

基于改进的LF-Transformer脓毒症早期预测方法

冯晓乐, 杨其宇, 韦力天

广东工业大学自动化学院, 广东 广州

收稿日期: 2025年2月23日; 录用日期: 2025年3月13日; 发布日期: 2025年3月21日

摘要

在ICU电子病历脓毒症早期预测任务中, 传统机器学习方法因难以捕捉稀疏表格数据中的动态特征交互而导致性能受限。为此, 本研究提出改进的LF-Transformer深度学习模型, 通过构建基于乘法算术块的特征交互增强方法, 结合交互候选生成器(ICG)的动态top-k特征选择机制, 有效提升了模型对稀疏医疗数据的表征能力。采用MIMIC-IV数据集验证模型性能, 实验结果表明改进模型在脓毒症预测任务中AUROC达到0.841, 特异度0.763, 敏感性0.759, 显著优于传统方法。该研究成果为开发ICU脓毒症智能预警系统提供了有效的算法支持, 具有一定的临床实践价值。

关键词

脓毒症, 特征交互增强, MIMIC-IV数据集, ICU智能预警系统, 深度学习

Early Sepsis Prediction Method Based on Improved LF-Transformer

Xiaole Feng, Qiyu Yang, Litian Wei

School of Automation, Guangdong University of Technology, Guangzhou Guangdong

Received: Feb. 23rd, 2025; accepted: Mar. 13th, 2025; published: Mar. 21st, 2025

Abstract

In the task of early sepsis prediction using ICU electronic medical records, traditional machine learning methods suffer from performance limitations due to their inability to capture dynamic feature interactions in sparse tabular data. To address this, we propose an improved LF-Transformer deep learning model. By constructing a feature interaction enhancement method based on multiplicative arithmetic blocks and integrating a dynamic top-k feature selection mechanism via an Interaction

Candidate Generator (ICG), the model significantly improves its representational capability for sparse medical data. Validated on the MIMIC-IV dataset, experimental results demonstrate that the enhanced model achieves an AUROC of 0.841, specificity of 0.763, and sensitivity of 0.759 in sepsis prediction, outperforming traditional methods significantly. This research provides effective algorithmic support for developing intelligent sepsis early warning systems in ICUs, offering practical clinical value.

Keywords

Sepsis, Feature Interaction Enhancement, MIMIC-IV Dataset, ICU Intelligent Early Warning System, Deep Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

脓毒症(Sepsis)是一种由感染引发的全身性炎症反应综合征(SIRS),其进展迅速且病死率高,是ICU患者死亡的主要原因之一[1]。根据世界卫生组织(WHO) [2]的报告,每年全球约有数千万例脓毒症病例,其中许多患者因治疗延误而失去生命。由于脓毒症发病机制涉及多器官功能障碍及免疫反应失代偿,早期识别对降低病死率具有关键作用。

传统机器学习方法在脓毒症预测中已取得初步成效,如Hu等人[3]使用7种机器学习方法进行脓毒症预测,并通过SHapley方法验证了特征解释的有效性,Srimedha等人[4]揭示了病理特征间非线性关系的重要性。然而,这些方法在捕捉稀疏表格数据中的动态特征交互方面存在局限。虽然Al-Mualemi等人[5]的Adaptive CNN和Zhou等人[6]的在线预测模型在性能上有所突破——前者搭建了脓毒症警报系统但泛化能力受限,后者突破传统历史数据训练框架却难以兼顾准确率与时效性——Dai等人[7]的深度强化学习框架虽取得一定效果,但受LSTM架构限制仍难以充分挖掘时序特征。

近期研究表明,基于注意力机制的模型在表格数据处理中展现出独特优势。Kwagtek等人[8]的LF-transformer通过改进位置编码增强了表格特征提取能力,而Cheng等人[9]的AMformer引入算术块(Arithmetic Block)有效改善了特征交互稀疏性问题。本文基于上述研究,提出融合动态top-k特征选择机制的改进LF-Transformer模型,通过构建交互候选生成器(ICG)优化特征交互过程,旨在提升ICU脓毒症早期预测的准确度和鲁棒性。

2. 数据集

2.1. 基础数据集

本研究采用MIMIC-IV (Medical Information Mart for Intensive Care IV)数据集,该数据库由麻省理工学院与波士顿贝斯以色列女执事医疗中心联合开发,包含2008~2019年间约40万住院患者的临床数据,其中部分患者接受重症监护治疗。如表1所示的关键特征统计信息主要涵盖人口统计学特征、生命体征时序数据及实验室检测指标(如乳酸、白细胞计数),同时该数据集还包含药物治疗记录与诊疗操作等临床参数。

该数据集呈现两个显著特征:其一,ICU患者的实验室指标与生命体征数据具有非均匀采样特性,其时间序列间隔呈现不规则分布;其二,高维稀疏特征矩阵与动态特征交互的复杂性对建模构成挑战。

