

- 医院,2022,26(4):5-7.
- [12] 王建,王敬毅. 基于 DRG 的某院神经内科住院医疗服务绩效评价[J]. 中国医院统计,2021,28(6):528-532.
- [13] 侯明,肖万超,刘静. DRG 评价指标在医院绩效管理体系中的应用[J]. 江苏卫生事业管理,2021,32(12):1562-1566.
- [14] 张静秋,陈伟红,张岩曦,等. 大型医院实施 DRG 付费的住院服务绩效分析[J]. 卫生经济研究,2021,38(12):66-68.
- [15] 袁磊,陈子华,黄耿文,等. 基于 DRG 精准制定科室平均住院日目标的实践与评价[J]. 中国医院管理,2021,41(6):34-38.
- [16] 朱旭敏,李祥飞,陈桢铤. 基于 DRG 付费的天津市三级公立医院医疗服务绩效评价研究[J]. 中国公共卫生管理,2024,40(3):329-333.
- [17] 薛灵敏,刘涛,李宏伟,等. 借力绩效国考助推医院高质量发展的探讨[J]. 中国城乡企业卫生,2024,39(6):223-225.
- [18] 周华川,韩秀艳,徐婧,等. 公立医院高质量发展综合绩效评价方法及其应用研究[J]. 卫生软科学,2024,38(6):23-28.
- [19] 于洪钊,赵茜倩,田园,等. 公立医院实施科室目
· 卫生管理 · doi:10.3969/j.issn.1671-8348.2025.06.037

网络首发 [https://link.cnki.net/urlid/50.1097.R.20250305.1813.007\(2025-03-05\)](https://link.cnki.net/urlid/50.1097.R.20250305.1813.007(2025-03-05))

标责任制管理的实践与思考[J]. 中国卫生标准管理,2024,15(6):39-43.

- [20] 王海燕. DRGs 评价指标在医疗绩效管理体系中的应用分析[J]. 中国卫生标准管理,2018,9(12):28-30.
- [21] 季新强. 基于 DRGs 和综合指数法的某肿瘤专科医院肿瘤外科住院医疗服务绩效评价分析[J]. 中国卫生统计,2016,33(5):817-819.
- [22] 李小莹,仇叶龙,刘春玲等. 北京地区某医院神经内科住院医疗绩效分析[J]. 中国病案,2014,15(12):49-52.
- [23] 丁婷. 新形势下公立医院绩效管理体系的优化策略[J]. 山西财税,2024(6):45-46.
- [24] 魏升,钟光辉,张晔,等. 国家医院绩效考核引领下的中医院发展策略探讨[J]. 医院管理论坛,2024,41(6):6-8.
- [25] 赵明,赵蓉,王爱荣,等. 以临床专科能力建设推动公立医院高质量发展[J]. 中国卫生质量管理,2024,31(3):1-4.

(收稿日期:2024-10-28 修回日期:2025-02-28)

(编辑:姚 雪)

基于机器学习算法预测缺血性心脏病住院费用的在线应用的构建^{*}

黄 越¹, 黄国海², 陈琳玲¹, 陈署明¹, 黄国鑫³

(1. 汕头大学医学院第二附属医院质量管理科, 广东汕头 515000; 2. 汕头市中心医院血液净化中心, 广东汕头 515000; 3. 汕头市中心医院临床医学研究中心, 广东汕头 515000)

[摘要] 目的 构建基于机器学习算法的缺血性心脏病患者住院费用在线预测应用。方法 选取 2022 年 1 月 1 日至 12 月 31 日在汕头大学医学院第二附属医院诊治并出院的 1 332 份缺血性心脏病患者住院病案进行研究。采用 logistic 回归、支持向量机、回归树、随机森林、神经网络、极限梯度提升(XGBoost)等统计学模型进行计算, 比较各个模型在建模组及验证组的效能指标, 以获取最优的统计学模型。使用 ShinyApp 工具将所选模型转化为在线应用, 以便于实际应用和推广。结果 重症监护治疗、疾病诊断数量>5 个、经皮冠状动脉介入治疗(PCI)、住院时间为缺血性心脏病患者住院费用的独立影响因素($P<0.05$)。验证组中, 6 个模型受试者工作特征(ROC)曲线的曲线下面积(AUC)比较, 由大到小依次为 XGBoost (AUC=0.932)、神经网络 (AUC=0.931)、随机森林 (AUC=0.928)、支持向量机 (AUC=0.928)、logistic 回归 (AUC=0.924)、回归树 (AUC=0.916)。结论 通过机器学习算法建立高精准度的预测模型, 可提供精准的医疗服务, 制订合理的诊疗路径, 提高诊疗服务效率。

