

引用格式: 周振. 成年肺癌患者住院费用预测模型及评分工具研究 [J]. 电子科技大学学报, 2025, 54(2): 305-310.

ZHOU Z. Prediction model and scoring tool for hospitalization costs in adult patients with lung cancer[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(2): 305-310.

成年肺癌患者住院费用预测模型及评分工具研究



周 振*

(复旦大学生命科学院, 上海 200433)

摘要: 分析和研究肺癌患者住院费用的影响因素有利于更好地理解肺癌住院支出及疾病负担, 也对优化医疗支付政策等工作有重要的参考意义。该研究共纳入 12 117 例 2020 年 1 月—2023 年 9 月间, 某省多家医院的成年肺癌患者住院记录数据, 首先利用 K-means 聚类将住院费用进行离散化预处理, 并采用单因素 logistic 回归从 42 个因素中筛选出 25 个潜在影响因素, 之后基于 CatBoost 和 XGBoost 分别构建成年肺癌患者住院费用预测模型并开展模型性能评估, 以变量的特征重要性评分为依据衡量其对住院费用的影响程度。该研究还使用基于多因素 logistic 回归的方法建立了高住院费用评分工具。结果显示, CatBoost 和 XGBoost 均具有良好的预测性能 ($AUC > 0.95$), CatBoost 表现略优于 XGBoost。基于 CatBoost 模型, 该研究明确了住院天数、手术级别、是否放疗、抢救次数、肺癌组织学分型、年龄、是否化疗、是否首次住院和中性粒细胞计数共 9 个影响肺癌住院费用的重要因素, 并根据赋分标准将其中 7 个因素纳入评分工具。评分工具的区分度和校准度在测试集上得到验证, 结果显示评分工具的 AUC 值达到 0.958, 表现出了卓越的性能。

关键词: 肺癌; 住院费用; 预测模型; 影响因素分析; 评分工具

中图分类号: TP399

文献标志码: A

DOI: 10.12178/1001-0548.2024054

Prediction model and scoring tool for hospitalization costs in adult patients with lung cancer

ZHOU Zhen*

(School of Life Sciences, Fudan University, Shanghai 200433, China)

Abstract: To predict the hospitalization costs of lung cancer patients and analyze its influence factors is conducive to better understanding the hospitalization expenses and economic burden of lung cancer patients, and has reference significance for optimizing medical payment policies. This study included records of 12 117 adult lung cancer patients hospitalized between Jan 2020 and Sep 2023 from multiple hospitals in a province. Firstly, K-means clustering was employed to categorize the hospitalization costs, and then 25 potentially influencing factors were screened out from 42 factors using single-factor logistic regression. After that, this study constructed and evaluated hospitalization costs prediction models based on CatBoost and XGBoost, respectively, and measured the influence of these factors on hospitalization costs based on the feature importance value. Furthermore, employing the significant factors identified by the prediction models, this study developed a high hospitalization costs scoring tool using a multi-variable logistic regression approach. Results show that both CatBoost and XGBoost have good predictive performance ($AUC > 0.95$), with CatBoost performing slightly better than XGBoost. Based on the CatBoost model, this study identified nine factors affecting the cost of hospitalization: length of hospital stay, type of surgery, radiotherapy, number of rescues, histological classification of lung cancer, age, chemotherapy, first hospitalization, and neutrophil count level, and seven of them were included in the scoring tool according to the assignment criteria. The differentiation and calibration of the scoring tool were validated on the test set, showing an excellent AUC of 0.958, indicating exceptional performance.

Key words: lung cancer; hospitalization costs; prediction model; factors analysis; scoring tool

收稿日期: 2024-03-11

基金项目: 国家自然科学基金重大研究计划 (91846302)

作者简介: 周振, 博士生, 高级工程师, 主要从事医疗大数据方面的研究。

*通信作者 E-mail: zhouzhen96@hotmail.com

肺癌是一种严重危及患者生命健康的常见恶性肿瘤^[1-2]。在中国,肺癌发病率均呈快速增长趋势,城市和农村地区的肺癌发病率均位列恶性肿瘤的第一位^[3]。

肺癌的治疗过程给患者及其家庭带来沉重负担,我国肺癌患者治疗负担为年人均可支配平均收入的 5-7 倍^[4-5]。为了减轻患者及其家庭的负担,我国自 2012 年开始将肺癌等 18 种重大疾病按照病种付费疾病。但截至 2022 年,我国公立医院肺癌住院患者人均住院费用仍高达 39 527.2 元^[2]。因此,分析和理解肺癌患者住院费用的影响因素有着重大现实意义。

