

基于生命体征时序数据的创伤致死性大出血伤情动态预测模型开发及验证

郭程娱^{1,2}, 龚明慧³, 沈翹楚³, 韩辉², 王若琳³, 张红亮², 王俊康², 李春平^{3*}, 黎檀实^{1,2*}

¹南开大学医学院, 天津 300071; ²解放军总医院第一医学中心急诊科, 北京 100853; ³清华大学软件学院, 北京 100083

[中图分类号] R641 [文献标志码] A [DOI] 10.11855/j.issn.0577-7402.2023.0427

[声明] 本文所有作者声明无利益冲突

[引用本文] 郭程娱, 龚明慧, 沈翹楚, 等. 基于生命体征时序数据的创伤致死性大出血伤情动态预测模型开发及验证[J]. 解放军医学杂志, 2024, 49(6): 629-635.

[收稿日期] 2022-06-06 [录用日期] 2022-12-28 [上线日期] 2023-04-27

[摘要] **目的** 基于生命体征时序数据和机器学习算法建立创伤致死性大出血伤情动态预测模型。**方法** 回顾性分析重症监护医疗信息(MIMIC-IV)数据库2008—2019年7522例创伤伤员的生命体征时序数据, 并按照创伤后是否发生致死性大出血事件分为致死性大出血组($n=283$)与非致死性大出血组($n=7239$)。采用逻辑回归(LR)、支持向量机(SVM)、随机森林(RF)、自适应提升(AdaBoost)、门控循环单元(GRU)、门控循环单元-D(GRU-D)共6种机器学习算法开发创伤致死性大出血伤情动态预测模型, 对创伤伤员未来T小时($T=1、2、3$)发生致死性大出血伤情的风险进行动态预测。通过准确率、敏感度、特异度、阳性预测值、阴性预测值、约登指数以及受试者工作特征(ROC)曲线下面积(AUC)评估模型性能。基于解放军总医院创伤数据库对模型进行外部验证。**结果** MIMIC-IV数据集中, 基于GRU-D算法开发的一组动态预测模型效果最优, 预测未来1、2和3h发生致死性大出血的AUC分别为 0.946 ± 0.029 、 0.940 ± 0.032 和 0.943 ± 0.034 , 且差异无统计学意义($P=0.905$)。创伤数据集中, GRU-D模型取得了最佳外部验证效果, 预测未来1、2和3h发生致死性大出血的AUC分别为 0.779 ± 0.013 、 0.780 ± 0.008 和 0.778 ± 0.009 , 且差异无统计学意义($P=0.181$)。该组模型已部署在公开的网页计算器和医院急诊科信息系统中, 便于公众和医护人员使用和验证。**结论** 成功开发并验证了一组动态预测模型, 可对创伤致死性大出血伤情进行早期诊断和动态预测。

[关键词] 创伤; 大出血; 机器学习; 辅助诊断

Development and validation of dynamic prediction models using vital signs time series data for fatal massive hemorrhage in trauma

Guo Cheng-Yu^{1,2}, Gong Ming-Hui³, Shen Qiao-Chu³, Han Hui², Wang Ruo-Lin³, Zhang Hong-Liang², Wang Jun-Kang², Li Chun-Ping^{3*}, Li Tan-Shi^{1,2*}

¹School of Medicine, Nankai University, Tianjin 300071, China

²Department of Emergency, the First Medical Center of Chinese PLA General Hospital, Beijing 100853, China

³School of Software, Tsinghua University, Beijing 100083, China

*Corresponding author. Li Tan-Shi, E-mail: lts301@163.com; Li Chun-Ping, E-mail: cli@tsinghua.edu.cn

This work was supported by the National Key Research and Development Program of China (2020YFC1512702)

[Abstract] **Objective** To establish a dynamic prediction model of fatal massive hemorrhage in trauma based on the vital signs time series data and machine learning algorithms. **Methods** Retrospectively analyze the vital signs time series data of 7522 patients with trauma in the Medical Information Mart for Intensive Care-IV (MIMIC-IV) database from 2008 to 2019. According to the occurrence of posttraumatic fatal massive hemorrhage, the patients were divided into two groups: fatal massive hemorrhage group

[基金项目] 国家重点研发计划(2020YFC1512702)

