

ICU 住院患者重度急性呼吸窘迫综合征早期预测模型的构建

武俊伟^{1,2}, 刘超³, 王雪⁴, 杜甲琚², 李姣¹, 谢菲⁵

¹中国医学科学院/北京协和医学院 医学信息研究所, 北京 100020; ²解放军总医院研究生院图书馆, 北京 100853; ³医渡云(北京)技术有限公司, 北京 100083; ⁴首都医科大学宣武医院 图书馆, 北京 100053; ⁵解放军总医院第八医学中心 呼吸与危重症医学部, 北京 100853

摘要: **背景** 急性呼吸窘迫综合征 (acute respiratory distress syndrome, ARDS) 发病率高, 10% 的 ICU 住院是由 ARDS 所致, 临床特征通常在诱发事件后 6~72 h 出现, 并迅速加重。其死亡率亦相当高且会随病情严重程度而增加。**目的** 构建一种便捷、无创的危重症 ARDS 早期预测模型。**方法** 采用麻省理工学院与飞利浦创建的 eICU 协作研究数据集, 从中检索诊断为 ARDS 患者的呼吸频率、体温、心跳三种生命体征数据以及氧合指数 ($\text{PaO}_2/\text{FiO}_2$)。 $\text{PaO}_2/\text{FiO}_2 \leq 100$ mmHg 为重度 ARDS。以每个氧合指数观测点 (ARDS 诊断时间点) 为原点, 96 h 为一个时间窗, 应用逻辑回归、随机森林及 LightGBM 构建预测模型, 分析诊断前 6~96 h、6~48 h 以及 6~24 h 的生命体征数据预测是否会发生重度 ARDS。通过 oob 评分、交叉验证以及校准曲线评估模型性能, 并选取 2014 年解放军总医院医院呼吸重症监护室的 ARDS 病例对模型进行外部验证。**结果** 从 eICU 数据库检索纳入 232 例 ARDS 患者共 3 140 次氧合指数测量记录, 其中 $\text{PaO}_2/\text{FiO}_2 \leq 100$ mmHg (1 mmHg=0.133 kPa) 共 1 042 次。以 6~96 h、6~48 h 以及 6~24 h 的生命体征数据各自采用逻辑回归、随机森林及 LightGBM 建立了 9 个预测模型。不同时间窗比较, 6~96 h 的预测准确率及 AUC 最高; 不同模型间比较随机森林模型的诊断性能最优; 6~96 h 随机森林模型准确率为 0.833, AUC 为 0.885; 6~48 h、6~24 h 时间窗口的 AUC 分别为 0.815、0.806; LightGBM、逻辑回归模型的 6~96 h 时间窗口 AUC 分为 0.868、0.634。各模型在解放军总医院 ARDS 患者中进行验证, 依然是 6~96 h 时间窗口的随机森林模型预测性能最佳, 准确率为 0.834, AUC 为 0.843。**结论** 基于随机森林构建的 ARDS 预测模型具有良好的预测能力, 通过无创且易获取的心率、体温、呼吸频率三种体征指标, 利用提前 6~96 h 时间窗数据对重度 ARDS 的发生进行预测, 可帮助医护人员更早地进行干预和治疗。

关键词: eICU 数据集; ICU; 机器学习; 随机森林; 急性呼吸窘迫综合征

中图分类号: R563.8 **文献标志码:** A **文章编号:** 2095-5227(2022)11-1146-05 **DOI:** 10.3969/j.issn.2095-5227.2022.11.008

网络出版时间: 2022-11-11 10:25 **网络出版地址:** <https://kns.cnki.net/kcms/detail/10.1117.R.20221110.0919.004.html>

引用本文: 武俊伟, 刘超, 王雪, 等. ICU 住院患者重度急性呼吸窘迫综合征早期预测模型的构建 [J]. 解放军医学院学报, 2022, 43 (11): 1146-1150.

