

机器学习算法对危重患者死亡率预测效能的比较^{*}

李惠萍¹, 胡安民^{2△}

深圳市人民医院¹呼吸与危重症医学科,²麻醉科(广东深圳 518020)

【摘要】 目的 检测支持向量机(SVM)和随机森林(RF)两种机器学习算法对单中心重症患者 30 d 死亡率预测性能。方法 使用美国重症医学数据库(MIMIC-III)作为数据来源,纳入波士顿贝斯以色列女执事医疗中心接受重症监护的成年患者,使用 SVM 和 RF 算法构建重症患者 30 d 死亡率的预测模型,通过与传统的简化急性生理评分-II(SAPS-II)模型比较,并使用十折交叉验证法对模型性能进行评估。结果 研究纳入 30 833 例患者,SAPS-II 评分为 33(25-43),其中 2 909 例(9.4%)患者在进入 ICU 后 30 d 内死亡。传统 SAPS-II 评分模型的曲线下面积为 0.73 ± 0.02 ,而机器学习算法的 SVM、RF 预测模型的 AUC 值分别为 0.75 ± 0.02 和 0.79 ± 0.02 。结论 SVM 模型对危重患者 30 d 全因死亡率预测与传统的 SAPS-II 评分效能相仿,而 RF 模型对危重患者的预后预测更加精准。

【关键词】 医疗大数据;重症患者;预测模型;机器学习

【中图分类号】 R181.2;O141.41

【文献标志码】 A

DOI: 10.13820/j.cnki.gdyx.20193897

Artificial intelligent learning for prediction of all-cause mortality in critically ill patients. LI Hui-ping^{*}, HU An-min. ^{*}Department of Respiratory and Critical Medicine, Shenzhen People's Hospital, Shenzhen 518020, Guangdong, China
Corresponding author: HU An-min. E-mail: eanmin@163.com

【Abstract】 Objective To investigate the utility of Support Vector Machines (SVM) and Random Forest (RF) algorithms in predicting 30-day mortality of critically ill patients at a single medical center. **Methods** We used the data from Medical Information Mart for Intensive Care III (MIMIC-III) database which includes adult patients admitted to an intensive care unit at the Beth Israel Deaconess Medical Centre, Boston, MA, USA. The predicted mortality of critically ill patients based on SVM and RF methods were performed and compared with Simplified Acute Physiology Score-II (SAPS-II). We calculated the predicting efficacy using 10-fold cross validation to avoid bias. The primary outcome was 30-day all-cause mortality. **Results** A total of 30 833 patients were included with median SAPS-II of 33 (IQR 25-43), and 2 909 (9.4%) patients died during the 30 days after ICU admission. The cross-validated area under the receiver operating characteristic curve (AUROC) for SAPS-II was 0.73 ± 0.03 , 0.75 ± 0.02 for SVM and 0.79 ± 0.02 for RF. **Conclusion** Compared with conventional SAPS-II scores, RF is more precise for predicting all-cause mortality in critically ill patients.

【Key words】 medical big data; critically ill patients; prediction model; machine learning

医疗信息化的发展促进了大数据技术的进步和在医疗领域中的应用,包括优化患者治疗、提高临床决策、改善临床试验设计、合理分配医疗资源以及实施个体化治疗等多个方面。数据科学使用机器学习(machine learning, ML)算法建立疾病预测和预后模型,尽早识别病情变化并进行干预,评估患者可能的临床终点,以利于指导临床决策。重症监护室

(ICU)患者病情重、病死率高,其中 ICU 老年患者的 30 d 死亡率为 11.6%,而合并真菌血症的 ICU 患者死亡率可达 40.6%^[1-2]。目前临床上常用的简化急性生理评分-II(simplified acute physiology score-II, SAPS-II)可在一定程度上反映患者的危重度,但此评分要求变量与对应结果呈线性关系,在病情复杂、存在众多影响预后因素的 ICU 患者上使用