[作者简介] 郭程娱, 医学博士, 主要从事创伤大出血预测模型方面的研究

[通信作者] 黎檀实, E-mail: lts301@163.com; 李春平, E-mail: cli@tsinghua.edu.cn

($n=283$) and non-fatal massive hemorrhage group ($n=7239$). Six machine learning algorithms, including logistic regression (LR), support vector machine (SVM), random forests (RF), adaptive boosting (AdaBoost), gated recurrent unit (GRU), and GRU-D were used to develop a dynamic prediction models of fatal massive hemorrhage in trauma. The probability of fatal massive hemorrhage in the following 1, 2, and 3 h was dynamically predicted. The performance of the models was evaluated by accuracy, sensitivity, specificity, positive predictive value, negative predictive value, Youden index, and area under receiver operating characteristic curve (AUC). The models were externally validated based on the trauma database of the Chinese PLA General Hospital. **Results** In the MIMIC-IV database, the set of dynamic prediction models based on the GRU-D algorithm was the best. The AUC for predicting fatal major bleeding in the next 1, 2, and 3 h were 0.946 ± 0.029 , 0.940 ± 0.032 , and 0.943 ± 0.034 , respectively, and there was no significant difference ($P=0.905$). In the trauma dataset, GRU-D model achieved the best external validation effect. The AUC for predicting fatal major bleeding in the next 1, 2, and 3 h were 0.779 ± 0.013 , 0.780 ± 0.008 , and 0.778 ± 0.009 , respectively, and there was no significant difference ($P=0.181$). This set of models was deployed in a public web calculator and hospital emergency department information system, which is convenient for the public and medical staff to use and validate the model. **Conclusion** A set of dynamic prediction models has been successfully developed and validated, which is greatly significant for the early diagnosis and dynamic prediction of fatal massive hemorrhage in trauma.

[Key words] wounds and injuries; massive hemorrhage; machine learning; assistant diagnosis

致死性大出血是由创伤引起的最严重的并发症之一,是创伤后可预防死亡的主要原因,占创伤相关死亡的40%^[1-2]。与显性大出血相比,隐性大出血易漏诊、误诊。若救护人员能够早期诊断,甚至提前预测致死性大出血病情,及时有效地进行干预并积极调整治疗策略,可显著降低伤员的伤残率和病死率,改善严重创伤的预后^[3]。目前,创伤大出血相关的预测模型大多为基于传统统计学方法构建的评分系统,如创伤相关严重出血(trauma-associated severe hemorrhage, TASH)评分^[4]、血液消耗评估(assessment of blood consumption, ABC)评分^[5]和威尔士亲王(Prince of Wales, PWH)评分^[6]等,其缺点为:(1)传统评分系统需要手动计算结果,较为耗时、复杂;(2)预测变量的多样性致使评分系统难以实现动态预测;(3)随着时间推移,以及应用于不同地区人群时,这些评分系统的准确率均出现不同程度地下降^[7]。近年来,医疗无创监护技术和设备的发展,实现了多维度生命体征信息的无创、实时、连续、动态监测,而长时程的连续生命体征信息蕴含着丰富的人体生理、病理状态,也是医疗大数据的重要组成部分^[8]。心率(heart rate, HR)、呼吸频率(respiratory rate, RR)、血压、血氧饱和度(peripheral oxygen saturation, SpO₂)是人体的基本生命体征,衡量人体最关键的功能,若对这些时间序列生命体征数据进行综合分析、利用机器学习技术学习其中蕴含的规律,则能够为疾病的预防、诊断和治疗提供重要的决策支持信息^[9-11]。本研究基于生命体征时间序列数据和机器学习算法构建一组临床预测模型,旨在动态预测创伤伤员发生致死性大出血的风险。

1 资料与方法

1.1 数据来源 本研究数据来源于重症监护医疗信

息(medical information mart for intensive care-IV, MIMIC-IV)数据库^[12]和解放军总医院创伤数据库(以下简称创伤数据库),其中MIMIC-IV数据库包含贝斯以色列女执事医疗中心(Beth Israel Deaconess Medical Center)2008—2019年创伤伤员住院期间去隐私化的真实医疗相关信息,由其机构审查委员会进行审查,并免除了后续审查;创伤数据库由2015—2020年进入解放军总医院急诊科“创伤”分诊通道伤员的去隐私化数据组成,包括护理记录、生命体征数据、检验结果、影像学结果以及治疗记录等。其中,分诊级别为I、II级的创伤伤员进入抢救间治疗,III、IV级的创伤伤员进入留观区治疗^[13]。相关数据的使用已获解放军总医院医学伦理委员会审批(S2021-466-01)。

纳入标准: (1)创伤伤员;(2)创伤后首次入院。**排除标准:** (1)年龄<16岁;(2)重型颅脑损伤,即格拉斯哥昏迷评分为3~8分;(3)临床数据不完整。临床数据完整的具体要求为:对于HR、RR、收缩压(systolic blood pressure, SBP)、舒张压(diastolic blood pressure, DBP)、SpO₂ 5项生命体征指标中的任1项,至少有一条数据记录。

1.2 结局变量 以致死性大出血为结局变量。由于致死性大出血的定义尚缺乏权威标准,通过查阅既往文献资料并结合临床实践,将其筛选标准确定为^[14-15]: (1)进行了大量输血,即入院24 h内某1 h内输注红细胞 ≥ 3 U;(2)入院24 h内进行了止血或栓塞手术;(3)入院24 h内死亡。若创伤伤员满足以上3项条件中的任1项,则纳入致死性大出血组,否则纳入非致死性大出血组。

