DOI:10.13210/j.cnki.jhmu.20241113.001
网络出版地址:https://link.cnki.net/urlid/46.1049.R.20241113.1618.002

多模态数据融合+动态机器学习构建ICU患者MDRO感染早期预警模型的研究

左 蝶[1],赵 佳[1],龙晓艳[2],马学先[3],刘 冰[3],王小燕[3],李 萍[3,4]✉
(1.新疆医科大学护理学院,新疆 乌鲁木齐 830091;2.中南大学湘雅医院,湖南 长沙 410028;3.新疆医科大学第二附属医院,新疆 乌鲁木齐 830063;4.新疆区域人群疾病与健康照护研究中心,新疆 乌鲁木齐 830063)

[**摘要**] **目的**:多模态数据融合+动态机器学习构建ICU患者多重耐药菌(multi-drug resistant organism,MDRO)感染风险预测模型,优选出最优预测模型为医院MDRO感染提供有效的评估工具。**方法**:选取某三甲医院ICU 2018年1月1日~2023年8月30日的1 200名患者,按照8∶2的比例随机分为训练集(n=960)和测试集(n=240),基于单因素分析将P<0.05的变量作为构建模型的纳入因素,运用随机森林(RF)、极度梯度提升(XGBoost)、决策树中的分类与回归树(classification and regression trees,CART)和logistic回归分别建立ICU患者MDRO感染风险预测模型,通过准确率、灵敏度、特异度、阳性预测值、阴性预测值、Kappa值、AUC值、决策曲线和校准曲线对4种模型的预测性能进行比较。**结果**:RF模型在训练集和测试集中表现最佳,其准确率、灵敏度、特异度、阳性预测值、阴性预测值及Kappa值均高于其他模型。AUC值从大到小顺序排列,训练集:RF>XGBoost>CRAT>logistic回归;测试集:RF>CRAT>logistic回归>XGBoost。本研究结果显示肺部感染、脑血管疾病、低蛋白血症及侵入性操作为4种模型的高风险预测因子是MDRO感染筛查及进行临床干预的重要理论依据。**结论**:基于RF算法建立的风险预测模型对ICU患者MDRO感染风险的预测性能优于其他三个机器算法构建的模型。

[**关键词**] ICU患者;MDRO感染风险;机器学习;预测模型
[**中图分类号**] R197.323 [**文献标志码**] A [**文章编号**] 2097-5953(2025)06-0421-12

Study on early warning model of MDRO infection in ICU patients based on multimodal data fusion and dynamic machine learning

ZUO Die[1], ZHAO Jia[1], LONG Xiaoyan[2], MA Xuexian[3], LIU Bing[3], WANG Xiaoyan[3], LI Ping[3,4]✉
(1. *School of Nursing, Xinjiang Medical University, Urumqi* 830091, *China*; 2. *Xiangya Hospital of Central South University, Changsha* 410028, *China*; 3. *The Second Affiliated Hospital of Xinjiang Medical University, Urumqi* 830063, *China*; 4. *Health Care Research Center for Xinjiang Regional Population, Urumqi* 830063, *China*)

[*Foundation Project*]: *This study was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region* (2023*D*01*C*121)

[*Author*]: *ZUO Die*, *E-mail*: 1425207135@qq.com.
[*Correspondence to*]: *LI Ping*, *Chief Nurse*, *E-mail*: 894088343@qq.com.

Received: 2024-09-17 *Revised*: 2024-11-03 *JHMU*, 2025;31(6):421-432

View from specialist: It is creative, and of certain scientific and educational value.

[Abstract] **Objective**: To build a risk prediction model of multi-drug resistant organism (MDRO) infection in ICU patients by multi-modal data fusion+dynamic machine learning and to select the optimal prediction model to provide an effective assessment tool for MDRO infection in hospitals. **Method**: A total of 1 200 patients from ICU of a tertiary hospital from between January 1, 2018 and August 30, 2023 were randomly divided into a training set (n=960) and a test set (n=240) according to the ratio of 8∶2. Based on univariate analysis, variables with P<0.05 were used as the inclusion factors in the model construction. Random forest

[基金项目]新疆维吾尔自治区自然科学基金(2023D01C121)
[作者简介]左蝶,E-mail:1425207135@qq.com。
[通讯作者]李萍,主任护师,E-mail:894088343@qq.com。
[收稿日期]2024-09-17 [修回日期]2024-11-03 网络出版时间:2024-11-14 10:01:03

(RF), eXtreme Gradient Boosting (XGBoost), classification and regression trees (CART) of the decision tree model, and logistic regression were used to establish MDRO infection risk prediction models for ICU patients. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, Kappa value, AUC value, decision curve, and calibration curve were used to compare the prediction performance of the four models. **Results**: The RF model performed the best in the training set and test set, and its accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and Kappa value were higher than the other models. AUC values were arranged in order from large to small, in the training set: RF>XGBoost>CRAT>logistic regression and in the test set: RF>CRAT>logistic regression>XGBoost. The results of this study showed that pulmonary infection, cerebrovascular disease, hypoproteinemia, and invasive operation were the risk predictors of the four models, which are important theoretical basis for MDRO infection screening and clinical intervention. **Conclusion**: The risk prediction model based on RF algorithm is better than the other three machine algorithms in predicting the risk of MDRO infection patients from ICU.

[**Key words**] ICU patients; MDRO infection risk; Machine learning; Prediction model

细菌耐药已成为全球公共卫生领域面临的一个巨大挑战,抗菌药物耐药性2019年被世界卫生组织列为影响全球健康十大威胁之一[1],对医院及社会经济均会产生重大影响[2]。多重耐药菌(multidrug resistant organism,MDRO)是指对3类及以上抗菌药物呈现耐药的细菌[3]。患者发生MDRO感染会伴随复杂且难治疗的特点,会延长患者住院时间、增加患者的痛苦和经济负担[4]。因此需要医护人员、检验科、药学及保洁等多部门协同做好防控和治疗工作[5]。重症监护室(intensive care unit,ICU)患者因自身免疫力下降、肠道菌群失调、抗生素使用时间长、侵入性操作等因素,是发生MDRO感染最多的人群[6]。研究指出[7],ICU患者发生感染的情况是普通科室的2~5倍,由MDRO引发的感染高达60%以上,其引起的死亡是普通感染的1.7倍[8]。尽早发现MDRO感染的高危人群对于感染的预防与控制工作尤为迫切。基于此构建风险评估预测模型,实现早期发现并实施规范化的干预措施极其重要[9]。

机器学习(machine learning,ML)是人工智能的核心技术,用于处理一些复杂变量之间的非线性关系,从而获得一个较为稳健的风险预测模型,模型输出结果主要是依靠算法来完成,与传统的统计分析方法相比,ML对临床诊疗指导意义更大,目前已在疾病风险预测方面被广泛使用[10]。本研究旨在结合多模态数据融合和动态ML技术,构建ICU患者MDRO感染早期预警模型,通过多模态数据融合技术,从患者的医疗记录、生理监测数据、影像资料等多个数据源中提取出对MDRO感染预测有价值的特征;利用动态ML算法,构建能够实时更新和优化的预测模型;通过临床验证和评估,验证模型的预测准确性和实用性。

1　资料与方法

1.1　一般资料

1.1.1　数据来源　采取单中心回顾性研究的方法,选取某三甲医院2018年1月1日~2023年8月30日ICU患者1 200例,发生MDRO感染有600例,按照1∶1的比例,选择相同时间内未发生感染的600例患者纳入。本研究通过医院医学伦理委员会审批(伦理审批号:KY2023112111)。

