基于机器学习的创伤伤员检伤分类预测模型构建及验证

张睿智[1,2]，罗瑞虹[3]，卢志林[3]，李春平[3]，卢 兵[1,2]，邢家溢[1,2]，黎檀实[2]
[1]解放军医学院，北京 100853；[2]解放军总医院第一医学中心急诊医学科，北京 100853；[3]清华大学软件学院，北京 100084

摘要：背景 创伤现场批量伤员的检伤分类是现场急救中的关键环节，探索如何更加高效准确地对伤员进行检伤分类具有重要意义。**目的** 基于生命体征数据和机器学习算法建立并验证创伤伤员检伤分类预测模型。**方法** 回顾性分析美国创伤数据库 2017—2019 年的院前急救创伤伤员数据，采用支持向量机 (support vector machine, SVM)、随机森林 (random forest, RF)、梯度提升决策树 (gradient boosting decision tree, GBDT)、极端梯度提升 (eXtreme gradient boosting, XGBoost) 和多层感知机 (multi-layer perceptron, MLP) 5 种机器学习算法开发创伤伤员检伤分类预测模型并验证。采用准确率、精准度、召回率、F1 值和 AUC 值 (ROC 曲线下面积) 进行结果评价，使用 ROC 曲线进行可视化，并在解放军总医院第一医学中心急诊创伤数据集中对最优模型结果进行验证。**结果** 共选取伤员数据 24 948 例，基于 ISS 分级标准分为轻伤 9 496 例，中等伤 9 532 例，重伤 5 496 例，危重伤 424 例。ROC 曲线分析显示，相较于其他四种模型，GBDT 算法预测上述 ISS 分级的效能最好，准确率为 82.63%，精确度为 68.21%，召回率为 60.92%，F1 值为 61.91%，AUC 为 90.38%。在解放军总医院第一医学中心急诊创伤数据集中验证 GBDT 模型，准确率为 83.15%，精确度为 77.38%，召回率为 59.89%，F1 值为 55.26%，AUC 为 90.38%。**结论** 本研究成功开发并验证了一组检伤分类机器学习预测模型，未来可应用于创伤伤员现场检伤分类辅助决策。

关键词： 创伤；机器学习；检伤分类；预测模型；急救医学
中图分类号： R641 **文献标志码：** A **文章编号：** 2095-5227(2024)03-0223-07
DOI： 10.12435/j.issn.2095-5227.2024.003
引用本文： 张睿智，罗瑞虹，卢志林，等. 基于机器学习的创伤伤员检伤分类预测模型构建及验证［J］. 解放军医学院学报，2024，45（3）：223-229.

Construction and validation of a machine learning-based predictive model for trauma casualty triage

ZHANG Ruizhi[1,2], LUO Ruihong[3], LU Zhilin[3], LI Chunping[3], LU Bing[1,2], XING Jiayi[1,2], LI Tanshi[2]
[1]Chinese PLA Medical School, Beijing 100853, China; [2]Department of Emergency, the First Medical Center, Chinese PLA General Hospital, Beijing 100853, China; [3]Tsinghua University Software School, Beijing 100084, China
Corresponding author: LI Tanshi. Email: lits301@sohu.com

Abstract: **Background** On-site triage of lot victims at the scene of the trauma is an essential link in first aid, and it is important to study how to triage casualty more effectively and accurately. **Objective** To develop and validate a predictive model for trauma casualty triage based on vital signs data and machine learning algorithms. **Methods** A retrospective analysis of pre-hospital emergency trauma casualty data from 2017 to 2019 in the National Trauma Data Bank (NTDB) was performed using five types of models, including Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP) to develop and validate the predictive model for trauma casualty detection and classification. The results were evaluated using Accuracy, Precision, Recall, F1 Score and AUC (area under the ROC curve), and visualized using the ROC curve. The results of the optimal model were also validated in the trauma database of the Emergency Department of the First Medical Centre of Chinese PLA General Hospital. **Results** A total of 24 948 records of the injured were collected, including 9 496 cases of mild injuries, 9 532 cases of moderate injuries, 5 496 cases of serious injuries, and 424 cases of critical injuries. Based on the ISS grading criteria, the ROC curve analysis showed that the GBDT algorithm was the most effective compared to the other four models, with an accuracy of 82.63%, a precision of 68.21%, a recall of 60.92%, an F1 value of 61.91%, and an AUC value of 90.38%. In the validation results in the trauma database of the Emergency Department of the First Medical Centre of Chinese PLA General Hospital, the accuracy reached 83.15%, the precision reached 77.38%, the recall reached 59.89%, the F1 value reached 55.26%, and the AUC value reached 90.38%. **Conclusion** We have successfully developed and validated a set of machine learning predictive models for triage of injuries, which can be applied to assist decision-making for on-site triage of trauma injuries in the future.