Establishing an early prediction model for severe acute respiratory distress syndrome

WU Junwei^{1,2}, LIU Chao³, WANG Xue⁴, DU Jiajun², LI Jiao¹, XIE Fei⁵

¹Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China;

²Library of Graduate School, Chinese PLA General Hospital, Beijing 100853, China; ³Yidu Cloud (Beijing) Technology Ltd, Beijing 100083, China; ⁴Medical Library, Xuanwu Hospital, Capital Medical University, Beijing 100053, China; ⁵College of

Pulmonary and Critical Care Medicine, Chinese PLA General Hospital, Beijing 100853, China

Corresponding authors: LI Jiao. Email: li.jiao@imicams.ac.cn; XIE Fei. Email: xiefei0522@163.com

Abstract: Background Acute respiratory distress syndrome (ARDS) is a disease with high morbidity and accounts for 10% of ICU admissions, with clinical features usually presenting within 6-72 hours of the pathogenesis and rapidly worsening. The mortality rate is also high and increases with the severity of the disease. **Objective** To establish a convenient, noninvasive early prediction model for severe ARDS. **Methods** The eICU Collaborative Research Database created by MIT and Philips was used to retrieve data on three vital signs (respiratory rate, temperature, and heart rate) and oxygenation index ($\text{PaO}_2/\text{FiO}_2$) of patients diagnosed with ARDS, and $\text{PaO}_2/\text{FiO}_2 \leq 100$ mmHg was considered as severe ARDS. 96 h was used as a time window, and logistic regression, random

收稿日期: 2022-02-18

基金项目: 国家自然科学基金项目 (61976223); 解放军总医院医学创新研究项目 (2019YFC0121703); 解放军总医院转化医学重点项目 (2018TM-03); 解放军总医院临床扶持基金 (2018FC-WJFWZX-2-04)

Supported by the National Natural Science Foundation of China (61976223)

作者简介: 武俊伟, 男, 本科, 馆员。Email: wujunwei0210@foxmail.com; 共同第一作者: 刘超, 男, 博士。Email: liuchao5186@gmail.com

通信作者: 李姣, 女, 博士, 研究员。Email: li.jiao@imicams.ac.cn; 谢菲, 女, 博士, 主任医师。Email: xiefei0522@163.com

forest and LightGBM were applied to establish a prediction model to analyze vital sign data from 6-96 h, 6-48 h and 6-24 h before diagnosis to predict whether severe ARDS would occur. Model performance was evaluated by oob score, cross-validation and calibration curve, and also ARDS patients from Respiratory Intensive Care Unit of Chinese PLA General Hospital were selected to validate the models independently. **Results** A total of 232 patients were retrieved from the eICU database with 3 140 oxygenation index measurements during hospitalization, including 1 042 with $\text{PaO}_2/\text{FiO}_2 \leq 100$ mmHg. The 6-96 h, 6-48 h, and 6-24 h vital sign data were respectively used to build 9 prediction models by using logistic regression, random forest, and LightGBM. Comparing different time windows, the highest prediction accuracy and AUC were obtained for 6-96 h; the best diagnostic performance was obtained for the random forest model compared among different models; the accuracy of the random forest model for 6-96 h was 0.833 and the AUC was 0.885; the AUCs for the 6-48 h and 6-24 h time windows were 0.815 and 0.806, respectively; the AUC of LightGBM, and logistic regression models of 6-96 h time window was 0.868 and 0.634, respectively. Each model was validated in ARDS patients in Chinese PLA General Hospital, and the random forest model with 6-96 h time window had the best prediction performance with an accuracy of 0.834 and AUC of 0.843. **Conclusion** The ARDS early prediction model based on random forest has good predictive ability. It can warn the occurrence of severe ARDS through non-invasive and three easy-to-obtain physical indicators of heart rate, body temperature and respiratory rate, and help medical staff to make earlier intervention and treatment, relieve the pressure of inadequate medical resources, and improve the success rate of treatment.

Keywords: eICU dataset; ICU; machine learning; random forest; acute respiratory distress syndrome

Cited as: Wu JW, Liu CH, Wang X, et al. Establishing an early prediction model for severe acute respiratory distress syndrome [J]. Acad J Chin PLA Med Sch, 2022, 43 (11) : 1146-1150.