* 基金项目:广东省高水平医院建设基金

△通信作者:胡安民 E-mail: eanmin@163.com

时,其准确性受到影响。机器学习算法是一门交叉学科,综合运用概率论、统计学、算法复杂度理论等多种知识,对大型数据库进行分析学习,并优化程序性能。机器学习包括监督学习、非监督学习和强化学习 3 种模式,其中监督学习中的支持向量机(support vector machines, SVM)和随机森林(random forest, RF)算法可对多个变量进行模式识别,构建预测模型,可应用于诊疗辅助、图像分析等,同时这两种算法还可通过非线性关系规则,构建危重患者预后的预测模型^[3-6]。本研究使用美国重症医学数据库(medical information mart for intensive care, MIMIC-III)的临床数据,分别使用 SVM 和 RF 算法构建预后模型,并与 SAPS-II 模型进行比较,评估这 3 种模型对危重患者 30 d 全因死亡率的预测效能。

1 资料与方法

1.1 一般资料 2018 年 1 月至 2019 年 6 月,本研究基于 MIMIC-III,收集库中 2002 年 6 月至 2012 年 10 月的重症患者信息。MIMIC-III 数据库由 Beth Israel Deaconess 医疗中心重症监护室中住院的 46 520 例患者的医疗数据组成。数据库根据患者的人口学信息、诊断编码、实验室检查、医疗干预事件进行分类整理^[7]。该数据库经过严格的去隐私处理,得到哈佛大学医学院伦理审查委员会与麻省理工学院共同批准。本研究数据由通过“保护人类受试者培训”的研究人员胡安民(证书编号:26450451)提取。下载的数据均已通过严格去隐私化。纳入标准:首次接受 ICU 治疗的成人重症患者。排除标准:年龄 <18 岁或 >89 岁;ICU 停留时间 <24 h。

1.2 研究结局与预测变量 以患者进入 ICU 接受治疗后 30 d 内的死亡率为主要研究结果。纳入患者的死亡时间,以及患者进入 ICU 首日可能影响预后的 103 项数据,包括性别、年龄、种族、体质指数、入 ICU 类型、时间、首日的液体出入量、尿量、生命体征(脉搏血氧饱和度、心率、收缩压、舒张压、体温)极大值和极小值、血气分析(酸碱度、碳酸氢根浓度、血钾离子浓度、血钙离子浓度、钠离子浓度、氯离子浓度、阴离子间隙、碱剩余、动脉血氧分压、动脉血二氧化碳分压、肺动脉-氧压力梯度、血乳酸)的极大值和极小值、血常规(血红蛋白、血细胞比容、血小板、白细胞计数、中性粒细胞百分比)的极大值和极小值、凝血功能检查(凝血酶原时间、活化部分凝血活酶时间、国际标准化比值)的极大值和极小值、肝肾功能检查(血清总胆红素、丙氨酸氨基转移酶、天冬氨酸氨基转移酶、血清白蛋白、血清肌酐、内生肌酐清除率、血清尿素氮)的极大值和极小值、C

反应蛋白极大值、心肌标志物(肌酸激酶同工酶质量、肌钙蛋白 T、肌钙蛋白 I)极大值、脑利钠肽极大值、GCS 评分极小值、是否存在合并症(高血压、糖尿病、肺循环疾病、慢性肺部疾病、充血性心力衰竭、心律失常、瓣膜疾病、糖尿病并发症、肝病、肾病、中风、其他神经系统疾病、外周血管疾病、消化性溃疡、艾滋病、淋巴瘤、实体肿瘤、转移性癌症、类风湿关节炎、凝血功能障碍、体重下降、电解质紊乱、失血性贫血、缺铁性贫血、酒精滥用、药物滥用、精神病、抑郁)。

以上变量经过特征变量筛选后用于构建预测模型。用于构建模型的原始数据均发布在 ZENODO 数据库中(<http://doi.org/10.5281/zenodo.3876289>)。