[关键词] 机器学习; 缺血性心脏病; 住院费用; 在线应用

[中图法分类号] R197

[文献标识码] B

[文章编号] 1671-8348(2025)06-1491-05

缺血性心脏病作为心血管疾病中最重要的临床

类型之一, 发病率与死亡率居高不下, 不仅会降低患

* 基金项目: 广东省医学科研基金项目(A2024749); 汕头市医疗卫生科技计划项目(221122176494998)。

者的生活质量,给家庭带来沉重的经济负担,也会对国家医疗保险基金产生巨大压力^[1]。鉴于影响住院费用的因素众多,寻找可控的影响因素有助于通过规范医疗管理,减少不必要的医疗费用。目前国内对住院费用的研究主要采用 2 种方法,即传统统计方法^[2-3]和机器学习方法^[4-5]。由于医疗数据通常具有多维性和复杂性,未经适当处理的数据难以适用于传统统计学方法。机器学习算法以其卓越的性能、灵活的扩展能力及对数据分布的高度适应性,在医学研究中受到越来越多的重视^[6]。然而,运用机器学习算法构建住院费用的在线预测应用尚属少见。本研究采用 6 种不同的机器学习算法构建缺血性心脏病住院费用的预测模型,并通过比较这些算法在预测性能上的差异,筛选出最优预测方法,并由此构建在线应用系统,现报道如下。

1 资料与方法

1.1 一般资料

选取 2022 年 1 月 1 日至 12 月 31 日在汕头大学医学院第二附属医院收治并出院的缺血性心脏病患者住院病案进行研究。排除标准:(1)患者年龄<18 岁或>90 岁;(2)住院时间<24 h;(3)在 1 年内重复住院且非首次住院的病案。经过筛选,共有 1 332 份病案纳入分析,研究人员逐一对病案进行审核和确认。纳入资料包括患者年龄、性别、婚姻状况、常住地、付费方式、入院途径、是否有转科、是否接受重症监护、是否抢救、病例分型、主要诊断分型、合并诊断数量、是否有并发症、病例分型、住院时间及住院费用等。本研究已通过汕头大学医学院第二附属医院伦理委员会审批(审批号:2022-114 号)。

1.2 方法

根据患者总住院费用的中位数将患者分成两组,根据变量的分布特征采用不同的统计方法对两组间的变量进行比较。将组间差异有统计学意义的变量纳入最小绝对收缩和选择算子(least absolute shrinkage and selection operator, Lasso)回归分析,进一步筛选变量。采用 logistic 回归、支持向量机、回归树、随机森林、神经网络、极限梯度提升(extreme gradient boosting, XGBoost)等统计学模型进行计算,比较各

个模型在建模组及验证组的效能指标,以获取最优的统计学模型。使用 ShinyApp 工具将所选模型转化为在线应用,便于实际应用和推广。

1.3 统计学处理

采用 SPSS21.0 及 R4.04 软件进行统计分析。符合正态分布和方差齐性检验的计量资料以 $\bar{x} \pm s$ 表示,组间比较采用 t 检验。不符合正态分布和方差齐性的计量资料以 $M(Q_1, Q_3)$ 表示,组间比较采用秩和检验。计数资料以例数或百分比表示,组间比较采用 χ^2 检验。采用受试者工作特征(receiver operating characteristic, ROC)曲线和曲线下面积(area under the curve, AUC)分析预测效能。以 $P < 0.05$ 为差异有统计学意义。

