

• 论 著 •

## 基于机器学习的耐碳青霉烯类肠杆菌感染风险因素分析\*

肖春海<sup>1</sup>, 梁爽<sup>1</sup>, 刘向禄<sup>1</sup>, 吴娟芳<sup>1</sup>, 马慧敏<sup>1</sup>, 钟杉<sup>2△</sup>  
上海市第六人民医院金山分院: 1. 检验科; 2. 呼吸科, 上海 201500

**摘要:**目的 探讨医院耐碳青霉烯类肠杆菌(CRE)感染的机器学习模型及风险因素分析。方法 回顾性收集 2018—2022 年在该院治疗的 451 例产超广谱  $\beta$  内酰胺酶(ESBL)的肠杆菌感染患者病例资料, 根据其耐碳青霉烯是否耐药分为 CRE 组(115 例)和敏感组(336 例)。用 Logistic 回归分析、随机森林、支持向量机、神经网络 4 种机器学习方法构建预测模型并绘制受试者工作特征曲线进行评估, 根据性能最好的预测模型分析 CRE 感染的风险因素。结果 随机森林模型性能最优, 其曲线下面积最大, 为 0.952 3。随机森林模型预测 CRE 感染的风险因素为发热超过 3 d、有脑损害、引流液标本、躯干手术、一级或特级护理、ICU 治疗、降钙素原、抗厌氧菌治疗、用 3 代头孢、年龄、前清蛋白、肌酐、白细胞计数和清蛋白 15 项临床资料。结论 该研究得出的 CRE 预测模型具有较好的预测价值, 其风险因素对于临床防治 CRE 早期感染有指导意义。

**关键词:** 风险因素; 耐碳青霉烯; 肠杆菌; 机器学习

DOI: 10.3969/j.issn.1673-4130.2024.01.015

中图法分类号: R446.5

文章编号: 1673-4130(2024)01-0079-05

文献标志码: A

**Risk factor analysis of carbapenem-resistant enterobacteriaceae infection based on machine learning\***XIAO Chunhai<sup>1</sup>, LIANG Shuang<sup>1</sup>, LIU Xianglu<sup>1</sup>, WU Juanfang<sup>1</sup>, MA Huimin<sup>1</sup>, ZHONG Shan<sup>2△</sup>

1. Department of Clinical Laboratory; 2. Department of Respiration, Jinshan Branch of Shanghai Sixth People's Hospital, Shanghai 201500, China

**Abstract: Objective** To explore the machine learning model and risk factor analysis for hospital infection caused by carbapenem-resistant enterobacteriaceae (CRE). **Methods** The clinical data of totally 451 patients infected with extended-spectrum  $\beta$ -lactamases (ESBL) producing Enterobacteriaceae treated in the hospital from 2018 to 2022 were retrospectively collected. The patients were divided into CRE group (115 cases) and sensitive group (336 cases) according to the susceptibility of carbapenem. Four machine learning methods including Logistic regression analysis, random forest, support vector machine, and neural network were used to build prediction models and receiver operating characteristic curve was used to evaluate. Based on the prediction model with the best performance, risk factors for CRE infection were analyzed. **Results** Random forest model had the best performance, with the area under the curve of 0.952 3. The risk factors for predicting CRE infection by the random forest model included 15 clinical data items, namely fever for more than 3 days, cerebral injury, drainage fluid sample, trunk surgery, first-level or special-level nursing, ICU treatment, procalcitonin, anti-anaerobic bacteria, the use of third-generation cephalosporins, age, pre-albumin, creatinine, white blood cell count, and albumin. **Conclusion** The CRE prediction model developed in this study has good predictive value and the risk factors have guiding significance for the early prevention and treatment of CRE infection in clinical practice.

