

基于随机森林法的重症肺炎进展为急性呼吸窘迫综合征风险预测模型的构建与评价

贾贵阳 王太珊 宋尔芹 肖玄 刘国跃

遵义医科大学第二附属医院重症医学科, 遵义 563000

通信作者: 刘国跃, Email: liuguoyue@zmu.edu.cn

【摘要】 **目的** 寻找重症肺炎进展为急性呼吸窘迫综合征(ARDS)的危险因素,并基于随机森林算法构建风险预测模型,为重症肺炎病情评估、早期干预、改善预后提供依据。**方法** 采用回顾性观察性研究方法,纳入2020年1月至2024年5月入住遵义医科大学第二附属医院重症监护病房(ICU)的重症肺炎患者,收集患者一般资料、入ICU 24 h内生命体征、血液检查结果、病情评估指标,以及结局指标等。根据患者是否进展为ARDS分为ARDS组和非ARDS组,采用单因素逻辑回归分析筛选重症肺炎进展为ARDS的危险因素,基于随机森林算法构建相关风险预测模型,并使用1 000次重抽样验证模型的效能及稳定性。**结果** 共纳入181例重症肺炎患者,其中73例进展为ARDS,发生率为40.3%。与非ARDS组相比,ARDS组最低收缩压、最低舒张压、最低氧合指数、pH值、白蛋白降低,最高活化部分凝血活酶时间、急性生理学及慢性健康状况评分II(APACHE II)、肺损伤预测评分(LIPS)升高(均 $P<0.05$),其他基线资料比较差异均无统计学意义(均 $P>0.05$)。单因素逻辑回归分析显示,pH值、最低收缩压、白蛋白、APACHE II评分、LIPS评分为重症肺炎患者进展为ARDS的影响因素[pH值:优势比(OR)=0.04,95%置信区间(95%CI)为0.00~0.96, $P=0.047$,截断值为7.34;最低收缩压:OR=0.98,95%CI为0.97~1.00, $P=0.044$,截断值为90 mmHg(1 mmHg=0.133 kPa);白蛋白:OR=0.94,95%CI为0.89~0.99, $P=0.032$,截断值为28.05 g/L;APACHE II评分:OR=1.08,95%CI为1.02~1.14, $P=0.008$,截断值为23分;LIPS评分:OR=1.37,95%CI为1.09~1.72, $P=0.007$,截断值为5分]。以上述危险因素构建随机森林模型,各指标的重要性从高到低排序依次为pH值、最低收缩压、白蛋白、APACHE II评分、LIPS评分(基尼指数分别为31.08、30.74、29.35、28.01、24.92)。经1 000次重抽样验证模型显示,模型的受试者工作特征曲线下面积(AUC)为0.909(95%CI为0.870~0.943),敏感度为0.823(95%CI为0.699~0.932),特异度为0.869(95%CI为0.741~0.963)。**结论** pH值 <7.34 、最低收缩压 <90 mmHg、白蛋白 <28.05 g/L、APACHE II >23 分、LIPS >5 分是重症肺炎患者进展为ARDS的危险因素,以此基于随机森林算法构建的模型可有效预测重症肺炎患者是否进展为ARDS。

【关键词】 重症肺炎;急性呼吸窘迫综合征;机器学习模型;随机森林法;预测模型

基金项目:贵州省科技计划项目(黔科合成果-LC[2024]077);贵州省高层次创新型人才项目(gzjwjr2024-023)

DOI: 10.3760/cma.j.cn121430-20250102-00004

Construction and evaluation of a risk prediction model for the progression of severe pneumonia to acute respiratory distress syndrome based on the random forest algorithm

Jia Guiyang, Wang Taishan, Song Erqin, Xiao Xuan, Liu Guoyue

Department of Intensive care Unit, the Second Affiliated Hospital of Zunyi Medical University, Zunyi 563000, China

Corresponding author: Liu Guoyue, Email: Liuguoyue@zmu.edu.cn

【Abstract】 **Objective** To identify the risk factors for the progression of severe pneumonia to acute respiratory distress syndrome (ARDS) and to construct a prediction model based on the random forest algorithm, providing a basis for disease assessment, early intervention, and prognosis improvement in severe pneumonia. **Methods** A retrospective observational study was conducted. Patients with severe pneumonia admitted to the intensive care unit (ICU) of the Second Affiliated Hospital of Zunyi Medical University from January 2020 to May 2024 were enrolled. Data including general patient information, vital signs, blood test results, disease assessment indicators within 24 hours of ICU admission, and outcome measures were collected. Patients were divided into ARDS group and non-ARDS group according to whether they progressed to ARDS. Univariate Logistic regression analysis was used to screen the risk factors for the progression of severe pneumonia to ARDS, and a random forest based prediction model was constructed. Model performance and stability were validated using 1 000 resampling iterations. **Results** A total of 181 severe pneumonia patients were included, of whom 73 progressed to ARDS, with an incidence rate of 40.3%. Compared to the non-ARDS group, the ARDS group had significantly lower lowest systolic blood pressure, lowest diastolic blood pressure, lowest oxygenation index, pH value, and albumin level, while showing significantly higher maximum activated partial thromboplastin time (APTT), Acute Physiology and Chronic Health Evaluation II (APACHE II), and Lung Injury Prediction Score (LIPS; all $P<0.05$). There were no statistically significant differences in other baseline data comparisons (all $P>0.05$). Univariate Logistic regression analysis showed that pH, lowest systolic blood pressure, albumin, APACHE II score, and LIPS score

were risk factors for the progression to ARDS in severe pneumonia patients [pH: odds ratio (OR)=0.04, 95% confidence interval (95%CI) was 0.00–0.96, $P=0.047$, cut-off was 7.34; lowest systolic blood pressure: OR=0.98, 95%CI was 0.97–1.00, $P=0.044$, cut-off was 90 mmHg (1 mmHg=0.133 kPa); albumin: OR=0.94, 95%CI was 0.89–0.99, $P=0.032$, cut-off was 28.05 g/L; APACHE II score: OR=1.08, 95%CI was 1.02–1.14, $P=0.008$, cut-off was 23; LIPS: OR=1.37, 95%CI was 1.09–1.72, $P=0.007$, cut-off was 5]. A random forest model constructed with these risk factors ranked the importance of the indicators from high to low as follows: pH, lowest systolic blood pressure, albumin, APACHE II score, and LIPS (with Gini Index of 31.08, 30.74, 29.35, 28.01, and 24.92, respectively). Validation with 1 000 bootstrap resamplings showed that the model had an area under the receiver operator characteristic curve (AUC) of 0.909 (95%CI was 0.870–0.943), a sensitivity of 0.823 (95%CI was 0.699–0.932), and a specificity of 0.869 (95%CI was 0.741–0.963).

