http://www.zjujournals.com/med
DOI:10.3785/j.issn.1008-9292.2019.12.02

·原 著·

决策树分析在急性心肌梗死事件预测中的应用

张 圣[1],胡振杰[2],叶 璐[3],郑亚如[4]
1. 浙江省人民医院 杭州医学院附属人民医院神经内科, 浙江 杭州 310014
2. 中国人民解放军联勤保障部队第九〇六医院呼吸与重症医学科, 浙江 宁波 315040
3. 浙江大学医学院精神卫生中心暨杭州市第七人民医院检验科, 浙江 杭州 310013
4. 浙江省人民医院 杭州医学院附属人民医院心血管内科, 浙江 杭州 310014

[摘 要] **目的**:评价和比较 Logistic 回归和决策树分析用于预测急性心肌梗死(AMI)事件的可行性和有效性。**方法**:回顾性分析 2018 年 10 月至 2019 年 4 月在浙江省人民医院因心绞痛或不明原因胸痛行选择性冠状动脉造影的 295 例患者的临床资料,其中 55 例诊断为 AMI。分别利用 Logistic 回归分析和决策树分析建立 AMI 事件预测模型,并在是否根据 Logistic 回归结果条件下建立决策树分析模型(决策树 1 和决策树 2),继而利用 ROC 曲线评估上述三组模型预测 AMI 的价值。**结果**:二元 Logistic 回归分析结果显示,冠心病史、冠状动脉多支病变、他汀类药物史和载脂蛋白 A1 是 AMI 发生的独立影响因素(均 $P<0.05$)。不根据 Logistic 回归分析结果建立的决策树模型(决策树 1)显示,冠状动脉多支病变为根节点,其后分别是冠心病史、载脂蛋白 A1 水平(以 1.314 g/L 作为分界点)和抗血小板聚集药物史作为子节点;而根据 Logistic 回归分析结果建立的决策树模型(决策树 2)显示,冠状动脉多支病变为根节点,其后是冠心病史和载脂蛋白 A1 作为子节点。在对 AMI 事件的预测中,Logistic 回归模型的 AUC 为 0.826,而决策树模型的 AUC 分别为 0.765(决策树 1)和 0.726(决策树 2)。三组模型间比较结果显示,Logistic 回归模型的 AUC 优于决策树 2(95% *CI*:0.041 ~ 0.145,$Z=3.534$,$P<0.01$),但与决策树 1 差异无统计学意义(95% *CI*: −0.014 ~ 0.121,$Z=-1.173$,$P>0.05$)。**结论**:在对 AMI 事件的预测分析中,不根据 Logistic 回归模型结果建立的决策树模型效力与 Logistic 回归模型相当,未来有望应用于 AMI 患者的防治工作。

[关键词] 心肌梗死; 急性病; Logistic 模型; 回归分析; 决策树; 预测
[中图分类号] R542.2+2 [文献标志码] A

Application of Logistic regression and decision tree analysis in prediction of acute myocardial infarction events

ZHANG Sheng[1], HU Zhenjie[2], YE Lu[3], ZHENG Yaru[4] (*1. Department of Neurology*,

收稿日期:2019-06-05 **接受日期**:2019-07-31
基金项目:国家自然科学基金(81801162);浙江省医学会临床科研资金项目(2017XYC-A02)
第一作者:张 圣(1986—),女,博士,主治医师,主要从事心脑血管疾病的临床研究;E-mail: xiaoxiaoqing_23@hotmail.com; https://orcid.org/0000-0003-0644-7930
通信作者:郑亚如(1990—),女,硕士,主治医师,主要从事冠心病方面的临床和基础研究;E-mail: zhengyaru@zjheart.com; https://orcid.org/0000-0003-2113-2435

Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou 310014, China; 2. Department of Respiratory and Critical Medicine, No. 906 Hospital of Chinese PLA, Ningbo 315040, China; 3. Clinical Laboratory, Mental Health Center of Zhejiang University School of Medicine, Hangzhou Seventh People's Hospital, Hangzhou 310013, China; 4. Department of Cardiology, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou 310014, China)
Corresponding author: ZHENG Yaru, E-mail: zhengyaru@zjheart.com, https://orcid.org/0000-0003-2113-2435

[**Abstract**] **Objective**: To evaluate the application of decision tree method and Logistic regression in the prediction of acute myocardial infarction (AMI) events. **Methods**: The clinical data of 295 patients, who underwent coronary angiography due to angina or chest pain with unidentified causes in Zhejiang provincial People's Hospital during October 2018 and April 2019, were retrospectively analyzed. Fifty five patients were identified as AMI. Logistic regression and decision tree methods were performed to establish predictive models for the occurrence of AMI, respectively; and the models created by decision tree analysis were divided into Logistic regression-independent model (Tree 1) and Logistic regression-dependent model (Tree 2). The performance of Logistic regression and decision tree models were compared using the area under the receiver operating characteristic (ROC) curve. **Results**: Logistic regression analysis showed that history of coronary artery disease, multi-vessel coronary artery disease, statin use and apolipoprotein (ApoA1) level were independent influencing factors of AMI events (all $P<0.05$). Logistic regression-independent decision tree model (Tree 1) showed that multi-vessel coronary artery disease was the root node, and history of coronary artery disease, ApoA1 level (the cutoff value: 1.314 g/L) and anti-platelet drug use were descendant nodes. In Logistic regression-dependent decision tree model (Tree 2), multi-vessel coronary artery disease was still the root node, but only followed by two descendant nodes including history of coronary artery disease and ApoA1 level. The area under the curve (AUC) of ROC of Logistic regression model was 0.826, and AUCs of decision tree models were 0.765 and 0.726, respectively. AUC of Logistic regression model was significantly higher than that of Tree 2 (95%CI = 0.041 – 0.145, Z = 3.534, $P<0.001$), but was not higher than that of Tree 1 (95%CI = –0.014 – 0.121, Z = –1.173, $P>0.05$). **Conclusion**: The predictive value for AMI event was comparable between Logistic regression-independent decision tree model and Logistic regression model, implying the data mining methods are feasible and effective in AMI prevention and control.