1.1.2　诊断标准　MDRO感染诊断依据医院感染的诊断标准[11],临床重要耐药菌感染诊断、治疗及防控指南[12],结合医院感染监测软件的提示信息、病原学诊断,由临床医生判断是否为医院感染,并由医院感染管理科专人审核,排除定植及污染患者,最终判断是否为MDRO医院感染。住院期间同一患者的相同菌种标本,均视为同一菌株,不重复记录。

1.1.3　研究对象　纳入标准:(1)ICU住院时间≥48 h;(2)患者感染部位为下呼吸道感染者;(3)临床标本中分离出的耐碳青霉烯类:鲍曼不动杆菌(Carbapenem-resistant *Acinetobacter baumannii*,CRAB)、大肠埃希菌(carbapenem-resistant *Enterobacteriaceae*,CRE)、肺炎克雷伯菌(carbapenem-resistant *Klebsiella pneumoniae*,CRKP)、铜绿假单胞菌(carbapenem-resistant *Pseudomonas aeruginosa*,CRPA)及耐甲氧西林金黄色葡萄球菌(methicillin-resistant *Staphylococcus aureus*,MRSA);(4)年龄≥18岁且病例资料完整者;(5)符合《医院感染诊断标准》和MDRO感染判定标准者。排除标准:(1)入住ICU前或24 h内感染者;(2)病历数据资料缺失≥20%者;(3)住院期间从未送检者及送检结果有误者;(4)妊娠期及产褥期妇女。本研究筛选感染患者流程图见(图1)。

1.1.4　样本量计算　逻辑回归模型参考构建风险预测模型样本量10倍原则,即阳性结局事件数量为

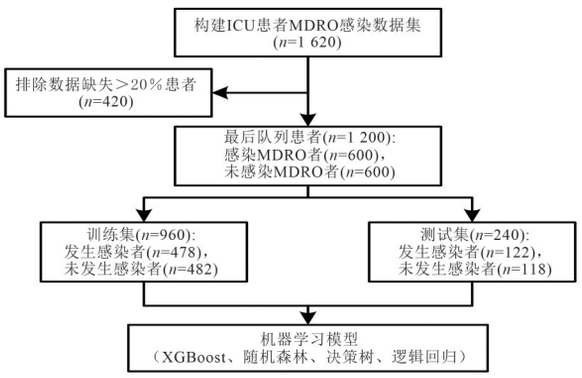


图1 本研究筛选感染患者流程图

Fig 1 Flowchart of screening infected patients in this study

最终进入模型变量数10倍以上[13]。然而目前ML模型还无相关针对性的样本估算公式，根据相关研究发现引起患者发生MDRO感染的影响因素约为26个，因此本研究基于逻辑回归模型的样本估算方式。本研究总计收集1 200例患者，对数据进行清洗后，使用python中函数将数据进行随机划分，即训练集960人，测试集240人。

1.2 方法

1.2.1 资料收集方法 调取医院信息系统中的ICU患者信息，根据纳入和排除标准回顾性获取患者的一般情况、基础疾病、实验室检查、抗生素使用等临床资料；结局变量MDRO感染通过医院感染信息系统获取；包括性别，年龄(＜60岁/≥60岁)、职业、婚姻情况、入院途径，基础疾病(冠心病、低蛋白血症、肺部感染等)、侵入性操作(气管插管/切开、吸痰、留置尿管、机械通气、中心静脉置管等)、实验室指标(白细胞计数、C-反应蛋白、白蛋白含量、降钙素及白介素-6)、急性生理与慢性健康评估(Acute Physiology and Chronic Health Evaluation Ⅱ，APACHE Ⅱ)和格拉斯哥昏迷评分法(Glasgow Coma Scale，GCS)评分情况(其中APACHE Ⅱ≥15分患者归属为重症，＜15分归属为非重症；GCS得分值越高，提示患者意识状态越好，≤8分为昏迷，昏迷程度越重者的昏迷指数分值越低，≤3分则提示脑死亡或预后极差)，共收集600例MDRO感染患者，包括CRAB 109例、CRE 119例、CRKP 275例、CRPA 42例、MRSA 55例，研究对象筛选过程(图1)。

1.2.2 统计学处理 使用SPSS26.0进行统计分析，计数资料用频数和百分比(%)描述。基于Python 3.9构建模型，本研究数据均采用卡方检验进行单因素分析来筛选影响因素，将$P<0.05$的变量纳入后续研究。本研究样本量按8∶2划分为训练集和测试集，利用训练集数据分别构建logistic回归、决策树中分类与回归树(classification and regression trees，CART)、极端梯度提升(eXtreme Gradient Boosting，XGBoost)和随机森林(random forest，RF)4种模型。采用5折交叉验证方法确定最优超参数，在测试集上进行独立验证，通过比较ROC曲线下面积(area under the receiver operating characteristic curve，AUC)、校准及决策曲线获得最佳模型，通过准确率、阳性预测值(positive predictive value，PPV)、阴性预测值(negative predictive value，NPV)及Kappa值等指标对模型预测性能进行评估并计算变量重要性。

2 结果

2.1 单因素分析

根据是否发生MDRO感染，进行单因素分析发现肺部感染、脑血管病、低蛋白血症、冠心病、慢性阻塞性肺病、高血压、手术史、住院天数、年龄、C-反应蛋白、白蛋白含量、抗生素使用情况等26个变量的差异具有统计学意义($P<0.05$)，性别、职业、婚姻情况等($P>0.05$)差异无统计学意义具有均衡可比性结果见(表1)。使用皮尔逊相关系数生成特征相关性图，显示各变量特征彼此间无高度相关，最大相关性值未超过0.5(图2)说明每个变量对MDRO感染风险的预测各有贡献。

表1 MDRO感染的单因素分析

Tab 1 Single factor analysis of MDRO infection

项目	分类	感染(%)		总计	χ^2	P
		未感染	感染			
一般资料						
性别	男	327 (54.50)	294 (49.00)	621 (51.75)	3.634	0.057
	女	273 (45.50)	306 (51.00)	579 (48.25)		
职业	无	249 (41.50)	216 (36.00)	465 (38.75)	3.824	0.051
	有	351 (58.50)	384 (64.00)	735 (61.25)		
年龄/岁	＜60	272 (45.33)	184 (30.67)	456 (38.00)	27.391	＜0.001**
	≥60	328 (54.67)	416 (69.33)	744 (62.00)		

