

论著·临床研究

基于机器学习结合非结构化 HER 数据建立
ICU 患者死亡风险预测模型陈琳¹, 何先玲², 费敏艳², 杨攀¹, 邱渝杰^{1△}

(南充市中心医院: 1. 门诊护理组; 2. 心内科, 四川南充 637000)

[摘要] **目的** 构建机器学习模型预测重症监护病房(ICU)患者全因死亡的风险。**方法** 基于重症监护医疗信息市场Ⅲ(MIMIC-Ⅲ)数据库, 使用机器学习法将电子病历(EHR)中的结构化和非结构化数据相整合, 创建 ICU 患者的死亡风险预测模型。**结果** 结合结构化和非结构化数据的机器学习模型提高了 ICU 患者临床结局预测的准确性, 优选出的梯度增强模型的受试者操作特征曲线下面积(AUROC)值为 0.88, 表明可以准确预测患者生命状态。**结论** 采用机器学习模型, 基于少量易于收集的结构化变量结合非结构化数据, 可以显著提高 ICU 患者死亡风险预测模型的预测性能。

[关键词] 非结构化数据; 机器学习; 重症监护病房; 电子病历

DOI: 10.3969/j.issn.1009-5519.2025.01.016

中图法分类号: R619

文章编号: 1009-5519(2025)01-0079-05

文献标识码: A

Mortality risk prediction model for ICU patients based on machine
learning combined with unstructured HER dataCHEN Lin¹, HE Xianling², FEI Minyan², YANG Pan¹, QIU Yujie^{1△}(1. Outpatient Nursing Group; 2. Department of Cardiology, Nanchong
Central Hospital, Nanchong, Sichuan 637000, China)

[Abstract] **Objective** To construct a machine learning model to predict the risk of all-cause mortality in intensive care unit(ICU) patients. **Methods** Based on the intensive care medical information market Ⅲ(MIMIC-Ⅲ) database, the machine learning method was used to integrate the structured and unstructured data in the electronic medical record(EHR) to create a mortality risk prediction model for ICU patients. **Results** The machine learning model combined with structured and unstructured data improved the accuracy of clinical outcome prediction of ICU patients. The AUROC value of the optimized gradient enhancement model was 0.88, indicating that the patient's life state could be accurately predicted. **Conclusion** Using machine learning models, based on a small number of easily collected structured variables combined with unstructured data, can significantly improve the prediction performance of ICU patients' mortality risk prediction models.

[Key words] Unstructured data; Machine learning; Intensive care unit; Electronic medical records

当新型冠状病毒感染大流行席卷世界各地时, 重症监护病房(ICU)的重要性得到了强调。面对日益稀缺的重症护理资源, 需要更积极地研究如何实现资源分配合理规划, 以提供更高效和有效的临床诊疗^[1]。在美国, 每年有超过 500 万患者入住 ICU, 其中 40% 的患者在住院期间死亡, 22% 的患者整个住院期间都在 ICU^[2]。预测 ICU 患者的死亡率是危重症监测和护理研究中最重要任务之一, 其不仅帮助卫生专业人员进行临床决策, 而且可以作为医院资源管理利用的决策基础。ICU 患者需要密切和持续监测, 以防止病情迅速恶化。通过 ICU 设备强化监测, 会持续产生大量医疗记录, 需要高效和准确的数据分析系统^[3]。

电子病历(EHR)是纸质病历的电子版本。过去, 许多研究人员利用 EHR 数据库数据来预测患者的死

亡率、入院时间、疾病诊断、发病等, 以在疾病早期阶段实现疾病预防和干预, 这对危重症护理至关重要。作为重要的风险评估工具, 多种预测模型已在医疗领域得到了开发和应用, 包括序贯器官衰竭评估(SOFA)、简明急性生理学评分(SAPS)和多器官功能障碍评分(MODS), 在临床实践中已被广泛用于预测死亡率^[4-6]。预测模型有助于早期识别死亡风险较高的患者, 并为那些最有可能从识别特定危险因素中获益的患者提供有效的干预措施。多项研究已经确定了数据分析和预测模型能够在诊治过程中帮助到医生和护士, 提高对患者病情进展的警觉性^[7-10]。这些研究中采用的现有统计预测模型主要采用 HER 中结构化数据(定量数据)进行模型构建, 但是非结构化数据(定性数据)占 EHR 数据的 80%^[11], 在对定性数据的