**Key words:** risk factors; carbapenem resistance; Enterobacteriaceae; machine learning

随着抗菌药物的广泛使用, 产超广谱  $\beta$  内酰胺酶(ESBL)的肠杆菌日益增多, 碳青霉烯类药物成为抗 ESBL 耐药菌的首选药物和最后一道防线<sup>[1]</sup>。但是耐碳青霉烯类肠杆菌(CRE)的数量却还是逐年递增<sup>[2-3]</sup>。CRE 传播能力强, 致死率高, 同时易携带多种耐药元器件, 已经成为全球公共卫生的重要挑战<sup>[4]</sup>。

一旦感染 CRE, 患者住院时间可能延长, 住院费用可能增加, 且患者的病情可能急剧恶化, 还会引发感染性休克、多器官脏器衰竭, 甚至死亡<sup>[5]</sup>。所以对于无药敏结果, 怀疑 ESBL 肠杆菌感染而进行经验性治疗 ESBL 耐药菌前, 分析是否可能感染 CRE 显得十分重要。本研究采用 4 种机器学习方法对 CRE 感染相对

\* 基金项目: 上海市金山区科委科技创新资助项目(2021-3-18)。

作者简介: 肖春海, 男, 副主任技师, 主要从事临床免疫、临床微生物方向的研究。△ 通信作者, E-mail: 18964864881@163.com。

于 ESBL 感染的影响因素进行分析, 以期为 CRE 防治提供科学依据。

### 1 资料与方法

**1.1 一般资料** 选取 2018—2022 年本院收治的 451 例 ESBL 肠杆菌感染住院患者作为研究对象, 其中男 246 例, 女 205 例。根据患者对碳青霉烯是否耐药分为 CRE 组(115 例)和敏感组(336 例)。本研究获本院伦理委员会批准(LLYJ-KY-2021-015-01)。ESBL 定义为对氨基青霉素耐药的肠杆菌科细菌。CRE 定义为对厄他培南、亚胺培南和美罗培南任一药物耐药的肠杆菌。

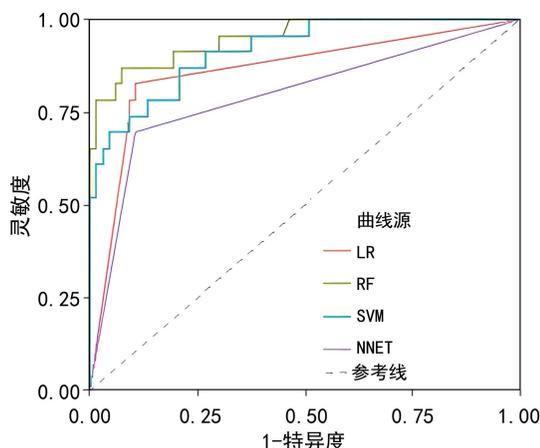
**1.2 方法** 收集患者的培养标本包括血、尿、痰、引流液(包括导管尖端)、伤口分泌物。采用生物梅里埃公司的 Vitek 2-Compact 全自动细菌鉴定药物敏感仪对本标本进行菌种鉴定、药敏试验。药敏试验结果的判读标准和质控均按美国临床和实验室标准委员会(CLSI)2021 版选药规则进行判断。

**1.3 统计学处理** 运用 R4.2.2 软件进行统计分析。符合正态分布的计量资料以  $\bar{x} \pm s$  表示, 组间比较采用 *t* 检验; 非正态分布的计量资料以  $M(P_{25}, P_{75})$  表示, 组间比较采用 Wilcoxon 秩和检验; 计数资料采用频数或百分率表示, 组间比较采用  $\chi^2$  检验。运用 Logistic 回归分析、随机森林、支持向量机、神经网络 4 种机器学习方法, 随机拆分 70% 的数据作为测试集, 进行风险因素的筛选, 分别构建预测模型, 剩余 30% 的数据作为验证集来评估模型的预测效果, 以受试者工作特征(ROC)曲线性能最优模型的风险因素重要性作为最终的风险因素。 $P < 0.05$  为差异有统计学意义。

## 2 结果

**2.1 两组患者基本临床资料比较** 与敏感组比较, CRE 组年龄较大( $P < 0.05$ ), 一级或特级护理、发热超过 3 d、用抗厌氧菌、使用喹诺酮、ICU 治疗、使用 3 代头孢、有脑损害、引流液标本、躯干手术比例更高( $P < 0.05$ ), 降钙素原、尿素氮水平较高( $P < 0.01$ ), 血红蛋白、淋巴细胞计数、清蛋白和前清蛋白水平较低( $P < 0.05$ )。见表 1。