Conclusions pH<7.34, lowest systolic blood pressure<90 mmHg, albumin<28.05 g/L, APACHE II >23, and LIPS>5 are risk factors for the progression of severe pneumonia to ARDS. The model constructed based on these factors using the random forest algorithm can effectively predict whether severe pneumonia patients will progress to ARDS.

【Key words】 Severe pneumonia; Acute respiratory distress syndrome; Machine learning model; Random forest algorithm; Prediction model

Fund program: Guizhou Science and Technology Plan Project (Qian Ke He Achievement–LC[2024]077); Guizhou High-Level Innovative Talent Program (gzwjrs2024–023)

DOI: 10.3760/cma.j.cn121430–20250102–00004

重症肺炎是重症监护病房(intensive care unit, ICU)中导致重要器官功能障碍甚至威胁生命的常见疾病,也是导致急性呼吸窘迫综合征(acute respiratory distress syndrome, ARDS)最重要的肺内病因^[1-2];同时,重症肺炎进展为 ARDS 也是导致患者死亡的主要原因,病死率可高达 13%~50%^[3],严重影响患者预后并造成社会负担。由于不同原因所致 ARDS 异质性较大,涉及的病理生理机制复杂,且目前尚缺乏针对性治疗方法,故仍以病因治疗及小潮气量、适当的呼气末正压和肺复张等呼吸支持治疗为主的综合措施来改善患者顽固性低氧血症,但其总体病死率未见明显降低^[4]。因此,早期识别重症肺炎进展为 ARDS 的高危患者及其影响因素,并采取有效干预措施对改善重症肺炎患者预后具有关键作用。

疾病预测模型是利用数学统计方法对既往数据进行分析,根据疾病危险因素,早期预测疾病发展的有效方法^[5],其已在众多疾病的早期诊断、预后预测及干预措施选择等多方面得到有效使用^[6-7]。目前常用的预测方法包括预测方程^[8]、列线图模型^[9-10]、机器学习模型^[11-12]等已广泛应用于医学领域。其中随机森林模型通过随机构建多个决策树,反复训练以提高模型预测精准度,同时还可计算基尼指数(Gini Index, GINI)以评估各变量的重要性,识别关键变量并预测事件的发生^[13-14]。

本研究中通过回顾性收集重症肺炎患者的相关数据,旨在探讨重症肺炎进展为 ARDS 的相关因素及其重要性,并基于机器学习中的随机森林算法构建重症肺炎进展为 ARDS 的风险预测模型,以期临床医生能更好地预测和评估重症肺炎进展为

ARDS 的风险,针对相关因素采取相应的干预措施,从而阻止重症肺炎患者进展为 ARDS 提供科学的理论依据。

1 资料与方法

1.1 研究人群:采用回顾性观察性研究方法,纳入 2020 年 1 月至 2024 年 5 月本院 ICU 收治的重症肺炎患者。

1.1.1 纳入标准:① 年龄≥18 岁;② 符合 2023 年社区获得性肺炎诊断标准^[15];③ 符合 2018 年医院获得性肺炎(hospital-acquired pneumonia, HAP)诊断标准^[16];④ 住院时间>24 h,且临床资料完整。

1.1.2 排除标准:① 入住 ICU 时已达 ARDS 诊断标准;② 妊娠期患者;③ 恶性肿瘤全身转移者。

1.2 伦理学:本研究符合医学伦理学标准,研究方案已经过医院伦理委员会审查批准(审批号: KYLL-2024-025)。

1.3 资料收集:收集患者一般资料,包括年龄、性别、个人史(吸烟、饮酒)和既往病史;入 ICU 24 h 内生命体征,包括最高体温、最高心率、最快呼吸频率、最低脉搏血氧饱和度、最高和最低舒张压及收缩压;血液检查结果,包括动脉血气指标(pH 值、最低氧合指数、最低和最高血二氧化碳分压、最高血乳酸)、血常规指标(最高中性粒细胞计数、最低血红蛋白、最低淋巴细胞计数)、血生化指标(最低白蛋白、最高总胆红素、最高血尿素氮、最高血清肌酐、最高血糖)、凝血指标(最高活化部分凝血活酶时间、最高 D-二聚体、最高超敏 C-反应蛋白);病情评估指标,包括急性生理学与慢性健康状况评分 II (Acute Physiology and Chronic Health Evaluation II, APACHE II)、肺损伤预测评分(Lung

Injury Prediction Score, LIPS); 结局指标(是否进展为 ARDS)。

1.4 分组: 根据观察结局(重症肺炎是否进展为 ARDS)将患者分为 ARDS 组和非 ARDS 组。ARDS 诊断以 2023 年 ARDS 全球新定义为标准^[17-18]。

1.5 统计学分析: 采用 SPSS 29.0 软件对数据进行统计分析。计量资料先进行正态性检验,符合正态分布的计量资料用均数 \pm 标准差($\bar{x}\pm s$)描述,方差不齐的变量选择独立样本 t 检验,方差不齐的变量采用 Tamhane's T_2 法;非正态分布的计量资料采用中位数(四分位数)[$M(Q_1, Q_3)$]描述,组间比较采用 Mann-Whitney U 检验。分类变量用百分比或率描述,组间比较采用 χ^2 检验。检验水准 α 值取双侧 0.05。