[**Key words**] Myocardial infarction; Acute disease; Logistic models; Regression analysis; Decision trees; Forecasting

[J Zhejiang Univ (Med Sci), 2019,48(6):594-602.]

急性心肌梗死(acute myocardial infarction, AMI)是危害我国居民生命和健康的重大疾病。近十年接受经皮冠状动脉介入治疗和冠状动脉旁路移植术治疗的病例数逐年增加,AMI 患者病死率呈现快速上升趋势[1-2]。1987 至 2014 年,因 AMI 死亡人数增加了 5.6 倍,其中城市人口 AMI 病死率从 2005 年的 11.3/10 万人增加到 2013 年的 51.46/10 万人,农村人口 AMI 病死率从 21.5/

10 万人上升到 66.62/10 万人[3]。目前我国至少有 250 万 AMI 患者,而且数量仍在不断攀升[4]。因此,AMI 早期防治极为重要。

研究结果表明,对公众进行早期高危人群筛查和健康管理能够明显减少 AMI 发病人数[5-6],如能有效提前预测 AMI 事件,将能有效控制 AMI 的发生率,从而达到一级预防的目的。既往大量研究采用 Logistic 回归模型建立了 AMI 的预测模型,但在过去数十年间,人工智能学习,尤其是决策树分析被广泛运用到医疗卫生的各个领域,并已证实对某些疾病的高危筛查有较好的效果[7-8]。比较两种分析方法,Logistic 回归对数据整体结构的分析优于决策树,而决策树对局部结构的分析优于 Logistic 回归[9]。因此,本研究比较两种分析方法建立模型预测 AMI 事件的效力,为未来防治 AMI 事件建立更为准确可靠的模型提供依据和新思路。

1 对象与方法

1.1 对 象

收集 2018 年 10 月至 2019 年 4 月因心绞痛或不明原因胸痛在浙江省人民医院行选择性冠状动脉造影患者的临床资料。综合临床症状和检查结果后,将患者分为 AMI 组和非 AMI 组。AMI 诊断依据第四版"全球心肌梗死定义"标准[10]。排除标准:入院后未能完成相关检查,临床及影像资料记录不完整者。最终 295 例患者纳入研究,平均年龄为 65 岁(IQR:55 ~ 72 岁),女性占 37.3%(110/295),AMI 组和非 AMI 组分别为 55 和 240 例。

本研究经浙江省人民医院伦理委员会批准(2015KY186),所有研究对象均知情同意。

1.2 检查方法及诊断标准

1.2.1 冠状动脉造影检查 采用贾金斯(Judkins)法进行冠状动脉造影。冠状动脉多支病变定义为累及左前降支、回旋支及右冠状动脉之中两支或两支以上狭窄程度超过 50%的病变(病变累及左主干时,视为同时累及前降支和回旋支)[11]。

1.2.2 颈动脉超声检查 采用美国 GE LOGIQ E9 超声诊断仪,M6-15 探头。所有患者均于入院 5 d 内进行颈动脉超声检查。患者取仰卧位,检查者从锁骨内侧端颈总动脉的起始段将探头沿血管走行向头部移动,对双侧颈总动脉、颈外、颈内动脉以及分叉部位依次进行检查,观察颈动脉血管壁有无斑块。不稳定斑块定义包括以下情况:①有不完全纤维帽或溃疡的斑块,根据斑块形态,纤维帽是一层比正常颈动脉内膜更厚但细胞成分更少的纤维结缔组织,而溃疡斑块的溃疡深度至少为 2 mm;②根据斑块的超声特点,将其分为高回声、等回声、低回声和混合回声斑块,其中低回声和混合回声斑块定义为不稳定斑块[12]。

颈内动脉狭窄诊断标准采用国际通用标准,狭窄程度分为:正常或少于 50%、50%~69%、70%~99%、闭塞。将颈内动脉狭窄程度 50%及以上定义为颈动脉狭窄[13]。

颈动脉斑块面积为血管纵切面图像上的最大面积。最大斑块面积的定义为左、右两侧颈总动脉和窦部 4 个血管段最大斑块的面积[14]。

1.3 统计学方法

采用 SPSS 20.0 软件进行数据分析。计数资料采用频数和百分率[n(%)]描述,计量资料采用均数 ± 标准差($\bar{x} \pm s$)或中位数和四分位数[M(IQR)]描述。两组计量资料比较采用 t 检验或 Mann-Whitney 非参数检验,两组计数资料比较采用 χ^2 检验。各变量间的相关性分析采用 Spearman 法。所有检验显著性水平均为 $\alpha = 0.05$,$P < 0.05$ 表示差异具有统计学意义。

1.3.1 Logistic 回归模型的建立 将单因素分析中 $P < 0.05$ 的变量作为自变量,以此次住院诊断新发 AMI 事件为因变量,建立二元 Logistic 回归分析模型,用优势比(OR)及 95%CI 表示该因素与 AMI 发生的联系强度。