续表

项目	分类	感染(%)		总计	χ^2	P
		未感染	感染			
婚姻	未婚	7 (1.17)	17 (2.83)	24 (2.00)	7.075	0.070
	已婚	544 (90.67)	518 (86.33)	1062 (88.50)		
	丧偶	42 (7.00)	55 (9.17)	97 (8.08)		
	离婚	7 (1.17)	10 (1.67)	17 (1.42)		
付费方式	医保	368 (61.33)	337 (56.17)	705 (58.75)	3.305	0.069
	自费	232 (38.67)	263 (43.83)	495 (41.25)		
住院天数/d	＜7	315 (52.50)	273 (45.50)	588 (49.00)	5.882	0.015*
	≥7	285 (47.50)	327 (54.50)	612 (51.00)		
住院次数/次	1	273 (45.50)	306 (51.00)	579 (48.25)	3.634	0.057
	＞1	327 (54.50)	294 (49.00)	621 (51.75)		
入院途径	急诊	220 (36.67)	213 (35.50)	433 (36.08)	0.188	0.911
	门诊	142 (23.67)	146 (24.33)	288 (24.00)		
	外院转入	238 (39.67)	241 (40.17)	479 (39.92)		
离院方式	出院	386 (64.33)	375 (62.50)	761 (63.42)	0.435	0.51
	死亡	214 (35.67)	225 (37.50)	439 (36.58)		
手术史	无	436 (72.67)	185 (30.83)	621 (51.75)	210.261	＜0.001**
	有	164 (27.33)	415 (69.17)	579 (48.25)		
外伤史	无	477 (79.50)	477 (79.50)	954 (79.50)	0	＞0.999
	有	123 (20.50)	123 (20.50)	246 (20.50)		
基础疾病						
肺部感染	否	447 (74.50)	159 (26.50)	606 (50.50)	276.508	＜0.001**
	有	153 (25.50)	441 (73.50)	594 (49.50)		
脑血管病	否	437 (72.83)	172 (28.67)	609 (50.75)	234.136	＜0.001**
	有	163 (27.17)	428 (71.33)	591 (49.25)		
低蛋白血症	否	419 (69.83)	196 (32.67)	615 (51.25)	165.867	＜0.001**
	有	181 (30.17)	404 (67.33)	585 (48.75)		
冠心病	否	403 (67.17)	182 (30.33)	585 (48.75)	162.905	＜0.001**
	有	197 (32.83)	418 (69.67)	615 (51.25)		
高血压	否	283 (47.17)	333 (55.50)	616 (51.33)	8.339	0.004**
	有	317 (52.83)	267 (44.50)	584 (48.67)		
慢性阻塞性肺病	否	293 (48.83)	334 (55.67)	627 (52.25)	5.615	0.018*
	有	307 (51.17)	266 (44.33)	573 (47.75)		
慢性支气管炎	否	289 (48.17)	322 (53.67)	611 (50.92)	3.631	0.057
	有	311 (51.83)	278 (46.33)	589 (49.08)		
呼吸衰竭	否	284 (47.33)	317 (52.83)	601 (50.08)	3.63	0.057
	有	316 (52.67)	283 (47.17)	599 (49.92)		
糖尿病	否	227 (37.83)	228 (38.00)	455 (37.92)	0.004	0.953
	有	373 (62.17)	372 (62.00)	745 (62.08)		
肿瘤	否	393 (65.50)	394 (65.67)	787 (65.58)	0.004	0.952
	有	207 (34.50)	206 (34.33)	413 (34.42)		
侵入性操作						
吸痰	无	332 (55.33)	275 (45.83)	607 (50.58)	10.831	0.001**
	有	268 (44.67)	325 (54.17)	593 (49.42)		
灌肠	无	276 (46.00)	320 (53.33)	596 (49.67)	6.454	0.011*
	有	324 (54.00)	280 (46.67)	604 (50.33)		
气管切开/插管	无	268 (44.67)	310 (51.67)	578 (48.17)	5.888	0.015*
	有	332 (55.33)	290 (48.33)	622 (51.83)		
机械通气	无	313 (52.17)	274 (45.67)	587 (48.92)	5.072	0.024*
	有	287 (47.83)	326 (54.33)	613 (51.08)		

续表

项目	分类	感染(%)		总计	χ^2	P
		未感染	感染			
导尿	无	321 (53.50)	282 (47.00)	603 (50.25)	5.07	0.024*
	有	279 (46.50)	318 (53.00)	597 (49.75)		
中心静脉置管	无	259 (43.17)	297 (49.50)	556 (46.33)	4.839	0.028*
	有	341 (56.83)	303 (50.50)	644 (53.67)		
引流管	无	273 (45.50)	311 (51.83)	584 (48.67)	4.817	0.028*
	有	327 (54.50)	289 (48.17)	616 (51.33)		
输血	无	504 (84.00)	489 (81.50)	993 (82.75)	1.314	0.252
	有	96 (16.00)	111 (18.50)	207 (17.25)		
鼻饲	无	134 (22.33)	122 (20.33)	256 (21.33)	0.715	0.398
	有	466 (77.67)	478 (79.67)	944 (78.67)		
血液透析	无	453 (75.50)	441 (73.50)	894 (74.50)	0.632	0.427
	有	147 (24.50)	159 (26.50)	306 (25.50)		
纤维支气管镜	无	263 (43.83)	269 (44.83)	532 (44.33)	0.122	0.727
	有	337 (56.17)	331 (55.17)	668 (55.67)		
呼吸机使用时间(d)	＜7	301 (50.17)	222 (37.00)	523 (43.58)	21.152	＜0.001**
	≥7	299 (49.83)	378 (63.00)	677 (56.42)		
尿管留置时间(d)	＜7	403 (67.17)	343 (57.17)	746 (62.17)	12.755	＜0.001**
	≥7	197 (32.83)	257 (42.83)	454 (37.83)		
实验室指标						
白蛋白(g/L)	正常	319 (53.17)	266 (44.33)	585 (48.75)	9.369	0.002**
	异常	281 (46.83)	334 (55.67)	615 (51.25)		
降钙素(ng/mL)	正常	323 (53.83)	282 (47.00)	605 (50.42)	5.604	0.018*
	异常	277 (46.17)	318 (53.00)	595 (49.58)		
C-反应蛋白(mg/L)	正常	324 (54.00)	289 (48.17)	613 (51.08)	4.085	0.043*
	异常	276 (46.00)	311 (51.83)	587 (48.92)		
白细胞计数×10^9/L	正常	253 (42.17)	246 (41.00)	499 (41.58)	0.168	0.682
	异常	347 (57.83)	354 (59.00)	701 (58.42)		
白介素-6(pg/mL)	正常	200 (33.33)	200 (33.33)	400 (33.33)	0	＞0.999
	异常	400 (66.67)	400 (66.67)	800 (66.67)		
抗生素使用情况						
联用情况	未联	202 (33.67)	201 (33.50)	403 (33.58)	9.95	0.007**
	≥2种	173 (28.83)	218 (36.33)	391 (32.58)		
	≥3种	225 (37.50)	181 (30.17)	406 (33.83)		
治疗时间/d	＜7	317 (52.83)	273 (45.50)	590 (49.17)	6.455	0.011*
	≥7	283 (47.17)	327 (54.50)	610 (50.83)		
治疗用药	无	315 (52.50)	278 (46.33)	593 (49.42)	4.564	0.033*
	是	285 (47.50)	322 (53.67)	607 (50.58)		
评分						
APACHE Ⅱ评分	＜15分	326 (54.33)	290 (48.33)	616 (51.33)	4.323	0.038*
	≥15分	274 (45.67)	310 (51.67)	584 (48.67)		
GCS评分	≤8分	301 (50.17)	335 (55.83)	636 (53.00)	3.867	0.049*
	＞8分	299 (49.83)	265 (44.17)	564 (47.00)		

注：*P＜0.05，**P＜0.01。

2.3 MDRO感染风险的4种预测模型

2.3.1 Logistic回归 将单因素分析中P＜0.05的26个变量作为自变量，构建逻辑回归模型，变量赋值(表2)使用5折交叉验证的方法确定最优模型。Logistic回归结果显示，肺部感染、脑血管病、低蛋白血症、高血压、冠心病、慢性阻塞性肺病、手术史、年龄、呼吸机使用时间、尿管留置时间、导尿、灌肠、气

