文章编号:1671-9352(2026)01-0026-10 DOI:10.6040/j.issn.1671-9352.8.2024.009

基于深度神经网络的重症监护室脓毒症患者死亡风险预测模型分析

余雷[1],孙懿[2],华金铭[2],李腊全[3]
(1.重庆医科大学第二附属医院急救部, 重庆 400010; 2.重庆邮电大学国际学院, 重庆 400065; 3.重庆邮电大学理学院, 重庆 400065)

摘要:提出基于变量选择网络(variable selection networks, VSN)和门控残差网络(gated residual networks, GRN)相结合的深度神经网络(deep neural network, DNN)模型,用于预测重症监护室(intensive care unit, ICU)脓毒症患者 30 d 内的死亡风险,并对模型的可解释性进行深入分析。在重症医学数据库中利用随机森林算法筛选 43 个重要特征,利用本文提出的模型评估死亡风险,并采用移除再训练(remove and retrain, ROAR)方法选出一种最佳的可解释性方法对结果进行解释。测试结果显示,本文提出的模型的预测性能优于其他机器学习模型,受试者工作特征曲线面积(area under receiver operating characteristic curve, AUROC)为 0.967。利用 ROAR 方法分析中,相关性分数逐层传播(layer-wise relevance propagation, LRP)方法的AUROC从 0.967 下降到 0.828。利用 LRP 方法对本文提出的模型进行可解释性分析后,确定查尔森合并症评分为最重要的特征,同时器官衰竭评分、年龄、呼吸频率也对重症监护室脓毒症患者的死亡风险有较大影响。

关键词:脓毒症;死亡风险预测;深度神经网络;移除再训练;相关性分数逐层传播
中图分类号:TP391 **文献标志码**:A
引用格式:余雷,孙懿,华金铭,等. 基于深度神经网络的重症监护室脓毒症患者死亡风险预测模型分析[J]. 山东大学学报(理学版),2026,61(1):26-35.

Analysis of the prediction model based on deep neural networks for mortality risk prediction for sepsis patients in intensive care units

YU Lei[1], SUN Yi[2], HUA Jinming[2], LI Laquan[3]
(1. Emergency Department, The Second Affiliated Hospital of Chongqing Medical University, Chongqing 400010, China; 2. International College, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 3. School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: A deep neural network (DNN) model is proposed by integrating variable selection networks (VSN) and gated residual networks (GRN) to predict the 30-day mortality risk of sepsis patients in the intensive care unit (ICU) and to conduct an in-depth interpretability analysis. In the critical care medical database, 43 significant features are selected using a random forest algorithm, and the proposed model is employed to evaluate mortality risk. The remove and retrain (ROAR) method is utilized to determine the optimal interpretability approach for explaining the results. Testing outcomes indicate that the proposed model outperforms other machine learning models, achieving an area under the receiver operating characteristic curve (AUROC) of 0.967. In the ROAR analysis, the AUROC of the layer-wise relevance propagation (LRP) method decreases from 0.967 to 0.828. Through interpretability analysis of the proposed model using LRP, the Charlson comorbidity score is identified as the most critical feature. In contrast, the organ failure score, age, and respiratory rate also have a pronounced impact on the mortality risk of ICU sepsis patients.

Key words: sepsis; mortality risk prediction; deep neural network; remove and retrain; layer-wise relevance propagation

收稿日期:2024-10-29; **网络出版时间**:2025-05-07
基金项目:国家自然科学基金资助项目(61902046,61901074,62076044); 中国博士后科学基金资助项目(2021M693771); 重庆市自然科学基金资助项目(CSTB2022NSCQ-MSX0145)
第一作者:余雷(1982—),男,副主任医师,讲师,博士研究生,主要研究方向为急诊危重症患者的救治、急性中毒救治. E-mail:yulei@cqmu.edu.cn

0　引言

脓毒症(全身性感染症)是一种严重的感染性疾病,其发病率和死亡率在临床中呈上升趋势。了解脓毒症的特点并预测重症监护室(intensive care unit, ICU)脓毒症患者 30 d 内的死亡风险,有助于医生在临床治疗中进行准确有效的诊断,从而提升患者的治疗水平和生存率。脓毒症特征是机体对感染的异常反应,导致全身性炎症反应综合征(systemic inflammatory response syndrome, SIRS)和器官功能障碍[1],诊断主要依据临床表现和实验室检查[2]。

本文的研究工作主要聚焦于提升对 ICU 脓毒症患者 30 d 内死亡风险的预测准确性。本文利用波士顿大学医学中心的重症医学数据集(medical information mart for intensive care-IV, MIMIC-IV),对 ICU 脓毒症患者的临床记录进行分析,提出基于变量选择网络(variable selection networks, VSN)和门残差神经网络(gated residual networks, GRN)的深度神经网络(deep neural network, DNN)模型。DNN-VSN-GRN 简称为 DVG 模型,用于预测重症监护室(intensive care unit, ICU)脓毒症患者 30 d 内的死亡风险。由于临床诊断时存在大量检测指标,为了提高 DNN 模型识别和筛选特征的能力,加入 GRN 模块,它使 DNN 模型能够自动学习到输入向量的特征重要性分数,通过 VSN 模块对不同特征分配权重,在高维数据集中筛选出重要特征,实现了 DVG 模型的预测高准确性。

