBDT算法模型对AMI患者PCI后住院期间发生HF有较好的预测价值,且优于传统Logistic回归模型,这可为AMI患者PCI后预后的评估及干预治疗提供参考依据。但本研究为单中心回顾性研究,可

能存在一定选择偏倚。未来若想进一步提高GBDT算法模型的预测准确度,一方面可选择前瞻性研究;另一方面,随着对ML算法的深入研究,可以进行如随机森林、人工神经网络等多种技术的对比探索,为优化预测模型提供参考。

作者贡献:胡文标进行文章的构思与设计,文章的可行性分析,文献/资料收集、整理,撰写论文,并对文章整体负责、监督管理;胡文标、刘礼敬、林晓峰、韦佳玲进行论文修订;胡文标、宋清华、陆元喜负责文章的质量控制及审校。

本文无利益冲突。

#### 参考文献

- [1] COLLET J P, THIELE H, BARBATO E, et al. 2020 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation [J]. *Eur Heart J*, 2021, 42 (14): 1289-1367. DOI: 10.1093/eurheartj/ehaa575.
- [2] ZHANG C H, LI T T, CHEN X. Acute myocardial infarction or not? [J]. *JAMA Intern Med*, 2022, 182 (6): 668-669. DOI: 10.1001/jamainternmed.2022.0519.
- [3] AUER J, BERENT R, GURTNER F. Complete revascularization with multivessel PCI for myocardial infarction [J]. *N Engl J Med*, 2020, 382 (16): 1569. DOI: 10.1056/NEJMc2000278.
- [4] HUANG D, GAO W, WU R D, et al. D-dimer level predicts in-hospital adverse outcomes after primary PCI for ST-segment elevation myocardial infarction [J]. *Int J Cardiol*, 2020, 305: 1-4. DOI: 10.1016/j.ijcard.2020.02.010.
- [5] 徐芳, 徐俊杰, 吴春苑, 等. 左室心肌做功对急性心肌梗死患者急诊PCI术后MACE发生的预测价值 [J]. *临床心血管病杂志*, 2022, 38 (12): 960-966. DOI: 10.13201/j.issn.1001-1439.2022.12.007.
- [6] 曾维斌, 吴爱梅, 蒋佳莉, 等. 急性心肌梗死患者PCI术后1年发生主要心血管事件的影响因素及预测模型构建 [J]. *中国老年学杂志*, 2022, 42 (20): 4905-4908. DOI: 10.3969/j.issn.1005-9202.2022.20.001.
- [7] 李沅洋, 张宇凡, 徐月, 等. 急性心肌梗死患者PCI术后心力衰竭风险预测模型建立与评估 [J]. *临床心血管病杂志*, 2019, 35 (10): 916-922. DOI: 10.13201/j.issn.1001-1439.2019.10.012.
- [8] NATTINO G, PENNELL M L, LEMESHOW S. Assessing the goodness of fit of logistic regression models in large samples: a modification of the Hosmer-Lemeshow test [J]. *Biometrics*, 2020, 76 (2): 549-560. DOI: 10.1111/biom.13249.
- [9] NGIAM K Y, KHOR I W. Big data and machine learning algorithms for health-care delivery [J]. *Lancet Oncol*, 2019, 20 (5): e262-273. DOI: 10.1016/S1470-2045(19)30149-4.
- [10] EATON J E, VESTERHUS M, MCCAULEY B M, et al. Primary sclerosing cholangitis risk estimate tool (PREsTo) predicts outcomes of the disease: a derivation and validation study using machine learning [J]. *Hepatology*, 2020, 71 (1): 214-224. DOI: 10.1002/hep.30085.
- [11] 中华医学会心血管病学分会, 中华心血管病杂志编辑委员会, 《中国循环杂志》编辑委员会. 急性心肌梗死诊断和治疗指南 [J]. *中华心血管病杂志*, 2001, 29 (12): 710-725. DOI: 10.3760/j.issn.0253-3758.2001.12.003.
- [12] 中华医学会心血管病学分会介入心脏病学组, 中国医师协会心