1.3.2 决策树模型的建立 分别在是否根据 Logistic 回归模型结果的条件下建立决策树模型,并将两个模型分别标记为决策树 1 和决策树 2。

采用 SPSS 20.0 软件决策树分析中的分类回归树(classification and regression tree, C&RT)法建立树模型。决策树生长"枝条"分割显著性检验水准定位 $\alpha_{merge} = \alpha_{aplit} = 0.05$。时限指定父节点上的最小样本量为 50,子节点上的最小样本量为 10,如节点上的样本量达不到此要求,则该节点为终末节点,不再进行分割。使用 10 倍交叉验证进行决策树计算效果的验证。采用准确度、敏感度、特异度和约登指数描述所建立的决策树模型内部验证的预测价值。

另采用随机抽取样本并按照 3∶1的比例将原

数据集切分成训练集和测试集,其中训练集用于建立筛查模型,然后将所得模型应用于测试集人群,以评价模型的实际应用效果。

1.3.3　Logistic 回归与决策树模型比较　以 Logistic 回归模型和决策树模型的预测概率为分析变量,有无发生 AMI 为分类变量,进行 ROC 曲线的绘制与分析。采用 Medcalc 15.0 软件对 Logistic 回归分析和决策树分析得到的 ROC 结果进行比较,$P<0.05$ 为差异具有统计学意义。

2　结　果

2.1　发生 AMI 影响因素的单因素分析结果

单因素分析结果显示,两组患者间年龄、冠心病史、抗血小板聚集和他汀类药物史、冠状动脉支架植入史、白细胞计数、低密度脂蛋白、载脂蛋白 A1、冠状动脉多支病变及冠状动脉各支狭窄程度差异有统计学意义(均 $P<0.05$),见表 1。

2.2　发生 AMI 影响因素的 Logistic 回归分析结果

根据单因素分析结果,将 $P<0.05$ 的变量纳入二元 Logistic 回归分析。结果显示,冠心病史、他汀类药物史和载脂蛋白 A1、冠状动脉多支病变是 AMI 发生的独立影响因素($P<0.05$ 或 $P<0.01$),见表 2。

2.3　发生 AMI 影响因素的决策树分析结果

在不根据 Logistic 回归分析结果的情况下建立决策树模型(决策树 1),结果显示,冠状动脉多支病变作为根节点,其后分别是冠心病史、载脂蛋白 A1 水平(以 1.314 g/L 作为分界点)和抗血小板聚集药物史作为子节点(图 1)。相关性分析结果显示,他汀类药物史与冠心病史(Rho = 0.368,$P<0.01$)及抗血小板聚集药物史(Rho = 0.761,$P<0.01$)呈正相关性。结果提示,在决策树分析中,他汀类药物史未能成为独立子节点的原因可能是其效应被另外两个相关因素所覆盖。

如将 Logistic 回归中有统计学意义的四个变量纳入决策树分析(决策树 2),结果显示,树模型以冠状动脉多支病变作为根节点,其后是冠心病史和载脂蛋白 A1 作为子节点(图 2)。

进一步将原数据集拆分为训练集和检验集,并重新进行决策树分析。结果显示,与拆分前相比,进入决策树 1 和决策树 2 模型的根节点与子节点均无变化(表 3)。

表 1　发生 AMI 影响因素的单因素分析结果

Table 1　Univariate analysis on predicting factors for AMI

[$M(IQR)$或$n(\%)$或$\bar{x}\pm s$]