住院费用包括患者住院期间产生的床位费、手术费、药品费、护理费以及化验费等,是医疗费用的主要组成部分。一些研究通过收集患者的临床特征、治疗方案等数据,构建基于统计学或机器学习方法的模型,探讨肺癌患者住院费用的影响因素。这些研究使用分位数回归模型^[6]、支持向量机^[7]、BP 神经网络模型^[8-10]、logistic 回归^[9, 11]、结构方程^[12]、多重线性回归^[13]、多元逐步回归^[14]、C5.0 算法^[15]等模型研究住院费用的影响因素。文献^[16-17]证实 XGBoost 在某些病种住院费用预测方面具有显著优势,但 CatBoost 在住院费用预测方面的应用仍需深入研究及验证。

基于此,本研究基于国家自然科学基金重大研究计划与某地健康医疗大数据试点工程合作,获得其书面知情同意,纳入了多家医院 2020 年 1 月—2023 年 9 月期间住院的成年肺癌病人的住院记录数据,建立住院费用预测模型并进一步构建评分工具,旨在帮助医保部门、医疗机构更好地理解肺癌住院支出及疾病负担,为优化医疗支付政策等工作提供参考价值。

1 材料与方法

1.1 数据来源

按照第 10 次修订的国际疾病分类版本,本研究纳入某省多家医院住院时间在 2020 年 1 月—2023 年 9 月之间,且诊断为肺癌(ICD-10 编码以 C34 开头)的成年患者住院记录作为研究对象,并将记录中的住院费用作为结局变量。

为确保数据的准确性以及分析结果的真实性和可靠性,本研究还采用了一定的排除标准,具体包括:1) 排除缺少性别、年龄或住院费用信息的记录。2) 将合理范围设为在 2.5%~97.5% 之间,使

用百分位法排除住院费用中的极端值。

在最终的数据集中,总共包括了 12 117 条成年患者住院记录,这些记录将用于研究分析,以深入了解肺癌患者的住院费用情况。

1.2 研究方法

本研究采用 R 软件(V4.2.2)进行数据统计分析。

1.2.1 数据预处理

1) 研究纳入变量

本研究首先基于既往相关文献研究选择潜在变量,如住院天数^[13-14]、手术级别^[15]、性别^[13]、年龄^[13-14]、离院方式^[15]等。并结合临床专家的意见,进一步纳入代表患者入院时身体状况的变量,包括患者人口统计学指标和患者入院早期实验室检验结果。同时,参考查尔森共患病指数^[18](Charlson-comorbidity index, CCI),选择与肺癌密切相关的 13 种合并症作为纳入变量,即冠心病、慢性阻塞性肺病、白血病、痴呆、恶性淋巴瘤、肝功能异常、高血压、脑血管疾病、偏瘫、肾功能不全、糖尿病、消化性溃疡和周围血管疾病。

最终,本研究选取了 42 个变量作为初步纳入研究的指标(如表 1)。其中,住院天数、年龄为连续变量,其他均为分类变量。所有纳入变量的缺失比例小于 30%。为保证数据的一致性和稳定性,对变量的缺失值和异常值进行处理。对于连续变量采用中位数填补缺失值,盖帽法填补异常值。对于分类变量,采用众数填补缺失值、替换异常值。

表 1 研究纳入变量

变量分类	变量
人口统计学	性别、年龄、籍贯、民族
肺癌组织学 分型	小细胞癌、腺癌、鳞癌、其他非小细胞癌、未知
合并症	冠心病、慢性阻塞性肺病、白血病、痴呆、恶性淋巴瘤、肝功能异常、高血压、脑血管疾病、偏瘫、肾功能不全、糖尿病、消化性溃疡、周围血管疾病
入院情况	入院途径、入院时病情、参保类型、是否首次住院、是否具有手术史
院内首次化 验结果	糖类肿瘤CA199、角蛋白抗原CY211、血小板计数、淋巴细胞计数、癌胚抗原、碱性磷酸酶、神经元特异性烯醇化酶、嗜酸性细胞计数、中性粒细胞计数、乳酸脱氢酶
院内诊疗	是否使用抗生素、是否放疗、是否化疗、是否靶向治疗、是否免疫治疗、抢救次数、手术级别、住院天数
离院情况	离院方式