1　相关工作

Wang 等[3]通过统计学家 David Cox 提出的 Cox 回归模型[4]和亚组分析发现,血尿素氮与血清白蛋白的比值是脓毒症患者死亡的重要预测因子。Dias 等[5]通过多因素分析发现,病房入住 ICU 的发热脓毒症患者死亡率较高。Hu 等[6]使用 Cox 风险模型和多因素回归分析指出,虽然白蛋白水平是评估脓毒症患者疾病严重程度的指标之一,但是低蛋白血症对脓毒症患者的死亡风险没有显著影响。

为了评估脓毒症患者的疾病严重程度和死亡风险,一些评分方法用于评估脓毒症患者的疾病严重程度,例如序贯器官衰竭评估(sequential organ failure assessment, SOFA)评分[7]是 ICU 中常用的严重程度评估工具。但是,一般的评分方法[8-10]对于预测脓毒症患者的死亡风险的评估不可靠、不准确,有学者利用一些机器学习方法预测脓毒症患者的死亡风险。Hou 等[11]使用极限提升树模型(extreme gradient boosting, XGBoost)预测 ICU 脓毒症患者 30 d 内死亡风险; Perng 等[12]利用具有卷积神经网络的软最大函数(softmax function)模型评估脓毒症患者死亡风险,该模型优于其他机器学习模型。

目前的深度学习模型往往更加注重预测的准确性,但在模型的可解释性方面存在明显不足,导致深度学习模型很少被用于临床实践[13-16]。尽管深度学习模型在特征提取和分类问题上表现出色,但其对特征的解释能力仍然是一个“黑盒子”。此外,机器学习模型在小样本场景下的学习能力有限,在一定程度上限制了对重要特征的准确提取。

Lim 等[17]提出了一种利用多种数据源并具备可解释性的时序多步预测模型,该模型通过变量选择网络模块和门残差神经网络模块有效优化了性能和可解释性。然而,该模型主要适用于时序预测问题,在处理静态或非时序数据时具有一定局限性。

为了解决这些问题,本文提出 DVG 模型。该模型从 MIMIC-IV 中提取相关临床数据集,通过随机森林算法剔除特征重要性最低的几组指标,降低了模型的复杂性。利用 DVG 模型进行训练,建立了针对脓毒症患者死亡风险的预测模型。同时,为了提升模型的可解释性,通过移除和重新训练(remove and retrain, ROAR)[18]方法对多种可解释性技术进行评估,最终选取相关性分数逐层传播(layer-wise relevance propagation, LRP)方法作为最优的可解释性方法,并得到了特征重要性的解释。

2　深度神经网络及可解释性方法

2.1　数据的收集和预处理

MIMIC-IV 是一个重要的临床医学数据库,来源于波士顿大学医学中心的电子健康记录数据,包括大量

的临床数据。该数据集规模庞大,涵盖了数十万患者的临床数据,包括多个科室的患者的生理参数、实验室检查结果、诊断信息、治疗方案等多种类型的医疗数据。研究人员必须通过测试才能使用数据集,该项目已获得批准从 MIMIC-IV 中提取数据并用于研究目的。

Meng 等[19]对 MIMIC-IV 进行深入研究,评估深度学习模型在可解释性和公平性方面的表现,并研究不同的可解释方法在深度学习模型在判断特征的重要时存在偏差,提供优化深度学习模型解释能力的思路。本文借鉴 Meng 研究中深度学习模型对特征重要性和可解释性评估的思路,应用于脓毒症死亡风险评估方面。本文提取与脓毒症相关的 72 个指标,使用 50 次多重插值处理特征的缺失值。这些医学指标分为 7 类:(1)人口统计学指标,包括性别、年龄、种族。(2)生理指标,包括体重、心率、收缩压、舒张压、平均血压、呼吸、体温、血氧饱和度。(3)实验室指标,包括血细胞比容、平均血红蛋白含量、平均血红蛋白浓度、平均红细胞体积、血小板、红细胞、红细胞分布宽度、白细胞、乳酸、酸碱度、氧分压、二氧化碳分压、碱剩余、总二氧化碳、血红蛋白、阴离子间隙、碳酸氢盐、血尿素氮、钙、氯、肌酐、钠、钾、国际标准化比值、凝血酶原时间、部分凝血活酶时间、葡萄糖。(4)输出和评分指标,包括尿量、序贯器官衰竭评估(sequential organ failure assessment, SOFA)评分、查尔森合并症评分、逻辑器官功能障碍评分。(5)其他健康状态和病史,包括心肌梗死、充血性心力衰竭、周围血管疾病、脑血管疾病、痴呆、慢性肺疾病、风湿病、消化性溃疡疾病、轻度肝病、无并发症的糖尿病、有并发症的糖尿病、截瘫、肾病、恶性癌症、重度肝病、转移性实体瘤、艾滋病。(6)特定条件,包括疑似感染、阳性培养、神经。(7)评分细分,包括年龄评分、肺的评分、凝血评分、肝脏评分、心血管评分、中枢神经系统评分。