标准化和利用方面至今仍存在明显的挑战。不可否认的是,忽视 EHR 中的定性数据可能会导致医生录入的诊断数据缺失,使得建模时存在关键信息遗漏^[12]。

机器学习是人工智能的一个子领域,其关注如何设计基于经验自动提高输出精度的计算机程序。最近,机器学习在临床医学领域的应用有所增加,包括临床前数据处理、床旁辅助诊断、医学影像辅助诊断、患者分层、治疗决策,以及一级和二级预防的早期预警等^[13]。

尽管非结构化数据是 HER 的重要组成部分,但处理非结构化数据至今仍是一项挑战,耗费着大量的时间和金钱。非结构化数据包含了患者在 ICU 的入院和诊断的信息。在患者入院后,医护人员会第一时间记录观察结果、首发症状和诊断相关数据,本研究结合了 8 个结构化变量和 10 个非结构化变量,以预测 ICU 患者的死亡率。结果表明,半结构化数据包含了许多有价值的信息,可以帮助临床医生做出关键的临床决策。

1 资料与方法

1.1 数据来源 数据来自重症监护医疗信息市场Ⅲ(MIMIC-Ⅲ)数据库,MIMIC-Ⅲ使用了来自 2001—2012 年美国马萨诸塞州波士顿贝斯以色列女执事医

疗中心住院患者的综合临床数据,其包含 46 520 例患者和 58 976 条入院相关数据信息,包括患者生命体征、药物、实验室测量值和观察记录。该数据库的建立和使用已得到美国机构伦理审查委员会的批准,个人身份信息已经被隐匿。

1.2 数据采集和预处理 使用了 MIMIC-Ⅲ数据库中 2010—2012 年从 ICU 收集的数据,针对“ICU 患者死亡风险预测”研究任务,本研究没有局限于特定病种,而是基于所有患者进行死亡风险预测。参考以往研究,选择成年 ICU 患者为研究对象,筛选条件为年龄大于 18 岁,删除年龄小于或等于 18 岁的患者记录;基于患者进入 ICU 后 24 h 的数据进行死亡风险预测,因此删除在 ICU 停留时间小于 24 h 的患者记录,最终共得到 3 814 条入院记录。

在具体的变量选择过程中,结构化数据使用了患者记录中的 8 个基本生命体征作为变量的预测因子,包括心率、呼吸频率、收缩压、舒张压、体温、血氧饱和度、格拉斯哥评分(GCS)评分、年龄。非结构化数据是从患者入院时医生做出的初始诊断中进行人工提取的。本研究中患者的平均年龄为 65 岁,其中 56.03%为男性,平均住院时间为 9.39 d,平均在 ICU 的住院时间为 4.98 d,见表 1。本研究选择的 8 个结构化变量的统计分析数据见表 1。