收稿日期： 2023-10-10
基金项目： 军队课题 (20224282077)
作者简介： 张睿智，男，在读硕士，医师。研究方向：急诊医疗大数据研究。Email: ifz0000@126.com
通信作者： 黎檀实，男，博士，主任医师。Email: lits301@sohu.com

Keywords: trauma; machine learning; triage; predictive modelling; emergency medicine
Cited as: Zhang RZH, Luo RH, Lu ZHL, et al. Construction and validation of a machine learning-based predictive model for trauma casualty triage［J］. Acad J Chin PLA Med Sch, 2024, 45（3）: 223-229.

目前，创伤现场检伤分类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 START (Simple Triage And Rapid Transport) 检伤分类方法[1]、SALT (Sort，Assess，Lifesaving intervention，Treatment/Transport) 检伤分类法[2]、MPTT (Modified Physiological Triage Tool) 检伤分类法[3]等。传统的各种检伤分类方法系统需要使用不同装置手动测量生命体征，需要的人工成本和时间成本较高，且生命体征的动态变化情况无法连续监测。近年来，随着便携式无创生命体征测量技术与设备的发展，对于伤员生命体征信息的连续监测有了极大提升[4]。心率 (heart rate，HR)、呼吸频率 (respiratory rate，RR)、收缩压 (systolic blood pressure，SBP)、舒张压 (diastolic blood pressure，DBP)、血氧饱和度 (peripheral oxygen saturation，SpO_2) 是人体最基本的生命体征，是可以基本判定个人生存状态的基本指标[5]。本研究旨在利用现场创伤伤员生命体征数据进行回顾性分析，通过机器学习方法进行模型构建并在不同算法中选取一组最优模型，将算法融入便携式生命体征监测设备并进行分析，对创伤现场医务人员起到辅助决策的作用，使得检伤分类变得更加高效智能。

1　资料与方法

1.1　数据来源

本研究数据来源于 2017—2019 年美国创伤数据库 (National Trauma Data Bank，NTDB)，NTDB 是一个美国最大的创伤数据登记中心，记录了去隐私化后的伤员基本信息、受伤信息、院前信息、急诊科信息、院内流程信息、现存情况等 10 大类信息[6-7]。纳入标准：(1) 创伤伤员；(2) 创伤后首次入院治疗。排除标准：(1) 年龄、性别、收缩压、心率、呼吸频率、血氧饱和度、体温、格拉斯评分中有任意一项数据缺失；(2) 异常记录数据 (收缩压＞250 mmHg 或＜40 mmHg；体温＜30℃ 或＞43℃；心率＞200/min 或＜20/min；呼吸频率＞60 次/min 或数据缺失；氧饱和度＞100% 或＜70%；身高＞220 cm 或＜140 cm；体质量＞250 kg 或＜40 kg)。

验证数据库来源于解放军总医院第一医学中心急诊创伤数据集，其中记录了 2015—2020 年进入解放军总医院第一医学中心急诊科“创伤”分诊通道伤员的去隐私化数据，包括护理记录、生命体征数据、检验结果、影像学结果和治疗记录等。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。相关数据的使用通过解放军总医院医学伦理委员会审查批准 (伦理批号：S2021-466-01)。本研究按 TRIPOD 规范进行预测模型报告。

1.2　相关定义

预测变量：选取 NTBD 数据库中医务人员到达创伤现场时所记录伤员的性别、年龄、身高、体质量、体温、收缩压、心率、呼吸频率、血氧饱和度和格拉斯哥昏迷评分 (Glasgow coma scale，GCS) 作为模型开发指标。生命体征数据在创伤现场能够快速便捷检测，对维持机体正常活动有重要意义[8]，GCS 评分在判断患者意识情况方面比较客观且被全世界所认可[9]。

结局变量：使用 ISS 分级规则作为结局变量，该规则以解剖学部位为依据，对多发伤的严重程度进行量化评估，作为多发伤严重程度的评判标准，≤16 分为轻伤，17～25 分为中等伤，26～50 分为重伤，＞50 分为危重伤[10-11]。选取美国 NTBD 数据库中所记录的伤员入院完善体格及影像学检查后进行评估的分数作为评分来源。

1.3　模型开发及验证

对数据进行整理清洗，去除明显错误数据及缺失数据，为了减小轻伤患者数据量，并且尽可能拟合轻伤患者的真实情况，将生命体征未在标准区间内 (收缩压 80～120 mmHg；体温 36～37.2℃；心率 60～100/min；呼吸频率 12～24 次/min；氧饱和度 95%～100%；GCS 评分 12～15 分) 的轻症伤员数据进行过滤，使特征更为典型，从而侧面提高对重伤的预测能力。通过这种筛选方式，不仅能够减少轻伤患者数据的数量，降低分析的复杂性，还能确保保留的数据能更好地反映轻伤患者的真实状况。更重要的是通过剔除异常值，可以提高数据集的整体质量，进而提升对重伤患者预测的准确性和可靠性。