使用 R-Studio 软件,采用单变量逻辑回归分析筛选重症肺炎进展为 ARDS 的相关变量,计算各变量的优势比(odds ratio, OR)及其 95% 置信区间(95% confidence interval, 95%CI),再通过计算最大约登指数寻找各指标的截断值、敏感度和特异度。随后纳入相关变量,采用网格化筛选袋外错误率最低的决策树和变量数构建重症肺炎进展为 ARDS 风险的随机森林预测模型,通过计算各变量的 GINI 评估其在模型中的重要性,GINI 值越高提示该指标越重要。最后,通过 1 000 次重抽样计算模型预测的敏感度、特异度及受试者工

作特征曲线下面积(area under the receiver operator characteristic curve, AUC),并采用 ggplot2 包绘制分布图,对所有结果进行可视化,评估模型的预测效能及稳定性。

2 结果

2.1 患者基线资料: 通过医院病历系统检索到重症肺炎住院患者 861 例,根据纳入、排除标准,最终纳入 181 例重症肺炎患者,其中 73 例进展为 ARDS,发生率为 40.3%。表 1 结果显示,ARDS 组与非 ARDS 组患者年龄、性别、个人史和既往病史等一般资料比较差异均无统计学意义(均 $P>0.05$)。在入 ICU 24 h 内各指标中,与非 ARDS 组相比,ARDS 组患者最低收缩压、最低舒张压、最低氧合指数、pH 值、最低白蛋白降低,最高活化部分凝血活酶时间、APACHE II 评分及 LIPS 评分升高,差异均有统计学意义(均 $P<0.05$);其余指标差异均无统计学意义(均 $P>0.05$)。

2.2 重症肺炎进展为 ARDS 的相关因素: 单因素逻辑回归分析结果显示(表 2),与重症肺炎进展为 ARDS 相关的变量包括 pH 值、最低收缩压、最低白蛋白、APACHE II 评分和 LIPS 评分(均 $P<0.05$),根据截断值提示, pH 值 <7.34 、最低收缩压 <90 mmHg(1 mmHg=0.133 kPa)、最低白蛋白 <28.05 g/L、APACHE II >23 分、LIPS >5 分是重症肺炎患者进展为 ARDS 的危险因素。

表 1 重症肺炎是否进展为 ARDS 两组患者基线资料比较

指标	非 ARDS 组 (108 例)	ARDS 组 (73 例)	统计量值	P 值	指标	非 ARDS 组 (108 例)	ARDS 组 (73 例)	统计量值	P 值
年龄(岁, $\bar{x}\pm s$)	64.29 \pm 14.99	65.81 \pm 14.63	$t=-0.676$	0.500	最低脉搏血氧饱和度 [$M(Q_1, Q_3)$]	0.90 (0.84, 0.95)	0.88 (0.82, 0.92)	$Z=-0.992$	0.321
性别[例(%)]			$\chi^2=0.005$	0.946	最高心率(次/min, $\bar{x}\pm s$)	116.38 \pm 22.77	120.92 \pm 21.39	$t=-1.348$	0.179
男性	69(63.9)	47(64.4)			最快呼吸频率[次/min, $M(Q_1, Q_3)$]	27(22, 33)	28(23, 36)	$Z=-1.510$	0.131
女性	39(36.1)	26(35.6)			最高体温[$^{\circ}C$, $M(Q_1, Q_3)$]	37.4 (37.0, 38.2)	37.50 (37.0, 38.5)	$Z=-1.139$	0.255
个人史[例(%)]			$\chi^2=2.464$	0.116	最高收缩压[mmHg, $M(Q_1, Q_3)$]	154 (136, 172)	150 (143, 166)	$Z=-0.236$	0.814
吸烟史	66(61.1)	36(49.3)	$\chi^2<0.001$	0.983	最低收缩压[mmHg, $M(Q_1, Q_3)$]	98(85, 110)	88(79, 104)	$Z=-2.347$	0.019
饮酒史	25(23.1)	17(23.3)			最高舒张压 (mmHg, $\bar{x}\pm s$)	84.65 \pm 11.63	83.22 \pm 11.24	$t=0.822$	0.412
既往病史[例(%)]			$\chi^2=1.309$	0.253	最低舒张压[mmHg, $M(Q_1, Q_3)$]	53(45, 60)	50(45, 57)	$Z=-1.981$	0.048
COPD	13(12.0)	5(6.8)	$\chi^2=0.183$	0.669	最高超敏 C-反应蛋白 [mg/L, $M(Q_1, Q_3)$]	77.38 (26.42, 145.27)	86.74 (43.67, 174.99)	$Z=-1.442$	0.149
高血压	52(48.1)	34(46.6)			最低血红蛋白 (g/L, $\bar{x}\pm s$)	106.16 \pm 28.42	104.12 \pm 31.64	$t=0.451$	0.652
糖尿病	22(20.4)	13(17.8)							
最低氧合指数[mmHg, $M(Q_1, Q_3)$]	145.35 (103.53, 198.75)	131.30 (77.15, 168.70)	$Z=-2.065$	0.039					
pH 值[$M(Q_1, Q_3)$]	7.35(7.32, 7.41)	7.33(7.27, 7.39)	$Z=-2.312$	0.021					
最低动脉血二氧化碳分压 [mmHg, $M(Q_1, Q_3)$]	30.65 (25.20, 35.78)	29.1 (24.65, 33.40)	$Z=-1.294$	0.196					
最高动脉血二氧化碳分压 [mmHg, $M(Q_1, Q_3)$]	40.15 (34.33, 51.45)	39.70 (34.30, 53.60)	$Z=-0.003$	0.998					
最高血乳酸[mmol/L, $M(Q_1, Q_3)$]	2.30 (1.70, 3.38)	2.10 (1.55, 3.20)	$Z=-0.715$	0.475					