2) 住院费用

住院费用为连续变量,在建立住院费用预测模型前,需将其转化为分类变量。对肺癌患者住院费

用做正态性检验，发现住院费用不服从正态分布 ($P < 0.001$)，故而本研究使用 K-means 聚类对住院费用进行离散化。

聚类时选择轮廓系数法来确定最佳的聚类个数，轮廓系数与聚类个数 (K 值) 的关系如图 1。从图 1 可以看出，随着 K 值的增大，轮廓系数逐渐降低，在 $K=2$ 时聚类效果最好，据此本研究将住院费用分为高费用组和低费用组。

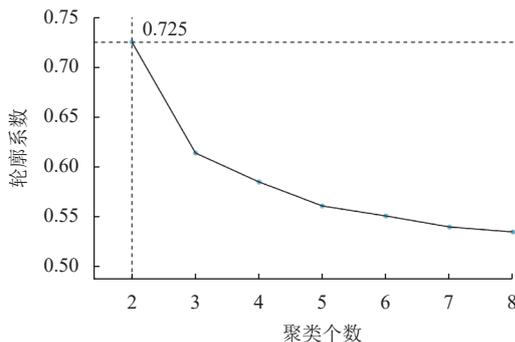


图 1 轮廓系数与聚类个数的关系

1.2.2 基于 CatBoost 和 XGBoost 的住院费用预测模型

为了降低机器学习模型的复杂性，本研究先使用单因素 logistic 回归模型初步筛选出与住院费用有显著相关性的变量 (检验水准 $\alpha = 0.05$)，再通过 CatBoost 和 XGBoost 两种基于决策树的分类机器学习方法建立住院费用预测模型。

1) 模型训练与评估

将数据按照 7:3 的比例划分为训练集和测试集。在训练集中采用 CatBoost 和 XGBoost 算法，进行住院费用预测模型训练，并利用十折交叉验证法进行模型调优。由于患者住院费用存在不均衡的问题，训练过程中本研究还使用混合采样平衡不同类别的训练样本数量，以改善模型对所有类别的学习能力，降低过拟合的风险。

模型的评估在测试集中进行，通过比较精确率、召回率、F1 值、Brier 值和曲线下面积 (area under the curve, AUC) 等评价指标，选择出较优模型。

2) 住院费用影响因素分析

CatBoost 和 XGBoost 输出的特征重要性可以作为衡量变量对住院费用的影响程度的依据。本研究以较优模型中变量的特征重要性为依据，对影响住院费用的变量进行分析和排序，进而筛选出可纳入评分工具的影响因素。

1.2.3 高住院费用评分工具的构建和验证

基于以上筛选出的影响因素，本研究沿用训练集和测试集的划分，采用文献 [19] 提出的基于多因素 logistic 回归的评分工具建立方法进行肺癌患者高住院费用评分工具的构建工作。

1) 离散化连续变量

通过绘制连续变量与住院费用的局部加权回归 (locally weighted regression, LOESS) 曲线，以 LOESS 曲线与高费用概率阶梯变化 (0.25、0.50、0.75、1.00) 的交点为分界点，将连续变量转换为分类变量。

2) 制定赋分规则

在训练集中基于以特征重要性筛选出的变量构建多因素 logistic 回归模型，并将各变量在模型中的回归系数值四舍五入后作为评分工具中各项的分值。

3) 评分总分和高住院费用概率对应曲线

在评分工具建立后，计算训练集中所有患者的评分总分，并将分数转换为发生高住院费用的概率，从而绘制分数和高住院费用概率的对应曲线。基于该曲线，可以通过单个患者的分数得到其发生高住院费用的概率。

评分工具的验证在测试集中进行，主要包括模型的区分度 (是否能正确分类) 和校准度 (预测与真实之间的差异) 两方面。使用 AUC 作为评分工具区分度评价指标，使用校准曲线 (calibration curve) 评估校准度，并进行不可靠性检验。