**2.2 机器学习模型的对比结果** 4 种模型的结果显示, 随机森林准确度、Kappa 值、特异度最高, 分别为 0.911 1、0.736 3、1.000 0; 曲线下面积(AUC)随机森林最大, 为 0.952 3, 支持向量机灵敏度最高, 为 0.958 1。本数据集的最优模型为随机森林。见表 2、图 1。



注: LR 为 Logistic 回归, RF 为随机森林, SVM 为支持向量机, NNET 为神经网络。

图 1 各个模型在测试集中的 ROC 曲线图

表 1 两组患者基本临床资料比较 [ $n(\%)$  或  $M(P_{25}, P_{75})$  或  $\bar{x} \pm s$ ]

因素	敏感组( $n=336$ )	CRE 组( $n=115$ )	$t/\chi^2$	$P$
性别			0.86	0.354
女	157(46.7)	48(41.7)		
男	179(53.3)	67(58.3)		
年龄(岁)	69(56,78)	79(66,86)	5.84	<0.001
一级或特级护理			73.10	<0.001
否	247(73.5)	33(28.7)		
是	89(26.5)	82(71.3)		
发热超过 3 d			102.50	<0.001
否	205(85.1)	22(26.2)		
是	36(14.9)	62(73.8)		
使用氨基糖苷治疗			3.85	0.073
否	325(96.7)	115(100.0)		
是	11(3.3)	0(0.0)		
抗厌氧菌治疗			19.99	<0.001
否	197(58.6)	94(81.7)		
是	139(41.4)	21(18.3)		
使用喹诺酮类抗菌药物治疗			60.53	<0.001
否	122(36.3)	90(78.3)		

续表 1 两组患者基本临床资料比较[ $n(\%)$ 或  $M(P_{25}, P_{75})$ 或  $\bar{x} \pm s$ ]