1.3 预测变量 选取HR、RR、SBP、DBP、SpO₂ 5项生命体征指标作为预测变量,基于5项指标的时间序列数据进行模型开发。对于致死性大出血组创

伤伤员，取其最早满足结局变量的时间点为研究截面时间；对于非致死性大出血组创伤伤员，取其入院后 24 h 这一时间点前的最后一条生命体征数据的记录时间为研究截面时间。取研究截面时间前 1~13 h、2~14 h、3~15 h 这 3 个时间区间为 3 个数据提取区间。提取研究对象在 3 个数据提取区间内的生命体征时序数据分别进行模型训练，以预测创伤伤员未来 1、2、3 h 发生致死性大出血的风险。

1.4 模型开发及验证 对数据进行清洗，去除明显错误的数据。采用前向填补或后向填补法对缺失值进行填补。使用 6 种机器学习算法进行模型开发及验证，包括 4 种传统的机器学习方法[逻辑回归(logistic regression, LR)、支持向量机(support vector machine, SVM)、随机森林(random forests, RF)、自适应提升(adaptive boosting, AdaBoost)]和两种深度学习方法[门控循环单元(gated recurrent unit, GRU)、门控循环单元-D(gated recurrent unit-D, GRU-D)]。6 种模型的参数设置详见附表 1(<https://dx.doi.org/10.11855/j.issn.0577-7402.1273.2023.0427FJ>)。采用早停机制避免过拟合，若当前最小验证误差连续 10 轮未更新则停止训练，最大轮数为 1000。深度学习模型使用交叉熵损失函数。使用 MIMIC-IV 数据库中的创伤伤员数据进行模型开发，十折交叉验证进行模型内部验证。使用创伤数据库中进入急诊抢救间的创伤伤员数据对开发的各项模型进行外部验证。最后，基于预测效果最好的模型开发相应的网页计算器，并部署在解放军总医院急诊科信息系统中。

通过准确率、敏感度、特异度、阳性预测值(positive predictive value, PPV)、阴性预测值(negative

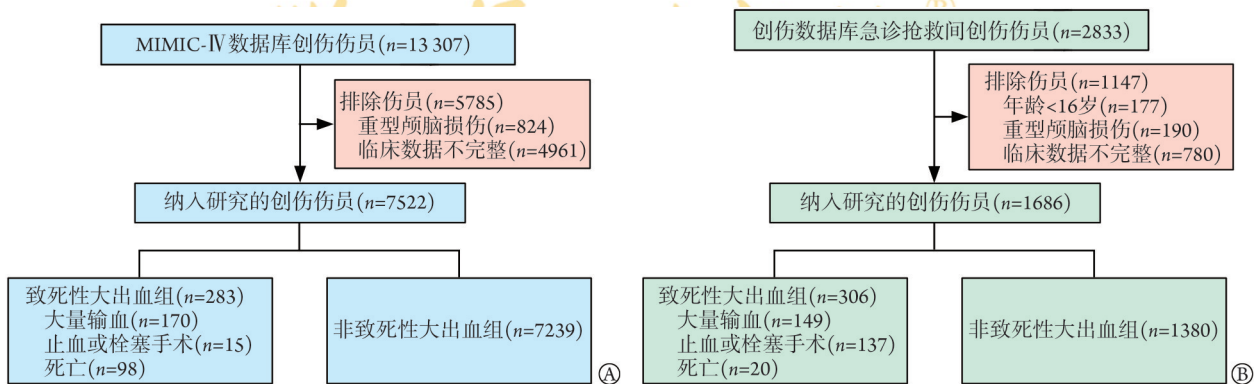
predictive value, NPV)、约登指数及受试者工作特征(receiver operating characteristic, ROC) 曲线下面积(area under the curve, AUC) 评估模型性能。上述指标取值越大，则模型预测效果越好。其中，AUC 是衡量模型综合性能的最佳指标。

1.5 统计学处理 使用 Python 3.8.5 软件进行模型开发及验证，Java JDK 1.8 软件进行网页计算器开发，SPSS 21 软件进行统计分析。计量资料符合正态分布者以 $\bar{x} \pm s$ 表示，不符合正态分布者以 $M(Q_1, Q_3)$ 表示，两组间比较采用 Mann-Whitney U 秩和检验，多组间比较采用 Kruskal Wallis H 秩和检验，进一步两两比较采用 Bonferroni 法；计数资料以例(%)表示，组间比较采用 χ^2 检验、连续校正法或 Fisher 确切概率检验。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 MIMIC-IV 数据库与创伤数据库中创伤伤员基线特征比较 提取 MIMIC-IV 数据库中创伤伤员 13 307 例，最终纳入 7522 例，其中致死性大出血组 283 例，占 3.8%(图 1A)。提取创伤数据库中急诊抢救间创伤伤员 2833 例，最终纳入 1686 例，其中致死性大出血组 306 例，占 18.1%(图 1B)。

MIMIC-IV 数据集中，致死性大出血组与非致死性大出血组创伤伤员的性别($P=0.727$)、年龄($P=0.368$) 差异无统计学意义，而致死性大出血组创伤伤员更易出现 HR($P < 0.001$)、RR($P < 0.001$) 加快，SBP($P < 0.001$)、DBP($P < 0.001$) 及 SpO₂($P < 0.001$) 降低(表 1)。创伤数据集中，致死性大出血组与非致死性大出血组创伤伤员也有类似规律(表 1)。