使用五种机器学习方法预测检伤分类效果，包括支持向量机 (support vector machine，SVM)[12]、随机森林 (random forest，RF)[13]、梯度提升决策树 (gradient boosting decision tree，GBDT)[14]、极致梯度提升 (eXtreme gradient boosting，XGBoost)[15] 和

多层感知机 (multi-layer perceptron，MLP)[16]，采用五折交叉验证进行试验。将数据集按照数据分成 5 等份，使用其中 4 份数据进行建模训练，剩余 1 份数据用于测试。试验重复 5 次，每份数据轮流作为测试集，得到 5 次测试结果。外部验证集选用解放军总医院第一医学中心急诊创伤数据集，因 2019 年之前较多数据缺失 GCS 评分，所以选取 2019—2020 年所有指标均记录完整且符合纳入标准的数据对效果最优模型进行验证。

1.4 评价指标

采用准确率、精准度、召回率、F1 值和 AUC 值 (ROC 曲线下面积) 进行分析，采用混淆矩阵和 ROC 曲线进行可视化。上述评价指标值使用宏平均 (Macro-average) 计算。指标取值越大，则模型预测效果越好。

1.5 统计学分析

使用 Python 3.7.11 软件进行模型开发及验证，SPSS 21.0 软件进行统计分析。计量资料各组均不符合正态性，以 *M*(*IQR*) 表示，选用非参数检验方法，对于单因素多水平设计计量资料的非参数检验的方法，采用 Kruskal-Wallis 检验，两两比较采用 Dunn 法。计数资料以例数 (百分比) 表示，采用 χ^2 检验、连续校正法或 Fisher 确切概率检验进行差异性分析，以 $P<0.05$ 为差异有统计学意义。ROC 曲线分析不同机器学习模型预测性能，AUC 值用于衡量分类器性能，AUC 值越接近 1，表示分类器性能越好。

2 结果

2.1 基线特征

提取 NTDB 数据库中创伤伤员 1 097 190 例，其中排除不完整数据 758 863 例，剩余指标完整数据 338 327 例，排除异常数据并基于规则的筛选后选出 24 948 例。其中轻伤 9 496 例，中等伤 9 532 例，重伤 5 496 例，危重伤 424 例。美国创伤数据库 (NTDB) 中创伤伤员 0 组表示轻伤伤员 (ISS≤16)，1 组表示中等伤伤员 (ISS 17 ~ 25)，2 组表示重伤伤员 (ISS 26 ~ 50)，3 组表示危重症伤员 (ISS＞50)。其中由 Kruskal-Wallis 检验的结果可知各组年龄、体温、收缩压、心率、血氧饱和度、GCS 评分、ISS 评分的平均值中任何 2 个平均值之间的差异均有统计学意义。呼吸频率 4 组平均值中 0 组分别与 1 组、2 组和 3 组的差异均有统计学意义 (表 1)。

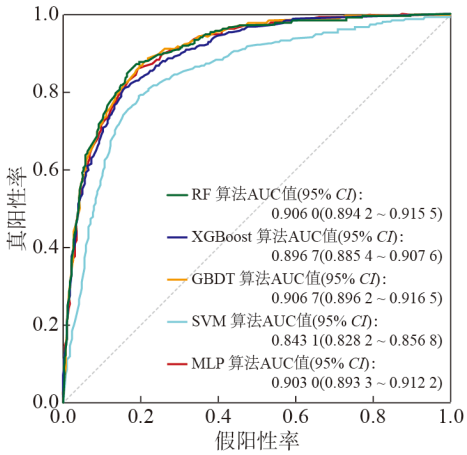
2.2 检伤分类模型开发

从美国 NTBD 数据库选取创伤病例，将伤员按照 ISS 分段划分为具有统计学差异的 4 组后，检伤分类转化为具有 4 个类别的分类变量。基于 ISS 分级标准使用 SVM、RF、XGBoost、GBDT 和 MLP 5 种算法进行试验。对于每一种机器学习模型，采用宏平均 (Macro-average) 方法绘制一条 ROC 曲线来评估四分类模型的性能，ROC 曲线的 X 轴表示模型分类结果的假阳性率，Y 轴表示模型分类结果的真阳性率。每种算法均使用五折交叉验证进行试验后 ROC 曲线结果见图 1，混淆矩阵见图 2，5 种算法结果见表 2。通过对比 5 种算法的试验结果，我们发现 GBDT 算法在准确率、精确度、召回率、F1 值和 AUC 值方面均表现出色。其准确率达 0.826 3，这意味着其能够正确分类 82.63% 的样本；精确度为 0.682 1，说明在所有被预测为正的样本中，真正为正的样本占比为 68.21%；召回率为 0.609 2，说明在所有真正的正样本中，GBDT 算法能够成功识别出 60.92% 的样本；F1 值为 0.619 1，这是精确度和召回率的调和

表 1 NTBD 中创伤伤员数据基线特征对比表

Tab. 1 Comparison table of baseline characteristics of trauma casualty data in NTBD