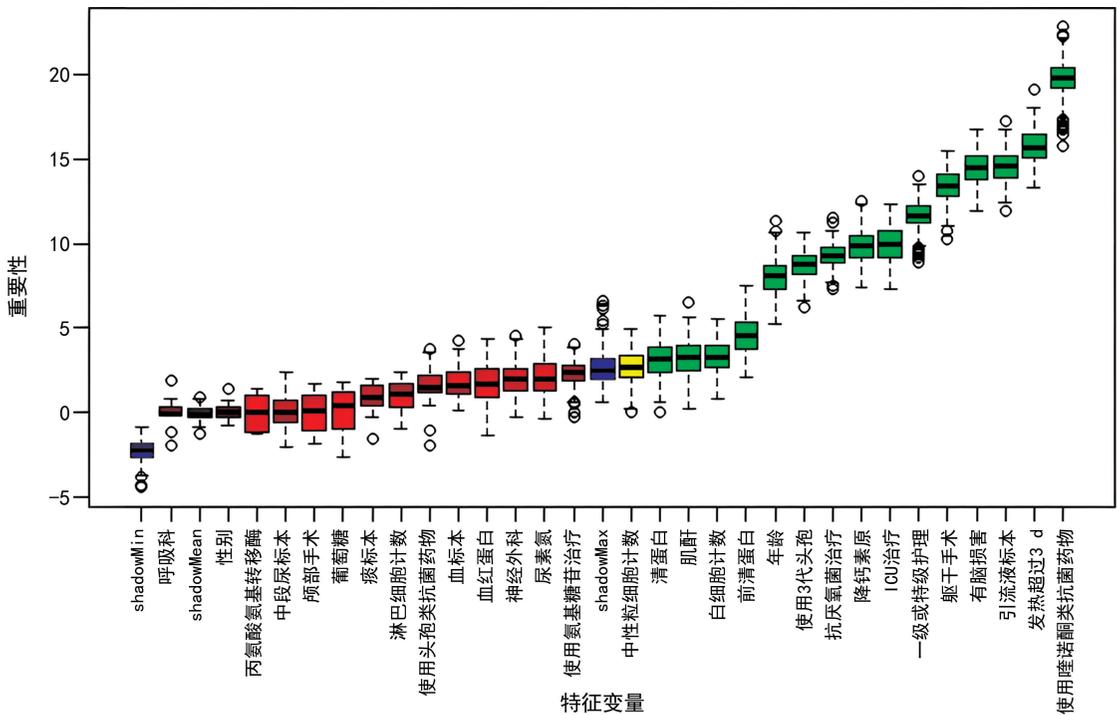
因素	敏感组( $n=336$ )	CRE 组( $n=115$ )	$t/\chi^2$	$P$
是	214(63.7)	25(21.7)		
有脑损害			49.91	<0.001
否	302(89.9)	70(60.9)		
是	34(10.1)	45(39.1)		
ICU 治疗			52.03	<0.001
否	266(79.2)	50(43.5)		
是	70(20.8)	65(56.5)		
使用 3 代头孢			9.27	0.002
否	175(52.1)	41(35.7)		
是	161(47.9)	74(64.3)		
引流液标本			59.33	<0.001
否	329(97.9)	87(75.7)		
是	7(2.1)	28(24.3)		
中段尿标本			0.02	0.885
否	175(52.1)	59(51.3)		
是	161(47.9)	56(48.7)		
痰标本			2.31	0.129
否	255(75.9)	79(68.7)		
是	81(24.1)	36(31.3)		
血标本			0.34	0.558
否	315(93.8)	106(92.2)		
是	21(6.2)	9(7.8)		
神经外科			0.01	0.914
否	305(90.8)	104(90.4)		
是	31(9.2)	11(9.6)		
呼吸科			1.41	0.234
否	290(86.3)	94(81.7)		
是	46(13.7)	21(18.3)		
颅部手术			0.04	1.000
否	326(97.0)	112(97.4)		
是	10(3.0)	3(2.6)		
躯干手术			50.51	<0.001
否	275(81.8)	55(47.8)		
是	61(18.2)	60(52.2)		
白细胞计数( $\times 10^9/L$ )	7.3(5.7,9.5)	7.9(5.8,10.9)	1.24	0.215
血红蛋白(g/L)	111.5 $\pm$ 21.5	106.5 $\pm$ 21.7	2.13	0.034
中性粒细胞计数( $\times 10^9/L$ )	5.0(3.6,7.0)	5.3(3.6,8.6)	1.18	0.240
淋巴细胞计数( $\times 10^9/L$ )	1.4(1.0,1.9)	1.2(0.8,1.8)	2.05	0.032
降钙素原(ng/mL)	0.1(0.0,0.5)	1.1(0.2,6.0)	6.06	<0.001
葡萄糖(mmol/L)	5.5(5.0,6.9)	5.7(4.9,6.9)	0.03	0.978
丙氨酸氨基转移酶(IU/L)	16.0(11.0,29.0)	15.0(9.2,26.8)	1.02	0.308
清蛋白(g/L)	35.5 $\pm$ 6.2	33.5 $\pm$ 6.3	2.89	0.004
前清蛋白(mg/L)	171.0(126.6,213.1)	155.9(107.4,199.0)	2.07	0.039
尿素氮( $\mu$ mol/L)	5.2(3.8,7.8)	6.2(4.2,10.7)	2.60	0.009
肌酐( $\mu$ mol/L)	68.0(55.5,87.0)	74.5(57.0,116.2)	1.84	0.066

表 2 4 种模型的对比结果表

机器学习模型	准确度(95%CI)	Kappa 值	特异度	灵敏度	AUC
Logistic 回归	0.866 7(0.778 7~0.929 2)	0.659 3	0.782 6	0.895 4	0.862 4
随机森林	0.911 1(0.832 3~0.960 8)	0.736 3	1.000 0	0.893 3	0.952 3
支持向量机	0.855 6(0.765 7~0.920 8)	0.549 7	0.478 3	0.958 1	0.902 7
神经网络	0.844 4(0.752 8~0.912 3)	0.591 2	0.695 7	0.895 5	0.795 6

2.3 随机森林的重要性变量 采用 Boruta 函数筛选出 15 个变量作为随机森林模型的重要性变量,分别为使用喹诺酮、发热超过 3 d、有脑损害、引流液标本、躯干手术、一级或特级护理、ICU 治疗、降钙素原、抗