1.3 模型评价 基于数据库采用结构化查询语言(structured query language, SQL)进行数据提取。将筛选出的原始数据通过 Rv. 3.6.0 软件进行数据分析、构建模型并评估。研究首先基于 SAPS-II 评分构建逻辑回归模型。将数据库患者的影响预后的特征变量资料通过 R 语言的 Boruta 包进行数据清洗后筛选^[8]。存在缺失值的特征变量通过 RF 多重插补算法依次进行建模插补^[9],分别构建 SVM 模型^[10]和 RF 模型^[11]。研究采用十折交叉的方法进行模型性能的验证,把数据随机分为 10 等份,依次将 10 等份中的每 1 份作为单独测试集,其余 9 份作为构建模型的训练集,进行效能验证。

在完成构建模型后,通过计算真阳性率(true positive rate, TPR)、阳性预测值(positive predictive value, PPV)、F 值、受试者工作特征(receiver operation characteristic, ROC)曲线下面积(area under the ROC curve, AUC)进行模型评估。TPR = 正确判断为阳性/所有实际阳性的样本 × 100%;PPV = 正确预测到的病例数/预测病例总数;F 值是 TPR 和 PPV 的加权调和平均值,综合反映模型整体性能。

1.4 统计学方法 采用 Rv. 3.6.0 统计软件,连续性计量资料的比较采用 *t* 检验和单因素方差分析,计数资料的比较采用 χ^2 检验。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 危重患者数据的基本特征 筛选出符合条件的危重患者 30 833 例,中位年龄 65 岁,住院的中位时间为 7.5 d。进入 ICU 后的 30 d 内死亡 2 909 例(9.4%)。与存活的患者比较,死亡患者年龄更大,女性占比多,急诊纳入 ICU 比例高,体温、pH 值和血红蛋白的最小值均更低, SAPS-II 评分更高, ICU 治疗时间和住院时间均更长,见表 1。

表 1 患者总体人口学特征及进入 ICU 30 d 后存活或死亡的患者特征 例(%)

项目	总例数 (<i>n</i> = 30 833)	生存 (<i>n</i> = 27 924)	死亡 (<i>n</i> = 2 909)	<i>P</i> 值
年龄(岁)*	65.0(52.4, 76.5)	64.5(51.9, 76)	69.7(57.1, 79.8)	<0.001
性别				<0.001
男	12 960(42.0)	16 266(58.3)	1 607(55.2)	
女	17 873(58.0)	11 658(41.7)	1 302(44.8)	
种族				<0.001
白种人	21 845(70.8)	19 871(71.2)	1 974(67.9)	
黑种人	2 362(7.7)	2 174(7.8)	188(6.5)	
拉丁美洲人	1 013(3.3)	940(3.4)	73(2.5)	
亚洲人	736(2.4)	668(2.4)	68(2.3)	
其他人种	4 877(15.8)	4 271(15.3)	606(20.8)	
入 ICU 类型				<0.001
择期	5 111(16.6)	5 029(18)	82(2.8)	
急诊	24 879(80.7)	22 134(79.3)	2 745(94.4)	
转入	843(2.7)	761(2.7)	82(2.8)	
体质指数(kg/m ²)*	27.4(24.0, 31.7)	27.5(24.1, 31.7)	26.6(23.0, 31.2)	0.161
收缩压最大值(mmHg)*	147(134, 164)	147(134, 163)	147(130, 166)	0.262
体温最小值(℃)*	36.1(35.7, 36.6)	36.2(35.7, 36.6)	36.0(35.4, 36.6)	<0.001
pH 值最小值*	7.3(7.3, 7.4)	7.3(7.3, 7.4)	7.3(7.2, 7.4)	<0.001
血红蛋白最小值(g/L)*	10.0(8.5, 11.6)	10.0(8.5, 11.6)	9.9(8.6, 11.5)	<0.001
血小板最小值($\times 10^9 \cdot L^{-1}$)*	184(130, 248)	185(132, 248)	177(105, 248)	0.317
患者入 ICU 当天的 SAPS-II 评分*	33(25, 43)	32(24, 41)	47(37, 58)	<0.001
ICU 住时间(d)*	2.5(1.6, 4.8)	2.4(1.6, 4.6)	3.3(1.9, 6)	<0.001
总住院时间(d)*	7.5(4.6, 12.8)	7.8(4.9, 13.5)	5(2.6, 8.1)	<0.001