指标	0组(n=9 496)	1组(n=9 532)	2组(n=5 496)	3组(n=424)	*P*值
性别(男/女)/例	5 048/4 448	6 504/3 028	4 001/1 495	305/119	＜0.001
年龄/[岁, *M*(*IQR*)]	47.00(29.00 ~ 65.00)	54.00(34.00 ~ 70.00)	53.00(31.00 ~ 70.00)	39.00(25.00 ~ 59.00)	＜0.001
身高/[cm, *M*(*IQR*)]	170.00(162.56 ~ 177.80)	172.72(165.10 ~ 180.30)	175.00(167.00 ~ 180.30)	175.26(167.00 ~ 182.00)	＜0.001
体质量/[kg, *M*(*IQR*)]	72.57(61.30 ~ 85.30)	79.38(68.00 ~ 93.00)	79.40(68.00 ~ 92.32)	78.50(66.53 ~ 91.00)	＜0.001
体温/[℃, *M*(*IQR*)]	36.70(36.40 ~ 36.90)	36.60(36.40 ~ 36.80)	36.50(36.20 ~ 36.80)	36.30(35.80 ~ 36.70)	＜0.001
收缩压/[mmHg, *M*(*IQR*)]	110.00(101.00 ~ 115.00)	137.00(122.00 ~ 151.00)	141.00(124.00 ~ 161.00)	115.00(92.00 ~ 138.00)	＜0.001
心率/[次/min, *M*(*IQR*)]	81.00(73.00 ~ 90.00)	88.00 (76.00 ~ 99.00)	90.00(76.00 ~ 110.00)	100.00(76.00 ~ 120.00)	＜0.001
呼吸频率/[次/min, *M*(*IQR*)]	18.00(16.00 ~ 19.00)	18.00 (16.00 ~ 20.00)	18.00(16.00 ~ 22.00)	18.00(14.00 ~ 24.00)	＜0.001
血氧饱和度/[%, *M*(*IQR*)]	98.00(97.00 ~ 99.00)	97.00 (95.00 ~ 98.00)	96.00(93.00 ~ 98.00)	95.00(88.75 ~ 98.00)	＜0.001
GCS评分/*M*(*IQR*)	15.00(15.00 ~ 15.00)	15.00(14.00 ~ 15.00)	13.00(6.00 ~ 15.00)	9.50(3.00 ~ 15.00)	＜0.001
ISS评分/*M*(*IQR*)	5.00(4.00 ~ 9.00)	18.00(17.00 ~ 22.00)	29.00(26.00 ~ 34.00)	66.00(57.00 ~ 75.00)	＜0.001



SMV：支持向量机；RF：随机森林；XGBoost：极致梯度提升；GBDT：梯度提升决策树；MLP：多层感知机。

图 1　基于宏平均的支持向量机、随机森林、极致梯度提升、梯度提升决策树、多层感知机算法 ROC 曲线

Fig.1　ROC curves of SVM, RF, XGBoost, GBDT and MLP based on macro-leverage

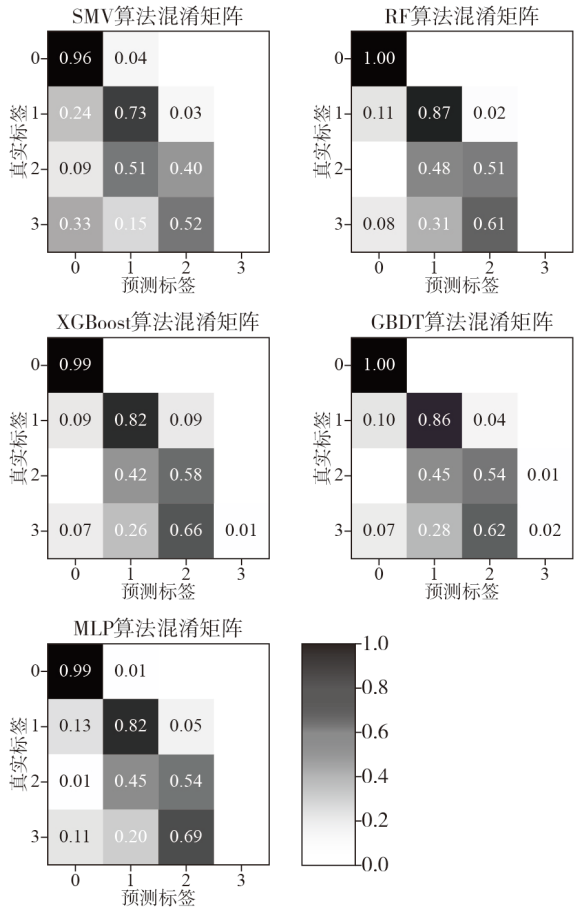


图 2　SVM、RF、XGBoost、GBDT、MLP 算法分别试验结果混淆矩阵

Fig.2　Confusion matrix of experimental results for SVM, RF, XGBoost, GBDT, MLP algorithms