2 结果

2.1 住院费用分布及 K-means 聚类结果

本研究共纳入 2020 年 1 月—2023 年 9 月之间的某省多家医院共计 12 117 例成年肺癌患者住院资料。其中，女性 4 287 例 (35.4%)，男性 7 830 例 (64.6%)。患者平均年龄 64.25 岁 (最小年龄 18.80 岁，最大年龄 98.30 岁，中位年龄为 65.00 岁)，平均住院日为 9.8 天。住院费用从 1 533.85 元到 63 473.15 元不等，对其正态性进行检验后，发现住院费用为明显的偏态 ($P < 0.001$)，如图 2 所示)。

表 2 为对住院费用进行 K-means 聚类的结果。住院费用被聚为 2 组，其中，低费用组的住院费用中心点为 10 778.2 元，包含有 10 041 条记录，高费用组的住院费用中心点为 41 593.5 元，包含有 2 076 条记录。

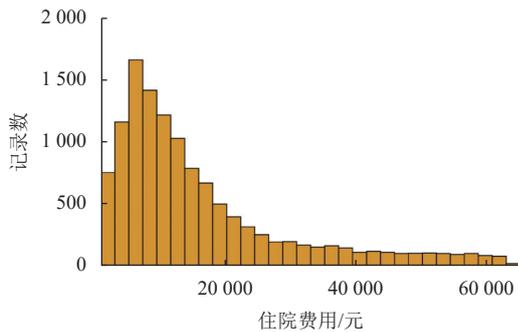


图2 住院费用分布

表2 K-means 聚类后的住院费用

分组	中心点/元	个数/例	最小值/元	最大值/元	标准差/元
低费用组	10 778.2	10 041	1 533.85	26 172.86	5 944.38
高费用组	41 593.5	2 076	26 202.52	63 473.15	10 690.12

2.2 CatBoost 和 XGBoost 住院费用预测模型

单因素 logistic 回归结果显示共有 25 个潜在影响变量 ($P < 0.05$), 分别是: 性别、年龄、肺癌组织学分型、冠心病、痴呆、脑血管疾病、偏瘫、周围血管疾病、入院途径、入院时病情、参保类型、是否首次住院、是否具有手术史、角蛋白抗原 CY211、淋巴细胞计数、神经元特异性烯醇化酶、嗜酸性细胞计数、中性粒细胞计数、是否使用抗生素、是否放疗、是否化疗、抢救次数、手术级别、住院天数和离院方式。

CatBoost 和 XGBoost 的预测性能 (精准率、召回率、F1 值、Brier 值、AUC) 比较见表 3。受试者操作特征曲线 (receiver operating characteristic curve, ROC) 见图 3。可以看出这 2 个模型的预测性能均表现较好, CatBoost 性能表现略优于 XGBoost。

表3 机器学习模型预测性能评价

模型	精准率	召回率	F1 值	Brier 值	AUC
CatBoost	0.915	0.917	0.947	0.060	0.974
XGBoost	0.914	0.915	0.947	0.632	0.972

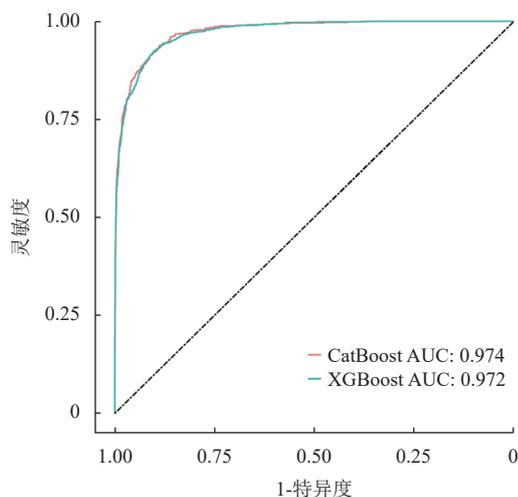


图3 机器学习模型的 ROC 曲线

2.3 住院费用影响因素分析

鉴于 CatBoost 的表现略优于 XGBoost, 本研究采用基于 CatBoost 建立的预测模型来分析影响住院费用的因素。表 4 显示了 CatBoost 模型中特征重要性大于 2% 的变量排序情况。住院费用的 9 个主要影响因素有: 住院天数、手术级别、是否放疗、抢救次数、肺癌组织学分型、年龄、是否化疗、是否首次住院和中性粒细胞计数。