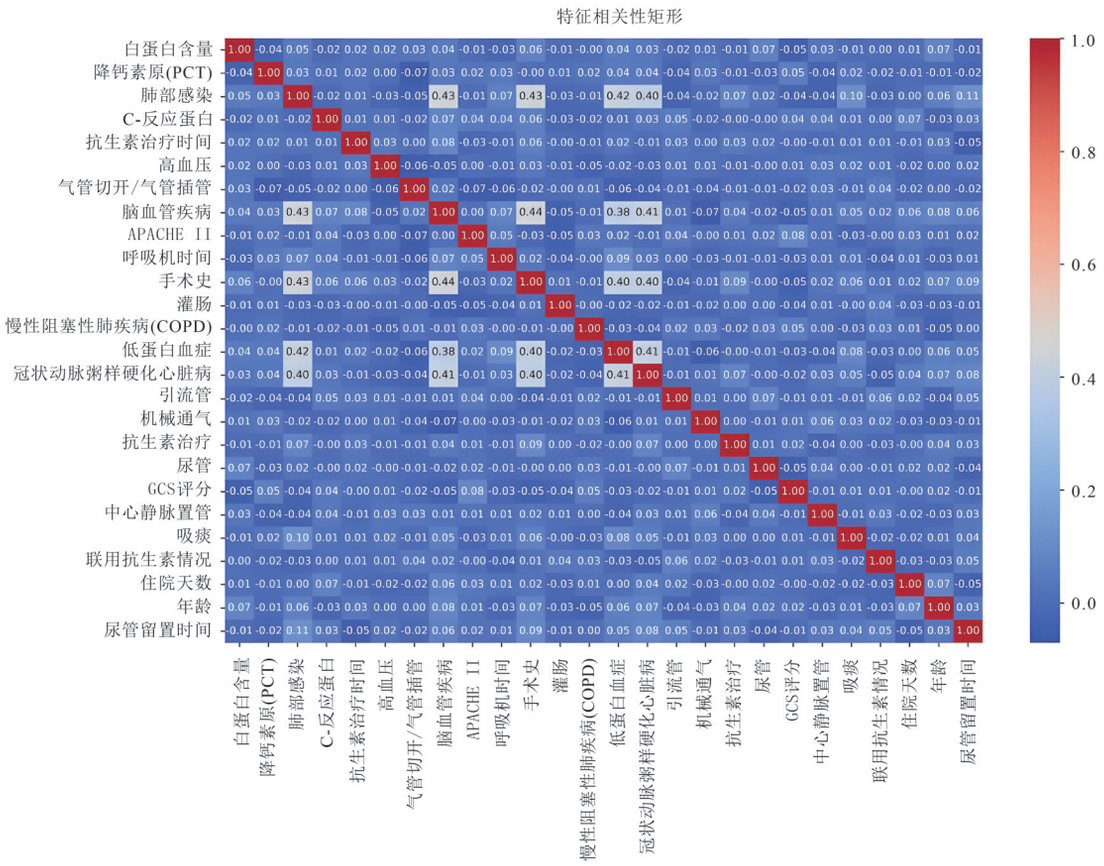


图2　特征相关性图

Fig 2　Feature correlation diagram

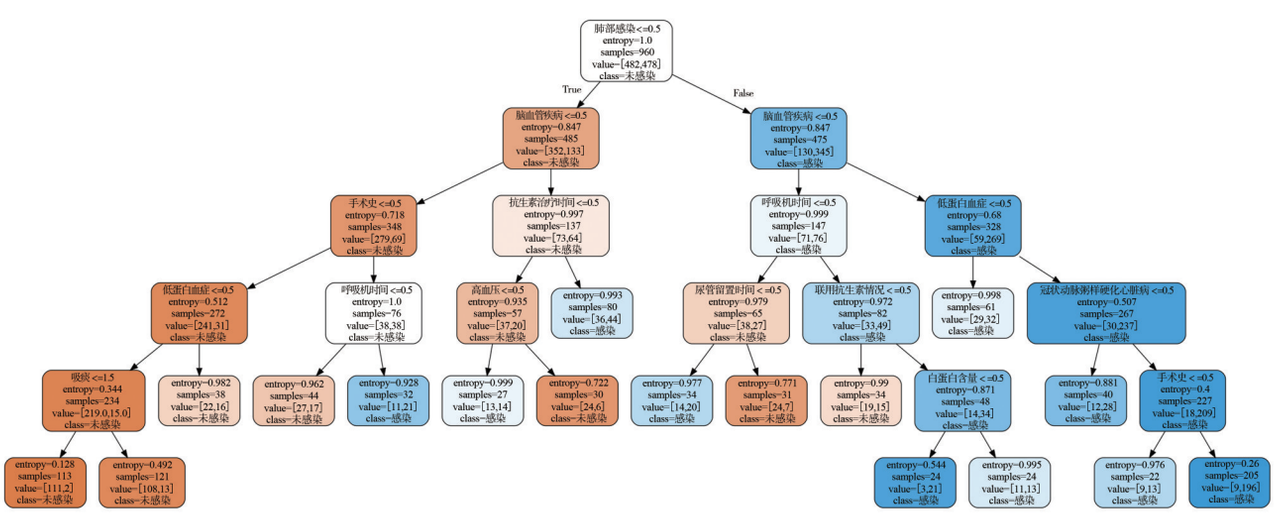


图3　决策树模型

Fig 3　Decision tree model

管切开/插管、中心静脉置管、引流管、白蛋白、降钙素原、APACHE Ⅱ评分18个变量是患者发生MDRO感染的重要危险因素(表3)。

2.3.2　决策树　基于单因素分析筛选出的25个变量作为预测因子纳入模型中，以信息增益为基础进行树的生长，选择最小代价复杂度剪枝法对决策树进行后剪枝，生成模型。该图包含6层，其中肺部感染、脑血管病、低蛋白血症、冠心病、高血压、灌肠、

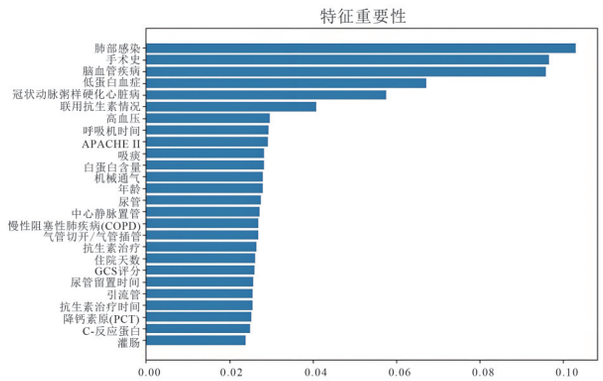


图4 随机森林模型

Fig 4 Random forest model

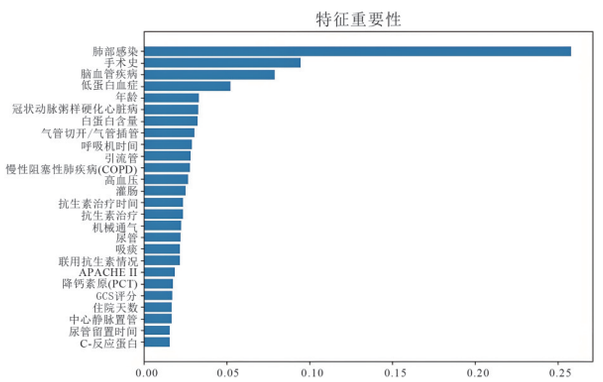


图5 XGBoost模型

Fig 5 XGBoost model

表2 赋值表

Tab 2 Assignment table

变量	赋值说明
感染(因变量)	无感染=0,感染=1
年龄	<60=1,≥60=2
住院天数/d	<7=1,≥7=2
手术史	无=0,有=1
高血压	无=0,有=1
冠心病	无=0,有=1
肺部感染	无=0,有=1
脑血管病	无=0,有=1
低蛋白血症	无=0,有=1
慢性阻塞性肺病	无=0,有=1
吸痰	无=0,有=1
灌肠	无=0,有=1
导尿	无=0,有=1
引流管	无=0,有=1
机械通气	无=0,有=1
气管切开/插管	无=0,有=1
中心静脉置管	无=0,有=1
呼吸机使用时间/d	<7=1,≥7=2
尿管留置时间/d	<7=1,≥7=2
抗生素治疗	无=0,有=1
联用抗生素	未联=0,2联=1,3联=2
抗生素治疗时间/d	<7=1,≥7=2
白蛋白	正常=0,异常=1
降钙素	正常=0,异常=1
C-反应蛋白	正常=0,异常=1
APACHE Ⅱ/分	<15=1,≥15=2
GCS/分	≤8=1,>8=2