注: * $M(P_{25}, P_{75})$

2.2 特征变量筛选和多重插补 筛选后得到特征变量 59 个,依次对含有缺失值的特征变量通过 RF 算法构建多重插补模型,然后使用该模型插补缺失值,补齐后的数据用于构建危重患者死亡率的预测模型。

2.3 模型评估 根据十折交叉验证结果,SAPS-II 模型的 TPR 值明显高于 SVM 模型,差异有统计学意义($P < 0.001$),但 SAPS-II 模型与 RF 模型之

间差异无统计学意义。SAPS-II、SVM、RF 的 PPV 值之间差异有统计学意义($P < 0.001$)。SVM 模型和 RF 模型的 F 值均显著高于 SAPS-II 模型($P < 0.001$)。3 个模型的 AUC 值逐渐升高,其中 RF 模型显著高于 SAPS-II 模型和 SVM 模型($P < 0.001$),RF 模型预测较 SAPS-II、SVM 模型更加精准。见表 2、图 1。

表 2 3 种模型的预测效能比较

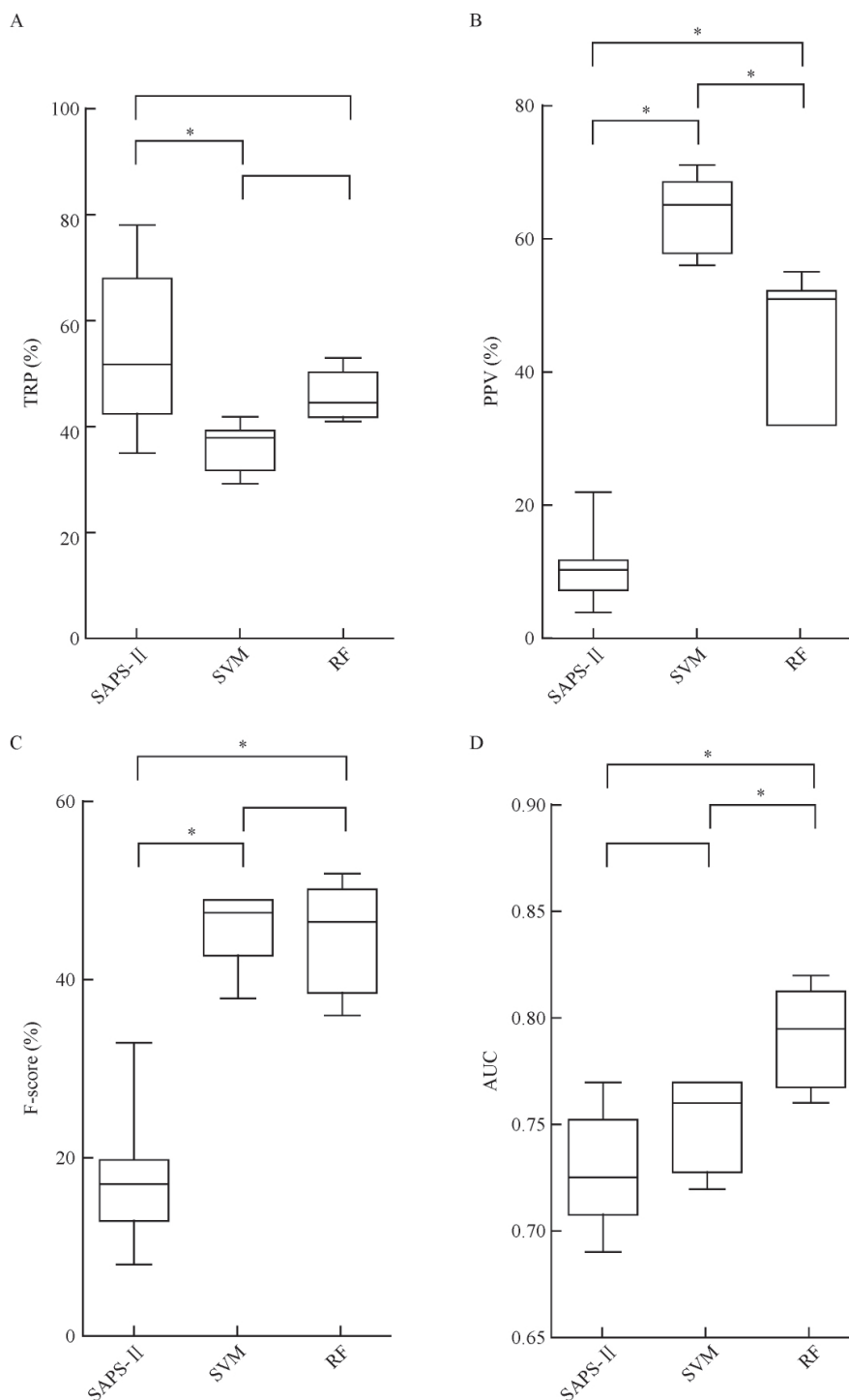
 $\bar{x} \pm s$

项目	TRP	PPV	F 值	AUC
SAPS-II	0.55 \pm 0.15	0.11 \pm 0.02	0.18 \pm 0.07	0.73 \pm 0.03
SVM	0.36 \pm 0.04	0.64 \pm 0.05	0.46 \pm 0.04	0.75 \pm 0.02
RF	0.46 \pm 0.05	0.45 \pm 0.10	0.45 \pm 0.06	0.80 \pm 0.02

3 讨论

危重患者的病情复杂,影响其预后的因素众多,预后评估方法繁琐,导致临床医生难以得到患者个体死亡风险结论^[12]。此外,临床工作中往往容易忽略个体患者的弱预测变量之间相互作用所导致的潜