本文挑选满足以下条件的患者:(1)患者入院时年龄在 18 岁以上;(2)患者已知的第一次 ICU 住院;(3) ICU 总住院时间为 30 d 之内,具体过程如图 1 所示。

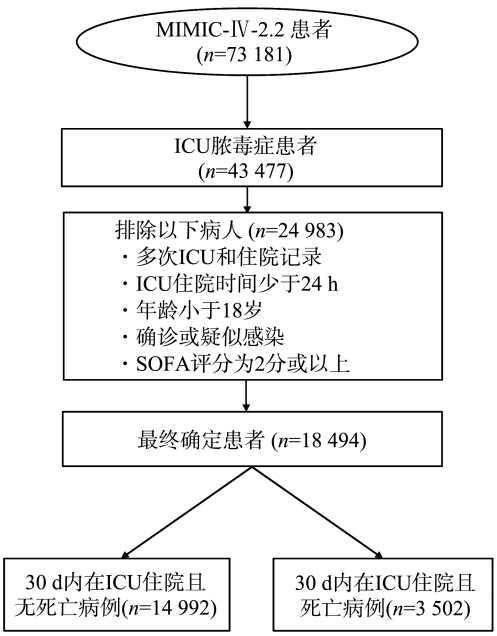

图 1 MIMIC-IV 中脓毒症数据的预处理和筛选

Fig.1 Sepsis data preparation and exclusion diagram in MIMIC-IV

2.2 统计分析

比较 30 d 内死亡和未在 30 d 内死亡的参与者的基线特征差异,本文将使用 Student t 检验和 Fisher 精确概率检验分析连续变量和分类变量之间的差异。

对于连续变量,本文使用 Student t 检验比较死亡组和非死亡组之间的均值差异;对于分类变量,本文使用 Fisher 精确概率检验比较死亡组和非死亡组之间的差异。通过分析可得,对于连续变量,死亡组和非死亡组之间的均值差异具有统计学意义($p<0.05$)。对于分类变量,死亡组和非死亡组之间的分布差异具有统计学意义($p<0.05$),因此,在 30 d 内死亡和未在 30 d 内死亡的脓毒症患者之间存在基线特征差异。这些差异可能与预后和治疗结果相关,并为进一步研究和临床实践提供了重要线索。

2.3　深度神经网络的结构

根据脓毒症患者的人口学特征和生活习惯数据,选取了 72 个特征,通过随机森林算法删除了特征重要性评分小于 0.01 的特征,保留了 43 个重要特征,如表 1 所示。

表 1　随机森林筛选出的 43 个重要特征

Table 1　43 important features selected by random forest

特征	特征重要性评分	特征	特征重要性评分
尿量	0.032	恶性癌症	0.021
呼吸频率	0.028	动脉二氧化碳分压	0.020
体温	0.027	平均红细胞血红蛋白	0.020
年龄	0.026	凝血酶原时间	0.020
血小板计数	0.026	钙	0.020
收缩压	0.026	血尿素氮	0.020
心率	0.025	氯化物	0.019
舒张压	0.024	钠	0.019
白细胞计数	0.024	血液酸碱度	0.019
血糖	0.024	钾	0.019
动脉氧分压	0.023	查尔森合并症评分	0.018
器官衰竭评分	0.023	平均红细胞体积	0.017
平均动脉压	0.023	碳酸氢盐	0.017
红细胞分布宽度	0.023	肌酐	0.017
血红蛋白	0.022	SOFA 评分	0.017
红细胞压积	0.022	阴离子间隙	0.017
体重	0.022	碱剩余	0.016
部分凝血活酶时间	0.022	总二氧化碳	0.016
血氧饱和度	0.022	心血管功能	0.014
红细胞计数	0.021	国际标准化比值	0.014
平均红细胞血红蛋白浓度	0.021	肝功能	0.013
乳酸	0.021		

输入特征与脓毒症患者死亡风险之间的精确关系未知,导致难以确定哪些变量与预测结果相关,以及变量非线性变化的程度该如何选择。本文将 DNN 模型与 VSN 模块和 GRN 模块相结合,如图 2 所示。门残差神经网络模块分为门控机制和残差连接两部分。首先门控线性单元(gate linear unit, GLU)能够控制非线性变化的程度,使得 DVG 模型能够选择性地通过信息。He 等[20]通过残差连接的方式将输入特征向量与经过非线性处理后的特征向量相加,发现残差神经网络能够缓解深层网络中的梯度消失问题,同时 DVG 模型能够充分学习到原始特征和经过非线性变化后的特征,使网络更好地学习复杂的特征。再通过变量选择网络模块,DVG 模型可以初步筛选出门残差神经网络关注到的特征,其核心计算式为