表 1 纳入模型的特征

项目	总计	存活患者	死亡患者
总例数[n(%)]	3 814(100.00)	3 373(88.44)	441(11.56)
性别[男,n(%)]	2 137(56.03)	1 905(89.14)	232(10.86)
住院时间[M(Q ₁ ,Q ₃),d]			
医院	9.39(5.36,12.42)	9.27(4.67,11.34)	10.98(4.11,11.54)
ICU	4.98(1.55,4.79)	4.39(1.98,4.20)	6.57(1.90,5.89)
结构化变量(入院后 24 h 内, $\bar{x}\pm s$)			
心率(X ₁ ,次/分)	84.93±13.82	84.18±13.59	90.11±15.95
呼吸频率(X ₂ ,次/分)	18.50±4.31	18.92±4.21	18.75±4.76
舒张压(X ₃ ,mm Hg)	60.10±11.58	60.11±12.46	57.39±13.40
收缩压(X ₄ ,mm Hg)	119.45±19.23	119.93±17.39	114.50±18.37
体温(X ₅ ,℃)	98.17±2.10	98.10±1.89	97.20±2.54
血氧饱和度(X ₆ ,%)	97.18±2.00	97.20±1.79	96.20±2.71
GCS 评分(X ₇ ,分)	12.09±3.10	12.66±2.89	9.71±3.89
年龄(X ₈ ,岁)	64.98±16.53	63.19±17.58	71.10±15.49

注:1 mm Hg=0.133 kPa。

针对非结构化数据,本研究利用 Doccano 工具进行数据标注,共有 5 名护士参与数据标注任务,其中 2 名为工作时间大于 6 年的高年资护士,均为本科学位。数据标注流程为先由高年资护士制定数据标注标准和质量控制规范,并标注 300 例示例数据;然后

经过培训的护士以“双人标注,双人核查”的方式标注剩余数据。对标注不一致的数据,由高年资护士进行最终决策,保证数据标注的准确性和规范性。本研究基于患者 EHR 记录中诊断栏进行关键词人工评估,将关键词分类为不同的 10 个主题,形成 10 个非结构化变

量,出现在关键词字集中的任何单词都与主题相关。主题建模选择的 10 个主题和对应的关键词见表 2。

1.3 机器学习 数据被分为两部分,其中 80%的数据用于训练集,其余 20%用于测试集。使用 8 种常用的机器学习算法建立 ICU 死亡率预测模型,包括自适应增强(AdaBoost)、装袋算法(Bagging)、梯度增强

(Gradient Boosting)、光梯度增强机(LightGBM)、逻辑回归(Logistic Regression)、多层感知器(MLP)、支持向量分类(SVC)、极限梯度增强(XGBoost)。所有数据挖掘任务都是使用 Python 编程语言(USA)进行的。8 个机器学习模型及其特定参数的设置见表 3。

表 2 数据集的诊断主题和关键词

非结构化变量主题	关键词
变量 1	冠状动脉疾病 (coronary artery disease) coronari,arteri,diseas,graft,bypass,sda,syndrom,effus,cath,avr,acut,etoh,pericardi,cerebr,pleural,choLANG,mvr,leav,vascular,angioplasty
变量 2	主动脉瓣置换术 (aortic valve replacement) aortic,sda, valv, replac, stenosi, mitral, cancer, subarachnoid, hemorrhag, esophag, procedur, ascend, regurgit,aorta,maze,airway,redo,bental,repair,invas,obstruct
变量 3	心力衰竭 (heart failure) failur,heart,congest,acut,infarct,myocardi,renal,cath,liver,pancreat,elev,cardiac,dehydr,rule,cholecyst,hyperkalemia,leukemia,lacer,hyperglycemia,block,chronic,implant
变量 4	肺炎 (pneumonia) pneumonia,telemetri,fractur,stroke,ischem,atrial,attack,transient,angina,dyspnea,fibril,hip,un-stabl,cath,chronic,segment,diseas,pulm,obst,ablat,cardiomyopathi,pelvic,septal
变量 5	激素依赖性哮喘 (steroid-dependent asthma) sda,right,aneurysm,leav,accid,motor,vehicl,short,breath,tachycardia,cellul,ventricular,abdomin,lung,injuri,hepat,bilater,metastat,perfor,colon,spinal
变量 6	胸痛 (chest pain) pain,chest,hemorrhag,intracrani,fever,abdomin,hypotens,fall,cath,telemetri,dissect,insuffici,stroke,cardiac,femur,strike,epidur,neck,pedestrian,skull,cervic
变量 7	出血性肿块 (bleed mass) bleed,upper,lower,mass,pulmonari,obstruct,bowel,head,brain,weak,emboli,bradycardia,hypertens,stemi,small,edema,hemoptysi,cirrholi,vomit
变量 8	脓毒症 (sepsis) sepsi,infect,asthma,exacerb,copd,sda,urinari,tract,tumor,brain,catheter,overdos,leav,anemia,pyelonephr,syncop,bscess,foot,ulcer,disord
变量 9	硬膜下血肿 (subdural hematoma) hematoma,subdur,respiratori,diabet,seizur,ketoacidosi,trauma,failur,blunt,sda,withdraw,distress,hyponatremia,wind,hernia,remot
变量 10	精神状态改变 (altered mental status) status,mental,alter,arrest,cardiac,carotid,hypoxia,chang,leg,angiogram,stenosi,transplant,kidney,chf,accid,extrem,cerebrovascular,fib,stent,thrombosi,ischemia