平均数，提供了一个综合评估模型性能的指标；AUC 值为 0.903 8，表明 GBDT 算法在分类任务中具有出色的性能。

表 2　五种机器学习算法基于 ISS 分级标准试验结果对比表

Tab. 2　Comparison of experimental results of ISS grading criteria of five machine learning methods

评价指标	机器学习模型				
	SVM	RF	GBDT	XGBoost	MLP
准确率	0.738 6	0.820 7	0.826 3	0.815 6	0.799 9
精准度	0.557 8	0.619 4	0.682 1	0.660 6	0.591 8
召回率	0.521 7	0.592 4	0.609 2	0.602 1	0.584 2
F1值	0.520 6	0.594 4	0.619 1	0.608 4	0.583 9
AUC值	0.843 1	0.906 0	0.906 7	0.896 7	0.903 0

与其他 4 种模型相比，GBDT 算法在 NTBD 数据库上的分类任务中展现出了更好的综合评价效果。这主要得益于 GBDT 算法在处理复杂数据时能够自动进行特征选择和数据降维，以及通过逐步拟合残差来不断提升模型的预测能力。因此，在未来的研究中，我们可以考虑将 GBDT 算法作为首选的机器学习模型之一。

2.3　GBDT 模型的外部验证

GBDT 是一种串行的集成算法，每次进行抽样都会调整训练样本的权重，在每次训练时都会减少上一次训练的残差，以此更新模型[17]。相较于 RF，GBDT 能灵活地处理连续型、离散型等多种类型数据，容易进行特征选择且模型参数少、调参时间短、预测准确度较高[18]。针对收缩压、心率等预测在 GBDT 模型中的重要程度，进行属性重要性分析 (图 3)，预测变量权重比前 5 名分别为收缩压、GCS 评分、血氧饱和度、心率、体温，其他指标对结局变量影响较小。

通过上述试验可知在检伤分类机器学习系统研究中 GBDT 模型效果最优，进而收集解放军总医院第一医学中心急诊创伤数据集中的数据进行

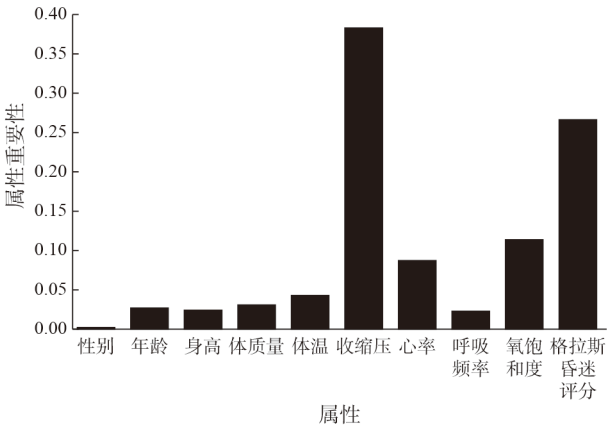


图 3　GBDT 的检伤分类模型中预测变量重要程度对比图

Fig.3　Comparison of the importance of predictor variables in the injury detection classification model of GBDT

进一步验证。纳入标准与排除标准同上，选取 1 098 例数据，其中轻伤 494 例，中等伤 485 例，重伤 55 例，危重伤 64 例，数据基线特征分析见表 3。由 Kruskal-Wallis 检验的结果可以得知，年龄、心率、呼吸频率、收缩压、血氧饱和度、GCS 评分的 P 值均＜0.001，可认为 4 组之间差异有统计学意义。性别与体温的 P 值＞0.05，可认为 4 组平均值之间差异无统计学意义。本研究使用 GBDT 在美国 NTBD 数据库中对 ISS 处于四个不同区间的训练数据进行四分类的模型训练后，按照相同的预处理流程和模型参数设置，基于解放军总医院第一医学中心急诊创伤数据集中的数据对 GBDT 模型进行了处理和分析。

该集使用的是“三区四级”分诊标准[19]，试验根据“四级”伤情级别作为分级标签进行分类，每级都分别考虑假阳性率和真阳性率 (即属于本级，还是属于这个类以外的其他三级)，分别对应 ROC 曲线的 X 轴和 Y 轴，绘制出混淆矩阵 (图 4) 和四个等级分别对应的 ROC 曲线，外加这四个等级的宏平均 (Macro-average) 和微平均 (Micro-average) ROC 曲线 (图 5)。

结果显示，GBDT 模型的准确率达 0.831 5，精确度达 0.773 8，召回率达 0.598 9，F1 值达 0.552 6，AUC 值达 0.903 8，这一结果与先前试验中的各项指标保持一致，尽管在某些指标上仍有提升空间，但整体来看，GBDT 模型展现出了较高的准确性和稳定性。