续表 1

指标	非 ARDS 组 (108 例)	ARDS 组 (73 例)	统计量值	P 值	指标	非 ARDS 组 (108 例)	ARDS 组 (73 例)	统计量值	P 值
最高中性粒细胞计数 〔×10 ⁹ /L, M(Q ₁ , Q ₃)〕	10.03 (7.05, 15.45)	8.98 (6.24, 12.75)	Z=-1.619	0.105	最高血糖〔mmol/L, M(Q ₁ , Q ₃)〕	10.5 (8.3, 14.1)	9.8 (7.7, 13.4)	Z=-0.714	0.475
最低淋巴细胞计数 〔×10 ⁹ /L, M(Q ₁ , Q ₃)〕	0.50 (0.28, 0.91)	0.47 (0.27, 0.83)	Z=-0.513	0.608	最高活化部分凝血活酶 时间〔s, M(Q ₁ , Q ₃)〕	30.90 (27.03, 36.65)	33.60 (28.85, 42.20)	Z=-2.052	0.040
最低白蛋白(g/L, $\bar{x} \pm s$)	30.20 ± 5.49	28.32 ± 5.90	t=2.192	0.030	最高 D-二聚体〔mg/L, M(Q ₁ , Q ₃)〕	1.79 (1.13, 3.00)	2.21 (1.26, 4.15)	Z=-1.907	0.056
最高总胆红素〔μmol/L, M(Q ₁ , Q ₃)〕	13.05 (9.80, 18.58)	13.70 (10.65, 23.85)	Z=-0.898	0.369	APACHE II 评分 〔分, M(Q ₁ , Q ₃)〕	20(16, 22)	21(17, 26)	Z=-2.391	0.017
最高血尿素氮〔mmol/L, M(Q ₁ , Q ₃)〕	9.27 (6.63, 12.54)	10.31 (6.42, 18.85)	Z=-1.183	0.237	LIPS 评分〔分, M(Q ₁ , Q ₃)〕	6(4, 6)	6(5, 7)	Z=-2.799	0.005
最高血清肌酐〔μmol/L, M(Q ₁ , Q ₃)〕	89.0 (72.0, 134.0)	107.0 (78.5, 219.0)	Z=-1.933	0.053					

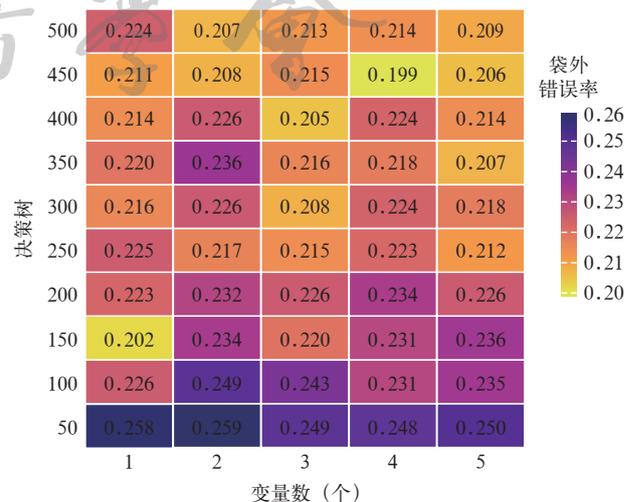
注：ARDS 为急性呼吸窘迫综合征，COPD 为慢性阻塞性肺疾病，APACHE II 为急性生理学及慢性健康状况评分 II，LIPS 评分为肺损伤预测评分；1 mmHg=0.133 kPa

表 2 重症肺炎患者进展为 ARDS 相关因素的单因素逻辑回归分析及变量截断值

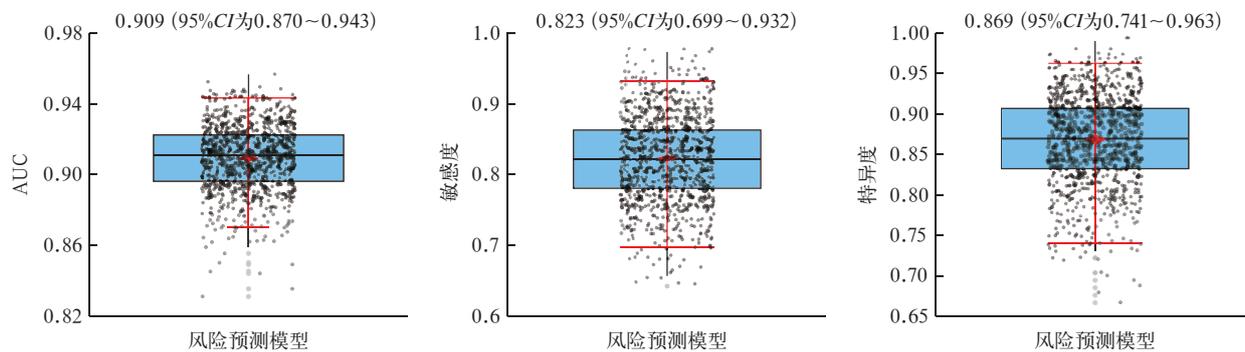
变量	OR 值	95%CI	P 值	约登 指数	截断 值	敏感 度	特异 度	变量	OR 值	95%CI	P 值	约登 指数	截断 值	敏感 度	特异 度
最低氧合指数	1.00	0.99 ~ 1.00	0.098					最高超敏 C-反应 蛋白	1.00	1.00 ~ 1.01	0.179				
pH 值	0.04	0.00 ~ 0.96	0.047	0.228	7.34	0.616	0.611	最低血红蛋白	1.00	0.99 ~ 1.01	0.650				
最低动脉血二氧化 化碳分压	0.98	0.94 ~ 1.01	0.193					最高中性粒 细胞计数	0.97	0.93 ~ 1.02	0.252				
最高动脉血二氧化 化碳分压	1.00	0.98 ~ 1.01	0.776					最低淋巴细胞计数	0.96	0.51 ~ 1.81	0.891				
最高血乳酸	1.07	0.94 ~ 1.22	0.289					最低白蛋白	0.94	0.89 ~ 0.99	0.032	0.173	28.05 g/L	0.534	0.639
最低脉搏血氧 饱和度	0.99	0.96 ~ 1.03	0.684					最高总胆红素	1.01	0.99 ~ 1.04	0.291				
最高心率	1.01	1.00 ~ 1.02	0.180					最高血尿素氮	1.03	1.00 ~ 1.07	0.086				
最快呼吸频率	1.03	0.99 ~ 1.08	0.103					最高血清肌酐	1.00	1.00 ~ 1.00	0.735				
最高体温	1.18	0.87 ~ 1.59	0.294					最高血糖	0.97	0.92 ~ 1.03	0.364				
最高收缩压	1.00	0.98 ~ 1.01	0.913					最高活化部分 凝血活酶时间	1.01	0.99 ~ 1.02	0.429				
最低收缩压	0.98	0.97 ~ 1.00	0.044	0.228	90 mmHg	0.562	0.667	最高 D-二聚体	1.03	0.98 ~ 1.09	0.275				
最高舒张压	0.99	0.96 ~ 1.02	0.410					APACHE II 评分	1.08	1.02 ~ 1.14	0.008	0.211	23 分	0.452	0.759
最低舒张压	0.97	0.94 ~ 1.00	0.063					LIPS 评分	1.37	1.09 ~ 1.72	0.007	0.221	5 分	0.740	0.481