MIMIC-IV. 重症监护医疗信息数据库；A. MIMIC-IV 数据库；B. 创伤数据库

图 1 创伤伤员筛选流程图

Fig.1 Flow chart for screening trauma victims

2.2 基于 MIMIC-IV 数据集的创伤致死性大出血伤情动态预测模型开发 ROC 曲线分析显示，与其余 5 种模型相比，GRU-D 模型预测创伤伤员未来 1、2、3 h 发生致死性大出血的准确率、敏感度、特异度、

PPV、NPV、约登指数、AUC 最高。GRU-D 模型预测创伤伤员未来 1、2、3 h 发生致死性大出血的 AUC 差异无统计学意义($P=0.905$, 表 2、图 2)。

2.3 基于创伤数据集的创伤致死性大出血伤情动态

表1 MIMIC-IV数据集与创伤数据集中创伤伤员基线特征比较

Tab.1 Comparison of baseline characteristics between MIMIC-IV and trauma datasets of traumatic patients

指标	MIMIC-IV数据集			创伤数据集		
	致死性大出血组 (n=283)	非致死性大出血组 (n=7239)	P	致死性大出血组 (n=306)	非致死性大出血组 (n=1380)	P
男/女(例)	165/118	4145/3094	0.727	240/66	1067/313	0.673
年龄[岁, M(Q ₁ , Q ₃)]	63.0(50.0, 77.5)	63.0(48.0, 76.0)	0.368	47.6(34.2, 58.5)	47.4(32.9, 58.2)	0.642
生命体征[M(Q ₁ , Q ₃)]						
HR(次/min)	94.00(81.00, 107.00)	87.00(74.00, 100.00)	<0.001	98.00(82.00, 115.00)	89.00(76.00, 103.00)	<0.001
RR(次/min)	20.00(16.00, 25.00)	19.00(16.00, 24.00)	<0.001	20.00(18.00, 21.00)	20.00(19.00, 20.00)	<0.001
SBP(mmHg)	117.00(104.00, 133.00)	120.00(106.00, 136.00)	<0.001	114.00(102.00, 127.00)	123.00(110.00, 137.00)	<0.001
DBP(mmHg)	62.00(52.00, 72.00)	64.00(55.00, 75.00)	<0.001	70.00(61.00, 79.00)	75.00(66.00, 83.00)	<0.001
SpO ₂ (%)	97.00(96.00, 100.00)	98.00(96.00, 100.00)	<0.001	97.00(96.00, 99.00)	98.00(97.00, 99.00)	0.002

MIMIC-IV. 重症监护医疗信息数据库; HR. 心率; RR. 呼吸频率; SBP. 收缩压; DBP. 舒张压; SpO₂. 外周血氧饱和度

表2 基于MIMIC-IV数据集开发的6种模型预测创伤伤员未来1、2、3 h发生致死性大出血的效果比较

Tab.2 Comparison of six models developed based on the MIMIC-IV data set for predicting fatal massive hemorrhage in traumatic patients in the next 1, 2, and 3 h