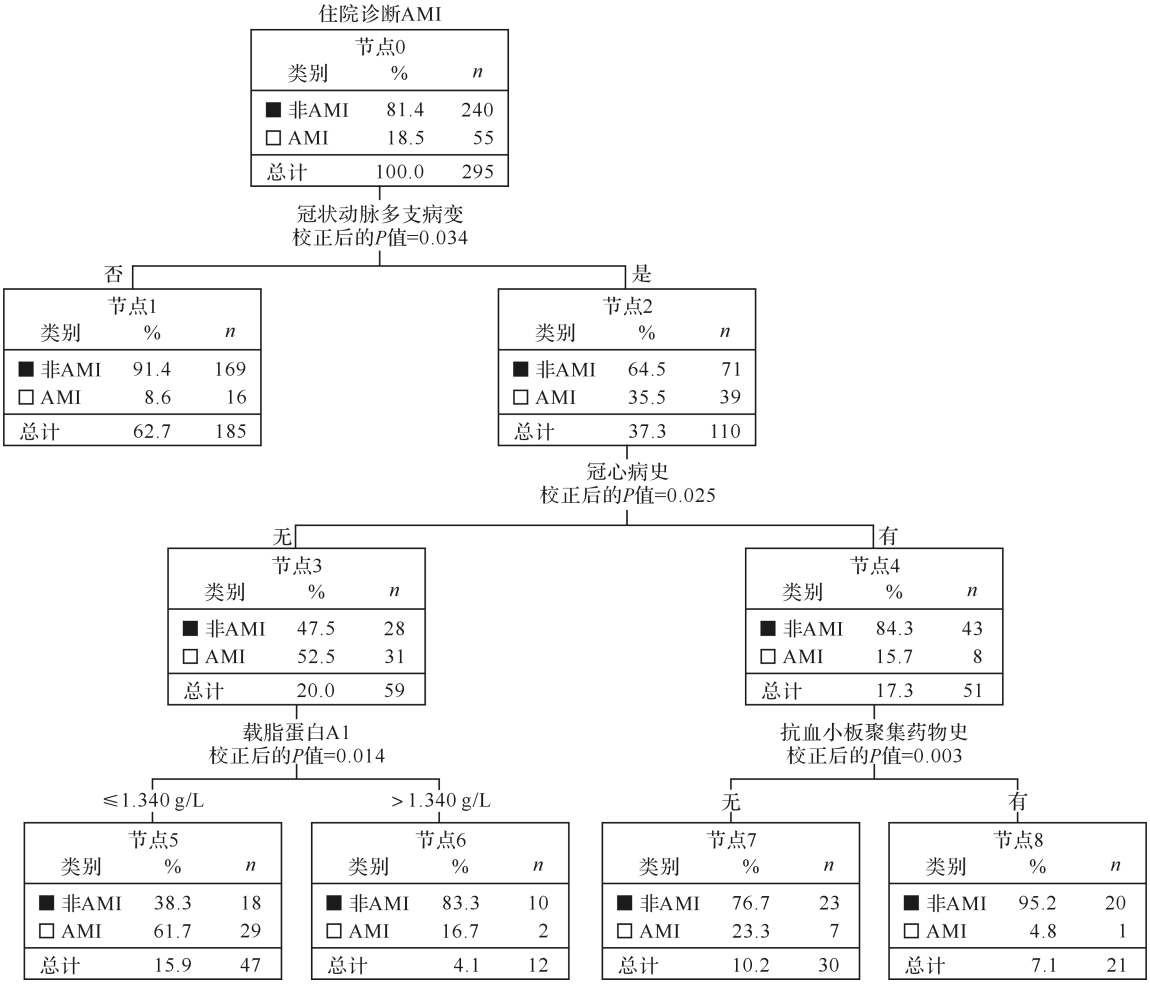
变　量	非 AMI 组(n=240)	AMI 组(n=55)	P 值
年龄(岁)	65(54~71)	66(60~76)	<0.05
女性	95(39.6)	15(27.3)	>0.05
高血压病史	153(36.3)	34(61.8)	>0.05
糖尿病史	49(20.4)	16(29.1)	>0.05
心房颤动史	15(6.3)	7(12.7)	>0.05
冠心病史	95(39.6)	11(20.0)	<0.01
心肌梗死史	10(4.2)	1(1.8)	>0.05
脑卒中史	11(4.6)	3(5.5)	>0.05
抗血小板聚集药物史	57(23.8)	3(5.5)	<0.01
他汀类药物史	59(24.6)	2(3.6)	<0.01
抗凝药物史	5(2.1)	0(0)	>0.05
冠状动脉支架植入史	45(18.8)	2(3.6)	<0.01
白细胞计数($\times10^9$/L)	6.2±1.8	8.2±3.0	<0.01
丙氨酸转氨酶(U/L)	19(14~27)	30(15~46)	>0.05
天冬氨酸转氨酶(U/L)	22(19~28)	60(25~201)	>0.05
肌酐(μmol/L)	85.7±40.0	87.0±26.1	>0.05
总胆固醇(mmol/L)	4.2±1.3	4.2±1.0	>0.05
三酰甘油(mmol/L)	1.6±1.7	1.1±0.5	>0.05
高密度脂蛋白(mmol/L)	1.1±0.4	1.1±0.2	>0.05
低密度脂蛋白(mmol/L)	2.2±0.9	2.6±0.8	<0.05
载脂蛋白 A1(g/L)	1.2±0.2	1.1±0.2	<0.01
载脂蛋白 B(g/L)	0.7±0.2	0.8±0.2	>0.05
左前降支狭窄程度(%)	42.3±34.3	77.0±27.6	<0.01
左回旋支狭窄程度(%)	0(0~50)	70(0~90)	<0.01
右冠状动脉狭窄程度(%)	30(0~50)	70(30~90)	<0.01
冠状动脉多支病变	71(29.6)	39(70.9)	<0.01
颈动脉斑块	104(43.3)	23(41.8)	>0.05
最大斑块面积(mm^2)	0(0~18.4)	0(0~37.1)	>0.05
低回声斑块	57(23.8)	11(20.0)	>0.05
高回声斑块	54(22.5)	8(14.5)	>0.05
混合回声斑块	33(13.8)	10(18.2)	>0.05
不稳定斑块	79(32.9)	19(34.5)	>0.05
颈动脉狭窄	13(5.4)	5(9.1)	>0.05

AMI:急性心肌梗死.

表 2　预测急性心肌梗死事件的二元 Logistic 回归分析结果

Table 2　Binary Logistic regression analysis for predicting acute myocardial infarction

变　量	OR 值	95% CI	P 值
冠心病史	0.280	0.116~0.673	<0.01
抗血小板聚集药物史	1.368	0.280~6.673	>0.05
他汀类药物史	0.060	0.006~0.638	<0.05
冠状动脉支架植入史	0.593	0.095~3.691	>0.05
低密度脂蛋白	0.945	0.664~1.345	>0.05
天冬氨酸转氨酶	1.000	0.999~1.002	>0.05
载脂蛋白 A1	0.112	0.020~0.626	<0.05
冠状动脉多支病变	8.981	4.216~19.128	<0.01



AMI:急性心肌梗死.