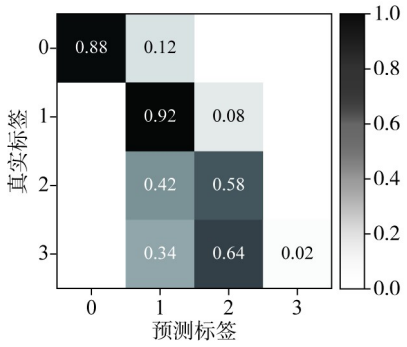


图 4 基于 GBDT 的检伤分类方法外部验证集混淆矩阵

Fig.4 Confusion matrix for external validation set of GBDT-based detection and injury classification methods

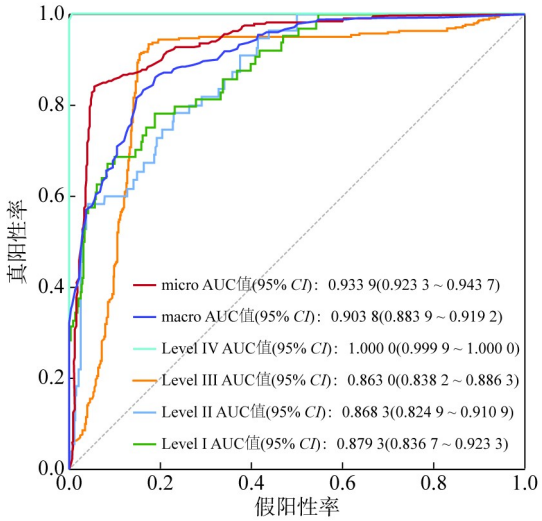


图 5 基于 GBDT 的四级检伤分类方法外部验证集 ROC 曲线

Fig.5 ROC curves for external validation set of GBDT-based detection and injury classification method

3 讨论

现有的创伤现场检伤分类方法大多采用如 START 法等国际上广泛认可的、基于统计学量表的传统手段。这些经过长期实践验证的方法在欧洲及美国等地显示出了一定的可靠性和有效性，为现场救治提供了重要的指导[20-22]。然而，随着科技的飞速发展，特别是人工智能和机器学习算法的崛起，传统的检伤分类方法开始面临新的挑战和机遇。

目前，利用真实数据驱动的机器学习算法构建检伤分类模型尚处于探索的初级阶段[23]。尽管

表 3 解放军总医院第一医学中心急诊创伤数据基线特征对比

Tab. 3 Comparison of baseline characteristics of emergency trauma data from the First Medical Centre of Chinese PLA General Hospital

指标	0组(n=494)	1组(n=485)	2组(n=55)	3组(n=64)	P值
性别(男/女)/例	303/191	300/185	38/17	43/21	0.579
年龄/[岁, $M(IQR)$]	59.00(42.80～71.00)	69.00(54.00～83.00)	61.00(51.00～77.00)	63.00(50.00～78.80)	＜0.001
GCS评分/$M(IQR)$	15.00(15.00～15.00)	15.00(15.00～15.00)	15.00(15.00～15.00)	11.00(3.00～15.00)	＜0.001
体温/[℃, $M(IQR)$]	36.50(36.30～36.80)	36.50(36.20～36.80)	36.50(36.40～36.60)	36.45(36.00～37.00)	0.587
心率/[次/min, $M(IQR)$]	84.00(74.00～93.00)	81.00(71.00～94.00)	81.00(72.00～95.00)	96.00(78.00～121.30)	＜0.001
呼吸频率/[次/min, $M(IQR)$]	20.00(19.00～20.00)	20.00(19.00～20.00)	20.00(19.00～20.00)	20.00(15.30～21.00)	＜0.001
收缩压/[mmHg, $M(IQR)$]	114.00(107.00～118.00)	138.00(128.00～153.00)	182.00(141.00～198.00)	135.00(108.00～150.00)	＜0.001
舒张压/[mmHg, $M(IQR)$]	68.00(64.00～75.00)	80.00(72.00～89.50)	98.00(80.00～107.00)	82.00(64.00～97.00)	＜0.001
血氧饱和度/[%, $M(IQR)$]	98.00(98.00～99.00)	98.00(97.00～98.00)	98.00(98.00～99.00)	95.00(78.50～98.00)	＜0.001

如此，这一新兴技术也展现出了巨大的潜力和优势。既往研究发现，机器学习模型通常能成功准确预测患者的预后，但即使在创伤医学这一单一领域中，模型开发、算法选择、模型验证、性能指标等研究也普遍存在差异[24]。还有研究表明，部分机器学习模型缺乏外部验证，而外部验证是开发有效、适用的机器学习模型的重要步骤[25-26]。相较于传统方法，基于机器学习的检伤分类模型能够在更短的时间内做出判断，且不需要对医护人员进行烦琐的专门培训。在创伤现场医护人员不足或缺乏专业经验的情况下，这类模型能够迅速发挥辅助决策的作用，对伤情进行初步分析，并确定救治的优先级和后送级别。这不仅极大地提高了救治效率，还有助于在关键时刻挽救更多伤员的生命。