急性呼吸窘迫综合征 (acute respiratory distress syndrome, ARDS) 是一种以急性双肺异常为特征的具有独特类型的低氧血症性呼吸衰竭。ARDS 的 ICU 死亡率和住院死亡率分别为 35.3% 和 40%，且死亡率会随病情严重程度而增加。有研究表明，轻、中、重度 ARDS 死亡率分别达到 34.9%、40.3%、46.1%^[1]。因此提前预测重度 ARDS 的发生变得尤为重要。现有研究报道呼吸频率 (respiratory rate, RR) 和外周血氧饱和度 (SpO_2) 是预测和识别 ARDS 的两个重要参数，并通过以下指标进行进一步评估： $\text{RR} \geq 30$ 次/min； $\text{SpO}_2 \leq 92\%$ ；氧合指数 (实际采用动脉血氧分压与吸入氧浓度分数的比值 $\text{PaO}_2/\text{FiO}_2$ ，下文均用 P/F 表示) $\text{P/F} \leq 300$ mmHg (1 mmHg=0.133 kPa)^[2]。但 SpO_2 和氧合指数都需要相应的医疗仪器及有创的检查来实现，这对于资源紧缺的医疗机构来说是一项难以解决的问题。以往研究报道了不同 ARDS 预测模型，如 Le 等^[3] 选取 22 种特征，提前 12 h、24 h、48 h 预测 ARDS 的发生，其中提前 12 h 预测 ARDS 发生的 AUC 最高 (AUC=0.827)。Singhal 等^[4] 选择 148 个统计特征对 COVID-19 患者提前 6~36 h 进行 ARDS 发生的预测，最佳预测时间为 ARDS 发生前 12 h，其 AUC 为 0.89。上述不同的预测模型存在样本量小、模型复杂、特征选取较多等问题。因此，有必要选取无创、快速的体征指标对重度 ARDS 的发生进行提前预测。我们重点关注的是重度 ARDS 预测模型的方法学，以期找到一种简单、高效、便捷、准确的预测方法。本次研究采用 eICU 协作研究数据库 (eICU Collaborative Research Database) 通过 RR、心率

(heart rate, HR)、体温三项生命体征及其相关统计指标，提前 6 h 预测患者是否会进展为重度 ARDS。

资料与方法

1 数据来源 eICU 协作研究数据库是一个由 MIT 计算生理学实验室与飞利浦 eICU 研究所合作创建的大型多中心公开数据库，该数据库包含 2014 - 2015 年美国各地重症监护病房超过 20 万例患者的 ICU 数据，且数据内容是经过去隐私化处理，包括生命体征监测、护理计划文档、疾病严重程度监测、诊治信息等^[5]。选取 eICU 数据库中诊断为 ARDS 的 1714 例患者的临床资料。研究的纳入标准：1) 年龄 ≥ 18 岁；2) 入院期间有至少一次 PaO_2 检测记录用于氧合指数 P/F 的计算 (图 1)；3) 被诊断为“pulmonary| respiratory failure|

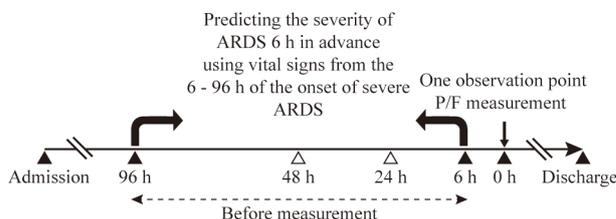


图 1 一个观测时间段 (时间窗口) 的详细信息

eICU 中 ARDS 患者的体温信息缺失较多，因此一个时间段内如果没有同时包含 HR、RR、体温三种体征信息，则不会纳入该时间窗口。体征信息较为完整的患者会有多个符合上述条件的观测点及时间窗口

Fig.1 Details of an observation period (a time window)

There is a mass of missing temperature data for ARDS patients in the eICU, so if a time period that does not contain HR, RR, and Temp, it will not be included in the time window. Patients with complete vital sign will have multiple observation points and time windows that meet the above criteria

ARDS”的患者, ICD 编码为 518.81, J80。排除标准: 1) 住院记录中没有测量 PaO₂; 2) 氧合指数观测时间段内体温、HR、RR 存在缺失值。

独立验证集来源 另外选取 2014 年解放军总医院呼吸重症监护室收治的 15 例因吸入性肺损伤所致的 ARDS 患者, 均为男性, 平均年龄 20.7(18~24) 岁, 平均住院时间 16.6 d。经上述相同的纳入排除标准, 最终纳入重度 ARDS 观测点 21 个, 轻中度 ARDS 观测点 62 个。