表4 变量特征重要性排序

变量名	特征重要性/%
住院天数	32.0
手术级别	29.8
是否放疗	6.2
抢救次数	5.6
肺癌组织学分型	4.4
年龄	3.7
是否化疗	3.6
是否首次住院	2.5
中性粒细胞计数	2.3

2.4 评分工具的构建及验证

由于住院天数和年龄为连续变量, 需先对其进行离散化处理, 本研究基于 LOESS 曲线将住院天数分为 [0,10]、(10,20]、(20,28]、(28,35]、(35,50]、(50,64] 的不同组, 将年龄分为 [18,38]、(38,98.3] 两组。在以上变量处理的基础上, 基于表 4 中的 9 个变量构建多因素 logistic 模型, 根据该模型中回归系数值构建评分工具, 结果见表 5。因年龄和是否首次住院的回归系数值四舍五入小于 1, 故不被纳入在评分工具中。从表 5 可以看出住院天数在 7 个评分指标中分值占比最大, 当住院天数超过 50 天时, 单项得分超过 17 分。

基于以上评分工具, 根据单次肺癌患者住院记录中肺癌组织学分型、中性粒细胞计数、是否放疗、是否化疗、抢救次数、手术级别和住院天数进行赋分后加总, 得到可预测当次住院发生高住院费用的概率的评分总分。图 4 的直方图表现了测试集中评分总分的每个分数段内高住院费用组 (蓝色) 和低住院费用组 (黄色) 的记录数分布。随着评分总分变高, 分段中的高住院费用记录占比变大。图中, 红色虚线体现了高住院费用概率与评分总分的对应关系, 通过这条曲线可以获得单个评分总分对应的具体高住院费用概率。若肺癌患者某次住院按照评分工具计算的评分总分越高, 则他该次住院费用越有可能被划分到高住院费用组。具体来说, 评分总分在 0~7 之间时, 发生高住院费用的概率较低

(<0.5)，评分总分在 8~23 之间时，发生高住院费用概率较高 (>0.5)。

表 5 肺癌患者高住院费用评分工具

变量	变量值	赋分
肺癌组织学分型	其他/未知	0
	腺癌	1
中性粒细胞计数	参考范围之间/低于参考范围	0
	高于参考范围	1
是否放疗	否	0
	是	3
是否化疗	否	0
	是	1
抢救次数	0次	0
	1次	1
	>1次	2
	未做手术	0
手术级别	1级	1
	2级	2
	3级	3
	4级	7
住院天数/d	[0,10]	0
	(10,20]	2
	(20,28]	4
	(28,35]	5
	(35,50]	7
	(50,64]	7
		17

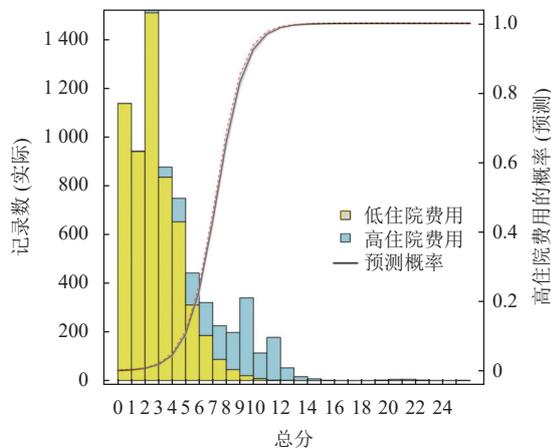


图 4 评分工具总分与当次住院发生高住院费用概率的关系

为了验证评分工具的区分度和校准度，本研究利用测试集数据开展了评分工具的区分度和校准度验证工作。图 5 展示了校准曲线和多个可用于评价评分工具性能的指标结果 (Emax 和 Eavg 分别是现有评分工具与理想评分工具的最大偏移量和平均偏移量，越小则说明评分工具与理想评分工具越接近； U 是不可靠性检验的统计量，其对应的 P 值为 $S:p$ ； $C(ROC)$ 是 ROC 曲线下的面积。结果显示，AUC 达到了 0.958，证明评分工具具有良好的区分能力；不可靠性检验得出的 $P=0.246 > 0.050$ ，Emax 为 0.058，Eavg 为 0.009，校准曲线