注：ARDS 为急性呼吸窘迫综合征，APACHE II 为急性生理学及慢性健康状况评分 II，LIPS 评分为肺损伤预测评分，OR 为优势比，95%CI 为 95% 可信区间；1 mmHg=0.133 kPa；空白代表无此项

2.3 随机森林预测模型的构建及验证：纳入相关危险因素，网格化筛选随机森林模型参数，结果显示(图 1)，在决策树为 450、变量数为 4 时，袋外错误率最低为 0.199；随后以决策树 450、变量数 4 构建随机森林模型，计算各指标 GINI(图 2)，重要性从高到低的指标依次为 pH 值(GINI=31.08)、最低收缩压(GINI=30.74)、最低白蛋白(GINI=29.35)、APACHE II 评分(GINI=28.01)、LIPS 评分(GINI=24.92)。经 1 000 次重抽样验证模型的 AUC、敏感度及特异度(图 3)，结果显示，模型的 AUC 为 0.909(95%CI 为 0.870 ~ 0.943)，敏感度为 0.823(95%CI 为 0.699 ~ 0.932)，特异度为 0.869(95%CI 为 0.741 ~ 0.963)，提示所构建的随机森林模型在重症肺炎进展为 ARDS 风险预测方面具有较好的效能及稳定性。

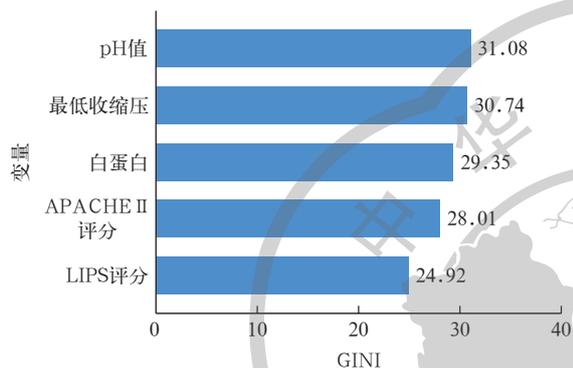


注：ARDS 为急性呼吸窘迫综合征
图 1 重症肺炎患者进展为 ARDS 风险预测随机森林模型参数筛选网格图



注：ARDS 为急性呼吸窘迫综合征，AUC 为受试者工作特征曲线下面积，95%CI 为 95% 置信区间

图 3 重抽样 1 000 次验证重症肺炎患者进展为 ARDS 风险预测模型的 AUC、敏感度和特异度



注：ARDS 为急性呼吸窘迫综合征，APACHE II 为急性生理学及慢性健康状况评分 II，LIPS 为肺损伤预测评分，GINI 为基尼指数

图 2 重症肺炎患者进展为 ARDS 风险预测模型中相关变量的重要性

3 讨论

重症肺炎是常见呼吸系统重症，其进展为 ARDS 可导致患者病情迅速加重，增加治疗难度^[19-20]；同时，ARDS 患者由于肺免疫防御和微生态环境改变，也会增加非典型病原体以及耐药菌感染的风险，造成肺部叠加感染，从而影响患者预后，甚至死亡^[21]。目前重症肺炎所致 ARDS 尚缺乏特异性治疗方法，仍以抗感染、抗炎、呼吸支持等治疗为主，但其整体病死率未见明显下降^[22]。本研究中通过逻辑回归分析筛选出与重症肺炎进展为 ARDS 相关的 5 个危险因素，即 pH 值 < 7.34、最低收缩压 < 90 mmHg、白蛋白 < 28.05 g/L、APACHE II > 23 分、LIPS > 5 分，对于存在上述危险因素的高风险患者，临床医生应提高警惕，早期识别及评估，并采取相应措施控制危险因素，以期改善重症肺炎患者的预后。

pH 值反映了患者血液中的酸碱度，与患者呼吸及氧合代谢密切相关，是常用于评估危重症患者全身代谢的重要指标^[23]。肺部感染患者由于肺组织炎症，可出现气体有效交换面积减少，进而导致呼吸性酸中毒；同时，由于组织氧供不足，无氧酵解增

多，也可导致酸性物质蓄积，表现为 pH 值降低及代谢性酸中毒。目前研究发现，在重症肺炎患者中，酸中毒会导致多种生理及免疫紊乱，感染发生发展，并与重症肺炎发病率及病死率增加相关^[24]。此外，血液酸碱度还会影响药物的吸收及体内分布等药代动力学过程，从而影响药物治疗效果，导致疾病进展^[25]。本研究显示，重症肺炎患者进展为 ARDS 与 pH 值显著相关，pH 值 < 7.34 是重症肺炎患者进展为 ARDS 的重要危险因素，结合 GINI 分析结果，pH 值是预测重症肺炎患者进展为 ARDS 最重要的预测指标。总之，pH 值一方面反映了重症肺炎患者呼吸功能状态，与肺部供氧及组织氧耗相关；另一方面 pH 值降低可导致药物治疗效果下降，增加疾病进展风险，可用于预测重症肺炎患者 ARDS 的发生。