吸痰、手术史、尿管留置时间、联用抗生素情况、白蛋白含量、呼吸机和抗生素治疗时间13个变量是模型的重要解释变量(图3)。

2.3.3 随机森林 通过Python实现算法拟合,经相应验证发现当树的数量为100时,总样本量误差最小,此时袋外数据out-of-bag (OOB)误差为0.256,此时该模型准确度较高,采用5折交叉验证的方式对变量进行筛选,结果显示当选取前23个变量时,OOB误差最小,即肺部感染、手术史及脑血管疾病等23个变量为模型的重要解释变量(图4)。

2.3.4 极度梯度提升树 通过Python实现算法拟合,经过模型验证使用网格搜索方式对每一种参数组合进行评估,使用5折交叉验证来评估不同特征子集性能,结果显示选择肺部感染、手术史、脑血管疾病及低蛋白血症前15个特征,交叉验证方法得到的评价指标性能最高为模型的重要解释变量(图5)。

2.4 模型预测性能比较

2.4.1 训练集比较 训练集中发现随机森林最佳,其准确率为97.08%,灵敏度为97.7%,特异度为96.47%,阳性预测值为96.49%,阴性预测值为97.69%,Kappa值为0.942,均高于其他三种模型(表4)。ROC曲线下面积AUC值由大到小:RF(99.50%)>XGBoost(95.20%)>CART(86.60%)>logistic回归(85.90%)(图6A)。校准曲线用于评估模型预测概率的准确性,其通过比较模型预测的概率与实际发生的概率来评估模型的校准度,横坐标是模型预测的概率平均预测概率,纵坐标是实际发生的概率,其中RF紧沿着45°对角线,表明其预测概率与实际概率有很好的一致性(图6B)。决策曲线分析通过比较不同阈值下模型净效益来评估模型在各种决策阈值上的表现;净效益是真阳性率和假阳性率之间的权衡,其中RF在大多数阈值概率下高于其他模型,阈值范围在0~1之间说明其具有较好的预测性能,在实际临床应用中能够有效地识别出需要早期干预的患者(图6C)。

2.4.2 测试集比较 测试集中显示随机森林模型最佳,其准确率为79.58%,特异度为84.75%,阳性预测值为83.49%,均优于其他三种(表5)。ROC曲线下面积AUC值由大到小:RF(90.50%)>CART(89.40%)>logistic回归(87.20%)>XGBoost

表3　Logistic回归分析结果

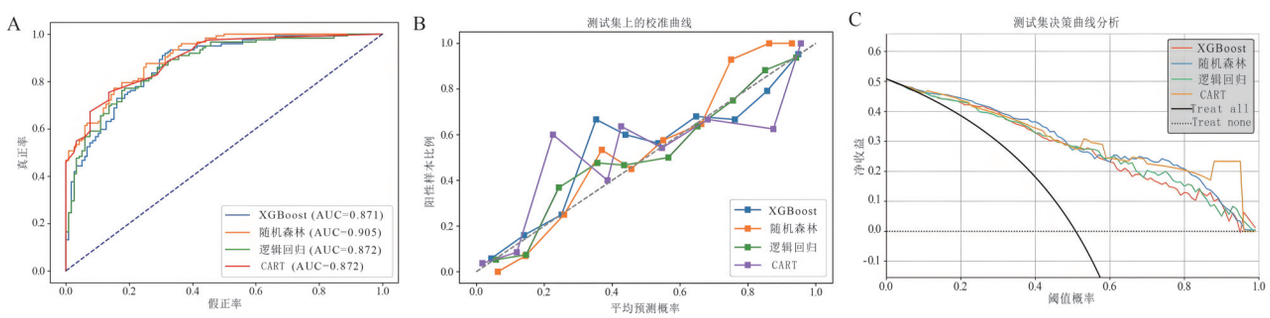
Tab 3　Logistic regression analysis results

常项	回归系数	标准误	*Z*	*Wald* χ^2	*P*	*OR*	*OR* 95% *CI*
住院天数	0.294	0.153	1.926	3.709	0.054	1.341	0.995～1.809
年龄	0.627	0.160	3.933	15.468	＜0.001	1.873	1.370～2.560
手术史	0.914	0.164	5.569	31.019	＜0.001	2.494	1.808～3.440
肺部感染	1.255	0.163	7.723	59.650	＜0.001	3.508	2.551～4.824
脑血管病	0.963	0.163	5.897	34.776	＜0.001	2.618	1.901～3.605
低蛋白血症	0.466	0.162	2.886	8.326	0.004	1.594	1.161～2.189
冠心病	0.551	0.162	3.390	11.495	0.001	1.734	1.261～2.384
高血压	－0.530	0.154	－3.450	11.903	0.001	0.588	0.435～0.795
慢性阻塞性肺病	－0.387	0.153	－2.536	6.432	0.011	0.679	0.504～0.916
吸痰	0.249	0.151	1.641	2.694	0.101	1.282	0.953～1.726
灌肠	－0.300	0.152	－1.978	3.913	0.048	0.741	0.550～0.997
气管切开/插管	－0.315	0.153	－2.053	4.216	0.040	0.73	0.540～0.986
机械通气	－0.241	0.153	－1.581	2.500	0.114	0.786	0.583～1.059
导尿	0.427	0.153	2.792	7.793	0.005	1.532	1.136～2.067
中心静脉置管	－0.369	0.153	－2.412	5.819	0.016	0.691	0.512～0.933
引流管	－0.338	0.153	－2.213	4.896	0.027	0.713	0.529～0.962
呼吸机时间	0.598	0.154	3.884	15.087	＜0.001	1.819	1.345～2.461
尿管留置时间	0.358	0.158	2.267	5.138	0.023	1.431	1.050～1.951
抗生素治疗	0.071	0.152	0.465	0.216	0.642	1.073	0.797～1.445
联用抗生素情	－0.092	0.092	－1.002	1.003	0.317	0.912	0.762～1.092
抗生素治疗时间	0.294	0.152	1.929	3.721	0.054	1.341	0.995～1.807
白蛋白	0.367	0.153	2.401	5.766	0.016	1.443	1.070～1.946
降钙素原	0.300	0.152	1.970	3.882	0.049	1.350	1.002～1.820
C-反应蛋白	0.254	0.153	1.659	2.751	0.097	1.290	0.955～1.742
APACHE Ⅱ	0.391	0.153	2.562	6.561	0.010	1.479	1.096～1.995
GCS	－0.200	0.153	－1.304	1.700	0.192	0.819	0.607～1.106