2 结 果

2.1 研究对象基本情况

本研究共纳入 1 332 例缺血性心脏病患者,男 946 例(71.0%),女 386 例(29.0%);患者平均年龄为 (64.1 ± 11.5) 岁,≥60 岁的老年患者 899 例(67.5%);1 120 例(84.1%)患者为门诊入院;急性心肌梗死 513 例(38.5%),心绞痛 369 例(27.7%),其他诊断分型 450 例(33.8%);863 例(64.8%)出现相关并发症;1 060 例(79.6%)行冠状动脉造影;689 例(51.7%)行经皮冠状动脉介入治疗(percutaneous-transluminal coronary intervention, PCI);患者总住院费用为 21 335.0(9 271.8,30 591.5)元。

2.2 不同住院费用分组比较

以总住院费用的中位数作为截断界值将患者分成高费用组和低费用组,两组性别、门诊入院、首次入院、转科、重症监护治疗、抢救、病例分型为 C/D、主要诊断类型、疾病诊断数量、并发症、冠状动脉造影、PCI、住院时间等变量比较,差异有统计学意义($P < 0.05$),见表 1。

2.3 Lasso 回归的变量筛选

运用 Lasso 回归对 13 个组间比较差异有统计学意义的变量做正则化分析,最终筛选获得 4 个变量,即重症监护治疗、疾病诊断数量≥5 个、PCI、住院时间,见图 1。

表 1 影响患者住院费用的单因素分析

项目	高费用组(n=666)	低费用组(n=666)	χ^2/Z	P
男/女(n/n)	512/154	434/232	22.193	<0.001
年龄[M(Q ₁ , Q ₃),岁]	65(57,72)	65(57,72)	0.019	0.985
老年(n)	447	452	0.086	0.770
已婚(n)	653	644	2.377	0.123
常住地为本市(n)	462	464	0.014	0.905
付费方式为自费(n)	47	56	0.852	0.356
门诊入院(n)	522	598	32.402	<0.001

续表 1 影响患者住院费用的单因素分析

项目	高费用组(<i>n</i> =666)	低费用组(<i>n</i> =666)	χ^2/Z	P
首次入院(<i>n</i>)	461	359	33.008	<0.001
转科(<i>n</i>)	494	98	476.805	<0.001
重症监护治疗(<i>n</i>)	506	85	539.092	<0.001
抢救(<i>n</i>)	520	133	449.928	<0.001
病例分型为 C/D(<i>n</i>)	663	634	24.677	<0.001
主要诊断类型(<i>n</i>)			298.798	<0.001
急性心肌梗死	401	112		
心绞痛	155	214		
其他缺血性心脏病	110	340		
疾病诊断数量>5 个(<i>n</i>)	414	355	10.710	0.001
并发症(<i>n</i>)	450	413	4.505	0.034
冠状动脉造影(<i>n</i>)	637	423	211.571	<0.001
PCI(<i>n</i>)	594	95	748.644	<0.001
住院时间[M(Q ₁ , Q ₃), d]	7(6,10)	5(3,7)	14.158	<0.001