血压是反映危重症患者循环状态的重要生命体征，循环功能障碍也是重症肺炎的主要诊断指标之一。目前研究发现，全身炎症反应导致的体内液体分布或出血是重症肺炎患者低血压及休克的主要原因之一，重症肺炎患者血压变化与全身炎症反应程度相关，与重症肺炎患者病死率也具有相关性^[26]。本研究中收集了患者最高、最低收缩压和舒张压，结果显示，与非 ARDS 组比较，ARDS 组患者最低收缩压显著降低；通过单因素逻辑回归分析及计算最大约登指数发现，最低收缩压 < 90 mmHg 是重症肺炎患者进展为 ARDS 的重要危险因素，且结合随机森林模型中 GINI 值分析，最低收缩压是模型中较为重要的指标，可较好地地区分重症肺炎患者是否会进展为 ARDS。此外，国内学者在关于老年急性肺损伤患者的研究中也发现，入院首日最低收缩压是老年急性肺损伤患者死亡的危险因素，联合其他指标构建的模型对老年急性肺损伤患者病死率具有一定的预测价值^[27]。

血清白蛋白水平是反映人体营养状态的重要指标之一,同时白蛋白还参与了免疫、内分泌等多系统之间的多向相互作用,在人体免疫功能的维持和对感染的免疫应答中具有比较重要的作用^[28]。目前研究发现白蛋白的氧化分解过程会影响生物活性脂质介质的相互作用,低白蛋白血症可通过影响脂质介质的相互作用影响抗菌和组织修复的过程,从而影响免疫反应;其与感染性疾病的发生及严重程度呈负相关,可用于预测相关疾病的发展^[29]。此外,也有研究发现,血清白蛋白水平还可对治疗药物药代动力学和药效学产生一定影响,从而影响治疗效果及患者病情变化^[30-31]。重症肺炎患者常常使用大量抗菌药物,而目前研究发现血清白蛋白可与部分药物结合,从而影响抗菌药物半衰期及分布容积;低白蛋白情况的存在,可导致一些抗菌药物在血液中的游离药物浓度升高,导致表观分布容积及清除效率增加,药物疗效下降,且副作用及不良反应增加,从而导致病情恶化^[32]。总之,在重症肺炎患者中,血清白蛋白水平降低一方面使患者营养状态及自身免疫防御力下降,感染进行性加重;另一方面血清白蛋白水平降低还可使抗菌药物疗效下降,副作用及不良反应增加,从而导致重症肺炎患者病情加重,进展为 ARDS。本研究结果显示,白蛋白水平是预测模型中较为重要的指标之一,白蛋白 < 28.05 g/L 可用于预测重症肺炎患者进展为 ARDS。

APACHE II 评分是入住 ICU 患者病情评估常用评分标准之一^[33-34],其不仅包含了患者入院 24 h 内生命体征、器官功能及生化指标等急性生理状态,还有年龄、基础病史等慢性健康状态,二者综合可更全面评估并预测患者病情变化情况。有研究发现,APACHE II 评分升高与重症社区获得性肺炎患者死亡相关,其可有效预测患者死亡风险^[35]。还有研究发现,APACHE II 评分升高与患者 30 d 死亡显著相关,诊断 HAP 后 24 h 内的 APACHE II 评分可有助于预测患者 30 d 病死率^[36]。在针对新型冠状病毒肺炎的研究中也发现,APACHE II 评分与患者无创呼吸机支持、有创呼吸机支持及病死率等指标显著相关,可用于预测患者病情进展和死亡的发生^[37]。本研究显示,APACHE II 评分与重症肺炎患者进展为 ARDS 存在相关性,APACHE II > 23 分是重症肺炎患者进展为 ARDS 的危险因素,结合建模结果也显示,使用 APACHE II 评分可有效预测重症肺炎患

者进展为 ARDS。

LIPS 评分是针对存在肺损伤风险患者常用的预测评分,其中包含了肺损伤诱因、高危手术及创伤以及危险因素等指标,可全面地预测患者发生肺损伤的风险。目前研究发现,LIPS 评分与 ICU 患者发生 ARDS 以及 ARDS 患者死亡之间存在正相关,联合使用 LIPS 评分可有效预测 ARDS 的发生和 ARDS 患者的病死率,便于临床医生加强对 ARDS 的管理,提高患者的救治成功率^[38-39]。LIPS 评分可在一定程度预测新型冠状病毒感染患者是否进展为 ARDS 以及是否需要创机械通气,且根据疾病特点结合其他相关指标可提高预测精准度^[40]。本研究也显示,LIPS 评分与重症肺炎患者进展为 ARDS 之间存在正相关,LIPS > 5 分是重症肺炎患者进展为 ARDS 的危险因素,通过结合其他具有相关性的指标联合建立预测模型,可更加准确地预测重症肺炎患者是否进展为 ARDS。

本研究也有不足之处,首先,本研究仅收集了单中心的病历资料,在患者筛选过程中可能存在一定的选择偏倚,在后续使用过程中仍需进一步收集多中心数据分析及验证;其次,本研究采用单因素分析作为变量筛选方法,虽可较好地筛选混杂因素,但也存在外部验证效能不稳定的风险^[41],尽管在模型验证中采用了 1 000 次重抽样充分评估模型的效能及稳定性,但仍有待进一步外部验证模型的泛化性;最后,本研究构建的预测模型尚未在临床治疗中实际应用,并对使用模型是否可改善患者预后进行评价,对模型用于指导临床诊疗改善重症肺炎患者预后的作用仍需进一步研究分析。

综上所述,本研究结果显示重症肺炎患者进展为 ARDS 的危险因素包括 pH 值、最低收缩压、白蛋白水平、APACHE II 评分、LIPS 评分,以上述危险因素构建随机森林模型预测重症肺炎是否进展为 ARDS 可为临床医生更好地评估重症肺炎患者提供可靠的依据。