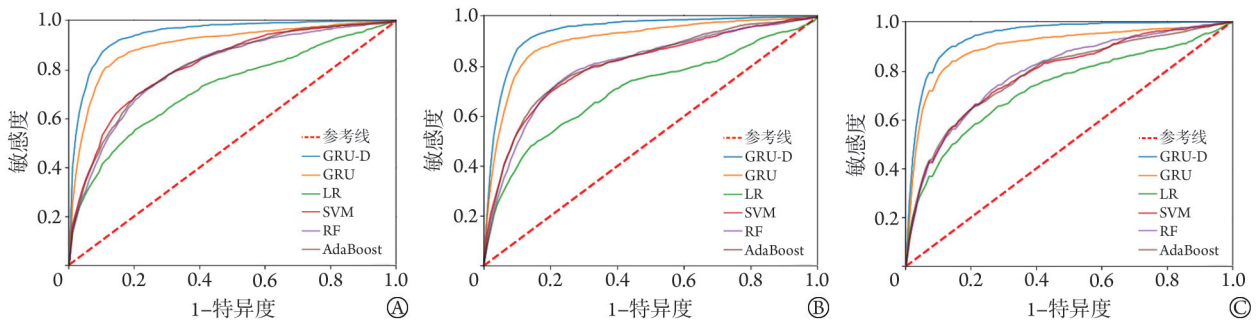
模型	准确率(%)	敏感度(%)	特异度(%)	PPV(%)	NPV(%)	约登指数	AUC
GRU-D							
1 h	89.2±4.0	90.0±6.9	89.1±4.3	59.2±10.1	98.2±1.3	0.790±0.089	0.946±0.029
2 h	89.2±3.7	88.9±7.7	89.2±3.7	59.0±9.6	98.0±1.4	0.782±0.097	0.940±0.032
3 h	88.5±4.2	89.4±8.0	88.3±4.3	57.3±10.0	98.0±1.5	0.777±0.101	0.943±0.034
GRU							
1 h	87.2±3.3	82.5±8.4	88.0±3.4	54.0±7.8	96.8±1.5	0.704±0.096	0.900±0.046
2 h	87.4±4.2	84.4±8.7	87.9±4.5	55.1±9.9	97.1±1.6	0.723±0.105	0.903±0.046
3 h	87.3±4.3	81.7±10.1	88.2±4.3	54.9±10.4	96.7±1.8	0.699±0.120	0.898±0.053
LR							
1 h	68.6±3.8	64.4±9.9	69.3±4.2	26.1±3.9	92.1±2.1	0.337±0.105	0.722±0.60
2 h	60.4±4.5	73.5±10.7	58.2±4.8	22.8±3.3	93.0±2.7	0.317±0.119	0.703±0.073
3 h	67.7±4.3	69.2±11.7	67.5±4.8	26.3±4.2	93.0±2.5	0.366±0.121	0.729±0.075
SVM							
1 h	78.6±3.5	68.3±9.8	80.3±3.9	37.0±5.7	93.9±1.8	0.486±0.101	0.816±0.044
2 h	79.1±3.5	68.3±10.9	80.9±4.2	37.7±5.5	93.9±1.9	0.492±0.106	0.803±0.053
3 h	74.7±4.5	67.3±12.7	75.9±5.3	32.1±5.9	93.4±2.4	0.432±0.127	0.788±0.067
RF							
1 h	79.2±3.4	65.5±11.7	81.5±3.8	37.4±6.1	93.5±2.1	0.470±0.117	0.804±0.055
2 h	80.5±3.8	65.8±12.5	83.0±3.9	39.6±7.1	93.6±2.2	0.488±0.130	0.802±0.067
3 h	78.1±3.6	62.1±11.9	80.8±4.3	35.4±6.0	92.8±2.1	0.429±0.114	0.796±0.063
AdaBoost							
1 h	79.4±3.2	68.0±11.2	81.3±3.6	38.0±5.7	93.9±2.0	0.493±0.111	0.811±0.052
2 h	78.4±3.8	71.1±10.6	79.6±4.3	37.2±6.1	94.3±2.0	0.507±0.109	0.809±0.061
3 h	77.7±4.2	63.0±13.4	80.2±4.6	35.0±7.3	92.9±2.4	0.432±0.137	0.787±0.067

MIMIC-IV. 重症监护医疗信息数据库; PPV. 阳性预测值; NPV. 阴性预测值; AUC. 曲线下面积; LR. 逻辑回归; SVM. 支持向量机; RF. 随机森林; AdaBoost. 自适应提升; GRU. 门控循环单元; GRU-D. 门控循环单元-D

预测模型外部验证 ROC曲线分析结果显示, 与其他5种模型比较, GRU-D模型预测创伤伤员未来1、2、3 h发生致死性大出血的敏感度、NPV、约登指数、AUC最高。GRU-D模型预测创伤伤员未来1、

2、3 h发生致死性大出血的AUC差异无统计学意义(P=0.181, 表3、图3)。

比较GRU-D模型在MIMIC-IV数据集与创伤数据集中的预测效果, 结果显示, 在MIMIC-IV数据集



MIMIC-IV. 重症监护医疗信息数据库; ROC. 受试者工作特征曲线; GRU-D. 门控循环单元-D; GRU. 门控循环单元; LR. 逻辑回归; SVM. 支持向量机; RF. 随机森林; AdaBoost. 自适应提升; A. 1 h; B. 2 h; C. 3 h

图2 基于MIMIC-IV数据集开发的6种模型预测创伤伤员未来1、2、3 h发生致死性大出血的ROC曲线

Fig.2 ROC curves developed with six models based on the MIMIC-IV data set to predict fatal massive bleeding in trauma patients in the next 1, 2, and 3 h

表3 基于创伤数据集开发的6种模型预测创伤伤员未来1、2、3 h发生致死性大出血的效果比较

Tab.3 Comparison of the effects of six models developed based on the trauma data set for predicting fatal massive hemorrhage in traumatic patients in the next 1, 2, and 3 h