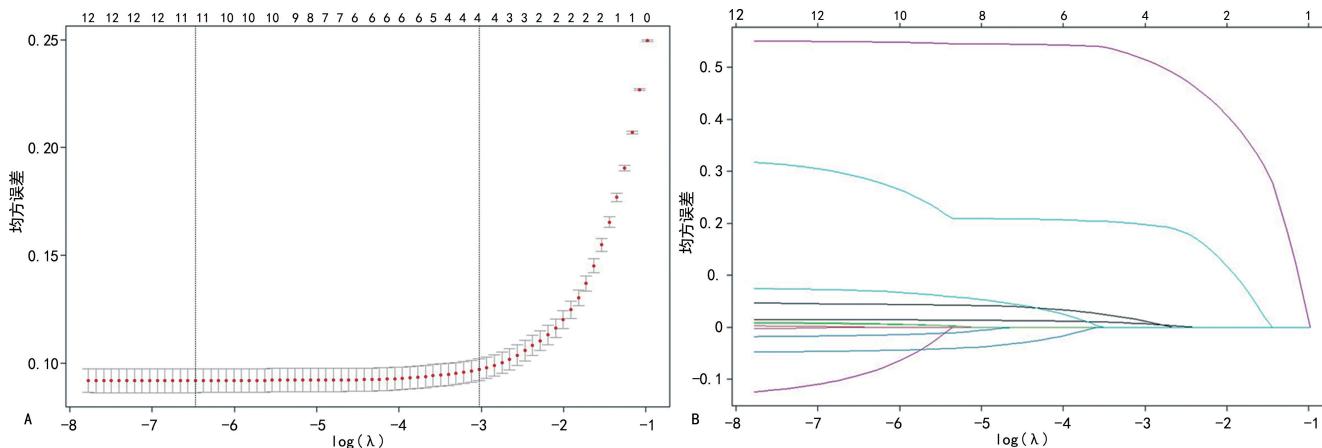


图 1 Lasso 回归分析变量筛选

2.4 多因素 logistic 回归分析

将 Lasso 回归分析筛选所得的 4 个变量纳入多因素 logistic 回归分析,结果显示 4 个变量均为缺血性心脏病患者住院费用的独立影响因素,见表 2。

表 2 多因素 logistic 回归分析

项目	β	OR	95%CI	P
重症监护治疗	1.734	5.661	3.891~8.236	<0.001
疾病诊断数量>5 个	0.540	1.717	1.167~2.527	0.006
PCI	3.481	32.504	21.978~48.071	<0.001
住院时间	0.168	1.182	1.120~1.249	<0.001

2.5 各模型在建模组的效能评价

将患者随机分成建模组(约占总患者数的 70%)和验证组(约占总患者数的 30%),两组间相关变量比较,差异无统计学意义($P > 0.05$)。建模组中,6 个模型 ROC 曲线的 AUC 比较,由大到小依次为 XGBoost ($AUC=0.961$)、神经网络 ($AUC=0.954$)、随机森林

($AUC=0.952$)、logistic 回归 ($AUC=0.949$)、支持向量机 ($AUC=0.947$)、回归树 ($AUC=0.929$)。验证组中,6 个模型 ROC 曲线的 AUC 比较,由大到小依次为 XGBoost ($AUC=0.932$)、神经网络 ($AUC=0.931$)、随机森林 ($AUC=0.928$)、支持向量机 ($AUC=0.928$)、logistic 回归 ($AUC=0.924$)、回归树 ($AUC=0.916$),见表 3、4。

表 3 各模型在建模组的效能评价

项目	AUC	准确度	精确度	灵敏度	特异度	截断值
logistic 回归	0.949	0.880	0.904	0.865	0.896	0.5
回归树	0.929	0.874	0.857	0.891	0.859	0.5
支持向量机	0.947	0.884	0.856	0.911	0.860	0.5
随机森林	0.952	0.897	0.881	0.913	0.881	0.5
神经网络	0.954	0.899	0.906	0.896	0.902	0.5
XGBoost	0.961	0.900	0.870	0.928	0.874	0.5

2.6 在线应用的构建和使用

成功构建在线预测应用并生成网页链接。终端设备都可通过访问网址 (https://hgh-163.shinyapps.io/shiny_IHD) 对缺血性心脏病患者的住院费用进行预测, 见图 2。

图 2 在线应用截图

表 4 各模型在验证组的效能评价

项目	AUC	准确度	精确度	灵敏度	特异度	截断值
logistic 回归	0.924	0.857	0.889	0.828	0.889	0.5
回归树	0.916	0.885	0.857	0.900	0.873	0.5
支持向量机	0.928	0.885	0.857	0.900	0.873	0.5
随机森林	0.928	0.883	0.862	0.891	0.876	0.5
神经网络	0.931	0.872	0.873	0.864	0.881	0.5
XGBoost	0.932	0.888	0.857	0.905	0.873	0.5