厌氧菌治疗、用 3 代头孢、年龄、前清蛋白、肌酐、白细胞计数、清蛋白。所有变量具体分布见图 2。中性粒细胞计数结合表 1 中的  $P$  值( $P > 0.05$ ),不作为重要性变量。



注:绿色方框的变量表示随机森林证实的重要性变量,黄色方框的变量表示可能重要的变量,红色方框的变量表示不重要的变量。

图 2 随机森林重要性变量选取

### 3 讨论

在当前的临床治疗感染中,ESBL 和 CRE 耐药机制的重要性是无法忽视的问题<sup>[6]</sup>,因为治疗方案是不同的。为了控制感染,早期的经验性用药在没有药敏结果的情况下起着关键作用<sup>[7]</sup>。为了解决这一问题,本文收集了 451 例 ESBL 肠杆菌感染患者的临床数据并使用 4 种机器学习方法(Logistic 回归、随机森林、支持向量机和神经网络)进行建模和验证。在随机森林模型的各项性能指标中,Kappa 值(0.736 3)和 AUC(0.952 3)均为最大值,表明随机森林模型具有最佳预测 CRE 的能力。

本研究使用随机森林模型,分析了 15 个重要性变量作为 CRE 感染的风险因素,可以大致分为 4 组。第 1 组为使用喹诺酮类抗菌药物。喹诺酮类抗菌药物是最大的风险因素,表明对于使用喹诺酮类抗感染治疗无效的患者,其 CRE 感染风险很高。喹诺酮类

抗菌药物因具有良好药代动力学特征,生物利用度高,半衰期长,血药浓度较高,组织分布较广的特点,而被广泛使用,但其具有诱导和促进微生物耐药的作用<sup>[8]</sup>。因此,本研究喹诺酮类抗菌药物是最大的风险因素结论与其他研究报道 CRE 耐喹诺酮类抗菌药物比例高的结论一致<sup>[9-10]</sup>。这提示对于重症感染患者,要慎重使用喹诺酮类抗菌药物,应避免长期使用,以减少 CRE 感染的风险。

第 2 组风险因素包括发热超过 3 d、脑损害、引流液标本、躯干手术和一级或特级护理 5 个因素。长时间的发热是人体无法控制感染的一种表现。脑损害后机体为避免损伤部位的炎症反应波及周围神经元,通过多种途径影响免疫细胞功能和促炎细胞因子的分泌,从而抑制免疫功能<sup>[11]</sup>。引流液标本送检细菌培养多为引流液未能短期结束引流,细菌容易通过引流导管逆行感染,也反映了导管护理的重要性。躯干手

术多为胸腹部手术,很多需要术后引流操作。护理等级越高,说明患者的自身情况越差,免疫功能越低下。综上所述,免疫功能低下患者的躯干部位引流手术是 CRE 感染的重要风险因素。

第 3 组风险因素包括在 ICU 治疗、降钙素原、抗厌氧菌治疗、使用 3 代头孢和年龄等因素。在 ICU 治疗的基本都是危重症患者,且多数免疫功能低下。降钙素原水平则是反映有无细菌感染的一项重要指标。抗厌氧菌抗菌药物有诱导细菌产生生物膜的作用<sup>[12]</sup>,可能促进了肠杆菌耐药。这表明 ICU 内年龄较大的患者抗厌氧菌治疗或 3 代头孢抗感染治疗失败后,CRE 感染的风险可能会增加<sup>[13-15]</sup>。

第 4 组风险因素为前清蛋白、肌酐、白细胞计数和清蛋白,清蛋白和前清蛋白都与患者的营养状况密切相关,营养状况是影响身体免疫功能重要因素之一,反映肾功能的肌酐同样与免疫功能密切相关,细菌感染刺激机体释放更多的中性粒细胞为主的白细胞,甚至是核左移。因此营养支持治疗,保护肾脏功能,提高免疫功能,可以更好地预防和治疗 CRE 感染<sup>[16]</sup>。