利益冲突 所有作者声明不存在利益冲突

作者贡献声明 贾贵阳: 酝酿和设计实验、实施研究、采集数据、分析/解释数据、起草文章、统计分析、获取研究经费;王太珊: 实施研究、采集数据、分析/解释数据、起草文章、统计分析;宋尔芹: 实施研究、采集数据、对文章的知识性内容作批评性审阅、技术支持;肖玄: 实施研究、采集数据、对文章的知识性内容作批评性审阅、支持性贡献;刘国跃: 酝酿和设计实验、实施研究、分析/解释数据、对文章的知识性内容作批评性审阅、统计分析、获取研究经费、指导

参考文献

- [1] Meyer NJ, Gattinoni L, Calfee CS. Acute respiratory distress syndrome[J]. *Lancet*, 2021, 398(10300): 622–637. DOI: 10.1016/S0140-6736(21)00439-6.
- [2] European Network for ICU-Related Respiratory Infections (ENIRRI) European Respiratory Society-Clinical Research Collaboration Investigators. European Network for ICU-Related Respiratory Infections (ENIRRI): a multinational, prospective, cohort study of nosocomial LRTI[J]. *Intensive Care Med*, 2023, 49(10): 1212–1222. DOI: 10.1007/s00134-023-07210-9.
- [3] Cillóniz C, Torres A, Niederman MS. Management of pneumonia in critically ill patients[J]. *BMJ*, 2021, 375: e065871. DOI: 10.1136/bmj-2021-065871.
- [4] Grothberg JC, Reynolds D, Kraft BD. Management of severe acute respiratory distress syndrome: a primer[J]. *Crit Care*, 2023, 27(1): 289. DOI: 10.1186/s13054-023-04572-w.
- [5] Frazer J, Notin P, Dias M, et al. Disease variant prediction with deep generative models of evolutionary data[J]. *Nature*, 2021, 599(7883): 91–95. DOI: 10.1038/s41586-021-04043-8. Erratum in: *Nature*, 2022, 601(7892): E7. DOI: 10.1038/s41586-021-04207-6.
- [6] 刘陈钰, 张策, 迟元卉, 等. 脑卒中患者院前无创筛查预测模型的构建与外部验证: 一项基于人工智能 DeepFM 算法的研究[J]. *中华危重病急救医学*, 2024, 36(11): 1163–1168. DOI: 10.3760/cma.j.cn121430-20240526-00461.
- [7] 刘军芳, 段小霞, 马志琴, 等. 急诊心肺复苏患者临床分析及院内自主循环恢复预测模型的建立[J]. *中华危重病急救医学*, 2024, 36(1): 40–43. DOI: 10.3760/cma.j.cn121430-20231005-00836.
- [8] American Heart Association. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: a scientific statement from the American Heart Association[J]. *Circulation*, 2023, 148(24): 1982–2004. DOI: 10.1161/CIR.0000000000001191.
- [9] Raghav K, Hwang H, Jácome AA, et al. Development and validation of a novel nomogram for individualized prediction of survival in cancer of unknown primary[J]. *Clin Cancer Res*, 2021, 27(12): 3414–3421. DOI: 10.1158/1078-0432.CCR-20-4117.
- [10] 孙水雅, 桂薇薇, 贾程芳, 等. 浙江省基层 2 型糖尿病患者糖尿病肾病危险因素列线图分析[J]. *中华内科杂志*, 2023, 62(2): 169–175. DOI: 10.3760/cma.j.cn12138-20220511-00360.
- [11] Kale M, Wankhede N, Pawar R, et al. AI-driven innovations in Alzheimer's disease: integrating early diagnosis, personalized treatment, and prognostic modelling[J]. *Ageing Res Rev*, 2024, 101: 102497. DOI: 10.1016/j.arr.2024.102497.
- [12] Tang R, Luo R, Tang SW, et al. Machine learning in predicting antimicrobial resistance: a systematic review and meta-analysis[J]. *Int J Antimicrob Agents*, 2022, 60(5–6): 106684. DOI: 10.1016/j.ijantimicag.2022.106684.
- [13] Wang MJ, Dai X, Yang X, et al. Serum protein fishing for machine learning-boosted diagnostic classification of small nodules of lung[J]. *ACS Nano*, 2024, 18(5): 4038–4055. DOI: 10.1021/acsnano.3c07217.
- [14] 谢政, 金晶, 刘东松, 等. 基于监督机器学习算法构建脓毒性休克患者死亡风险的预测模型[J]. *中华危重病急救医学*, 2024, 36(4): 345–352. DOI: 10.3760/cma.j.cn121430-20230930-00832.
- [15] Martin-Loeches I, Torres A, Nagavci B, et al. ERS/ESICM/ESCMID/ALAT guidelines for the management of severe community-acquired pneumonia[J]. *Intensive Care Med*, 2023, 49(6): 615–632. DOI: 10.1007/s00134-023-07033-8. Erratum in: *Intensive Care Med*, 2023, 49(8): 1040–1041. DOI: 10.1007/s00134-023-07082-z.
- [16] 中华医学会呼吸病学分会感染学组. 中国成人医院获得性肺炎与呼吸机相关性肺炎诊断和治疗指南(2018 年版)[J]. *中华结核和呼吸杂志*, 2018, 41(4): 255–280. DOI: 10.3760/cma.j.issn.1001-0939.2018.04.006.
- [17] European Society of Intensive Care Medicine Taskforce on ARDS. ESICM guidelines on acute respiratory distress syndrome: definition, phenotyping and respiratory support strategies[J]. *Intensive Care Med*, 2023, 49(7): 727–759. DOI: 10.1007/s00134-023-07050-7.
- [18] Matthay MA, Arabi Y, Arroliga AC, et al. A new global definition of acute respiratory distress syndrome[J]. *Am J Respir Crit Care Med*, 2024, 209(1): 37–47. DOI: 10.1164/rccm.202303-0558WS.
- [19] Long ME, Mallampalli RK, Horowitz JC. Pathogenesis of pneumonia and acute lung injury[J]. *Clin Sci (Lond)*, 2022, 136(10): 747–769. DOI: 10.1042/CS20210879.
- [20] 吴嘉荔, 肖红科, 李雪, 等. 序贯器官衰竭评分对重症肺炎 ARDS 患者预后的评估价值[J]. *中华危重病急救医学*, 2021, 33(9): 1057–1062. DOI: 10.3760/cma.j.cn121430-20210115-00076.