3 讨 论

Lasso 回归通过引入 B 的绝对值之和作为惩罚项, 有效地减少了模型的复杂性和过拟合风险。其能够将某些特征变量的 B 压缩至零, 从而实现变量的选择和缩减^[7]。这种方法不仅增强了模型的稳定性和鲁棒性, 也有助于解决自变量之间的多重共线性问题, 在临床和基础医学研究中得到了广泛应用^[8]。本研究通过 Lasso 回归分析, 确定了影响缺血性心脏病住院费用的 4 个关键因素。PCI 是影响住院费用的最主要因素, 这与李洪坤等^[9]、黄果等^[10]的研究结果一致。重症监护治疗和疾病诊断数量 >5 个表明患者病情更为严重, 需要更多的医疗资源, 从而增加住院费用。此外, 住院时间的增加也会导致住院费用的增高, 这与此前相关研究结论一致^[11-12]。

机器学习是应用统计分析算法对训练数据进行分析, 利用获得的经验改进并分析算法, 创建模型以实现对结果的预测^[13-14]。相较于传统统计学处理, 机器学习算法构建的模型具有更高的模型拟合度及准确度, 近年来已被应用于多种疾病的临床研究分析^[15]。本研究通过比较 6 种机器学习算法构建预测模型的各项效能指标, 发现 logistic 回归、神经网络、随机森林、回归树、支持向量机、XGBoost 模型在验证组的准确度分别为 0.857、0.872、0.883、0.885、0.885、0.888, AUC 分别为 0.924、0.931、0.928、0.916、0.928、0.932, XGBoost 模型的准确度、灵敏度、AUC 均为最高, 是本研究中预测效能最优的统计

yapps.io/shiny_IHD) 对缺血性心脏病患者的住院费用进行预测, 见图 2。

学模型。XGBoost 是基于梯度提升的集成学习算法^[16], 其目标是通过不断学习将弱分类器通过某种策略进行融合以得到更加强大的分类器^[17]。作为一种新型高效的算法, XGBoost 的关注度和应用价值正不断提升, 已广泛地运用于医药领域的数据分析项目^[18]。

随着社会经济的不断进步, 医疗需求持续增长, 不可避免地导致医疗支出上升。2020 年, 中国的卫生总费用占国内生产总值的 7.12%^[19]。在医疗费用增长中, 住院费用的上升尤为明显, 其对医疗秩序产生了不利影响。缺血性心脏病作为一种常见的心血管疾病, 发病率和死亡率较高, 但随着医疗技术的提升, 已经可以通过 PCI 进行较好的治疗^[20]。医院可以通过规范临床路径管理, 严控相关治疗及手术指征, 减少过度医疗行为, 明确耗材使用数据的合理性指标, 监测医用耗材的合理使用, 从而控制住院费用的增长。同时, 提高医疗技术水平, 优化资源利用, 实现医疗质量全程管理, 可以在保证治疗效果的前提下, 尽量减少低效和无效的住院时间^[21]。此外, 还可以利用机器学习算法的强大功能, 实现医院管理的智能化和自动化, 提高医院绩效管理、医疗质量和费用管控的工作效率, 积极探索基于机器学习算法在医院信息化系统建设中的应用^[22-23]。机器学习算法模型的临床应用较为困难^[24], 而本研究将构建的模型转换成在线应用^[25], 用户可以通过访问特定网址来使用这一模型。这种在线应用不仅具备临床操作的便捷性, 还具有良好的可访问性, 临床实用价值和便携性较高, 适合在临床实践中推广使用。

综上所述, 通过机器学习算法建立高精准度的预测模型, 可提供精准的医疗服务, 制订合理的诊疗路径, 提高诊疗服务效率, 加强成本管控能力, 有助于控制医疗费用的不合理增长。