图 1　不根据 Logistic 回归分析结果建立的预测 AMI 事件的决策树模型

Figure 1　Logistic regression-independent decision tree analysis for predicting AMI

2.4　Logistic 回归模型与决策树模型预测 AMI 效力比较

Logistic 回归模型预测准确度为 86.2%,AUC 为 0.826(95%*CI*:0.762 ~ 0.889)。根据总体样本(未拆分数据集)得到的决策树分析结果,决策树 1 的预测准确度为 85.4%,AUC 为 0.765(95% *CI*:0.645 ~ 0.816);决策树 2 的预测准确度为 85.1%,AUC 为0.726(95%*CI*:0.641 ~ 0.812),见表 4 和图 2。结果提示,Logistic 回归模型和决策树 1 模型均对 AMI 具有良好的预测价值,而决策树 2 对 AMI 事件的预测能力中等。

三组模型间比较结果显示,Logistic 回归模型的 AUC 优于决策树 2(95%*CI*:0.041 ~ 0.145,Z = 3.534,P < 0.01),但与决策树 1 差异无统计学意义(95% *CI*:0.014 ~ 0.121,Z = −1.173,P > 0.05)。

表 3　拆分数据集后进入决策树模型的变量

Table 3　Variables selected by decision tree models after splitting the dataset

决策树		级别*	P 值#
决策树 1	冠状动脉多支病变	1	0.031
	冠心病史	2	0.045
	载脂蛋白 A1	3	0.020
	抗血小板聚集药物史	3	0.001
决策树 2	冠状动脉多支病变	1	0.035
	冠心病史	2	0.027
	载脂蛋白 A1	3	0.004

*各变量在决策树中所处的节点级别(如级别 1 为根节点,2 和 3 为级别递降的子节点);#基于各变量所在节点的数据拆分后比较分析得出.

2.5　预测冠状动脉多支病变的相关因素分析

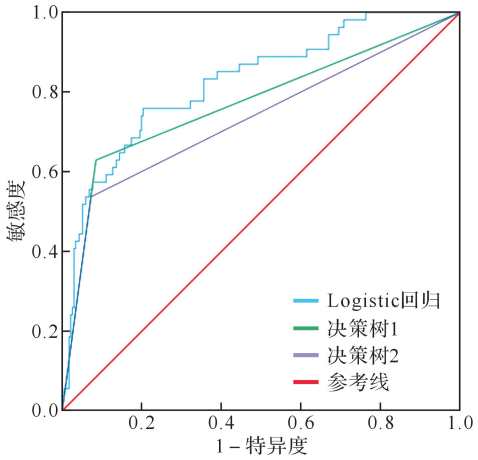
单因素分析结果显示,冠状动脉多支病变与

表 4　Logistic 回归和决策树模型预测 AMI 的 ROC 曲线分析结果

Table 4　Comparison between Logistic regression model and decision tree model in predicting AMI

模　型	AUC	标准误	*P* 值	95%*CI*	准确度(%)	敏感度(%)	特异度(%)	约登指数
Logistic 回归	0.826	0.032	<0.01	0.762 ~ 0.889	86.2	75.9	79.7	0.56
决策树 1	0.765	0.041	<0.01	0.684 ~ 0.846	85.4	61.8	91.2	0.53
决策树 2	0.726	0.044	<0.01	0.641 ~ 0.812	85.1	52.7	92.5	0.45

AMI:急性心肌梗死.



AMI:急性心肌梗死.

图 2　Logistic 回归和决策树模型预测 AMI 的 ROC 曲线

Figure 2　ROCs of Logistic regression model and decision tree model in predicting AMI

颈动脉狭窄、不稳定斑块、颈动脉最大斑块面积、年龄、女性、冠心病史和糖尿病史相关(均 $P<0.05$),见表 5。将上述因素纳入二元 Logistic 回归分析显示,女性($OR=0.463$,95% CI:0.266 ~ 0.809,$P<0.01$)、颈动脉最大斑块面积($OR=1.013$,95%CI:1.001 ~ 1.027,$P<0.05$)、冠心病史($OR=1.800$,95%CI:1.065 ~ 3.044,$P<0.05$)和糖尿病史($OR=2.795$,95%CI:1.544 ~ 5.060,$P<0.01$)与冠状动脉多支病变独立相关(表 6)。根据 ROC 曲线分析结果,颈动脉最大斑块面积预测冠状动脉多支病变的最佳切点为 15.6 mm^2(敏感度为 73.6%,特异度为 78.4%),见图 3。

表 5　预测冠状动脉多支病变的单因素分析结果

Table 5　Univariate analysis on predicting factors for multi-vessel coronary artery disease

[$M(IQR)$或 n(%)或 $\bar{x}\pm s$]