显示评分工具的校准度较好。

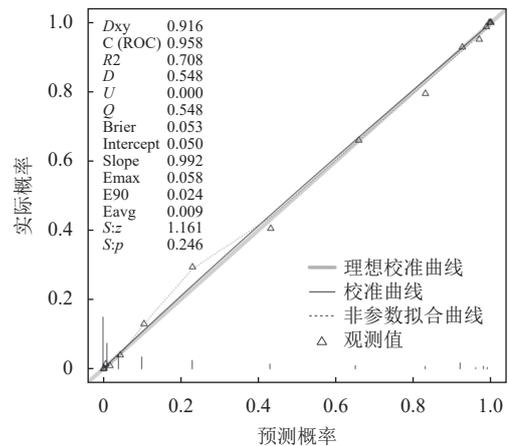


图 5 评分工具的区分度和校准度验证

3 结束语

本研究建立了成年肺癌患者住院费用预测模型，构建了高住院费用评分工具，旨在更好地理解肺癌住院费用的相关影响因素，为优化医疗支付政策等工作提供参考价值。

本研究针对住院费用非正态分布的特点和规律，通过 K-means 聚类进行离散化处理；使用 CatBoost 和 XGBoost 建立住院费用预测模型，并验证了其优异性能。同时，本研究基于多因素 logistic 回归模型建立评分工具，相较于机器学习，多因素 logistic 回归具有较强的可解释性，其模型参数能够直观地反映各因素对结局的影响程度，这在需要向非专业人员解释模型结果的场景中更加直接便捷。在研究方法推广方面，本研究可为其他疾病住院费用预测研究的方法选择和设计提供参考，但应充分考虑疾病的特异性，比如本研究中的肺癌组织学分型变量不一定适用于其他疾病的类似研究。

本研究采用数据局限于某单一省份，而肺癌患者住院费用的影响因素展现出地区差异性，未来研究可以考虑纳入更多地区的肺癌患者住院费用数据。此外，针对不同组织学类型的肺癌患者，可以尝试发展个性化预测模型。

参考文献

[1] BRAY F, FERLAY J, SOERJOMATARAM I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2018, 68(6): 394-424.

[2] 中华人民共和国统计局. 中国卫生健康统计年鉴[M]. 北京: 中国统计出版社, 2022.