表 3 机器学习模型及参数设置

模型	参数
自适应增强	base_estimatOR = DecisionTreeClassifier, random_state = 1, n_estimators = 50, learning_rate = 1.0, algorithm = ‘SAMME.R’
装袋算法	base_estimatOR=None,n_estimators=500,max_samples=100,bootstrap=True,bootstrap_features=False,oob_score=False,warm_start=False,n_jobs=1,random_state=None,verbose=0
梯度增强	n_estimators=100,learning_rate=1.0,max_depth=1,random_state=0
光梯度增强机	boosting_type=‘gbdt’,num_leaves=31,max_depth=-1,learning_rate=0.1,n_estimators=100,subsample_for_bin=200,000,min_child_samples=20,subsample=1.0,subsample_freq=0,colsample_bytree=1.0,reg_alpha=0.0,reg_lambda=0.0 n_jobs=-1,importance_type=‘split’
逻辑回归	solver=‘sag’,penalty=‘l2’,max_iter=‘max_iter’
多层感知器	solver=‘adam’,alpha=1e-5,hidden_layer_sizes=(13,13,13),max_iter=1 000
支持向量分类	C=1.0,kernel=‘rbf’,degree=3,gamma=‘auto’,coef0=0.0,shrinking=True,probability=False,tol=0.001,cache_size=200,class_weight=None,verbose=False,max_iter=-1
极限梯度增强	n_estimators=100,booster=‘gbtree’,eta=0.3,min_child_weight=1,max_depth=3,gamma=0,max_delta_step=0,subsample=1,colsample_bytree=1,colsample_bylevel=1,lambda=1,learning_rate=0.1,n_jobs=1,base_score=0.5,max_delta_step=0,min_child_weight=1

2 结 果

2.1 ICU 患者死亡率预测 使用 k 折交叉验证方法来评估训练后模型的性能,使用 10 折交叉验证构建。根据入院后 24 h 内收集的数据,预测患者入院后 3、30、365 d 的死亡率。预测 3 d 死亡率的受试者操作特征曲线下面积(AUROC)可超过 85%,30、365 d 的死亡率可超过 75%。本研究中的最佳 AUROC 为 87.90%,可以在入院 24 h 时准确预测患者入院后 3 d 的死亡。梯度增强是预测所有时间段 ICU 患者死亡率的最佳模型。见表 4。

本研究还总结了 4 种不同指标(精准度、召回率、F1 分数、准确度)的结果,以更全面地了解不同机器学习方法构建模型的预测准确性差异。在所使用的算法中,3 d 死亡率预测模型极限梯度增强的预测准

确度最高(97.13%),其次是光梯度增强机(96.61%),装袋算法的预测召回率最高(95.23%)。见表 5。

表 4 不同机器学习模型的 AUROC($\bar{x} \pm s$)