- [21] Luyt CE, Bouadma L, Morris AC, et al. Pulmonary infections complicating ARDS[J]. *Intensive Care Med*, 2020, 46(12): 2168–2183. DOI: 10.1007/s00134-020-06292-z.
- [22] Qadir N, Chang SY. Pharmacologic treatments for acute respiratory distress syndrome[J]. *Crit Care Clin*, 2021, 37(4): 877–893. DOI: 10.1016/j.ccc.2021.05.009.
- [23] 方可, 笪欢欢, 孙瑞祥, 等. 体外循环手术停机困难转静脉-动脉体外膜肺氧合的建立配合与转运管理[J]. *中华危重病急救医学*, 2024, 36(4): 415–421. DOI: 10.3760/cma.j.cn121430-20230829-00714.
- [24] Elsebai MF, Habib EE. Blood pH and COVID-19[J]. *Arch Pharm (Weinheim)*, 2023, 356(5): e2200558. DOI: 10.1002/ardp.202200558.
- [25] Quade BN, Parker MD, Occhipinti R. The therapeutic importance of acid-base balance[J]. *Biochem Pharmacol*, 2021, 183: 114278. DOI: 10.1016/j.bcp.2020.114278.
- [26] Xie K, Guan SN, Kong XX, et al. Predictors of mortality in severe pneumonia patients: a systematic review and meta-analysis[J]. *Syst Rev*, 2024, 13(1): 210. DOI: 10.1186/s13643-024-02621-1.
- [27] 陈婵, 梁飞宇, 张静, 等. 老年急性肺损伤患者列线图预后模型的构建[J]. *温州医科大学学报*, 2021, 51(6): 449–453. DOI: 10.3969/j.issn.2095-9400.2021.06.004.
- [28] Daujat-Chavanieu M, Kot M. Albumin is a secret factor involved in multidirectional interactions among the serotonergic, immune and endocrine systems that supervises the mechanism of CYP1A and CYP3A regulation in the liver[J]. *Pharmacol Ther*, 2020, 215: 107616. DOI: 10.1016/j.pharmthera.2020.107616.
- [29] Wiedermann CJ. Hypoalbuminemia as surrogate and culprit of infections[J]. *Int J Mol Sci*, 2021, 22(9): 4496. DOI: 10.3390/ijms22094496.
- [30] Klinkmann G, Klammt S, Jäschke M, et al. Impact of albumin binding function on pharmacokinetics and pharmacodynamics of furosemide[J]. *Medicina (Kaunas)*, 2022, 58(12): 1780. DOI: 10.3390/medicina58121780.
- [31] Huang YZ, Yang J, Xie JF, et al. Association between pathophysiology and volume of distribution among patients with sepsis or septic shock treated with imipenem: a prospective cohort study[J]. *J Infect Dis*, 2020, 221(Suppl 2): S272–S278. DOI: 10.1093/infdis/jiz651.
- [32] 武阿龙, 徐丙发. 抗菌药物游离浓度测定的研究进展[J]. *中国新药杂志*, 2020, 29(23): 2672–2676.
- [33] Hurley JC. Trends in ICU mortality and underlying risk over three decades among mechanically ventilated patients. A group level analysis of cohorts from infection prevention studies[J]. *Ann Intensive Care*, 2023, 13(1): 62. DOI: 10.1186/s13613-023-01159-0.
- [34] Zhao LN, Su FH, Zhang NN, et al. The impact of the new acute respiratory distress syndrome (ARDS) criteria on Berlin criteria ARDS patients: a multicenter cohort study[J]. *BMC Med*, 2023, 21(1): 456. DOI: 10.1186/s12916-023-03144-7. Erratum in: *BMC Med*, 2024, 22(1): 515. DOI: 10.1186/s12916-024-03750-z.
- [35] Pan JJ, Bu W, Guo T, et al. Development and validation of an in-hospital mortality risk prediction model for patients with severe community-acquired pneumonia in the intensive care unit[J]. *BMC Pulm Med*, 2023, 23(1): 303. DOI: 10.1186/s12890-023-02567-5.
- [36] Wen JN, Li N, Guo CX, et al. Performance and comparison of assessment models to predict 30-day mortality in patients with hospital-acquired pneumonia[J]. *Chin Med J (Engl)*, 2020, 133(24): 2947–2952. DOI: 10.1097/CM9.0000000000001252.
- [37] Cheng PF, Wu H, Yang JZ, et al. Pneumonia scoring systems for severe COVID-19: which one is better[J]. *Virology*, 2021, 18(1): 33. DOI: 10.1186/s12985-021-01502-6.
- [38] 赵峰, 沈子渊, 杨翠, 等. LIPS 评分联合 APACHE II 评分和氧合指数预测 ARDS 发生模型的建立和验证[J]. *中华危重病急救医学*, 2022, 34(10): 1048–1054. DOI: 10.3760/cma.j.cn121430-20220930-00883.
- [39] Kim BK, Kim S, Kim CY, et al. Predictive role of lung injury prediction score in the development of acute respiratory distress syndrome in Korea[J]. *Yonsei Med J*, 2021, 62(5): 417–423. DOI: 10.3349/ymj.2021.62.5.417.
- [40] Society of Critical Care Medicine Discovery Viral Infection and Respiratory Illness Universal Study (VIRUS): COVID-19 Registry Investigator Group. Development and validation of an acute respiratory distress syndrome prediction model in coronavirus disease 2019: updated lung injury prediction score[J]. *Mayo Clin Proc*, 2023, 98(5): 736–747. DOI: 10.1016/j.mayocp.2022.11.021.
- [41] 聂朦, 鄂娜, 袁志权, 等. 临床流行病学研究中多因素分析时的变量选择方法[J]. *中华内科杂志*, 2024, 63(5): 462–467. DOI: 10.3760/cma.j.cn12138-20230828-00093.