2 ARDS 严重程度界定 根据 ARDS 的柏林定义^[6], 轻度 ARDS: 呼吸机设置为呼气末正压(positive end expiratory pressure, PEEP) 或持续气道正压(continuous positive airway pressure, CPAP) ≥ 5 cmH₂O (1 cmH₂O=0.098 kPa) 时, 200 mmHg < P/F ≤ 300 mmHg; 中度 ARDS: 呼吸机设置为 PEEP ≥ 5 cmH₂O 时, 100 mmHg < P/F ≤ 200 mmHg; 重度 ARDS: 呼吸机设置为 PEEP ≥ 5 cmH₂O 时, P/F ≤ 100 mmHg。

3 预测变量选择及处理 心率、呼吸和体温是机体在应急状态下最早出现变化的生命体征, 国际上已有临床评分指标采用了体征及临床检验数据对急诊患者的病情进行评估预测, 如改良早期预警评分(MEWS)^[7] 及全身炎症反应综合征评分等^[8]。本次研究将探讨仅采用 HR、RR、体温三种体征信息来预测重度 ARDS 发生的可行性。

eICU 的 lab 表中包含患者在 ICU 期间全部的实验室检查的结果, 其中 PaO₂ 和 FiO₂ 在同一时刻分别记录了二者实际的数值, 可以计算出不同时刻的氧合指数 P/F 的值, 判断 ARDS 严重程度。模型以 ARDS 观测点为原点, 分析倒推 6~96 h 时间段内的 HR、RR、体温体征信息及其在该时间段内的波动值(极值、极差和平均值) 患者是否会进展为重度 ARDS(P/F ≤ 100 mmHg)。本研究亦对 6~48 h 及 6~24 h 两个不同观测时间段的数据进行分析处理, 用于比较不同时间窗口对模型预测能力的影响。

4 算法选择 研究选择三种不同的算法建立预测模型。1) 逻辑回归(logistic regression, LR) 是从线性回归模型推广的一种广义的分析模型, 具有模型简单、训练快速等特点。2) 随机森林(random forest, RF) 是利用 bootstrap 重抽样方法从原始样本中有放回的选择多个样本, 对于每个 Bootstrap 都构建一个决策树模型, 多棵决策树通过投票的方式最终给出预测结果, 大量的研究证实 RF 具有很高的预测准确性且不容易出现过拟合^[9], 且决策

树的绘制与临床路径的制定原理相似, 具有一定的模型可解释性。3) LightGBM 是 Microsoft 开发的基于决策树的一种梯度提升框架, 具有快速、分布式及高性能等特点。可用于排序、分类、回归及相关机器学习领域。它采用新的一直分裂叶子节点的策略(按最大增益的节点或叶子明智, Leaf-wise) 进行展开^[10]。

5 数据处理 使用 PostgreSQL 数据库筛选并提取符合纳入标准的数据信息, 应用 Python 3.7 进行进一步的数据清洗, 去除患者体温 < 30℃、HR、RR 为 0 的测量值, 然后对数据进行统计分析并采用 Python 建立预测模型。

6 模型评价指标 应用 Python 的 sklearn 库对三种不同算法的预测模型及三个不同时间窗口的受试者工作特征曲线下面积(area under the receiver operation characteristic curve, AUC) 进行评价, AUC 用于衡量机器学习的泛化能力(性能), 值越大表示模型的性能越好, 预测能力越强。采用 oob error(袋外错误率, out-of-bag error) 来评价随机森林的预测性能误差, oob score(1-oob error) 越高则模型预测性能越好。应用 sklearn.calibration 库中的校准曲线(calibration curve) 比较分类器概率预测的校准效果, 最理想情况下, 校准曲线是一条对角线(以虚线表示), 模型越贴近校准曲线, 说明预测概率越接近于真实概率。

结果

1 纳入人群的人口学特征 eICU 队列 eICU 中共有 200 859 例住院患者, 最终纳入 232 例患者的 3 140 个氧合指数观测点, 全部氧合指数观测点包括 1 042 次 P/F ≤ 100 mmHg(重度 ARDS), 2 098 次 P/F > 100 mmHg(轻中度 ARDS)。232 例 ARDS 患者, 男性 115 例, 女性 117 例。年龄 19~88 (56.1 \pm 16.1) 岁。身高 (168.0 \pm 14.2) cm; 体质量 (90.0 \pm 31.4) kg。死亡 101 例 (43.5%)。人种分布: 非裔美籍 25 例, 亚裔 8 例, 白种人 165 例, 西班牙裔 15 例, 印第安人 3 例, 其他或未知 16 例。平均住院时间 13.4 d。见表 1。此队列数据按照 3: 1 的比例分为测试集和验证集。