变　量	非冠脉多支病变($n=185$)	冠脉多支病变($n=110$)	*P* 值
年龄(岁)	62 ± 12.	67 ± 12	<0.01
女性	82(44.3)	28(25.5)	<0.01
高血压病史	116(62.7)	71(64.5)	>0.05
糖尿病史	28(15.1)	37(33.6)	<0.01
心房颤动史	14(7.6)	8(7.3)	>0.05
冠心病史	55(29.7)	51(46.4)	<0.01
心肌梗死史	6(3.2)	5(4.5)	>0.05
脑卒中史	7(3.8)	7(6.4)	>0.05
抗血小板聚集药物史	37(20.0)	23(20.9)	>0.05
他汀类药物史	41(22.2)	20(18.2)	>0.05
抗凝药物史	4(2.2)	1(0.9)	>0.05
冠状动脉支架植入史	27(14.6)	20(18.2)	>0.05
白细胞计数($\times10^9$/L)	6.3 ± 2.0	6.7 ± 2.6	>0.05
丙氨酸转氨酶(U/L)	19(14 ~ 29)	21(15 ~ 38)	>0.05
天冬氨酸转氨酶(U/L)	22(19 ~ 29)	24.5(21 ~ 53)	>0.05
肌酐(μmol/L)	82 ± 26	82 ± 51	>0.05
总胆固醇(mmol/L)	4.0 ± 1.1	4.2 ± 1.4	>0.05
三酰甘油(mmol/L)	1.4 ± 0.9	1.6 ± 2.2	>0.05
高密度脂蛋白(mmol/L)	1.1 ± 0.4	1.1 ± 0.3	>0.05
低密度脂蛋白(mmol/L)	2.2 ± 0.8	2.3 ± 1.0	>0.05
载脂蛋白 A1(g/L)	1.2 ± 0.2	0.8 ± 0.2	>0.05
载脂蛋白 B(g/L)	0.7 ± 0.2	0.8 ± 0.2	<0.05
颈动脉斑块	74(40.0)	53(48.2)	>0.05
最大斑块面积(mm^2)	0(0 ~ 13)	2(0 ~ 42)	<0.01
低回声斑块	41(22.2)	27(24.5)	>0.05
高回声斑块	39(21.1)	23(20.9)	>0.05
混合回声斑块	15(8.1)	28(25.5)	<0.01
不稳定斑块	51(27.6)	47(42.7)	<0.01
颈动脉狭窄	7(3.8)	11(10.0)	<0.05

3　讨　论

本研究发现,Logistic 回归模型显示,冠心病史、他汀类药物史、冠状动脉多支病变以及载脂蛋白 A1 是影响 AMI 发生的独立预测因子;而通过决策树模型(决策树 1)证实,冠状动脉多支病变、冠心病史、载脂蛋白 A1 以及抗血小板聚集药物史参与 AMI 事件发生。对比两种模型后发现,决策树分析的有效性不劣于 Logistic 回归模型。

在本研究中,Logistic 回归模型中有意义的变量与进入决策树模型的节点变量不同,决策树分析没有体现出他汀类药物史这一因素的作用,但是 Logistic 回归中显示这一因素的主效应有统计学意义。造成这一差异的原因可能为:他汀类药

表 6　预测冠状动脉多支病变的二元 Logistic 回归分析结果

Table 6　Binary Logistic regression analysis for predicting multi-vessel coronary artery disease

变　量	*OR* 值	95% *CI*	*P* 值
颈动脉狭窄	0.858	0.236 ~ 3.124	>0.05
不稳定斑块	1.097	0.579 ~ 2.077	>0.05
最大斑块面积	1.013	1.001 ~ 1.027	<0.05
年龄	1.016	0.993 ~ 1.040	>0.05
女性	0.463	0.266 ~ 0.809	<0.01
冠心病	1.800	1.065 ~ 3.044	<0.05
糖尿病	2.795	1.544 ~ 5.060	<0.01

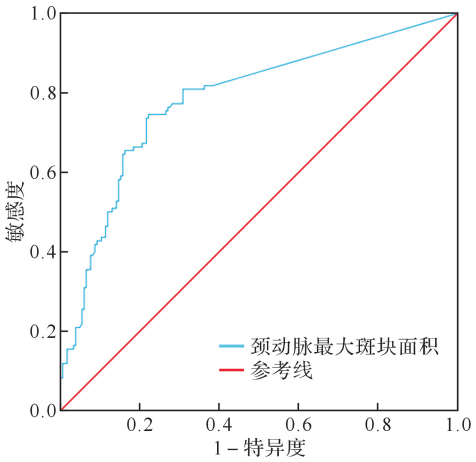


图 3　颈动脉最大斑块面积预测冠状动脉多支病变的 ROC 曲线

Figure 3　ROC of maximum plaque area of carotid artery predicting multi-vessel coronary artery disease

物史与冠心病史和抗血小板聚集药物史两个变量间的相关系数具有统计学意义,因此决策树在纳入冠心病史和抗血小板聚集药物史这两个因素后,可能就忽略了他汀类药物史的效应,导致他汀类药物史这一变量没有进入最终的树模型。但 Logistic 回归模型表明他汀类药物史的这一作用是不可忽视的。因此,目前采用决策树 C&RT 算法对变量属性的考虑仅限于单属性,而在实际的分类系统中,类的划分往往与多种属性(属性集)有关[15],如果未来能够将上述算法扩充到考虑多属性则有可能填补决策树忽略主效应这一缺陷。

在两种模型结果的比较中,Logistic 回归模型的特异度、约登指数高于决策树模型,决策树模型的特异度高于 Logistic 回归模型,但两种模型的 ROC 曲线分析结果统计学上显示决策树 1 模型并不劣于 Logistic 回归分析模型。Logistic 回归的优势在于表现某变量的主效应,在自变量对因变量变化关系方面的信息比决策树模型更充分。与 Logistic 回归相比,决策树展现的图形结构直观且可视化效果好,能够展示对分类或预测有意义的变量,并产生特定规则为决策提供依据。目前认为,实际应用中模型的选择不应只限于研究两种数据挖掘方法的优劣,而应最大程度发挥决策树与 Logistic 回归两种方法的优越性。有研究表明,通过 Logistic 回归筛选出有意义的主效应变量,再采用决策树模型进一步分析变量间的交互作用可获得较好的结果[16-17]。但在本研究中我们发现,根据 Logistic 回归结果改良的决策树 2 模型并不优于决策树 1,也进一步证实:由于部分因素间存在明显交互作用(本研究中他汀类药物史与另外两个变量具有显著相关性),则这些因素在进入 Logistic 回归方程中后可能显示无统计学意义,因此在模型的构建上未能优于未经 Logistic 回归筛选的决策树模型(即决策树 1)。另外,由于 Logistic 回归无法对样本进行切割,从而无法找到一个发生目标结果可能性最大的亚组,亦无法针对自变量的重要程度去构建方程[18-19]。因此根据 Logistic 回归产生的变量来构建决策树方程时存在局限性。