$$g(\boldsymbol{x})=l(\boldsymbol{x}+h(\boldsymbol{z})), \tag{1}$$

式中,$\boldsymbol{x}$ 是输入的特征向量,$g(\boldsymbol{x})$为门残差神经网络的输出,即经过层归一化和残差连接计算后的结果,l 是层归一化函数 LayerNorm 的替代符号,$\boldsymbol{z}$ 是通过输入特征向量 $\boldsymbol{x}$ 得到的中间表示,$\boldsymbol{z}=e(\boldsymbol{W}_1\boldsymbol{x}+\boldsymbol{b}_1)$,$e$ 是 ELU 激活函数的替代符号。$h(\boldsymbol{z})$是门控线性单元 GLU 的输出,$h(\boldsymbol{z})=\sigma(\boldsymbol{W}_2\boldsymbol{z}+\boldsymbol{b}_2)\odot(\boldsymbol{W}_3\boldsymbol{z}+\boldsymbol{b}_3)$,表示门控机制计算的结果,$\odot$表示哈达玛乘积,$\sigma$ 是 Sigmoid 激活函数,参数 $\boldsymbol{W}_1$、$\boldsymbol{W}_2$、$\boldsymbol{W}_3$ 分别表示每层网络的权重,$\boldsymbol{b}_1$、$\boldsymbol{b}_2$、$\boldsymbol{b}_3$ 分别表示每层网络的偏置。

利用门残差神经网络模块将每个特征 $\boldsymbol{x}_i$ 生成并输出 $\boldsymbol{n}_i$,计算特征权重表示为 $\boldsymbol{v}_i$,即

$$\boldsymbol{n}_i=g(\boldsymbol{x}_i), \tag{2}$$

$$\boldsymbol{v}_i=S(\boldsymbol{\epsilon}_i), \tag{3}$$

式中,$\forall i=1,2,\cdots,D$,$D\in\mathbf{N}$,S 是 Softmax 函数的替代符号,用于归一化特征权重,$\boldsymbol{\epsilon}_i=(n_1,n_2,\cdots,n_D)$,$D$ 为输入的特征维度。变量选择网络模块将输入的特征向量 $\boldsymbol{x}$ 与权重 $\boldsymbol{v}_i^{\mathrm{T}}$ 相乘得到筛选后的特征,将筛选后的特征输入到 2 层线性层,最后经过 Sigmoid 激活函数输出,变量选择网络模块的结果 $\hat{\boldsymbol{x}}$ 为

$$\hat{\boldsymbol{x}}=\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{x}。 \tag{4}$$

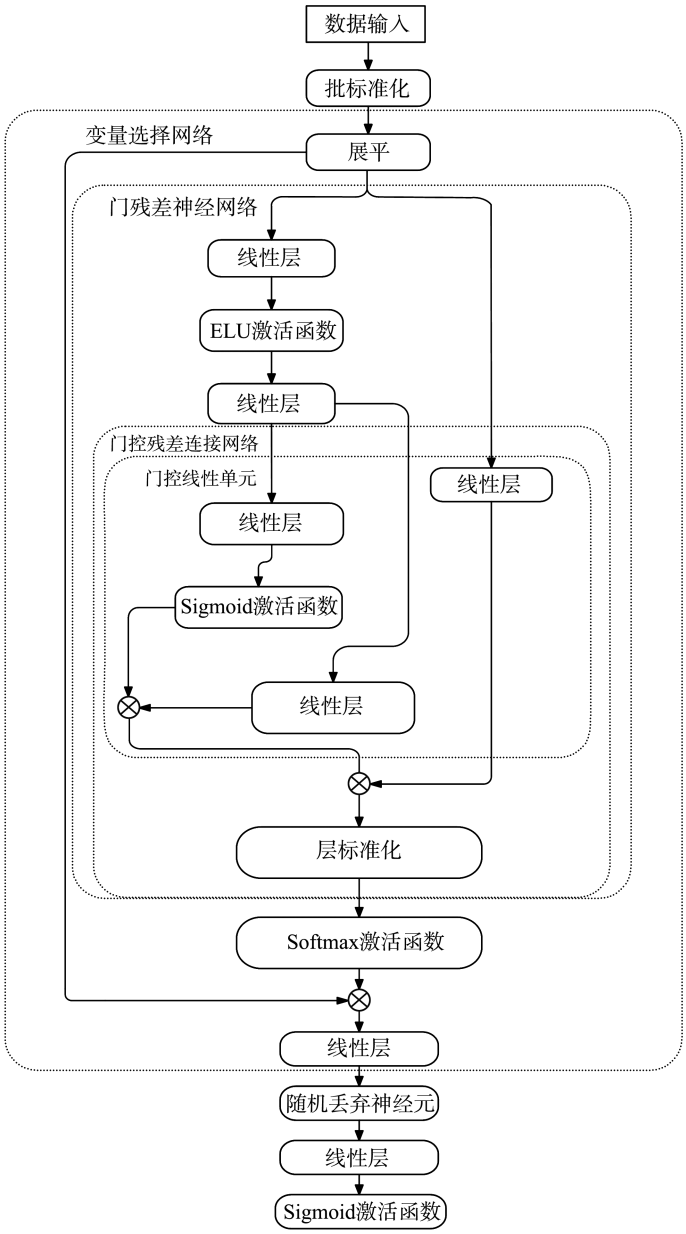

图 2 DVG 模型框架图

Fig.2 The frame of DVG model

对于线性层参数和随机丢弃神经元(dropout)的概率,本文分别选择(128,256,512,1 024)、(0.1,0.2,0.4,0.5,0.6)进行实验,最终发现当第 1 层线性层参数为 128,第 2 层线性层参数为 64,dropout 率为0.2时,模型表现出最佳的性能。线性层表示为

$$\boldsymbol{z}=\sigma\left(\sum \boldsymbol{w}_i \boldsymbol{t}_i+\boldsymbol{b}\right), \tag{5}$$