表4　训练集4种模型预测性能

Tab 4　Four models predict performance of the training set

模型	灵敏度(%)	特异度(%)	PPV(%)	NPV(%)	*Kappa*值	准确率(%)	AUC
XGBoost	87.87	85.27	85.54	87.63	0.731	86.56	0.952
RF	97.70	96.47	96.49	97.69	0.942	97.08	0.995
Logistic	76.15	75.31	75.36	76.10	0.515	75.73	0.859
CART	84.10	69.50	73.22	81.51	0.536	76.77	0.866



A:ROC曲线;B:校准曲线;C: 决策曲线

图6　效能评估曲线

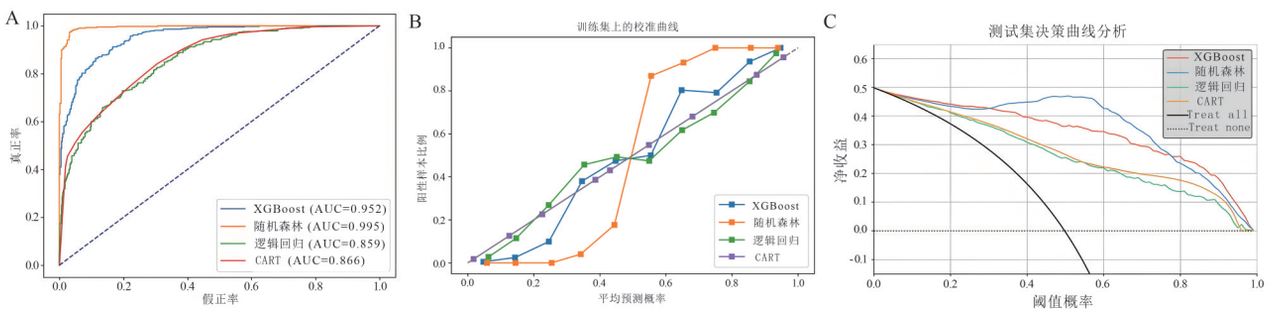
Fig 6　Performance evaluation curve

(87.10%)(图7A)。校准曲线图显示RF接近45°对角线,校准度最佳(图7B)。决策曲线图显示RF在大多数阈值概率下高于其他模型,其阈值范围在0～1之间说明能够保持较好的预测性能(图7C)。

表5 测试集4种模型的预测性能

Tab 5 Four models predict performance of the test set

模型	灵敏度(%)	特异度(%)	PPV(%)	NPV(%)	*Kappa*值	准确率(%)	AUC
XGBoost	76.23	77.97	78.15	76.03	0.542	77.08	0.871
RF	74.59	84.75	83.49	76.34	0.592	79.58	0.905
Logistic	77.05	77.97	78.33	76.67	0.550	77.50	0.872
CART	82.79	71.19	74.81	80.00	0.541	77.08	0.894



A:ROC曲线;B:校准曲线;C:决策曲线

图7 效能评估曲线

Fig 7 Performance evaluation curve

3 讨论

3.1 4种算法构建的预测模型的优势与不足

目前MDRO感染风险预测模型大多基于logistic回归[14,15],其是一种线性模型,但引起MDRO感染相关危险因素错综复杂,每个变量间会有所关联,仅用logistic回归对感染风险预测不太具有代表性。因此本研究将基于4种算法构建ICU患者MDRO感染风险预测模型,研究结果显示RF是最优模型,其在训练集和测试集均呈现最佳表现。Logistic回归因可解释性常用于研究危险因素,对样本量及数据类型要求较高,易受缺失数据影响,能提供一种直观方式来解释每个特征对预测结果的影响程度[16],通过系数可直接了解特征重要性,计算成本较低且训练速度快,有助于临床工作者进行风险分层。CART解释性强且模型结构直观,对缺失变量及数据分布类型不敏感,但容易产生复杂结构过拟合和泛化能力弱等问题[17]。XGBoost自动分析预测因子间交互作用和非线性效应[18],对缺失值和异常值有较好处理能力,能有效提高预测精度从而获得更准确稳定结果,可构建出预测精度较高模型。RF通过组合多个决策树进行分析,降低单个决策树过拟合风险,提高模型泛化能力[19],每棵树不同分类边界相互补充,降低整体预测错误率从而达到预测效果,对异常值具有较高容忍度,不易受到这些因素影响。因此与其他算法相比,RF对于预测ICU患者发生MDRO感染风险具有较强优势,可筛选出可解释性关键危险因素,能有效为感染预防和干预提供一定参考价值。

本研究将静态算法转换为动态算法,主要基于在线学习框架[20],将RF、XGBoost、CRAT和logistic回归4种算法作为基础,对数据进行流处理和模型更新,能有效处理单个数据点并在接收新数据时进行即时调整模型参数。主要先基于静态算法初始化方法设置初始模型参数来逐个处理数据点,计算预测值与真实值之间损失,根据损失函数对模型参数进行优化,根据模型表现动态调整学习率,为防止过拟合风险,更新过程中进行模型正则化,实现快速、有效模型更新机制,并能有效适应数据分布变化[21,22]。但在实际临床应用中,需根据具体问题调整策略和参数,以达到模型最佳效果。

3.2 共同预测因子是MDRO预防与控制的关键因素

特征重要性结果显示,年龄是影响患者发生MDRO感染的重要危险因素,与以往的研究结果[23]一致,老年患者因身体器官发生退行性变化、免疫功能减弱等,使感染反复发作引起病原菌难以消灭,MDRO感染风险相对较大[24]。ICU患者身体机能衰退引起免疫力下降,手术会导致患者极易发生MDRO感染[25]。白蛋白含量的下降会减弱患者抵抗病原菌能力[26],导致身体免疫屏障受损,诱发组织器官感染,甚至引起耐药菌感染。原发疾病对感染也有重要的影响,心脑血管病患者可能会因误吸,将口咽部及呼吸道分泌物带至肺部,导致病菌进入机体发生感染[27]。慢性阻塞性肺病患者在治疗中会使用过多抗菌药物及激素,会增加感染风险[28];肺部感染患者伴有侵入性操作,导致呼吸道对病原菌防御功能下降[29]。侵入性操作会破坏机

体生理屏障和防御机制，导致病菌入侵机体引起感染，其中气管切开及插管操作会削弱患者呼吸道的保护机制，导致细菌易于入侵患者机体而引起耐药菌感染[30]。危重症患者使用有创机械通气≥7 d[31]，会增加患者发生MDRO感染的风险。抗菌药物的选择性压力是诱导细菌产生耐药的重要原因之一，细菌中由于多种耐药质粒可协同作用，对不同抗菌药物产生耐药从而导致患者发生MDRO感染[32]。抗菌药物时间使用过长是感染的独立危险因素[33]，长时间使用抗菌药物会将诱导药物主动外排、细胞膜通透性降低及药物作用靶位点改变[34]进而诱发危重症患者发生感染。

综上所述，肺部感染、脑血管病、低蛋白血症、冠心病、慢性阻塞性肺病、高血压、年龄、手术史、白蛋白含量、气管切开/气管插管、呼吸机时间、引流管、抗生素治疗等因素对于MDRO感染的预防有着极其重要的作用。