模型	3 d	30 d	365 d
自适应增强	0.818 6±0.007 3	0.753 1±0.006 6	0.761 9±0.011 2
装袋算法	0.871 3±0.005 9	0.734 4±0.009 8	0.762 5±0.004 8
梯度增强	0.879 0±0.011 9	0.781 5±0.007 3	0.735 4±0.009 1
光梯度增强机	0.846 1±0.014 3	0.778 0±0.011 8	0.770 5±0.003 5
逻辑回归	0.829 8±0.010 9	0.731 8±0.010 2	0.740 2±0.005 0
多层感知器	0.867 9±0.016 8	0.739 3±0.011 2	0.752 0±0.004 3
支持向量分类	0.814 2±0.008 2	0.751 8±0.011 0	0.757 2±0.004 5
极限梯度增强	0.735 5±0.017 4	0.767 9±0.006 7	0.732 5±0.008 4

表 5 不同模型的精准度、召回率、F1 分数、准确度($\bar{x} \pm s$)

死亡率	模型	评估指标			
		精准度	召回率	F1 分数	准确度
3 d 死亡率	自适应增强	0.175 2±0.013 5	0.875 2±0.018 8	0.292 0±0.013 1	0.862 4±0.013 7
	装袋算法	0.131 4±0.013 3	0.952 3±0.015 7	0.231 9±0.012 3	0.808 5±0.013 6
	梯度增强	0.186 3±0.016 0	0.859 4±0.022 0	0.316 3±0.019 3	0.873 7±0.014 6
	光梯度增强机	0.378 8±0.019 1	0.697 5±0.029 9	0.507 3±0.013 1	0.966 1±0.013 8
	逻辑回归	0.126 4±0.013 3	0.832 4±0.021 3	0.219 4±0.012 7	0.810 8±0.013 4
	多层感知器	0.156 8±0.018 0	0.860 1±0.052 0	0.264 8±0.018 8	0.852 3±0.018 1
	支持向量分类	0.121 3±0.013 8	0.829 7±0.013 7	0.213 7±0.013 2	0.803 6±0.014 8
	极限梯度增强	0.578 6±0.012 1	0.524 3±0.036 4	0.560 1±0.015 7	0.971 3±0.010 5
30 d 死亡率	自适应增强	0.266 5±0.012 6	0.727 5±0.014 5	0.413 1±0.015 0	0.774 0±0.013 2
	装袋算法	0.267 3±0.013 3	0.753 1±0.026 4	0.415 8±0.014 9	0.769 6±0.013 2
	梯度增强	0.292 6±0.010 7	0.733 5±0.019 0	0.418 5±0.013 3	0.778 0±0.012 1
	光梯度增强机	0.341 3±0.013 4	0.675 2±0.026 9	0.453 2±0.017 8	0.822 6±0.012 1
	逻辑回归	0.266 4±0.014 6	0.729 1±0.020 5	0.370 2±0.017 7	0.756 8±0.013 7
	多层感知器	0.268 9±0.019 7	0.752 9±0.049 2	0.375 8±0.015 9	0.768 6±0.021 0
	支持向量分类	0.268 5±0.014 6	0.725 0±0.023 6	0.371 8±0.018 3	0.764 9±0.013 1
	极限梯度增强	0.381 6±0.016 1	0.523 5±0.013 7	0.448 3±0.018 2	0.852 3±0.013 5
365 d 死亡率	自适应增强	0.311 8±0.019 8	0.746 2±0.019 7	0.429 8±0.013 2	0.775 8±0.015 3
	装袋算法	0.315 2±0.015 7	0.768 1±0.013 8	0.437 0±0.016 6	0.776 0±0.012 4
	梯度增强	0.316 5±0.018 3	0.746 3±0.014 1	0.434 5±0.010 7	0.780 1±0.015 7
	光梯度增强机	0.360 2±0.015 6	0.699 7±0.016 8	0.475 7±0.016 1	0.825 3±0.013 2
	逻辑回归	0.262 2±0.015 7	0.738 1±0.019 4	0.408 3±0.017 1	0.757 8±0.012 0
	多层感知器	0.262 6±0.013 7	0.751 8±0.016 5	0.410 7±0.014 6	0.765 7±0.017 9
	支持向量分类	0.266 6±0.015 8	0.731 6±0.015 7	0.413 7±0.016 8	0.769 6±0.014 2
	极限梯度增强	0.374 5±0.013 3	0.565 9±0.013 2	0.464 9±0.016 0	0.852 6±0.013 7