National Bureau of Statistics of China. China health

- statistical yearbook[M]. Beijing: China Statistics Press, 2022.
- [3] HE S Y, LI H, CAO M M, et al. Trends and risk factors of lung cancer in China[J]. *Chinese Journal of Cancer Research*, 2020, 32(6): 683-94.
- [4] 石春雷, 娄培安, 石菊芳, 等. 中国 1996—2014 年肺癌经济负担研究系统评价[J]. *中国公共卫生*, 2017, 33(12): 1767-1774.
- SHI C L, LOU P A, SHI J F, et al. Economic burden of lung cancer in mainland China, 1996-2014: A systematic review[J]. *Chinese Journal of Public Health*, 2017, 33(12): 1767-1774.
- [5] 宋佳芳, 官海静, 刘国恩. 中国肺癌患者直接医疗费用研究的系统评价[J]. *中国循证医学杂志*, 2019, 19(1): 44-53.
- SONG J F, GUAN H J, LIU G E. Direct medical cost of lung cancer in China: A systematic review[J]. *Chinese Journal of Evidence-Based Medicine*, 2019, 19(1): 44-53.
- [6] 李月, 党媛媛, 汤榕. 基于 BP 神经网络模型的恶性肿瘤患者住院费用及影响因素[J]. *中国药物经济学*, 2022, 17(4): 10-13.
- LI Y, DANG Y Y, TANG R. Analysis of hospitalization expenses and influencing factors of malignant tumor patients based on BP neural network model[J]. *China Journal of Pharmaceutical Economics*, 2022, 17(4): 10-13.
- [7] 王婷婷, 于丽华, 郎婧婧等. 基于结构方程模型的肺癌手术患者住院费用影响因素分析[J]. *中国卫生经济*, 2019, 38(6): 61-63.
- WANG T T, YU L H, LANG J J, et al. Hospitalization costs influencing factors analysis of patients with lung cancer operation based on the method of structural equation model[J]. *Chinese Health Economics*, 2019, 38(6): 61-63.
- [8] 孙霖, 祁爱琴, 徐天和, 等. BP 神经网络模型与 logistic 回归方法在肺癌病人住院费用影响因素分析中的比较[J]. *中国医院统计*, 2015, 22(3): 173-175.
- SUN L, QI A Q, XU T H, et al. Comparison of BP neural network with logistic regression in influencing factor analysis of lung cancer hospitalization costs[J]. *Chinese Journal of Hospital Statistics*, 2015, 22(3): 173-175.
- [9] YU T Z, HE Z, ZHOU Q H, et al. Analysis of the factors influencing lung cancer hospitalization expenses using data mining[J]. *Thoracic Cancer*, 2015, 6(3): 338-345.
- [10] 张颖, 李利杰, 刘海容, 等. 支持向量机模型在肺癌病人住院费用影响因素分析中的应用[J]. *中国医院*, 2014, 18(10): 30-32.
- ZHANG Y, LI L J, LIU H R, et al. Application of the support vector machine model in the analysis of impact factors for hospitalization expenses[J]. *Chinese Hospitals*, 2014, 18(10): 30-32.
- [11] 黄利娟, 梁学柱, 查君敬. 基于 BP 神经网络的肺癌手术患者住院费用影响因素分析[J]. *中国病案*, 2014, 15(6): 53-55.
- HUANG L J, LIANG X Z, ZHA J J. Hospitalization costs influencing factors analysis of patients with lung cancer operation based on the BP neural network[J]. *Chinese Medical Record*, 2014, 15(6): 53-55.
- [12] 李文瑾, 田立启, 李晓雨, 等. 基于分位数回归模型的肺癌手术患者住院费用影响因素分析[J]. *中国卫生经济*, 2021, 40(9): 58-61.
- LI W J, TIAN L Q, LI X Y, et al. Analysis of influencing factors of hospitalization expenses of patients with lung cancer surgery based on quantile regression model[J]. *Chinese Health Economics*, 2021, 40(9): 58-61.
- [13] 赵少峰, 吕红亮, 彭映姝, 等. 肺癌患者住院费用影响因素分析——以四川省某三甲医院为例[J]. *预防医学情报杂志*, 2014, 30(5): 342-345.
- ZHAO S F, LU H L, PENG Y S, et al. Study on factors influencing hospitalization expenses of lung cancer in a tertiary hospital of Sichuan[J]. *Journal of Preventive Medicine Information*, 2014, 30(5): 342-345.
- [14] 陈曦, 刘晓雪, 兰勇兵, 等. 2018—2020 年武汉市肺癌患者住院费用及其影响因素分析[J]. *公共卫生与预防医学*, 2023, 34(4): 63-66.
- CHEN X, LIU X X, LAN Y B, et al. Influencing factors of hospitalization costs for lung cancer patients in Wuhan in 2018-2020[J]. *Journal of Public Health and Preventive Medicine*, 2023, 34(4): 63-66.
- [15] 吕红亮, 赵少峰, 谢小萍, 等. 四川省 16866 例肺癌患者住院费用影响因素分析[J]. *中国循证医学杂志*, 2013, 13(11): 1283-1287.
- LYU H L, ZHAO S F, XIE X P, et al. Influencing factors of hospital costs of 16866 cases of patients with lung cancer in Sichuan[J]. *Chinese Journal of Evidence-Based Medicine*, 2013, 13(11): 1283-1287.
- [16] LUO L, LI J L, LIAN S H, et al. Using machine learning approaches to predict high-cost chronic obstructive pulmonary disease patients in China[J]. *Health Informatics Journal*, 2020, 26(3): 1577-1598.
- [17] GOPUKUMAR D, GHOSHAL A, ZHAO H M. Predicting readmission charges billed by hospitals: machine learning approach[J]. *JMIR Medical Informatics*, 2022, 10(8): e37578.
- [18] CHARLSON M E, POMPEI P, ALES K L, et al. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation[J]. *Journal of Chronic Diseases*, 1987, 40(5): 373-383.
- [19] ZHANG Z H, ZHANG H Y, KHANAL M K. Development of scoring system for risk stratification in clinical medicine: A step-by-step tutorial[J]. *Annals of Translational Medicine*, 2017, 5(21): 436.