本文基于 NTDB 数据库采用了多种机器学习方法进行建模和比较，NTBD 数据库作为美国国家创伤数据库，具有一定的可靠性[27]。其中数据库所使用的 ISS 分级标准也被各国用作创伤中心评审的依据，现已发展为全球公认的创伤评分系统[28]。在此前提下我们开发针对创伤现场的检伤分类模型，试验结果显示，使用 GBDT 算法构建的模型在综合评价方面表现最佳，其性能优于 SVM、RF、XGBoost 和 MLP等其他常用模型。为了进一步验证模型的稳定性和可靠性，我们将模型应用于解放军总医院第一医学中心急诊创伤数据集并进行了外部验证。验证结果显示，模型在外部数据集上的表现与在测试集上的表现相差不大，这表明模型具有较好的泛化能力和实际应用价值。

此外，开发的模型能够结合可穿戴式设备采集的收缩压、心率、呼吸频率、体温等关键生命体征数据，对伤员的伤情进行快速而准确的评估。生命体征数据在院现场即可通过专业的采集设备快速获取，这意味着在尚未获得试验室检查结果或影像学结果之前，医护人员就能够对伤员的情况进行初步的分析和判断。这有助于提高伤情评估的时效性，使得救治工作能够更迅速、有效地开展，且监测过程无创、易于重复进行，这使得模型能够在不同时间点和不同情境下对伤员的伤情进行持续的监测和评估，从而实现更加准确和个性化的预测[29]。

为了方便医护人员和公众在实际场景中使用和验证这些模型，我们还将模型输出为易于部署的 exe 格式，并使其能够兼容多种设备终端。通过简单的方式手动输入生命体征数据，用户即可获得模型对伤情的初步分析结果和建议。这一创新举措不仅极大地提高了模型的实用性和可及性，还为现场救治工作提供了强有力的技术支持。

本研究存在以下局限性。(1) 数据来源的局限性：本研究基于美国创伤数据库，研究人群为成年创伤伤员，未基于年龄、性别对研究人群进行分层，可能存在地域和文化偏见。虽然进一步采用解放军总医院第一医学中心急诊创伤数据集进行验证，但未来研究仍可以考虑多中心、多国际化的数据来源，以提高模型的泛化能力。(2) 算法选择的局限性：虽然 GBDT 在本研究中表现最佳，但机器学习领域不断发展，未来可能有更先进的算法出现。研究可以持续关注算法进展，并尝试将新算法应用于检伤分类问题。智能化检伤分类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。(3) 模型应用的局限性：模型目前主要基于回顾性数据构建，其在实际前瞻性应用中的效果尚需验证。此外，模型的应用场景和使用者也需要进一步明确和界定。

综上所述，本研究在美国创伤数据库中创伤伤员生命体征数据资源支持下，结合 GBDT 机器学习算法，成功开发并验证了一组高效、准确的预测模型。这些模型未来有望在创伤检伤分类领域展现出强大的辅助决策能力，为医护人员提供更加科学、便捷的决策支持。为进一步推动模型的实际应用，我们还基于这组模型精心设计了相应的软件程序。这一创新性的软件程序不仅简化了模型的使用流程，还提高了模型的实用性和可操作性，使得更多的医护人员能够轻松地将这些先进的预测模型应用于实际的创伤救治工作中。通过这一系列的努力，我们期望能够为创伤救治领域带来更加精准、高效的辅助工具，从而提升救治水平，挽救更多的生命。

作者贡献 张睿智：论文构思，收集数据，撰写初稿；罗瑞虹、卢志林：论文构思，模型构建，论文修订；卢兵、邢家溢：收集、整理数据，临床问题咨询；黎檀实：资源提供，论文审阅修订；李春平：论文框架，项目管理及监督指导，论文审阅修订。

利益冲突 所有作者声明无利益冲突。

数据共享声明 本论文相关数据可依据合理理由从作者处获取。Email：ifz0000@126.com。

参考文献

1 Schultz CH，Koenig KL，Noji EK. A medical disaster response to reduce immediate mortality after an earthquake［J］. N Engl J Med，1996，334（7）：438-444.

2 SALT mass casualty triage：concept endorsed by the American College of Emergency Physicians，American College of Surgeons Committee on Trauma，American Trauma Society，National Association of EMS Physicians，National Disaster Life Support Education Consortium，and State and Territorial Injury Prevention Directors Association［J］. Disaster Med Public Health Prep，2008，2（4）：245-246.

3 Vassallo J，Smith J，Bouamra O，et al. The civilian validation of the Modified Physiological Triage Tool (MPTT)：an evidence-based approach to primary major incident triage［J］. Emerg Med J，2017，34（12）：810-815.

4 Chen H，Lundberg SM，Erion G，et al. Forecasting adverse surgical events using self-supervised transfer learning for physiological signals［J］. NPJ Digit Med，2021，4（1）：167.