式中:σ 是激活函数,$\boldsymbol{t}_i$ 为门残差神经网络模块的输出,$\boldsymbol{w}_i$ 和 $\boldsymbol{b}$ 分别为权重和偏置,$\boldsymbol{z}$ 为线性层计算得到的中间表示。二分类的目标函数定义为 binary-cross entropy, 对于预测值 $\hat{y}$ 和真实值 y 使用以下公式计算损失函数:

$$L=-(\hat{y}\ln(y)+(1-\hat{y})\ln(1-y)), \tag{6}$$

每 1 个线性层的输出使用 Relu 激活函数进行非线性变换。

2.4 实验设置

本文采用 70/10/20 的训练/验证/测试数据分割,对于 DVG 模型,使用批量为 1 024 的 Adam 优化器,在 TensorFlow 2. 9. 0 版本中完成所有实验。

2.5　可解释性

本文采用多种解释性方法解释 DVG 模型做出的决策。从可解释性的角度来看,分为与模型无关的和特定于模型的可解释性方法。与模型无关的可解释性方法通过分析特征的输入和输出解释模型的决策,应用于任何模型;特定于模型的可解释性方法则针对特定模型,基于该模型的内部结构进行解释。

2.5.1　独立于模型的解释

夏普利采样(Shapley sampling)方法[21]利用博弈论中的 Shapley 值原理,计算了每个特征在不同特征组合中的平均边际贡献,即第 j 个特征的 Shapley 值

$$\varnothing_j=\sum_{A\subseteq M\backslash j}\frac{|A|!\ (|M|-|A|-1)!}{|M|!}[\nu(A\cup\{j\})-\nu(A)], \tag{7}$$

式中:M 是特征集合,A 是不包含特征 j 的特征子集,$A\subseteq M\backslash j$ 表示特征子集 A 是由特征集合 M 中除了第 j 个特征以外的所有特征构成,$\nu(A\cup\{j\})$ 表示包含特征 j 的特征子集的预测结果,$\nu(A)$ 表示不包含特征 j 的特征子集的预测结果。

显著性分析(saliency analysis)方法[22]通过计算输出相对于输入特征的梯度的绝对值衡量每个输入特征对于预测的重要性,即第 p 个输入特征的显著性评分

$$S_p=\left|\frac{\partial f(X)}{\partial x_p}\right|, \tag{8}$$

式中:x_p 表示输入向量 $\boldsymbol{x}$ 中的第 p 个特征,$f(X)$ 表示模型的输出。

2.5.2　基于模型的解释

深度学习特征归因(deep learning important features, DeepLift)方法[23]计算每个特征对模型输出的贡献,获得对模型中所有样本的特征重要性的分析。模型的最终输出 $f(x)$,即基准输出 $f(x_0)$ 与所有输入特征贡献值 C_{x_qy} 的总和:

$$f(x_0)+\sum_{q=1}^{P}C_{x_q}y=f(x), \tag{9}$$

式中:x_0 是参考输入,输入 x_q 对模型输出 y 的影响用 C_{x_qy} 表示,P 是特征的数量。

积分梯度(integrated gradients)方法[24]通过衡量输入特征对模型预测结果的贡献,评估特征的重要性。模型的最终输出 f 是 u 个输入特征对模型预测结果的归因值,即输入特征对模型输出的重要性评分:

$$f=(x-x')\times\int_{a=0}^{1}\frac{\partial F(x'+\alpha\times(x-x'))}{\partial x_u}\mathrm{d}\alpha, \tag{10}$$

式中:x 是输入,x'是基线值,F 是函数映射,α 是积分路径上的插值系数。

逐层相关性传播(layer-wise relevance propagation, LRP)方法[25]通过神经网络进行反向传播来识别重要特征。它将输出的“相关性”沿着每一层向后传播到输入,从而确定每个输入特征的重要性,第 l 层的第 m 个神经元的相关性通过第 l+1 层的第 k 个神经元向后传播得到,即

$$R_{m\leftarrow k}^{(l,l+1)}=R_k^{(l+1)}\frac{a_m w_{mk}}{\sum_h a_h w_{hk}}, \tag{11}$$

式中:a_m 表示第 l 层第 m 个神经元的激活值,w_{mk}表示连接第 l 层的第 m 个神经元和第 l+1 层的第 k 个神经元的权重,$\sum_h a_h w_{hk}$ 表示第 l+1 层第 k 个神经元的输入加权和。

遮挡(occulsion)方法[26]是将数据按一定的窗口进行遮挡,将需要遮挡的数据替换为高斯噪声,计算数据替换前后的深度学习模型或者机器学习模型性能评分的数值差异进行累加,从而评估每个特征的重要性。