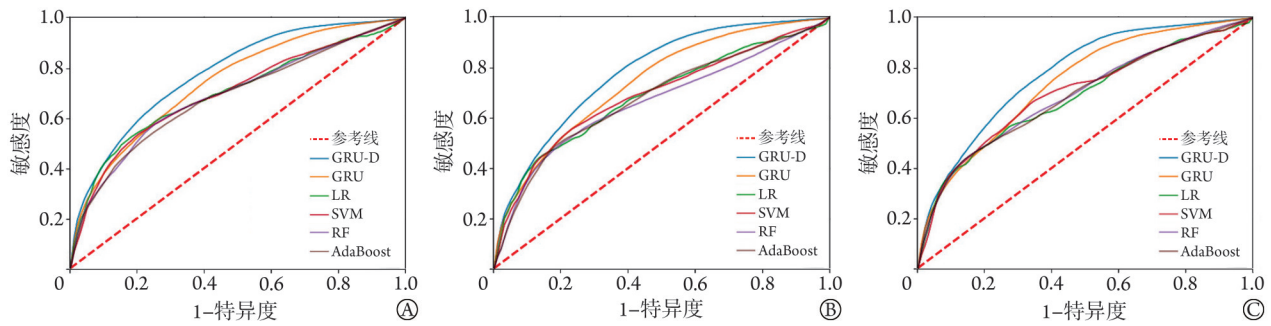
模型	准确率(%)	敏感度(%)	特异度(%)	PPV(%)	NPV(%)	约登指数	AUC
GRU-D							
1 h	58.7±4.3	84.4±5.4	52.6±6.5	29.8±2.1	93.6±1.2	0.370±0.021	0.779±0.013
2 h	58.2±3.9	87.1±4.1	51.6±5.7	29.4±1.6	94.7±1.1	0.387±0.020	0.780±0.008
3 h	57.9±4.8	87.4±5.8	51.7±6.9	27.7±2.3	95.3±1.3	0.391±0.020	0.778±0.009
GRU							
1 h	56.1±3.2	82.4±3.4	49.9±4.7	28.0±1.2	92.4±0.8	0.323±0.022	0.744±0.011
2 h	55.9±3.3	83.5±3.2	49.6±4.8	27.7±1.3	92.9±0.8	0.330±0.024	0.741±0.011
3 h	57.7±4.1	81.8±6.0	52.7±6.1	26.7±1.7	93.4±1.1	0.345±0.022	0.741±0.011
LR							
1 h	67.8±0.8	62.1±0.7	69.2±1.1	32.2±0.6	88.6±0.1	0.312±0.007	0.701±0.002
2 h	60.3±0.7	68.0±0.9	58.6±0.9	27.4±0.4	88.8±0.3	0.265±0.010	0.687±0.003
3 h	63.6±1.0	60.6±1.2	64.3±1.3	26.1±0.5	88.7±0.2	0.249±0.009	0.684±0.004
SVM							
1 h	72.3±0.9	56.7±1.0	76.0±1.3	35.8±1.0	88.2±0.2	0.327±0.009	0.698±0.002
2 h	72.6±0.7	55.8±1.0	76.5±1.1	35.3±0.8	88.3±0.2	0.323±0.007	0.690±0.003
3 h	71.1±0.8	56.3±1.4	74.1±1.2	31.3±0.6	89.1±0.2	0.305±0.007	0.699±0.002
RF							
1 h	74.4±1.7	49.3±5.8	80.3±3.2	37.2±2.0	87.1±1.0	0.296±0.036	0.690±0.004
2 h	75.7±0.6	47.0±2.4	82.3±1.2	38.0±0.9	87.1±0.4	0.293±0.016	0.664±0.006
3 h	76.7±1.0	46.1±1.7	83.1±1.4	36.4±1.6	88.1±0.3	0.292±0.014	0.691±0.004
AdaBoost							
1 h	72.6±2.2	51.3±4.1	77.6±3.5	35.3±2.2	87.2±0.6	0.289±0.020	0.684±0.004
2 h	72.6±0.7	52.5±1.1	77.2±1.0	34.7±0.9	87.6±0.2	0.298±0.010	0.679±0.005
3 h	74.4±1.9	49.1±2.5	79.7±2.8	33.7±2.1	88.2±0.3	0.288±0.015	0.687±0.005

PPV. 阳性预测值; NPV. 阴性预测值; AUC. 曲线下面积; GRU-D. 门控循环单元-D; GRU. 门控循环单元; LR. 逻辑回归; SVM. 支持向量机; RF. 随机森林; AdaBoost. 自适应提升

中, GRU-D模型预测创伤伤员未来1、2、3 h发生致死性大出血的AUC分别为0.946±0.029、0.940±0.032、0.943±0.034; 在创伤数据集中, GRU-D模型预测创伤伤员未来1、2、3 h发生致死性大出血的AUC分别为0.779±0.013、0.780±0.008和0.778±0.009。由此可

见, 虽然GRU-D模型在创伤数据集中的AUC略有下降, 但仍保持了较好的区分度, 表明GRU-D模型具有较好的泛化能力。

据此开发了一个公开的网页计算器(<http://82.156.217.249:5006/>), 集成了效果最优的GRU-D模



ROC. 受试者工作特征曲线; GRU-D. 门控循环单元-D; GRU. 门控循环单元; LR. 逻辑回归; SVM. 支持向量机; RF. 随机森林; AdaBoost. 自适应提升; A. 1 h; B. 2 h; C. 3 h

图3 基于创伤数据集开发的6种模型预测创伤伤员未来1、2、3 h发生致死性大出血的ROC曲线

Fig.3 ROC curves developed with six models based on the trauma data set to predict fatal massive hemorrhage in traumatic patients in the next 1, 2, and 3 h

型, 包括数据输入页面(附图1A)及预测结果页面(附图1B)(<https://dx.doi.org/10.11855/j.issn.0577-7402.1273.2023.0427FJ>), 提供给各地区的公众和医护人员使用。为了进一步促进模型的实际应用, 将GRU-D模型部署在解放军总医院急诊科信息系统中, 实现了对于5项生命体征时序数据的自动提取, 以及对于创伤致死性大出血病情的高效、动态、智能预测。