5 Guo CY，Tian ML，Gong MH，et al. Development and validation of a dynamic prediction model for massive hemorrhage in trauma［J/OL］. https://doi.org/10.1155/2022/9438159.

6 Hartka T，Weaver A，Sochor M. Breadth of use of The Abbreviated Injury Scale in The National Trauma Data Bank bank［J］. Traffic Inj Prev，2022，23（sup1）：S219-S222.

7 Tsiklidis EJ，Sims C，Sinno T，et al. Using the National Trauma Data Bank (NTDB) and machine learning to predict trauma patient mortality at admission［J］. PLoS One，2020，15（11）：e0242166.

8 Areia C，Biggs C，Santos M，et al. The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients：a systematic review and meta-analysis［J］. Crit Care，2021，25（1）：351.

9 Bodien YG，Barra A，Temkin NR，et al. Diagnosing level of consciousness：the limits of the Glasgow Coma scale total score［J］. J Neurotrauma，2021，38（23）：3295-3305.

10 Alam A，Gupta A，Gupta N，et al. Evaluation of ISS，RTS，CASS and TRISS scoring systems for predicting outcomes of blunt trauma abdomen［J］. Pol Przegl Chir，2021，93（2）：9-15.

11 Lovely R，Trecartin A，Ologun G，et al. Injury Severity Score alone predicts mortality when compared to EMS scene time and transport time for motor vehicle trauma patients who arrive alive to hospital［J］. Traffic Inj Prev，2018，19（sup2）：S167-S168.

12 Huang S，Cai N，Pacheco PP，et al. Applications of support vector machine (SVM) learning in cancer genomics［J］. Cancer Genom Proteom，2018，15（1）：41-51.

13 Zheng JX，Xia S，Lv S，et al. Infestation risk of the intermediate snail host of Schistosoma japonicum in the Yangtze River Basin：improved results by spatial reassessment and a random forest approach［J］. Infect Dis Poverty，2021，10（1）：74.

14 Seto H，Oyama A，Kitora S，et al. Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data［J］. Sci Rep，2022，12（1）：15889.

15 Wang RR，Zhang J，Shan BY，et al. XGBoost machine learning algorithm for prediction of outcome in aneurysmal subarachnoid hemorrhage［J］. Neuropsychiatr Dis Treat，2022，18：659-667.

16 Zhang RZ，Wang LJ，Cheng SL，et al. MLP-based classification of COVID-19 and skin diseases［J］. Expert Syst Appl，2023，228：120389.

17 Hancock JT，Khoshgoftaar TM. CatBoost for big data:an interdisciplinary review［J］. J Big Data，2020，7（1）：94.

18 Sense F，Wood R，Collins MG，et al. Cognition-enhanced machine learning for better predictions with limited data［J］. Top Cogn Sci，2022，14（4）：739-755.

19 刘文书. 2018 版专家共识之急诊预检分诊分级标准对急诊危重症患者预后评估价值的研究［D］. 沈阳：中国医科大学，2020.

20 Mehralian G，Pazokian M，Akbari Shahrestanaki Y，et al. Development and validation of SALT Triage method to facilitate the identification and classification of patients in Mass Casualty Incidents［J］. J Inj Violence Res，2023，15（2）：137-146.

21 Tan YT，Shin CKJ，Park B，et al. Pediatric emergency medicine didactics and simulation：JumpSTART secondary triage for mass casualty incidents［J］. Cureus，2023，15（6）：e40009.

22 Bazyar J，Farrokhi M，Salari A，et al. Accuracy of triage systems in disasters and mass casualty incidents；a systematic review［J］. Arch Acad Emerg Med，2022，10（1）：e32.

23 Xu YW，Malik N，Chernbumroong S，et al. Triage in major incidents:development and external validation of novel machine learning-derived primary and secondary triage tools［J］. Emerg Med J，2024，41（3）：176-183.

24 Liu NT，Salinas J. Machine learning for predicting outcomes in trauma［J］. Shock，2017，48（5）：504-510.

25 Zhang T，Nikouline A，Lightfoot D，et al. Machine learning in the prediction of trauma outcomes：a systematic review［J］. Ann Emerg Med，2022，80（5）：440-455.

26 Zadeh FA，Ardalani MV，Salehi AR，et al. An analysis of new feature extraction methods based on machine learning methods for classification radiological images［J/OL］. https://doi.org/10.1155/2022/3035426.

27 Haider AH，Saleem T，Leow JJ，et al. Influence of the National Trauma Data Bank on the study of trauma outcomes：is it time to set research best practices to further enhance its impact?［J］. J Am Coll Surg，2012，214（5）：756-768.

28 Liu JB，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.

29 黄金锋. 基于 PPG 和 ECG 信号的连续无创血压测量研究［D］. 重庆：重庆理工大学，2021.

（责任编辑：施晓亚）