2.6　ROAR 分析

本文使用 Hooker 等[18]提出的 ROAR(移除和重新训练)进行可解释性方法的选取。对于每种可解释性方法,将每个数据样本的某些部分的最重要特征替换为固定的无信息值,本文在训练集和测试集都进行了这种操作。然后用修改后的训练集对模型进行重新训练,并在修改后的测试集上评估其分类性能。在特征被删除的数据集上对模型进行再训练,保证训练数据和测试数据来自相似的分布,减少数据分布差异对模型性能的影响,性能的下降的原因是由于信息的删除而不是数据分布的移动引起的。

对于序列输入 $\boldsymbol{X}\in R^{\wedge}(T\times F)$,$T$ 为样本个数,每个样本有 F 个特征。本文将序列输入扁平化,并给出

$T\times F$个特征的特征重要性评分。本文使用每个特征在训练集中的平均值作为该特征的非信息值。用了 5 个特征比例表示从原始特征向量中删除一定比例的特征来评估每种可解释性方法,特征比例分别为 0%、25%、50%、75%、100%。

3 实验结果与分析

在实验中,本文的 DVG 模型对脓毒症患者入住 ICU 后 30 d 内的死亡风险做出预测,首先比较了 DVG 模型、其他机器学习模型和一些深度学习模型在预测中不同性能,使用多种可解释方法对模型进行解释,并通过 ROAR 方法对这些可解释方法做出对比,最后选出了性能最好的可解释方法。

3.1 DVG 模型和机器学习以及深度学习方法的比较

本文选取 4 种用于脓毒症预测研究的主流机器学习模型:逻辑回归(logistics regression, LR)模型[27]、极致梯度提升树(extreme gradient boosting, XGBoost)模型[28]、随机森林(random forest, RF)模型[29]、梯度提升树(gradient boosted trees, GB)模型[30]。关于深度学习模型,本文选择了多层前馈神经网络(multilayer perceptron, MLP)模型和卷积神经网络(convolutional neural network, CNN)模型,MLP 模型采用了 4 层网络结构,每层隐藏神经元的个数分别为 128、256、128 和 64;CNN 模型采用了 4 层卷积层,每层的卷积核大小为 3×3,步长为 1。各模型的性能指标精确率-召回率曲线下面积(area under the precision versus recall curve, AUPRC)和受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)如表 2、3 所示。

表 2 置信区间为 95%时不同模型的 AUPRC

Table 2 AUPRC of different models at a 95% confidence interval

指标范围	LR[27]	XGBoost[28]	RF[29]	GB[30]	CNN	MLP	**DVG**
0.656~0.676	0.668						
0.919~0.932		0.928					
0.956~0.964			0.961				
0.924~0.933				0.929			
0.948~0.957					0.953		
0.871~0.889						0.882	
0.964~0.971							**0.967**

表 3 置信区间为 95%时不同模型的 AUROC

Table 3 AUROC of different models at a 95% confidence interval

指标范围	LR[27]	XGBoost[28]	RF[29]	GB[30]	CNN	MLP	**DVG**
0.630~0.661	0.647						
0.942~0.952		0.948					
0.962~0.970			0.962				
0.944~0.952				0.949			
0.941~0.955					0.949		
0.885~0.898						0.895	
0.962~0.971							**0.966**

本文的 DVG 模型的 AUPRC 为 0.967, AUROC 为 0.966,均大于其他机器学习和深度学习模型。在脓毒症患者死亡风险预测中,本文的 DVG 模型能够处理复杂、高维度的医疗数据,捕捉高维和非线性数据中的潜在模式,筛选出重要的输入特征。门残差神经网络模块和变量选择网络模块使 DVG 模型能够在预测准确性方面显著超越传统的机器学习和深度学习模型。

3.2 ROAR 方法分析

本文利用 ROAR 方法对各模型的可解释性方法进行评估,主要通过 AUROC 和 AUPRC 进行分析。随着特征移除的比例增加,AUROC 和 AUPRC 越小,表示预测性能下降越快,特征重要性越高。

根据表 4、5 的定量结果,LRP 方法的 AUROC 和 AUPRC 下降幅度最大,这些表明其特征重要性估计的稳定性优于其他方法。当特征比例从 0%增加到 75%,LRP 方法的 AUROC 从 0.967 下降到了 0.828,

AUPRC 从0.96下降到 0.79,当特征比例为 75%时,LRP 方法的 AUPRC 和 AUROC 都是最小。因此,对于入住 ICU 30 d内的患者死亡率预测,LRP 方法给出的特征重要性评分排序是本文考虑的所有可解释性方法的结果中效果最好的。

表 4　采用不同特征比例时不同方法的 AUROC

Table 4　AUROC of different methods with varying feature proportions

特征比例/%	IntegratedGrads	Saliency	Deeplift	LRP	Occlusion	Shapley_sampling
0	0.967	0.967	0.967	0.967	0.967	0.967
25	0.943	0.946	0.950	0.939	0.947	0.946
50	0.912	0.925	0.922	0.920	0.947	0.929
75	0.845	0.847	0.844	0.828	0.843	0.843

表 5　采用不同特征比例时不同方法的 AUPRC

Table 5　AUPRC of different methods with varying feature proportions

特征比例/%	IntegratedGrads	Saliency	Deeplift	LRP	Occlusion	Shapley_sampling
0	0.966	0.966	0.966	0.966	0.966	0.966
25	0.938	0.944	0.947	0.933	0.943	0.938
50	0.905	0.920	0.915	0.914	0.917	0.923
75	0.827	0.829	0.834	0.799	0.823	0.825