- 血管内科医师分会血栓防治专业委员会, 中华心血管病杂志编辑委员会. 中国经皮冠状动脉介入治疗指南(2016) [J]. 中华心血管病杂志, 2016, 44(5): 382-400. DOI: 10.3760/ema.j.issn.0253-3758.2016.05.006.
- [13] 中国医师协会心血管内科医师分会, 中国心血管健康联盟, 心肌梗死后心力衰竭防治专家共识工作组. 2020心肌梗死后心力衰竭防治专家共识 [J]. 中国循环杂志, 2020, 35(12): 1166-1180. DOI: 10.3969/j.issn.1000-3614.2020.12.002.
- [14] 卢永昕. HFpEF精准医疗时代到来了吗? [J]. 临床心血管病杂志, 2019, 35(4): 295-296. DOI: 10.13201/j.issn.1001-1439.2019.04.001.
- [15] 杨洋, 徐彧, 袁锦霞, 等. 血清miR-145、miR-150表达水平与急性心肌梗死患者PCI术后心力衰竭的关系 [J]. 国际检验医学杂志, 2022, 43(4): 414-419. DOI: 10.3969/j.issn.1673-4130.2022.04.007.
- [16] 屈俊超, 薛孟欣, 孙慧平, 等. 急性心肌梗死患者PCI后住院期间发生心力衰竭影响因素的Meta分析 [J]. 实用心脑血管病杂志, 2023, 31(1): 77-84. DOI: 10.12114/j.issn.1008-5971.2023.00.001.
- [17] WANG X Y, ZHANG F, ZHANG C, et al. The biomarkers for acute myocardial infarction and heart failure [J]. Biomed Res Int, 2020, 2020: 2018035. DOI: 10.1155/2020/2018035.
- [18] PAYROVNAZIRI S N, BARRETT L A, BIS D, et al. Enhancing prediction models for one-year mortality in patients with acute myocardial infarction and post myocardial infarction syndrome [J]. Stud Health Technol Inform, 2019, 264: 273-277. DOI: 10.3233/SHTI190226.
- [19] 范焯, 季顾惟, 许正刚, 等. 梯度提升机模型在肝内胆管癌手术预后预测中应用价值研究 [J]. 中国实用外科杂志, 2022, 42(2): 172-178. DOI: 10.19538/j.cjps.issn1005-2208.2022.02.10.
- [20] ASKARI G, ROUHANI M H, SATTARI M. Prediction of length of hospital stay of COVID-19 patients using gradient boosting decision tree [J]. Int J Biomater, 2022, 2022: 6474883. DOI: 10.1155/2022/6474883.
- [21] LIANG Y, MA X. iACP-GE: accurate identification of anticancer peptides by using gradient boosting decision tree and extra tree [J]. SAR QSAR Environ Res, 2023, 34(1): 1-19. DOI: 10.1080/1062936X.2022.2160011.
- [22] 张萌, 许艳, 郑红艳. 急性心肌梗死患者经皮冠状动脉介入术后1年内再入院风险预测列线图模型构建与评估 [J]. 实用心脑血管病杂志, 2022, 30(3): 17-21. DOI: 10.12114/j.issn.1008-5971.2022.00.058.
- [23] MCDONAGH T A, 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. DOI: 10.1093/eurheartj/ehab368.
- [24] DUNLAY S M, GIVERTZ M M, AGUILAR D, et al. Type 2 diabetes mellitus and heart failure: a scientific statement from the American Heart Association and the Heart Failure Society of America: this statement does not represent an update of the 2017 ACC/AHA/HFSA heart failure guideline update [J]. Circulation, 2019, 140(7): e294-324. DOI: 10.1161/CIR.0000000000000691.
- [25] 戚德青, 刘朵, 蒲强, 等. 不同年龄急性心肌梗死患者经皮冠状动脉介入治疗后的临床特点及预后 [J]. 贵州医科大学学报, 2020, 45(9): 1082-1087. DOI: 10.19367/j.cnki.2096-8388.2020.09.018.
- [26] HEIDENREICH P A, BOZKURT B, AGUILAR D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines [J]. J Am Coll Cardiol, 2022, 79(17): 1757-1780. DOI: 10.1016/j.jacc.2021.12.011.
- [27] 马军, 彭毅, 樊光辉. 急性非ST段抬高型心肌梗死患者合并心力衰竭的危险因素分析 [J]. 华南国防医学杂志, 2018, 32(4): 234-237. DOI: 10.13730/j.issn.1009-2595.2018.04.005.
- [28] KAWABE A, YASU T, MORIMOTO T, et al. WBC count predicts heart failure in diabetes and coronary artery disease patients: a retrospective cohort study [J]. ESC Heart Fail, 2021, 8(5): 3748-3759. DOI: 10.1002/ehf2.13513.
- [29] VULESEVIC B, SIROIS M G, ALLEN B G, et al. Subclinical inflammation in heart failure: a neutrophil perspective [J]. Can J Cardiol, 2018, 34(6): 717-725. DOI: 10.1016/j.cjca.2018.01.018.
- [30] 孙梦涵, 王薇, 李勇, 等. ST段抬高型心肌梗死患者直接PCI后住院期间发生心力衰竭的相关因素分析 [J]. 慢性病学杂志, 2020, 21(6): 807-812. DOI: 10.16440/j.cnki.1674-8166.2020.06.004.
- [31] LI K, SHI Q W, LIU S R, et al. Predicting in-hospital mortality in ICU patients with sepsis using gradient boosting decision tree [J]. Medicine, 2021, 100(19): e25813. DOI: 10.1097/MD.00000000000025813.
- [32] SUFRIYANA H, HUSNAYAIN A, CHEN Y L, et al. Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: systematic review and meta-analysis [J]. JMIR Med Inform, 2020, 8(11): e16503. DOI: 10.2196/16503.
- [33] SETO H, OYAMA A, KITORA S, et al. Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data [J]. Sci Rep, 2022, 12(1): 15889. DOI: 10.1038/s41598-022-20149-z.

(收稿日期: 2023-03-06; 修回日期: 2023-04-26)

(本文编辑: 崔丽红)