2 不同模型预测性能比较 模型以 ARDS 患者 6~96 h、6~48 h、6~24 h 三个时间段内的 HR、RR、体温体征信息及其在该时间段内的波动值(极值、极差和平均值) 为变量, 预测 6 h 后患者是否会进展为重度 ARDS (P/F ≤ 100 mmHg)。随机森林模型 6~96 h 时间窗口的预测准确率为 0.833,

AUC 为 0.885, oob 评分 0.809(oob error 为 0.191)。LightGBM 6~96 h 时间窗口的准确率为 0.805, AUC 为 0.868。逻辑回归 6~96 h 时间窗口准确率为 0.702, AUC 为 0.634。6~48 h、6~24 h 时间窗口的详细指标见表 2、图 2。校准曲线显示, 96 h 时间窗口中, 随机森林和 LightGBM 的预测能力要远优于逻辑回归的预测能力(图 3)。

3 模型预测性能的外部验证 为了进一步比较模型在独立数据集中的表现, 我们提取了来自解放军总医院医院(图 2 中以 PLAGH 表示)15 例吸入性肺损伤所致的 ARDS 患者, 其中重度 ARDS 观测点 21 个, 轻中度 ARDS 观测点 62 个, 随机森林解放军总医院医院 96 h 时间窗口的准确率为 0.834, AUC 为 0.843(各模型的 ROC 曲线及 AUC 值见表 2、图 2)。

表 1 eICU 中 232 例 ARDS 患者人口学特征

Tab. 1 Demographic characteristics of 232 patients with ARDS in the eICU

Characteristic	Total	Severe ARDS (n=137)	Mild/moderate ARDS (n=95)	P
Age/(yrs, range)	19-88 (56.1 ± 16.1)	19-88 (53.8 ± 15.9)	23-88 (59.4 ± 16.1)	0.009
Male/(n, %)	115(49.6)	63(46.0)	52(55.2)	0.180
Height/cm	168.0 ± 14.2	167.2 ± 15.7	169.2 ± 11.5	0.300
Weight/kg	90.0 ± 31.4	90.4 ± 32.6	90.5 ± 29.6	0.980
Alive/(n, %)	134(56.5)	66(48.0)	68(71.6)	0.770
LOS/h	322.6 ± 247.8	340.0 ± 265.9	297.6 ± 218.1	0.200

LOS: length of stay.

表 2 不同模型各个时间窗口的预测性能指标

Tab. 2 Prediction performance metrics of different models for each time window

Model	Time window/h	Accuracy	AUC	Sensitivity	Specificity
eICU					
Random forest	6-24	0.760	0.806	0.759	0.520
	6-48	0.779	0.815	0.766	0.549
	6-96	0.833	0.885	0.779	0.686
LightGBM	6-24	0.771	0.820	0.747	0.575
	6-48	0.809	0.863	0.789	0.652
	6-96	0.805	0.868	0.762	0.602
Logistic regression	6-24	0.675	0.612	0.589	0.239
	6-48	0.678	0.615	0.593	0.144
	6-96	0.702	0.634	0.731	0.077
PLAGH ^a					
Random forest	6-96	0.843	0.834	0.750	0.571
LightGBM	6-96	0.783	0.772	0.588	0.476
Logistic regression	6-96	0.783	0.833	0.000 ^b	0.000 ^b

^aThe performance of the three algorithms for 6-96 h in eICU is higher than that in 6-48 h and 6-24 h time windows, so only the performance for 6-96 h is compared in PLAGH dataset; ^bAll labels with 1 are predicted incorrectly, so true positive/TP=0, then $S\bar{x}$ and $Sp=0$.