3 讨论

本研究基于大规模公共数据库中创伤伤员生命体征时序数据开发并验证一组深度学习模型, 以动态预测创伤伤员未来3个时间点发生致死性大出血的风险, 并在解放军总医院创伤数据库中进一步验证。此外, 本研究提供了一个开放和可访问的数据接口, 以便医护人员和公众使用和验证模型。最后, 将模型部署在医院信息系统中, 辅助临床医师早期诊断、动态预测创伤致死性大出血。

采用生命体征时序数据开发创伤致死性大出血病情动态预测模型, 具备以下优势: 首先, 由于生命体征数据在院前和院内环境中均容易快速获取, 医护人员可将数据定期记录, 并输入模型进行风险预测, 而不必等待血常规、血凝等实验室检查结果或超声、CT等影像学结果, 有助于提高模型的时效性^[16]。其次, 生命体征指标可无创监测, 且容易重复测量, 使模型可实现实时、连续预测^[9]。最后, 同时对创伤伤员未来3个时间点出现致死性大出血的风险进行预测, 实现了动态预测, 能够提供更多病情变化信息, 便于医疗人员及时调整治疗计划^[10]。

目前, 文献报道中创伤大出血相关的预测模型主要采用大量输血作为结局变量^[17-19]。然而, 大量输血并不能解释所有与大出血相关的临床重要结局。例如, 存在大出血的伤员很快被发现, 并在达到大量输血的标准之前即进行止血或栓塞手术干预; 存

在大出血伤员的创伤情况严重, 并在达到大量输血的标准之前即发生大出血相关死亡。因此, 当单独使用大量输血作为预测模型的结局变量时, 容易出现相互竞争的风险偏倚和生存偏倚^[15]。此外, 大量输血的传统定义为24 h内输注 ≥ 10 U浓缩红细胞, 目前认为该定义已经过时。关键管理阈值(critical administration threshold, CAT), 即受伤后24 h内的任何1 h内输注 ≥ 3 U红细胞, 能够更准确地识别大出血伤员, 最大限度地减少幸存者偏差^[20]。因此, 笔者采用致死性大出血作为结局变量, 将入院后24 h内进行大量输血, 止血、栓塞手术, 或死亡作为结局变量的筛选条件, 强调出血的严重程度而不是输血量, 并采用CAT这一现代定义代替大量输血的传统定义。

基于MIMIC-IV数据集, GRU-D模型预测创伤伤员未来1、2、3 h发生致死性大出血的效果优于GRU模型, GRU模型优于AdaBoost、RF及SVM模型, AdaBoost、RF及SVM模型之间效果无明显差异, 但优于LR模型。造成上述6种模型预测效果差异的原因可能是: LR、SVM、RF和AdaBoost模型属于传统机器学习方法, 输入模型的均为5项生命体征指标在入院后的首次记录, 即一个5维向量。GRU^[21]和GRU-D模型^[22]属于深度学习模型, 输入数据是5项生命体征指标组成的5维向量的序列, 序列长度是创伤伤员生命体征数据的条数。此外, GRU-D模型是基于GRU的一种变体, 其输入除了时序数据本身, 还包括缺失标记和时间间隔信息。GRU-D算法通过学习创伤伤员生命体征时序数据中蕴含的信息, 分析时间间隔信息中隐藏的疾病变化规律, 获得了相较采用单次生命体征数据的传统机器学习方法更好的模型效果。深度学习在临床中的作用迅速增长, 提供了新的工具来增强患者和临床医师的体验, 并可能成为常规临床路径中无处不在且不可或缺的一

部分。

为了便于医护人员使用模型，笔者开发了一个网页计算器，提供了一个用户友好的界面，输入预测变量后，对创伤伤员未来3个时间点发生致死性大出血的概率进行动态预测。该结果可帮助临床决策者了解伤员状况，并及时干预和调整治疗策略。然而，网页计算器虽然可以满足不同国家、地区公众的访问需要，但对于工作负担重、节奏快的急诊医护人员，手动输入5项生命体征指标时序数据仍然较为耗时、复杂，特别是在时间序列较长时，容易产生人为误差，这在很大程度上限制了模型的实际应用。为了降低模型使用的时间、人力成本，笔者将GRU-D模型部署在解放军总医院急诊科信息系统中，能够自动提取5项生命体征指标的时序数据，自动输入模型计算并返回预测结果，最后将预测结果在系统页面中进行可视化展示。创伤致死性大出血伤情动态预测模型作为急诊科计算机辅助临床决策支持系统的一部分，实现了对于创伤致死性大出血的自动、连续、高效、准确、动态预测。

但本研究仍存在一定局限性：首先，研究人群为成年创伤伤员，没有基于年龄对研究人群进行分层；第二，创伤致死性大出血动态预测模型只能指导医疗人员的临床决策过程，不能取代其临床判断和其他诊断试验；最后，本研究为回顾性观察性研究，未来仍需进行前瞻性研究加以验证。未来可开展亚组研究，进一步明确使用该组模型是否可减少实施大规模输血方案或手术干预前的等待时间。