在风险,从而影响对患者预后评估。因此建立能够精准预测患者预后的模型十分重要。传统的 SAPS-II 评分预测模型要求预测变量与相应结果之间呈线性关系^[2],对多元共线性数据、缺失数据和非平衡数据敏感性差,难以实现高阶交互作用^[13]。此外未得



注: A: TRP; B: PPV; C: F 值; D: AUC; * $P < 0.001$

图1 SAPS-II、SVM 与 RF 模型的预测效能比较

到充分校准的 SAPS-II 评分模型无法准确预测患者的死亡率^[14-15]。因此 SAPS-II 评分预测模型的预测效能并不理想。

本研究使用机器学习的 SVM 和 RF 算法构建危重患者 30 d 死亡率的预测模型,并评估其预测效能。SVM 具有分类思想简单、分类效果好、通过使用核函数可以向高维空间进行映射来解决非线性的特点^[10]。RF 模型对分类不平衡的数据和缺失数据

具有很强的抗干扰能力,抗过拟合能力比较强^[9],因此可用于构建死亡预测模型。本研究结果显示,与传统的 SAPS-II 预测模型比较,SVM 和 RF 算法构建的模型均优于 SAPS-II 模型,而 RF 模型比 SVM 模型在危重患者预后的预测效能更好。

在这项研究中,我们证明了机器学习算法可通过提供全面的危重患者临床数据而深度整合,具有重要的临床意义。先前的大规模研究已经证明了

SAPS - II 评分系统选择变量有限,对危重患者的预后预测存在缺陷^[14-15],而机器学习算法能够在大量数据的基础上在很多领域进行准确预测和决策,同样机器学习算法也可以为危重患者预后风险分层提供更好的预测模型。SVM 和 RF 模型各有特点,但 RF 模型比 SVM 模型对危重患者 30 d 死亡率的预测效能更好,从而为建立新的危重患者预后预测系统提供了重要依据。