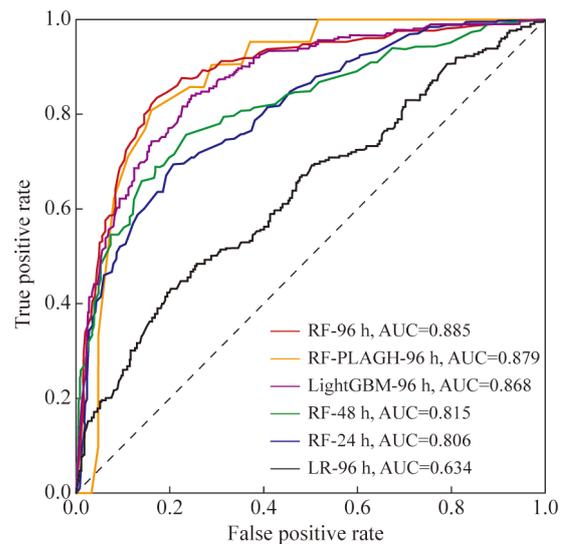


图 2 不同模型及时间窗口的 ROC 曲线

Fig.2 ROC curves for different models and time windows

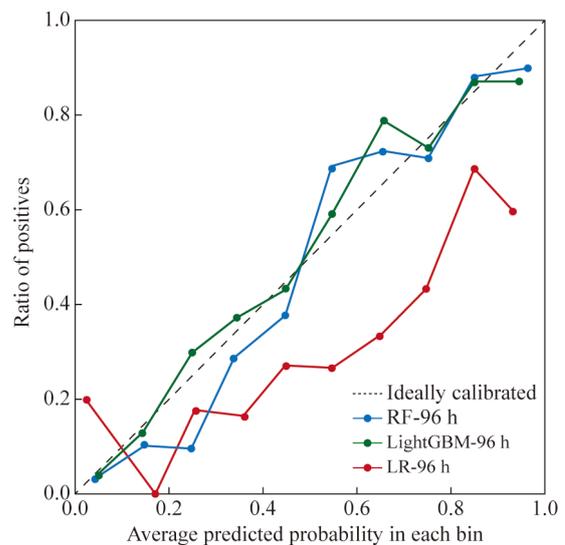


图 3 不同算法的校准曲线

Fig.3 Calibration curves for different algorithms

讨论

随着 ARDS 严重程度的增加, 相应的临床干预措施也随之增强。以往的研究多为 ARDS 的风险预测^[11]、预后评估^[12]、死亡率预测^[13]等, 应用机器学习对 ARDS 的发生进行早期预警的模型报道较少, 采用体征指标建立早期预警模型尤甚。Taoum 等^[14]通过 MIMIC- II 中 40 例 ARDS 患者持续的 HR、RR、SpO₂、平均动脉血压 (ABP_{Mean}) 四种动态生理信号监测来建立模型, 预测 1~180 h 后 ARDS 的发生, 测试集中的敏感度和特异性分别为 62.89% 和 72.67%。刘京铭等^[15]通过 MIMIC- III 中四种连续的基本生理特征 SpO₂、RR、HR、MAP 的时序数据, 采用 11 个包括 SpO₂ 的

平均值、标准差、倾斜度和样本熵, RR 的平均值和标准差, HR 的平均值和标准差, MAP 的平均值、倾斜度和峰度在内的特征, 提前预测 ARDS 的发生, 最佳 AUC 为 0.94。上述两种模型存在着仅使用单中心数据进行预测及验证、模型性能差异较大, 预测变量处理十分复杂、对几种生理指标的时序性信息关注度较高等问题。本研究采用多中心的 eICU 数据集, 使用更为简单的心率、呼吸频率及体温三种体征信息, 应用解释性强、快速高效的随机森林算法建立预测模型, 旨在对模型预测的方法学进行简化和改进, 对重度 ARDS 的发生予以简单高效、快速无创的早期预测。

本次研究选取三种基础的体征信息来自 232 例患者的 88 万余次临床监测数值, 大量的数据保证了建模后模型的可靠性, 同时在进行模型预测时, 通过同步计算三种体征数据的最大值、最小值、极差值和平均值, 可反映出患者在该时间段内体征数据波动情况。此种方式的优势是不需要对连续大量的体征信息进行存储, 对于上述的预测变量进行实时更新替换即可, 极大地减少了 ARDS 患者基数大、采集体征数据信息密集造成存储压力、访问获取压力和分析处理的计算压力等问题。本研究的模型应用到临床后, 可以更早对潜在的重症 ARDS 患者进行干预。纳入本次研究的患者年龄分布广泛, 为 19~88(56.1±16.1) 岁, 身高、体质量等均没有特异性, 在构建模型时, 排除了入住 ICU 类型等患者相关的全部静态信息, 增加了模型应用的普适性。为进一步验证模型在独立验证集中的表现, 研究还选取 15 例来自真实世界的 ARDS 患者进行独立验证。独立验证集的患者体征数据采集间隔远高于 eICU(1 h/次 vs 5 min/次), 造成独立验证集的准确性降低, 但 AUC 的值没有过多下降 (AUC=0.834), 说明模型在体征数据测量间隔 1 h 以上依旧表现良好。