值得注意的是,尽管决策树 1 模型的 AUC 水平在数值上低于 Logistic 回归模型,但 AUC 大于 0.75,提示该模型对 AMI 事件的预测力良好。此外,与拆分数据集处理前相比,经过拆分后获得的决策树 1 模型并没有随着样本量的下降而导致变量被剪枝化处理,进入模型的节点恒定亦提示该树模型结构稳定,进一步支持其临床应用价值。

根据决策树 1 模型,我们可以建立 AMI 事件筛查的策略为:首先考虑的因素为是否存在冠状动脉多支病变,并据此分组,经评估无冠状动脉多支病变者,则诊断 AMI 的概率较低(8.6%);而在发现冠状动脉多支病变的患者中,是否存在冠心病史应优先考虑;否认冠心病史者中,如载脂蛋白 A1 低于 1.340 g/L,则发生 AMI 的概率较高(61.7%),如高于 1.340 g/L,则概率较低(16.7%);而在有冠心病史的患者中,入院前规律服用抗血小板聚集药物的患者发生 AMI 概率极低(4.8%),而未规律服药或未服药者发生 AMI 的概率则相对较高(23.3%)。

但是在现实中,大多数患者未能在 AMI 发病前获知冠状动脉病变情况(本研究中既往诊断冠心病史者为 106 例,占 35.9%,冠状动脉支架植入术史者为 47 例,占 15.9%),且大多数患者在此次发病入院前可能并无症状,因此对预知冠状动脉病变造成困难。本研究发现,颈动脉最大斑块面积是冠状动脉多支病变的危险因素,因此定期进行颈动脉超声检查评估最大斑块面积或将有助于早期发现冠状动脉多支病变。既往研究亦支持颈动脉最大斑块面积可较好地反映冠状动脉病变情况[14,20]。建议对于颈动脉超声检查结果提示最大斑块面积超过 15.6 mm^2时可选择进一步完善冠状动脉 CT 血管造影评估冠状动脉情况。如能早期发现冠状动脉多支病变或诊断冠心病,则通过监测载脂蛋白 A1 水平及规律服用抗血小板聚集药物或可达到控制 AMI 发病的目的。因此,该筛查方法为未来开展 AMI 高危患者筛查制定了可行计划,并只需根据分类树追寻其终末子集即可预测该人群发生 AMI 事件的情况,对筛查 AMI 患者有较强的实用性。

综上,采用决策树分析结果能更为直观、形象地反映 AMI 患者的特征,相比 Logistic 回归模型,决策树模型不仅可筛选出有统计学意义的因素,还能直观比较各种因素对 AMI 发生的影响强度。应用这些因素对 AMI 患者分类可快速找到对 AMI 事件影响最大的因素组合,方便指导临床工作。本研究旨在评估 AMI 事件并提供一种新颖的辅助预测工具,未来将通过进一步充实患者的临床信息和扩大样本量以完善对 AMI 事件的预测评估,从而为 AMI 的防治提供更有价值的临床指导方案。

参考文献

[1] GAO R, PATEL A, GAO W, et al. Prospective observational study of acute coronary syndromes in China: practice patterns and outcomes[J]. **Heart**, 2008,94(5):554-560.

[2] 张啸飞,胡大一,丁荣晶,等.中国心脑血管疾病死亡现况及流行趋势[J].**中华心血管病杂志**,2012,40(3):179-187.
ZHANG Xiaofei, HU Dayi, DING Rongjin, et al. Status and trend of cardio-cerebral-vascular diseases mortality in China: data from national disease surveillance system between 2004 and 2008[J]. **Chinese Journal of Cardiology**, 2012, 40(3): 179-187. (in Chinese)

[3] CHANG J, LIU X, SUN Y. Mortality due to acute myocardial infarction in China from 1987 to 2014: Secular trends and age-period-cohort effects[J]. **Int J Cardiol**,2017,227:229-238.

[4] 陈伟伟,高润霖,刘力生,等.中国心血管病报告 2013 概要[J].**中国循环杂志**,2014,8(7):487-491.
CHEN Weiwei, GAO Runlin, LIU Lisheng, et al. China cardiovascular diseases report 2013: A summary[J]. **Chinese Circulation Journal**,2014,8(7):487-491. (in Chinese)

[5] KITAMURA A, YAMAGISHI K, IMANO H, et al. Impact of hypertension and subclinical organ damage on the incidence of cardiovascular disease among Japanese residents at the population and individual levels- the circulatory risk in communities study (CIRCS)[J]. **Circ J**,2017,81(7):1022-1028.

[6] BHATIA R S, DORIAN P. Screening for cardiovascular disease risk with electrocardiography[J]. **JAMA Intern Med**,2018,178(9):1163-1164.