但本研究尚存在不足之处:首先,机器学习方法是通过使用更多的变量和交互作用来预测风险,对明确能够降低该个体患者风险的具体治疗目标指导作用相对比较弱。目前临床需要根据危重患者的病理改变进行对应的治疗,如果无法确定具体的风险因素或相互作用,则可能减弱医生治疗方案调整的及时性。另外,本研究只讨论机器学习算法的可行性和预测效能,未能讨论其具体实现的过程。本研究采用的机器学习方法具备处理大量多维度信息的能力,但只有在将这些变量实际输入到机器学习算法中时才具有优势。目前这一过程相对耗时,随着计算机计算能力的提高,这一问题也会逐渐解决。目前我国很多医院可以自动收集危重患者临床数据,为医院构建多变量的预后模型提供便利^[16],因此在对模型的实际应用中不会给医护人员带来额外负担,从而可以进行即时风险评分计算。

本研究证实,基于机器学习算法的 SVM 和 RF 模型优于传统的 SAPS - II 模型,但在危重患者 30 d 全因死亡率的预测效能上,RF 模型整体预测性能优于 SVM 模型,能更准确地协助临床医生进行决策。本研究为今后危重患者预后的预测系统在临床使用提供坚实的数据支持。

参考文献

- [1] Nielsson MS, Christiansen CF, Johansen MB, et al. Mortality in Elderly ICU Patients: A Cohort Study [J]. *Acta Anaesthesiol Scand*, 2014, 58(1): 19-26.
- [2] Paiva JA, Pereira JM, Tabah A, et al. Characteristics and risk factors for 28 - day mortality of hospital acquired fungemias in ICUs: data from the EURO-BACT Study [J]. *Crit Care*, 2016, 20: 53.
- [3] Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists [J]. *JAMA*, 2016, 316(22): 2353-2354.
- [4] Ambale - Venkatesh B, Yang X, Wu CO, et al. Cardiovascular Event Prediction by Machine Learning: The Multi - Ethnic Study of Atherosclerosis, Machine Learning Prediction Models for In - Hospital Mortality After Transcatheter Aortic Valve Replacement [J]. *Circ Res*, 2017, 121(9): 1092-1101.
- [5] Angraal S, Martazavi BJ, Gupta A, et al. Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction [J]. *JACC Heart Failure*, 2020, 8(1): 12-21.
- [6] Dybowski R, Weller P, Chang R, et al. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm [J]. *Lancet*, 1996, 347(9009): 1146-1150.
- [7] Johnson AE, Pollard TJ, Shen L, et al. MIMIC - III, a freely accessible critical care database [J]. *Sci Data*, 2016, 3: 160035.
- [8] Kursa MB, Rudnicki WR. Feature selection with the boruta package [J]. *J Stat Software*, 2010, 36(11): 1-13.
- [9] Tang F, Ishwaran H. Random Forest Missing Data Algorithms [J]. *Stat Anal Data Min*, 2017, 10(6): 363-377.
- [10] Platt J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines [J]. *Adv Kernel Methods Support Vector Learn*, 1998, 208: 1-21.
- [11] Liaw A, Wiener M. Classification and Regression by randomForest [J]. *R News*, 2002, 2(3): 18-22.
- [12] Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE - acute physiology and chronic health evaluation: a physiologically based classification system [J]. *Crit Care Med*, 1981, 9(8): 591-597.
- [13] Liang C, Li Y, Luo J. A Novel Method to Detect Functional microRNA Regulatory Modules by Bicliques Merging [J]. *IEEE/ACM Trans Comput Biol Bioinform*, 2016, 13(3): 549-556.
- [14] Poole D, Rossi C, Latronico N, et al. Comparison between SAPS II and SAPS 3 in predicting hospital mortality in a cohort of 103 Italian ICUs. Is new always better? [J]. *Intensive Care Med*, 2012, 38(8): 1280-1288.
- [15] Nassar AP Jr, Mocelin AO, Nunes AL, et al. Caution when using prognostic models: a prospective comparison of 3 recent prognostic models [J]. *J Crit Care*, 2012, 27(4): 423. e1-e7.
- [16] Zhang Z. Big data and clinical research: focusing on the area of critical care medicine in mainland China [J]. *Quant Imaging Med Surg*, 2014, 4(5): 426-429.

(收稿日期: 2019 - 11 - 29 编辑: 罗劲娜)