综上所述，本研究基于生命体征时序数据和GRU-D深度学习算法开发并验证了一组动态预测模型，以辅助创伤致死性大出血伤情的早期诊断和动态预测。基于该组模型开发了相应的网页计算器，并将其部署在解放军总医院急诊科信息系统中，促进了模型的临床应用。

【参考文献】

- [1] Rhee P, Joseph B, Pandit V, *et al.* Increasing trauma deaths in the United States[J]. *Ann Surg*, 2014, 260(1): 13-21.
- [2] Spahn DR, Bouillon B, Cerny V, *et al.* The European guideline on management of major bleeding and coagulopathy following trauma: fifth edition[J]. *Crit Care*, 2019, 23(1): 98.
- [3] Zhao Y, Jia L, Jia R, *et al.* A new time-window prediction model for traumatic hemorrhagic shock based on interpretable machine learning[J]. *Shock*, 2022, 57(1): 48-56.
- [4] Yucel N, Lefering R, Maegele M, *et al.* Trauma associated severe hemorrhage (TASH)-score: probability of mass transfusion as surrogate for life threatening hemorrhage after multiple trauma[J]. *J Trauma*, 2006, 60(6): 1228-1236.
- [5] Nunez TC, Voskresensky IV, Dossett LA, *et al.* Early prediction of massive transfusion in trauma: simple as ABC (assessment of blood consumption)?[J]. *J Trauma*, 2009, 66(2): 346-352.
- [6] Rainer TH, Ho AMH, Yeung JHH, *et al.* Early risk stratification of patients with major trauma requiring massive blood transfusion[J]. *Resuscitation*, 2011, 82(6): 724-729.
- [7] Maegele M, Lefering R, Wafaisade A, *et al.* Revalidation and update of the TASH-Score: a scoring system to predict the probability for massive transfusion as a surrogate for life-threatening haemorrhage after severe injury[J]. *Vox Sang*, 2011, 100(2): 231-238.
- [8] Orphanidou C. A review of big data applications of physiological signal data[J]. *Biophys Rev*, 2019, 11(1): 83-87.
- [9] Convertino VA, Moulton SL, Grudic GZ, *et al.* Use of advanced machine-learning techniques for noninvasive monitoring of hemorrhage[J]. *J Trauma*, 2011, 71(1 Suppl): S25-S32.
- [10] Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards[J]. *Resuscitation*, 2016, 102: 1-5.
- [11] Barton C, Chettipally U, Zhou Y, *et al.* Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs[J]. *Comput Biol Med*, 2019, 109: 79-84.
- [12] Johnson A, Bulgarelli L, Pollard T, *et al.* MIMIC-IV (version 0.4) [J]. *PhysioNet*, 2020. doi: 10.13026/a3wn-hq05.
- [13] 中华护理学会急诊专业委员会, 浙江省急诊医学质量控制中心. 急诊预检分级分诊标准[J]. *中华急诊医学杂志*, 2016, 25(4): 415-417.
- [14] Cantle PM, Cotton BA. Prediction of massive transfusion in trauma [J]. *Crit Care Clin*, 2017, 33(1): 71-84.
- [15] Tran A, Matar M, Steyerberg EW, *et al.* Early identification of patients requiring massive transfusion, embolization, or hemostatic surgery for traumatic hemorrhage: a systematic review protocol[J]. *Syst Rev*, 2017, 6(1): 80.
- [16] Liu J, Khitrov MY, Gates JD, *et al.* Automated analysis of vital signs to identify patients with substantial bleeding before hospital arrival: a feasibility study[J]. *Shock*, 2015, 43(5): 429-436.
- [17] El-Menyar A, Mekkodathil A, Abdelrahman H, *et al.* Review of existing scoring systems for massive blood transfusion in trauma patients: where do we stand?[J]. *Shock*, 2019, 52(3): 288-299.
- [18] Sharma A, Naga Satish U, Tevatia MS, *et al.* Prehospital shock index, modified shock index, and pulse pressure heart rate ratio as predictors of massive blood transfusions in modern warfare injuries: a retrospective analysis[J]. *Med J Armed Forces India*, 2019, 75(2): 171-175.
- [19] Shih AW, Al Khan S, Wang AYH, *et al.* Systematic reviews of scores and predictors to trigger activation of massive transfusion protocols [J]. *J Trauma Acute Care Surg*, 2019, 87(3): 717-729.
- [20] Savage SA, Sumislawski JJ, Zarzar BL, *et al.* The new metric to define large-volume hemorrhage: results of a prospective study of the critical administration threshold[J]. *J Trauma Acute Care Surg*, 2015, 78(2): 224-229.
- [21] Gao S, Zheng Y, Guo X. Gated recurrent unit-based heart sound analysis for heart failure screening[J]. *Biomed Eng Online*, 2020, 19(1): 3.
- [22] Che Z, Purushotham S, Cho K, *et al.* Recurrent neural networks for multivariate time series with missing values[J]. *Sci Rep*, 2018, 8(1): 6085.

(责任编辑: 纪方方)