本研究提出了一种基于随机森林模型的 CRE 风险因素预测方法,同时分析了各个因素影响 CRE 治疗的可能性原因,为临床治疗提供了有用参考。该方法需要对不同风险因素采取有针对性的治疗策略,帮助医生根据患者的个体差异制订更为精细的治疗方案<sup>[17]</sup>。尽管随机森林模型在预测 CRE 风险因素方面表现良好,但仍存在局限性。本研究仅分析了 4 种机器学习方法,未涉及更先进的深度学习模型。今后可以尝试引入更复杂的深度学习模型,以提高预测准确性。

## 参考文献

- [1] 武明芬,赵志刚.青霉烯类及碳青霉烯类抗生素的结构特点与分级管理建议[J].药学服务与研究,2021,21(3):193-199.
- [2] 金桂林,杨小军,肖亮,等.247 株耐碳青霉烯类肠杆菌目细菌的临床分布及耐药特性分析[J].江西中医药大学学报,2021,33(5):25-29.
- [3] BONOMO R A, BURD E M, CONLY J, et al. Carbapenemase-producing organisms: a global scourge[J]. Clin Infect Dis, 2018, 66(8): 1290-1297.
- [4] LOGAN L K, WEINSTEIN R A. The epidemiology of Carbapenem-resistant Enterobacteriaceae: the impact and evolution of a global menace[J]. J Infect Dis, 2017, 215 (Suppl\_1): S28-S36.
- [5] 中国碳青霉烯耐药药肠杆菌科细菌感染诊治与防控专家共识编写组,中国医药教育协会感染疾病专业委员会,中华医学会细菌感染与耐药防控专业委员会,等.中国碳青霉烯耐药药肠杆菌科细菌感染诊治与防控专家共识[J].中华医学杂志,2021,101(36):2850-2860.
- [6] DARLOW C A, HOPE W. Flomoxef for neonates: extending options for treatment of neonatal sepsis caused by ESBL-producing Enterobacteriales-authors' response[J]. J Antimicrob Chemother, 2022, 77(7): 2047-2048.
- [7] TAMMA P D, AITKEN S L, BONOMO R A, et al. Infectious diseases society of america antimicrobial resistant treatment guidance: gram-negative bacterial infections [J]. Clin Infect Dis, 2020, 72(7): e169-e183.
- [8] LUNDBORG C S, TAMHANKAR A J. Antibiotic residues in the environment of South East Asia[J]. BMJ, 2017, 358: 2440.
- [9] 陈虹虹,尤晓萍,蔡媛媛,等.血流感染中多重耐药菌的耐药性及感染危险因素分析[J].中国热带医学,2023,23(2):176-180.
- [10] 刘佳,周浩泉.某三甲医院儿童耐碳青霉烯类非发酵菌感染的危险因素及耐药性分析[J].安徽医学,2023,44(2):131-135.
- [11] 姜道明,滕晓华.脑损伤后免疫细胞及促炎因子与免疫抑制关系的研究进展[J].中国临床神经外科杂志,2022,27(2):138-140.
- [12] VUOTTO C, MOURA I, BARBANTI F, et al. Subinhibitory concentrations of metronidazole increase biofilm formation in Clostridium difficile strains [J]. Pathog Dis, 2016, 74(2): 114.
- [13] WANG Q, ZHANG Y, YAO X, et al. Risk factors and clinical outcomes for carbapenem-resistant Enterobacteriaceae nosocomial infections [J]. Eur J Clin Microbiol Infect Dis, 2016, 35(10): 1679-1689.
- [14] 刘洋,叶永青,王璐.老年患者感染耐碳青霉烯类肺炎克雷伯菌的危险因素及预测模型的建立[J].临床肺科杂志,2023,28(2):205-210.
- [15] 杨梅.医院耐碳青霉烯类肠杆菌科细菌临床特征和耐药分析[J].中国科技期刊数据库医药,2023,15(4):41-43.
- [16] 陈晓娟,王晓娟.耐碳青霉烯类β-内酰胺酶肠杆菌科细菌感染的诊断与治疗进展[J].中国医药导报,2022,19(1):1-4.
- [17] TOMPKINS K, VAN DUIN D. Treatment for carbapenem-resistant Enterobacteriales infections: recent advances and future directions [J]. Eur J Clin Microbiol Infect Dis, 2021, 40(10): 2053-2068.

(收稿日期:2023-03-16 修回日期:2023-09-25)