研究的不足之处在于没有应用真实世界中 COVID-19 所致的 ARDS 患者的体征数据进行验证, 在接下来的研究中, 我们将对美国国家卫生研究所组建的国家 COVID 协作研究组织 (N3C) 提供的超过 300 万 COVID-19 阳性患者的临床数据进行验证^[16]。本研究虽然是侧重于预测模型的方法学的可行性研究, 但并未对模型的参数进行优

化以提高模型的性能。因此我们还将继续在算法参数优化、模型的可解释性上做进一步的研究, 并尝试探讨体征数据的不同测量间隔对预测准确性的影响。

参考文献

- 1 Bellani G, Laffey JG, Pham T, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries [J]. *JAMA*, 2016, 315 (8): 788-800.
- 2 Gibson PG, Qin L, Pua SH. COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS [J]. *Med J Aust*, 2020, 213 (2): 54-56.
- 3 Le S, Pellegrini E, Green-Saxena A, et al. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS) [J]. *J Crit Care*, 2020, 60: 96-102.
- 4 Singhal L, Garg Y, Yang P, et al. eARDS: a multi-center validation of an interpretable machine learning algorithm of early onset Acute Respiratory Distress Syndrome (ARDS) among critically ill adults with COVID-19 [J]. *PLoS One*, 2021, 16 (9): e0257056.
- 5 Pollard TJ, Johnson AEW, Raffa JD, et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research [J]. *Sci Data*, 2018, 5: 180178.
- 6 Force ARDSDT, Ranieri VM, Rubenfeld GD, et al. Acute respiratory distress syndrome: the Berlin Definition [J]. *JAMA*, 2012, 307 (23): 2526-2533.
- 7 陈湘, 雷光锋, 张雪晴, 等. 急诊科患者 30 d 死亡风险预测模型的构建与验证 [J]. *中华危重病急救医学*, 2022 (4): 421-425.
- 8 王盛标, 李涛, 李云峰, 等. 4 种评分系统对脓毒症患者预后的评估价值: 附 311 例回顾性分析 [J]. *中华危重病急救医学*, 2017, 29 (2): 133-138.
- 9 杨丰春, 郑思, 李姣. 可解释机器学习方法在疾病预测中的应用: 脓毒症患者死亡风险研究 [J]. *首都医科大学学报*, 2022, 43 (4): 610-617.
- 10 张渊, 冯聪, 李开源, 等. ICU 患者急性肾损伤发生风险的 LightGBM 预测模型 [J]. *解放军医学院学报*, 2019, 40 (4): 316-320.
- 11 荆晨晨, 孙淑青, 秦德春. 急性呼吸窘迫综合征患者早期风险预测模型的建立 [J]. *中华护理杂志*, 2020, 55 (9): 1285-1291.
- 12 程俐, 黄健. IL-8、IL-12 评估急性呼吸窘迫综合征患者预后的价值 [J]. *国际医药卫生导报*, 2021, 27 (24): 3850-3854.
- 13 方伯梁, 钱素云. 一项简单有效的急性呼吸窘迫综合征患儿病死率风险床旁模型 [J]. *中国小儿急救医学*, 2016, 23 (8): 515-515.
- 14 Taoum A, Mourad-Chehade F, Amoud H. Early-warning of ARDS using novelty detection and data fusion [J]. *Comput Biol Med*, 2018, 102: 191-199.
- 15 刘京铭, 王若琳, 郭伟, 等. 基于连续生理指标的急性呼吸窘迫综合征预警模型 [J]. *中国急救医学*, 2020, 40 (10): 956-960.
- 16 National COVID Cohort Collaborative (N3C) [DB/OL]. <https://covid.cd2h.org/n3c>.