2.2 特征重要性 特征重要性得分反映了在决策树

的构建过程中每个特征所获得的信息^[14]。根据上述

AUROC 分析结果,梯度增强是最佳预测模型,分析了该模型的特征重要性得分。GCS 评分(X_7)、年龄(X_8)和心率(X_1)是 ICU 患者入院 24 h 内所用结构化数据预测死亡率的相对重要变量。针对初始临床诊断记录(非结构化数据),胸痛(变量 6)和冠状动脉疾病(变量 1)也比较显著。在构建模型过程中,出血量(变量 7)是预测 3 d 死亡率的一个相对重要的变量,精神状态改变(变量 10)是预测 30 d 死亡率时的重要变量,脓毒症(变量 8)是预测 365 d 死亡率时的重要变量。无论预测哪个时间段的 ICU 患者死亡率,均需要考虑的重要变量包括 GCS 评分(X_7)、胸痛(变量 6)和冠状动脉疾病(变量 1)。见表 6。

表 6 梯度增强的重要变量

变量重要性	3 d	30 d	365 d
1	变量 6	X_7	X_7
2	X_7	变量 1	变量 1
3	变量 1	变量 6	变量 6
4	变量 7	X_8	X_8
5	变量 8	变量 10	变量 8

3 讨 论

先前的研究侧重于使用 EHR 数据中的定量变量建立预测模型,以预测 ICU 患者的死亡率、住院时间和疾病诊断。然而,由于数据标准化和数据利用方面的挑战,这些研究在很大程度上忽视了定性数据的潜在价值。忽略 EHR 中的非结构化数据,临床医生可能会错过最初观察时提供的关键信息和线索。为了充分利用非结构化数据,本研究采用了人工标注的方法来分 析临床记录。本研究将结构数据(如基本生命体征)与非结构化数据相结合,以预测患者死亡率,这些数据来源于 ICU 入院时医生的初始临床诊断。希望本次分析结果能够增强医务人员对患者状况的了解,优化诊疗和护理资源配置,为患者、家属和医护人员提供更多信息,以便做出明智的决策。

在以往的研究中,研究人员通过使用过多的特征来获得较高的准确性。例如,GONG 等^[9]使用 50 个特征获得了 0.85 的 AUROC,LIU 等^[15]使用 99 个特征获得了 0.78 的 AUROC。然而,在本研究中,仅使用 16 个变量,即 GCS、年龄和患者入住 ICU 时的初步临床记录和诊断,建立的死亡率预测模型就实现了 97% 的准确率和 0.88 的 AUROC。本研究的 AUROC 分析结果显示,ICU 患者入院后的 3 d 死亡率可超过 85%,30 d 和 365 d 的死亡率可超过 75%。本研究还发现,梯度增强算法提供了最准确的预测模型;极限梯度增强具有最高的预测精准度,表明本研究提出的方法可以很好地预测 ICU 患者的死亡率。本研究结果表明,患者入住 ICU 时的临床观察和诊断的初

始书面记录包含了大量有用的信息,可以帮助 ICU 医生和护士做出关键的临床决策。此外,本研究仅使用了 ICU 患者入院 24 h 内的结构化和非结构化数据,这使得预测模型更适合预测短期死亡率,因为建立的 3 d 死亡率预测模型的准确度优于 30 d 和 365 d 死亡率预测模型。