3.3　基于 LRP 方法的特征重要性排序

本文利用随机森林算法筛选出 43 个重要特征,通过 LRP 方法评估这些特征对 DVG 模型和其他对比模型预测结果。表 6 为脓毒症患者入住 ICU 30 d 内 LRP 方法死亡率预测的特征重要性排序结果。特征重要性评分越大表示该特征对预测结果的影响越大。从网络的内在结构出发,LRP 方法根据最终的分类结果,通过反向传播过程中激活函数的输出和神经网络的权值计算各层神经元的相关性系数,相关性越大的神经元对输出的影响越大,因为网络结构参数量少、网络层数较少、计算复杂度较低,特征之间多为线性关系,说明解释性较高。但是 LRP 方法的缺点在于无法解决梯度饱和的问题。

表 6　不同特征重要性评分排序

Table 6　Different feature importance rankings

特征	特征重要性评分	特征	特征重要性评分
查尔森合并症评分	1.452 766	红细胞分布宽度	0.980 646
器官衰竭评分	1.444 654	血糖	0.956 025
年龄	1.379 947	舒张压	0.940 862
呼吸频率	1.373 930	红细胞计数	0.937 352
恶性癌症	1.337 036	平均红细胞血红蛋白	0.927 330
血红蛋白	1.303 254	总二氧化碳	0.909 867
SOFA 评分	1.244 695	氯化物	0.884 516
平均红细胞血红蛋白浓度	1.215 579	动脉二氧化碳分压	0.882 826
阴离子间隙	1.199 471	肝功能	0.854 934
白细胞计数	1.193 252	动脉氧分压	0.840 603
尿量	1.178 063	部分凝血活酶时间	0.833 690
钾	1.157 106	血氧饱和度	0.826 395
心率	1.151 866	钙	0.813 371
体温	1.129 805	红细胞压积	0.811 334
收缩压	1.112 333	乳酸	0.804 634
血尿素氮	1.090 198	碱剩余	0.800 547
血液酸碱度	1.075 291	体重	0.782 599
血小板计数	1.066 778	肌酐	0.762 512
平均动脉压	1.061 299	平均红细胞体积	0.752 026
心血管功能	1.046 365	国际标准化比值	0.509 818
钠	1.035 003	凝血酶原时间	0.492 904
碳酸氢盐	1.027 718		

分析结果表明,查尔森合并症评分在特征重要性排序中,特征重要性评分最高,为 1.452 766,同时器官衰竭评分、年龄、呼吸频率以及是否患有恶性癌症等特征对重症监护室脓毒症患者死亡的相关性同样较大。表明患者的年龄、既往疾病负担和多系统功能受损程度对于评估入住 ICU 30 d 内脓毒症患者的死亡风险具有关键性影响。平均红细胞体积、国际标准化比值、凝血酶原时间等特征评分较低,表明这些指标对于重症监护室脓毒症患者死亡的影响较低。此外,在临床上常用的 SOFA 评分[6](sofa score)也具有显著的重要性,进一步验证了实验结果的可靠性和临床适用性。

4 结论

本文提出一种新型的深度神经网络模型,DVG 模型结合变量选择网络模块和门控残差网络模块,在提升脓毒症患者死亡风险预测准确性的同时,提高了模型的可解释性。实验结果表明,与传统机器学习和深度学习模型相比,本文的 DVG 模型在脓毒症患者死亡风险评估中具有显著优势。在众多可解释性方法中,LRP 方法给出最佳的特征重要性排序。在特征重要性排序中,查尔森合并症评分是影响死亡率预测的最重要特征,器官衰竭评分、年龄、呼吸频率以及是否患有恶性癌症等对重症监护室脓毒症患者死亡的影响较大。本文的 DVG 模型有望成为监测 ICU 脓毒症患者实时风险的有用决策支持工具,帮助医生更早识别和干预患者的病情,从而降低死亡率和并发症发生率。下一步工作将聚焦于优化 DVG 模型结构、验证 DVG 模型的泛化能力,开发能够融入临床流程的实时风险评估系统。

参考文献:

[1] 邱海波,杜斌,刘大为,等. 全身炎症反应综合征与多器官功能障碍综合征的临床研究[J]. 中华外科杂志,1997,35(7):402-405.
QIU Haibo, DU Bin, LIU Dawei, et al. Clinical study on systemic inflammatory response syndrome and multiple organ dysfunction syndrome [J]. Chinese Journal of Surgery, 1997, 35(7):402-405.

[2] GAUER R, FORBES D, BOYER N. Sepsis:diagnosis and management[J]. American Family Physician, 2020, 101(7):409-418.

[3] WANG Yuhe, SHAN Gao, LEI Hong, et al. Prognostic impact of blood urea nitrogen to albumin ratio on patients with sepsis: a retrospective cohort study[J]. Scientific Reports, 2023, 13(1):10013.

[4] COX D R. Regression models and life-tables[J]. Journal of the Royal Statistical Society(Series B: Methodological), 1972, 34(2):187-220.