3.3　基于随机森林模型的预测性能最高

RF算法在国内外医疗领域中被广泛应用，常用于解决各类临床预测、诊断等问题。吕永楠等[35]利用RF建立急性心肌梗死诊断模型，结果指出其能较好区分急性心肌梗死与心绞痛，能为临床工作者提供一种新思路。盛晓欣等[36]使用RF方法研究癫痫患者正常脑电和癫痫脑电信号，建立识别与预测模型，显示该预测模型总预测误差仅为2%左右，为预测癫痫发作提供较好参考依据。Song等[37]基于4种算法构建患者发生压疮风险预测模型中指出RF模型准确度最高。本研究经分析发现，在预测MDRO感染发生情况中，随机森林模型在训练集和测试集上均呈现最佳表现，其次是XGBoost模型，然后是决策树模型，最后是逻辑回归模型。训练集和测试集中随机森林模型准确率、灵敏度、特异度、阳性预测值、阴性预测值、AUC值、Kappa值均优于其他三者。决策曲线分析(decision curve analysis，DCA)是一种用于评估预测模型在临床决策中应用价值的工具。DCA图指出随机森林模型的净效益曲线在大多数阈值概率下高于其他模型，这表明随机森林模型在这些阈值下能够提供更好的预测性能，这意味着在实际临床应用中，随机森林模型可能更能够有效地识别出需要治疗的真阳性患者，同时减少不必要的假阳性患者治疗情况。XGBoost模型虽然在净效益上略低于随机森林模型，但仍然表现出了良好的预测能力。CART模型和逻辑回归模型的净效益曲线也显示出一定的预测价值，但相对于前两个模型来说，它们的表现略显逊色。本研究发现RF模型在患者MDRO感染风险预测中具有较高准确率，但因“黑箱”特性使模型可解释性较差，本研究主要通过特征重要性评估方法来解决该问题[38]，帮助了解特征对预测结果影响程度，从而在一定程度上解释模型决策过程，提高可解释性，为ICU患者MDRO感染防控工作提供有力支持。

本研究通过多模态数据融合与动态ML技术，构建了针对ICU患者MDRO感染的早期预警模型。该模型利用多种数据源(如医疗记录、生理监测、影像资料等)的综合信息，提升了预测的准确性。动态ML算法的应用使得模型能够根据患者的实时病情变化进行快速更新和优化，提高了预警的时效性和准确性。不足之处：(1)本研究是一项单中心回顾性研究，数据选择可能具有偏倚情况；(2)模型仅进行内部验证，数据仅来源于一家医院，可能存在地域性局限性，同时无法检验模型的泛化能力；(3)MDRO感染变量除文中提及外，可能与医护人员监测方式和频率、医疗环境消毒、相关人员手卫生依从性等外部因素有关，因该项数据收集较为困难，本研究未将该因素纳入，后期将进一步纳入这些因素并进行分析，在今后研究中将收集感染年份靠前患者并扩大样本量去开展多中心研究；在临床应用中去检验模型实用性和可靠性，对模型进行相应完善，使本研究结果能够为临床MDRO感染预防工作提供更可靠理论与实践依据。

作者贡献度说明：

左蝶：文章设计、采集数据、数据分析和起草文章；赵佳：采集数据和收集相关文献；王小燕、李萍：提供文章修改意见和方法指导；龙晓艳、马学先、刘冰：写作指导和对文章内容作批评性审阅。

所有作者声明不存在利益冲突关系。

参考文献

1　EClinicalMedicine. Antimicrobial resistance: A top ten global public health threat[J]. EClinicalMedicine, 2021, 41: 101221.

2　Feng S, Wu Z, Liang W, et al. Prediction of antibiotic resistance evolution by growth measurement of all proximal mutants of beta-lactamase [J]. Mol Biol Evol, 2022, 39(5): msac086.

3　杨启文，吴安华，胡必杰，等. 临床重要耐药菌感染传播防控策略专家共识[J]. 中国感染控制杂志，2021，20(1)：1-14.
Yang Q, Wu A, Hu B, et al. Expert consensus on strategies for the prevention and control of spread of clinically important antimicrobial-resistant organisms[J]. Chin J Infect Control, 2021, 20(1): 1-14.

4　董宁，吴元浩，孙文秀，等. 临床试验框架下噬菌体治

疗多重耐药菌感染患者护理专家共识[J].军事护理，2024，41(7)：1-5.
Dong N，Wu Y，Sun W，et al. Expert consensus on nursing practices of phage therapy for multi-drug resistant organisms infection within the framework of clinical trials[J]. Mil Nurs，2024，41(7)：1-5.

5 Larrosa MN，Almirante B. Isolation strategy for controlling the spread of multidrug-resistant organisms：Is this still an essential option in hospitals?[J]. Enferm Infecc Microbiol Clin（Engl Ed），2021，39(9)：425-428.

6 Tabah A，Buetti N，Staiquly Q，et al. Epidemiology and outcomes of hospital-acquired bloodstream infections in intensive care unit patients：The EUROBACT-2 international cohort study[J]. Intensive Care Med，2023，49(2)：178-190.

7 Cuénod A，Wüthrich D，Seth-Smith HM，et al. Whole-genome sequence-informed MALDI-TOF MS diagnostics reveal importance of *Klebsiella oxytoca* group in invasive infections：A retrospective clinical study[J]. Genome Med，2021，13：150.

8 刘阳，吴明华，令娟，等.重症监护病房多重耐药菌感染患者疾病谱分析[J].中国感染控制杂志，2022，21(3)：273-279.
Liu Y，Wu M，Ling J，et al. Disease spectrum of patients with multidrug-resistant organism infection in intensive care unit[J]. Chin J Infect Control，2022，21(3)：273-279.

9 邹倩，耿苗苗，祝延红.基于循证理论构建重症监护病房患者多重耐药菌感染风险预测模型及外部验证研究[J].中国全科医学，2022，25(12)：1441-1448.
Zou Q，Geng M，Zhu Y. Development and external validation of an evidence-based risk prediction model for multidrug-resistant bacterial infections in ICU patients[J]. Chin Gen Pract，2022，25(12)：1441-1448.

10 Haug CJ，Drazen JM. Artificial intelligence and machine learning in clinical medicine，2023[J]. N Engl J Med，2023，388(13)：1201-1208.

11 王燕，孙博群，谷钰峰，等.医院感染常见病原体快速检测方法的研究进展[J].中华医院感染学杂志，2022，32(6)：955-960.
Wang Y，Sun B，Gu Y，et al. Progress of research on rapid detection of common species of pathogens causing nosocomial infection[J]. Chin J Nosocomiol，2022，32(6)：955-960.

12 曾玫，夏君，宗志勇，等.碳青霉烯类耐药革兰阴性菌感染的诊断、治疗及防控指南[J].中国感染与化疗杂志，2024，24(2)：135-151.
Zeng M，Xia J，Zong Z，et al. Guidelines for the diagnosis，treatment，prevention and control of infections caused by carbapenem-resistant gram-negative bacilli[J]. Chin J Infect Chemother，2024，24(2)：135-151.

13 Riley RD，Ensor J，Snell KI，et al. Calculating the sample size required for developing a clinical prediction model[J]. BMJ，2020，368：m441.

14 俞岚，吴卫江，郑洁，等.神经外科重症患者多药耐药菌感染危险因素及其预测模型构建[J].中华医院感染学杂志，2023，33(18)：2790-2794.
Yu L，Wu W，Zheng J，et al. Risk factors for multidrug-resistant organisms infections in critically ill patients of neurosurgery department and establishment of prediction model[J]. Chin J Nosocomiol，2023，33(18)：2790-2794.

15 Wang S，Li J，Dai J，et al. Establishment and validation of models for the risk of multi-drug-resistant bacteria infection and prognosis in elderly patients with pulmonary infection：A multicenter retrospective study[J]. Infect Drug Resist，2023，16：6549-6566.