本研究存在一定的局限性,首先,本文中使用的 所有数据都来自一个大型回顾性临床数据库,而这些数据是针对全部患者群体而非特定人群的。为了确保彻底收集相关数据,本研究仅考虑了 MIMIC-III 数据库。未来的研究应该评估从更大地理区域的更多医疗机构收集的数据。由于这项研究仅限于大城市大型医疗机构积累的 ICU 医疗数据,因此这些发现无法安全地应用于小型医疗机构的 ICU 患者。通过将 这些结果与从农村或其他小型医疗机构获得的数据进行比较,可以获得更全面的结果和验证。其次,MIMIC-III 数据库中医生输入的信息包含许多拼写错误、标点符号不一致和语序不规范,使得现有的自然语言分析工具很难处理,也加大了本研究中人工评估 诊断结果的难度。此外,由于本研究是回顾性的,因此无法在真实临床环境中验证预测算法的性能。未来的研究可以更深入地评估和分析这些制约因素,使 这类研究更加客观和彻底。

本研究中,不仅使用了 ICU 患者入院后 24 h 内的结构化数据(包括生命体征测量、GCS 评分和患者入院时年龄),还重点关注了入院时初始状态观察和诊断的非结构化数据。本研究讨论了使用不同的机器学习技术预测 ICU 患者死亡率的有效性,这些非结构化数据包含丰富的信息,可以有效地帮助后期的临床决策。然而,本研究中开发的模型主要侧重于预测 ICU 患者的死亡率,有必要进一步研究其他临床任务,如住院时间、并发症和疾病预测。

参考文献

[1] CAICEDO-TORRES W,GUTIERREZ J. ISeeU: visually interpretable deep learning for mortality prediction inside the ICU[J]. J Biomed Inform,2019,98:103269.

[2] ROMANO' M. The role of palliative care in the cardiac intensive care unit[J]. Healthcare,2019,7(1):30.

[3] 徐良辰,郭崇慧. 基于时间序列特征表示与信息融合的 ICU 患者死亡风险预测[J]. 系统工程理论与实践,2022,42(10):2815-2828.

[4] VINCENT J L,MORENO R,TAKALA J,et al. The SOFA(sepsis-related organ failure assessment) score to describe organ dysfunction/failure, on behalf of the working group on sepsis-related problems of the European society of intensive care medicine[J]. Intensive Care Med,1996,22(7):707-710.

(下转第 87 页)

biol,2016,59(1):41-46.

[4] 孙颖昊,李佳宁,何昆,等.原发性胰腺淋巴瘤临床特征分析[J].中华胰腺病杂志,2024,24(1):52-56.

[5] SADOT E,YAHALOM J,DO R K G,et al. Clinical features and outcome of primary pancreatic lymphoma[J]. Ann Surg Oncol,2015,22(4):1176-1184.

[6] NISHIMURA R,TAKAKUWA T,HOSHIDA Y,et al. Primary pancreatic lymphoma:clinicopathological analysis of 19 cases from Japan and review of the literature[J]. Oncology,2001,60(4):322-329.

[7] 郑杰,杨琰昭.原发性胰腺弥漫性大B细胞淋巴瘤的影像学表现[J].浙江实用医学,2022,27(1):62-65.

[8] TAKAHASHI H,TOMITA N,YOKOYAMA M,et al. Prognostic impact of extranodal involvement in diffuse large B-cell lymphoma in the rituximab era[J]. Cancer, 2012,118(17):4166-4172.

[9] CASTILLO J J,WINER E S,OLSZEWSKI A J. Sites of extranodal involvement are prognostic in patients with diffuse large B-cell lymphoma in the rituximab era: an analysis of the Surveillance, Epidemiology and End Results database[J]. Am J Hematol,2014,89(3):310-314.

[10] MAJOR A,SMITH S M. DA-R-EPOCH vs R-CHOP in DLBCL:how do we choose? [J]. Clin Adv Hematol Oncol,2021,19(11):698-709.

[11] GISSELBRECHT C,GLASS B,MOUNIER N,et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era[J]. J Clin Oncol,2010,28(27):4184-4190.