[5] DIAS A, GOMEZ V C, VIOLA L R, et al. Fever is associated with earlier antibiotic onset and reduced mortality in patients with sepsis admitted to the ICU[J]. Scientific Reports, 2021, 11(1):23949.

[6] HU Jing, LÜ Chenwei, HU Xinxing, et al. Effect of hypoproteinemia on the mortality of sepsis patients in the ICU: a retrospective cohort study[J]. Scientific Reports, 2021, 11(1):24379.

[7] VINCENT J L, MENDONCA A, CANTRAINE F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study[J]. Critical Care Medicine, 1998, 26(11):1793-1800.

[8] FORD D W, GOODWIN A J, SIMPSON A N, et al. A severe sepsis mortality prediction model and score for use with administrative data[J]. Critical Care Medicine, 2016, 44(2):319-327.

[9] CARRARA M, BASELLA G, FERRARIO M. Mortality prediction model of septic shock patients based on routinely recorded data[J]. Computational and Mathematical Methods in Medicine, 2015, 2015(1):761435.

[10] KHWANNIMIT B, BHURAVANONTACHAI R, VATTANVANIT V. Validation of the sepsis severity score compared with updated severity scores in predicting hospital mortality in sepsis patients[J]. Shock, 2017, 47(6):720-725.

[11] HOU Nianzong, LI Mingzhe, HE Lu, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost[J]. Journal of Translational Medicine, 2020, 18(1):462-462.

[12] PERNG J W, KAO I H, KUNG C T, et al. Mortality prediction of septic patients in the emergency department based on machine learning[J]. Journal of Clinical Medicine, 2019, 8(11):1906.

[13] 化盈盈,张岱墀,葛仕明. 深度学习模型可解释性的研究进展[J]. 信息安全学报,2020,5(3):1-12.
HUA Yingying, ZHANG Daichi, GE Shiming. Research progress in the interpretability of deep learning models[J]. Journal

of Information Security, 2020, 5(3):1-12.
[14] 陈珂锐,孟小峰. 机器学习的可解释性[J]. 计算机研究与发展,2020,57(9):1971-1986.
CHEN Kerui, MENG Xiaofeng. Interpretability of machine learning[J]. Journal of Computer Research and Development, 2020, 57(9):1971-1986.
[15] 曾春艳,严康,王志锋,等. 深度学习模型可解释性研究综述[J]. 计算机工程与应用,2021,57(8):1-9.
ZENG Chunyan, YAN Kang, WANG Zhifeng, et al. A review of research on the interpretability of deep learning models[J]. Computer Engineering and Applications, 2021, 57(8):1-9.
[16] RAIKOMAR A, DEAN J, KOHANE I. Machine learning in medicine[J]. New England Journal of Medicine, 2019, 380(14):1347-1358.
[17] LIM B, ARIK S O, LOEFF N, et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting[J]. International Journal of Forecasting, 2021, 37(4):1748-1764.
[18] HOOKER S, ERHAN D, KINDERMANS P J, et al. A benchmark for interpretability methods in deep neural networks[C]// Advances in Neural Information Processing Systems. Vancouver: Curran Associates Incorporated, 2019.
[19] MENG C, TRINH L, XU N, et al. Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset[J]. Scientific Reports, 2022, 12:7166.
[20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vega: Curran Associates Incorporated, 2016:770-778.
[21] MITCHELL R, COOPER J, FRANK E, et al. Sampling permutations for Shapley value estimation[J]. Journal of Machine Learning Research, 2022, 23(43):1-46.
[22] HE Shengfeng, ZHE Huang, LIU Wenxi, et al. SuperCNN: a superpixelwise convolutional neural network for salient object detection[J]. International Journal of Computer Vision, 2015, 115(3):330-344.
[23] ZAFEIRIOU A, KALLIPOLITIS A, MAGLOGIANNIS I. Ensembling to leverage the interpretability of medical image analysis systems[J]. IEEE Access, 2023,11:76437-76447.
[24] SUNDARARAJAN M, TALY A, YAN Q. Axiomatic attribution for deep networks[C]// International Conference on Machine Learning. Sydney: PMLR, 2017:3319-3328.
[25] BACH S, BINDER A, MONTAVON G, et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation[J]. Public Library of Science One, 2015, 10(7):e0130140.
[26] HOCHULI J, HELBLING A, SKAIST T, et al. Visualizing convolutional neural network protein-ligand scoring[J]. Journal of Molecular Graphics & Modelling, 2018, 84:96-108.
[27] TOLLES J, MEURER W J. Logistic regression: relating patient characteristics to outcomes[J]. The Journal of the American Medical Association, 2016, 316(5):533-534.
[28] OSMAN A I A, AHMED A N, CHOW M F, et al. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in selangor malaysia[J]. Ain Shams Engineering Journal, 2021, 12(2):1545-1556.
[29] MASCARO J, ASNER G P, KNAPP D E, et al. A tale of two "forests": random forest machine learning AIDS tropical forest carbon mapping[J]. Public Library of Science One, 2014, 9(1):E85993.
[30] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. The Annals of statistics, 2001, 29(5): 1189-1232.

(编辑:陈丽萍)