16 张张，白轩，廖志斌，等.基于逻辑回归算法的列车单端双网无线接入方案研究[J].现代城市轨道交通，2023，5：42-47.
Zhang Z，Bai X，Liao Z，et al. TAU Study on wireless access scheme of "dual network on one terminal" of the train based on logistic regression[J]. Modern Urban Transit，2023，5：42-47.

17 申明尧，韩萌，杜诗语，等.数据流决策树集成分类算法综述[J].计算机应用与软件，2022，39(9)：1-10.
Shen M，Han M，Du S，et al. A survey of decision tree ensemble classification algorithms of data streams[J]. Comput Appl Softw，2022，39(9)：1-10.

18 王艳琴，谢卓峰，韩国鹏，等.基于极端梯度提升的PEMFC长短期老化趋势预测[J].太阳能学报，2024，45(7)：232-239.
Wang Y，Xie Z，Han G，et al. Short-and long-term degradation prediction for proton exchange membrane fuel cell based on extreme gradient boosting[J]. Acta Energiae Solaris Sin，2024，45(7)：232-239.

19 刘闻博，王丛虎.公共管理研究中的机器学习方法——原理、应用及挑战[J].公共管理与政策评论，2024，13(5)：152-168.
Liu W，Wang C. Machine learning in public administration research：Tutorial，application and challenges[J]. Public Manag Policy Rev，2024，13(5)：152-168.

20 Mazhar O，Ramdani S，Cherubini A. A deep learning framework for recognizing both static and dynamic gestures[J]. Sensors，2021，21(6)：2227.

21 Celik B，Singh P，Vanschoren J. Online AutoML：An adaptive AutoML framework for online learning[J]. Mach Learn，2023，112：1897-1921.

22 Yuan J，Wan J，Zhang X，et al. A second-order dynamic and static ship path planning model based on reinforcement learning and heuristic search algorithms[J]. J Wireless Com Network，2022，2022：128.

23 叶永强，何兰兰，刘桂玲，等. 重症颅脑损伤患者合并多重耐药菌肺部感染病原菌分布、影像学特征以及风险预测模型的建立与验证[J]. 中国医学科学院学报，2022，44(4)：636-642.
Ye Y，He L，Liu G，et al. Pathogen distribution，imaging characteristics，and establishment and verification of risk prediction model of pulmonary infection with multi-drug resistant organism in patients with severe craniocerebral injury[J]. Acta Acad Med Sin，2022，44(4)：636-642.

24 安霞，高英华，孟玲，等. 老年患者多药耐药铜绿假单胞菌肺部感染危险因素和辅助性T细胞17及调节性T细胞表达水平[J]. 中华医院感染学杂志，2023，33(9)：1294-1297.
An X，Gao Y，Meng L，et al. Risk factors for multidrug-resistant *Pseudomonas aeruginosa* pulmonary infection and expressions of helper T cell 17，regulatory T cell in elderly patients[J]. Chin J Nosocomiol，2023，33(9)：1294-1297.

25 丁梦媛，李文进，耿苗苗，等. 综合重症监护病房患者多重耐药菌医院感染风险评估模型构建[J]. 中国卫生资源，2020，23(4)：384-387，399.
Ding M，Li W，Geng M，et al. Risk assessment model construction of nosocomial infection of multi-drug resistant organism in comprehensive intensive care unit patients[J]. Chin Health Resour，2020，23(4)：384-387，399.

26 Matin S，Fouladi N，Pahlevan Y，et al. The sufficient vitamin D and albumin level have a protective effect on COVID-19 infection[J]. Arch Microbiol，2021，203(8)：5153-5162.

27 谢美英，唐知己，刘晓兰，等. 重症颅脑损伤多重耐药菌肺部感染危险因素及预测模型建立[J]. 中华医院感染学杂志，2024，34(6)：861-865.
Xie M，Tang Z，Liu X，et al. Risk factors for multidrug-resistant organisms pulmonary infections in patients with severe craniocerebral injury and establishment of prediction model[J]. Chin J Nosocomiol，2024，34(6)：861-865.

28 于佳，黄艾弥，李莉珊，等. 胸外科老年患者术后肺部多重耐药菌感染临床特征及危险因素[J]. 中国感染控制杂志，2023，22(8)：932-938.
Yu J，Huang A，Li L，et al. Clinical features and risk factors of postoperative pulmonary multidrug-resistant organism infections in elderly patients undergoing cardiothoracic surgery[J]. Chin J Infect Control，2023，22(8)：932-938.

29 谢朝云，蒙桂鸾，熊芸，等. 新生儿重症监护病房中重症肺炎多种菌感染相关因素分析[J]. 临床儿科杂志，2020，38(4)：260-263.
Xie C，Meng G，Xiong Y，et al. Risk factors of multiple bacterial infections in NICU neonates with severe pneumonia[J]. J Clin Pediatr，2020，38(4)：260-263.

30 李舒鑫，史素玲，陈媛，等. 非急性期脑卒中气管切开患者肺部感染现状及影响因素[J]. 护理学杂志，2023，38(8)：30-32，43.
Li S，Shi S，Chen Y，et al. Pulmonary infection in non-acute stroke patients undergoing tracheotomy：Incidence and influencing factors[J]. J Nurs Sci，2023，38(8)：30-32，43.

31 Hu J，Hu S，Li Z，et al. Risk factors of multidrug-resist antbacteria infection in patients with ventilator-associated pneumonia：A systematic review and meta-analysis[J]. J Infect Chemother，2023，29(10)：942-947.

32 陈晓旭，李松琴，刘娟，等. 三级医院住院患者抗菌药物治疗前病原学送检率管理现状调查[J]. 中国感染控制杂志，2023，22(5)：532-538.
Chen X，Li S，Liu J，et al. Current management status of pathogen examination rate before antimicrobial treatment for hospitalized patients in tertiary hospitals[J]. Chin J Infect Control，2023，22(5)：532-538.

33 陈惠刚，李桂艳，马微. 老年下呼吸道感染铜绿假单胞菌的病原学特征分析[J]. 中国病原生物学杂志，2024，19(1)：74-78.
Chen H，Li G，Ma W. Etiological characteristics of lower respiratory tract infection in the elderly[J]. J Pathogen Biol，2024，19(1)：74-78.

34 Wu Z，Zeng M，Li C，et al. Time-dependence of NIHSS in predicting functional outcome of patients with acute ischemic stroke treated with intravenous thrombolysis[J]. Postgrad Med J，2019，95(1122)：181-186.

35 吕永楠，李迪，李艳. 基于随机森林的男性急性心肌梗死诊断模型建立及验证[J]. 中华诊断学电子杂志，2019，7(4)：233-238.
Lv Y，Li D，Li Y. Establishment and validation of diagnosis model for acute myocardial infarction based on random forest classification in men[J]. Chin J Diagn，2019，7(4)：233-238.

36 盛晓欣，田翔华，周毅. 基于随机森林癫痫患者脑电数据的分析研究[J]. 中国数字医学，2020，15(1)：41-43.
Sheng X，Tian X，Zhou Y. Analysis of EEG pattern recognition of epileptic patients based on random forest[J]. Chin Digit Med，2020，15(1)：41-43.

37 Song J，Gao Y，Yin P，et al. The random forest model has the best accuracy among the four pressure ulcer prediction models using machine learning algorithms[J]. Risk Manag Healthc Policy，2021，14：1175-1187.

38 Zhang Y，Zhang X，Razbek J，et al. Opening the black box：Interpretable machine learning for predictor finding of metabolic syndrome[J]. BMC Endocr Disord，2022，22：214.

［本文编辑］雷 燕 宋睿璞 陈思润