参考文献

- [1] 国家心血管病中心.《中国心血管健康与疾病报

- 告 2022》[J]. 中国心血管病研究, 2023, 21(7): 577-600.
- [2] 尚迪, 邹芳. 2018—2020 年某三级医院急性心肌梗死患者住院费用影响因素分析[J]. 中国医院统计, 2022, 29(1): 47-50.
- [3] 黎珍君, 覃华凌, 陆丹艳, 等. 冠脉支架降价前后患者住院费用及影响因素分析[J]. 中国病案, 2023, 24(7): 76-78.
- [4] 余浪, 易东. B-P 神经网络模型在缺血性心脏病住院费用影响因素分析中的应用[J]. 中国病案, 2018, 19(3): 78-82.
- [5] 张晔, 迟卫军, 徐云伟, 等. 基于随机森林的冠心病患者住院费用影响因素的分析[J]. 中国病案, 2019, 20(2): 53-57.
- [6] 孙冲. 多种机器学习算法预测呼吸系统疾病住院费用效果比较分析[D]. 长春: 吉林大学, 2020.
- [7] TAY J K, NARASIMHAN B, HASTIE T. Elastic net regularization paths for all generalized linear models[J]. J Stat Softw, 2023, 106: 1.
- [8] 车前子, 王晶, 白卫国, 等. 基于 LASSO 回归的骨质疏松肾阳虚状态辨识模型研究[J]. 中华中医药杂志, 2022, 37(10): 5928-5933.
- [9] 李洪坤, 王新生, 李江峰, 等. 冠心病经皮冠状动脉介入治疗患者住院费用影响因素分析[J]. 精准医学杂志, 2022, 37(4): 332-336.
- [10] 黄果, 董来东. 山东省某三甲医院不同基本医疗保险类型冠心病患者住院费用比较[J]. 医学与社会, 2022, 35(3): 67-71.
- [11] 李妮妮, 陆瑾, 蒙钟经, 等. 慢性肾功能不全患者住院费用影响因素分析[J]. 中国病案, 2023, 24(2): 37-40.
- [12] 唐涛, 王赛, 张雨篷, 等. 湖南省脑出血患者住院费用分析[J]. 中国卫生统计, 2023, 40(1): 99-102.
- [13] LOPEZ E, ETXEBARRIA-ELEZGARAI J, AMIGO J M, et al. The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples[J]. Anal Chim Acta, 2023, 1275: 341532.
- [14] KRAJCER Z. Artificial intelligence in cardiovascular medicine: historical overview, current status, and future directions [J]. Tex Heart Inst J, 2022, 49(2): e207527.
- [15] UDDIN S, KHAN A, HOSSAIN M E, et al. Co-mparing different supervised machine learning algorithms for disease prediction[J]. BMC Med Inform Decis Mak, 2019, 19 (1): 281.
- [16] QIAO Z, SUN N, LI X, et al. Using machine learning approaches for emergency room visit prediction based on electronic health record data[J]. Stud Health Technol Inform, 2018, 247: 111-115.
- [17] 卢娅欣, 黄月, 李康. XGBoost 算法在二分类非平衡高维数据分析中的应用[J]. 中国卫生统计, 2021, 38(1): 21-24.
- [18] 刘艳, 陈霁晖, 刘昕竹, 等. 机器学习 XGBoost 算法在医学领域的应用研究进展[J]. 分子影像学杂志, 2021, 44(5): 856-862.
- [19] 高峰, 刘颜. DRG 实施的难点及解决对策[J]. 医学信息, 2022, 35(11): 1-4.
- [20] 段同庆, 鲁瑞, 史新军, 等. 偏最小二乘回归在探索 PCI 治疗冠心病患者预后影响因素中的应用[J]. 中国卫生统计, 2019, 36(6): 824-828.
- [21] 刘笑, 方鹏骞. 医疗费用控制背景下公立医院医疗质量保障对策分析[J]. 中国医院管理, 2019, 39(11): 26-28.
- [22] 俞高. 基于 XGBoost 算法的病例分型系统设计[J]. 中国数字医学, 2018, 13(3): 69-71.
- [23] YU Y, PENG C, ZHANG Z, et al. Machine learning methods for predicting long-term mortality in patients after cardiac surgery[J]. Front Cardiovasc Med, 2022, 9: 831390.
- [24] AKBILGIC O, OBI Y, POTUKUCHI P K, et al. Machine learning to identify dialysis patients at high death risk[J]. Kidney Int Rep, 2019, 4(9): 1219-1229.
- [25] ZHU J, ZHENG J, LI L, et al. Application of machine learning algorithms to predict central lymph node metastasis in T1—T2, non-invasive, and clinically node negative papillary thyroid carcinoma[J]. Front Med, 2021, 8: 635771.

(收稿日期: 2024-07-18 修回日期: 2024-12-27)

(编辑: 张兢捷)