[12] BEHRNS K E,SARR M G,STRICKLER J G. Pancreatic lymphoma:is it a surgical disease? [J]. Pancreas,1994,9 (5):662-667.

[13] BATTULA N,SRINIVASAN P,PRACHALIAS A,et al. Primary pancreatic lymphoma: diagnostic and therapeutic dilemma[J]. Pancreas,2006,33(2):192-194.

[14] KAWAMOTO K,MIYOSHI H,YOSHIDA N,et al. MYC translocation and/or BCL 2 protein expression are associated with poor prognosis in diffuse large B-cell lymphoma[J]. Cancer Sci,2016,107(6):853-861.

[15] NIITSU N,OKAMOTO M,MIURA I,et al. Clinical significance of 8q24/c-MYC translocation in diffuse large B-cell lymphoma[J]. Cancer Sci,2009,100(2):233-237.

[16] YOON S O,JEON Y K,PAIK J H,et al. MYC translocation and an increased copy number predict poor prognosis in adult diffuse large B-cell lymphoma(DLBCL), especially in germinal centre-like B cell(GCB) type[J]. Histo-pathology,2008,53(2):205-217.

[17] QUESADA A E,MEDEIROS L J,DESAI P A,et al. Increased MYC copy number is an Independent prognostic factor in patients with diffuse large B-cell lymphoma[J]. Mod Pathol,2017,30(12):1688-1697.

[18] SCHIEPPATI F,BALZARINI P,FISOJNI S,et al. An increase in MYC copy number has a progressive negative prognostic impact in patients with diffuse large B-cell and high-grade lymphoma, who May benefit from intensified treatment regimens [J]. Haematologica, 2020, 105 (5): 1369-1378.

(收稿日期:2024-05-21 修回日期:2024-11-10)

(上接第 83 页)

[5] LE G J R. LEMESHOW S,SAULNIER F. A new simplified acute physiology score(SAPS II) based on a European/North American multicenter study[J]. JAMA,1993, 270(24):2957-2963.

[6] BAUE A E,DURHAM R,FAIST E. Systemic inflammatory response syndrome(SIRS), multiple organ dysfunction syndrome(MODS), multiple organ failure(MOF): are we winning the battle? [J]. Shock, 1998, 10 (2): 79-89.

[7] IBRAHIM Z M,WU H,HAMOUD A,et al. On classifying sepsis heterogeneity in the ICU: insight using machine learning[J]. J Am Med Inform Assoc,2020,27(3): 437-443.

[8] DARABI S,KACHUEE M,FAZELI S Y,et al. TAPER: time-aware patient EHR representation[J]. IEEE J Biomed Health Inform,2020,24(11):3268-3275.

[9] GONG M G,PAN K,XIE Y,et al. Preserving differential privacy in deep neural networks with relevance-based adaptive noise imposition[J]. Neural Netw, 2020, 125: 131-141.

[10] LORETO M,LISBOA T,MOREIRA V P. Early prediction of ICU readmissions using classification algorithms [J]. Comput Biol Med,2020,118:103636.

[11] HASHIR M,SAWHNEY R. Towards unstructured mortality prediction with free-text clinical notes[J]. J Biomed Inform,2020,108:103489.

[12] TOOTOONI M S,PASUPATHY K S,HEATON H A, et al. CCMapper: an adaptive NLP-based free-text chief complaint mapping algorithm [J]. Comput Biol Med, 2019,113:103398.

[13] ADLUNG L,COHEN Y,MOR U,et al. Machine learning in clinical decision making[J]. MED,2021,2(6):642-665.

[14] UPADHYAY D,MANERO J,ZAMAN M,et al. Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids[J]. IEEE Transact Network Service Manag, 2021, 18 (1): 1104-1116.

[15] LIU D Y,WU Y L,LI X,et al. Medi-Care AI: predicting medications from billing codes via robust recurrent neural networks[J]. Neural Networks,2020,124:109-116.

(收稿日期:2024-07-01 修回日期:2024-11-03)