[7] 陈振明,纪双斌,史湘铃,等.Markov 决策树模型在优化 15~49 岁女性戊型肝炎免疫接种策略中的应用[J].**中华流行病学杂志**,2017,38(2):267-271.
CHEN Zhengmin, JI Shuangbin, SHI Xiangling, et al. Use the Markov-decision tree model to optimize vaccination strategies of hepatitis E among women aged 15 to 49[J]. **Chinese Journal of Epidemiology**, 2017,38(2):267-271. (in Chinese)

[8] LE RAY I, LEE B, WIKMAN A, et al. Evaluation of a decision tree for efficient antenatal red blood cell antibody screening[J]. **Epidemiology**,2018,29(3):453-457.

[9] 帅　健,李丽萍,陈业群.决策树模型与 Logistic 回归模型在伤害发生影响因素分析中的作用[J].**中华疾病控制杂志**,2015,19(2):185-189.
SHUAI Jian, LI Liping, CHEN Yequn. The role of Decision tree model and Logistic regression in injury influencing factors analysis[J]. **Chinese Journal of Disease Control & Prevention**, 2015, 19(2): 185-189. (in Chinese)

[10] THYGESEN K, ALPERT J S, JAFFE A S, et al. Fourth universal definition of myocardial infarction (2018)[J]. **Eur Heart J**,2019,40(3):237-269.

[11] ROBERTS J K, RAO S V, SHAW L K, et al. Comparative efficacy of coronary revascularization procedures for multivessel coronary artery disease in patients with chronic kidney disease[J]. **Am J Cardiol**,2017,119(9):1344-1351.

[12] XU T, ZUO P, CAO L, et al. Omentin-1 is

associated with carotid plaque instability among ischemic stroke patients [J]. **J Atheroscler Thromb**, 2018, 25(6): 505-511.

[13] 华 扬,刘蓓蓓,凌 晨,等. 超声检查对颈动脉狭窄50%~69%和70%~99%诊断准确性的评估[J]. **中国脑血管病杂志**, 2006, 3(5): 211-218.
HUA Yang, LIU Beibei, LING Chen, et al. Accurate assessment of the diagnosis between 50 – 69% and 70 – 99% carotid stenoses with ultrasonography[J]. **Chinese Journal of Cerebrovascular Diseases**, 2006, 3(5): 211-218. (in Chinese)

[14] HE J, CHEN P, LUO Y, et al. Relationship between the maximum carotid plaque area and the severity of coronary atherosclerosis[J]. **Int Angiol**, 2018, 37(4): 300-309.

[15] 何 跃,邓唯茹,刘司寰. 基于组合决策树的急诊等待时间预测[J]. **统计与决策**, 2016, 1(6): 72-74.
HE Yue, DENG Weiru, LIU Sihuan. Emergency waiting time prediction based on combined decision tree[J]. **Statistics and Decision**, 2016, 1(6): 72-74. (in Chinese)

[16] 赵自强,郑 明. 应用分类树模型筛选logistic回归中的交互因素[J]. **中国卫生统计**, 2007, 24(2): 114-116.
ZHAO Ziqiang, ZHENG Ming. Apply classification tree to automatically screen some potential interaction factors in Logistic regression[J]. **Chinese Journal of Health Statistics**, 2007, 24(2): 114-116. (in Chinese)

[17] 薛允莲. Logistic回归结合决策树技术在冠心病患者住院费用组合分析中的应用[J]. **中国卫生统计**, 2015, 32(6): 988-989.
XUE Yunlian. The application of logistic regression combined with decision tree technology in the combination analysis of hospitalization expenses of patients with coronary heart disease[J]. **Chinese Journal of Health Statistics**, 2015, 32(6): 988-989. (in Chinese)

[18] 黄晓霞,严玉洁,尉敏琦,等. logistic回归、决策树和神经网络在脑卒中高危筛查中的性能比较[J]. **中国慢性病预防与控制**, 2016, 24(6): 412-415.
HUANG Xiaoxia, YAN Yujie, WEI Minqi, et al. Comparison of screening group with high risk of stroke among logistic regression, decision trees and neural networks[J]. **Chinese Journal of Prevention and Control of Chronic Non-Communicable Diseases**, 2016, 24(6): 412-415. (in Chinese)

[19] 张娴静,陈 政,赵耐青,等. 上海市嘉定区农村居民就诊单位选择的影响因素分析——决策树和多分类无序反应变量的logistic回归相结合的方法[J]. **中国卫生统计**, 2005, 22(2): 80-84.
ZHANG Xianjing, CHEN Zheng, ZHAO Naiqing, et al. Researches on the factors Influencing the outpatients' choice of selecting care providers in Jiading district of Shanghai: a method of combining decision tree model with multinomial Logistic regression [J]. **Chinese Journal of Health Statistics**, 2005, 22(2): 80-84. (in Chinese)

[20] 王 梦,谢高强,王 浩,等. 颈动脉最大斑块面积的进展速率与新发缺血性心血管事件的关系[J]. **中国循环杂志**, 2014, 29(7): 532-536.
WANG Meng, XIE Gaoqiang, WANG Hao, et al. Relationship between the progression pate of corotid maximal plaque area and the risk of new ischemic cardiovascular disease [J]. **Chinese Circulation Journal**, 2014, 29(7): 532-536. (in Chinese)

[本文编辑 沈 敏 余 方]

· 读者 · 作者 · 编者 ·

本刊2020年专题报道征稿

本刊2020年将就芳香化酶抑制剂在儿科应用、睡眠医学、前列腺癌和乳腺癌的基础与临床、肺纤维化机制、中药药理研究等主题组织专题报道,欢迎垂询和投稿。

《浙江大学学报(医学版)》编辑部
2019年12月