这些特性为研究重症监护环境下的时序数据建模提供了典型范例。

Table 1. Descriptive statistics of key features from MIMIC-IV database

表 1. MIMIC-IV 部分关键特征统计信息

特征类别	特征名称	描述	均值 \pm 标准差	缺失率(%)
人口统计特征	年龄(Age)	入院时患者年龄	62.09 \pm 17.33	0.0
	性别(Gender)	男性比例	男性: 54.15%	0.0
生命体征	收缩压(SBP)	收缩压的动态变化	121.89 \pm 21.64	26.84
	血氧饱和度(SpO ₂)	血氧饱和度的动态变化	96.57 \pm 2.98	24.80
	心率(Heart rate)	每分钟心拍数	83.14 \pm 17.71	22.77
	呼吸率(Respiratory rate)	每分钟呼吸次数	19.04 \pm 5.28	23.97
实验室检测	体温(Temperature)	患者体温	36.85 \pm 0.56	78.19
	白细胞计数(WBC)	白细胞总数	11.11 \pm 7.17	94.91
	血清乳酸(Lactate)	血清乳酸水平	2.36 \pm 2.20	97.87
	肌酐(Creatinine)	肾功能指标	1.35 \pm 1.50	94.55
	血糖(Glucose)	血液中糖分	145.96 \pm 60.89	84.69
	总胆红素(Total Bilirubin)	肝功能评估	2.05 \pm 4.71	98.86

2.2. 数据集预处理

2.2.1. 样本筛选

本研究基于 MIMIC-IV 数据集中的 364,627 例患者病例进行分析。本研究的目的是针对于患者的生理指标，预测往后 6 小时内是否发生脓毒症。我们采用严格的纳入排除标准对原始数据进行筛选。排除标准包括：(1) 无 ICU 住院记录的病例(n = 287,696)；(2) ICU 住院时长不足 8 小时的病例(n = 1724)，考虑到这两类患者发生脓毒症的风险相对较低，且观察时间过短，不适合进行脓毒症预测研究；(3) 脓毒症发作前监测时长少于 8 小时的病例(n = 28,692)，因其观察窗口过短，统计价值有限。经过上述筛选流程，最终纳入 46,516 例符合条件的病例用于脓毒症预测分析。

2.2.2. 特征处理

针对医疗数据的高维稀疏特性，采用分层填充策略：

(1) 时序特征：对生命体征与实验室指标等非均匀采样数据，采用向前填充与线性插值组合方法，确保动态特征交互建模的连续性。

(2) 静态特征：通过最邻近插值保持局部相似性，对连续变量采用中位数填充增强鲁棒性。

(3) 异常值处理：基于临床阈值法识别真异常(如心率 < 30 或 > 250 次/分)，采用中位数替换；对数据脱敏导致的伪异常(如年龄 > 90 岁记录为 300 岁)，依据数据字典进行校正。

3. 方法

3.1. 特征提取

为增强模型对稀疏医疗数据的动态特征交互建模能力，本研究基于原数据集特征采用以下方法做进一步提取。

(1) 基于滑动窗口的数值分析特征

临床数据由多个动态的生理和生化指标组成，这些指标以小时为单位记录，其数值的变化能够反映

患者的健康状态，是判断是否存在脓毒症的重要依据。同时，基于滑动窗口的数值分析方法可以有效捕捉时间序列数据中的局部特征。在窗口内的数据可以用于计算诸如平均值、方差、最大值和最小值等统计特征，从而揭示该窗口内数据的特性。考虑到本数据集的特点，我们确定滑动窗口的大小为 6 小时，而且选择了缺失率较低、数据质量较高的生命体征指标，因此选取了心率(HR)、血氧饱和度(O₂Sat)、收缩压(SBP)、平均动脉压(MAP)和呼吸率(Resp)进行分析。

(2) 临床专家评分特征

在临床实践中，专家共识制定的评分系统为患者状态评估提供了重要依据。本研究整合了多个具有临床指导价值的评分指标，以增强数据集的临床表征能力。

针对感染性休克患者的识别，本研究提出感染性休克指数(Septic_shock)，其定义基于脓毒性休克的典型临床特征：患者需在充分容量复苏后仍依赖血管加压药物维持平均动脉压 ≥ 65 mmHg，同时伴随血清乳酸水平 > 2 mmol/L。若患者平均动脉压 < 65 mmHg，指数增加幅度为 1；若血清乳酸水平 < 2 mmol/L，指数同样增加 1。该指数通过整合血流动力学与代谢指标，为早期识别危重患者提供了量化依据。

在肾功能评估方面，血尿素氮(BUN)与肌酐(Creatinine)比值被纳入分析体系。血尿素氮/肌酐比值可敏感反映肾功能变化趋势，特别是在肌酐水平正常而血尿素氮异常升高时，能够提示潜在的肾功能减退。此外，该指标对区分肾前性与肾性急性肾损伤具有重要鉴别价值。

氧合功能的评估采用动脉血氧饱和度(SaO₂)与吸入氧浓度(FiO₂)比值作为核心指标。相较于需要动脉血气分析的 PaO₂/FiO₂ 比值，SaO₂/FiO₂ 通过无创方式反映患者氧合状态，为临床动态监测提供了便利。

同时引入休克指数(心率/收缩压)作为循环状态评估。该指标通过量化心率和血压的平衡关系，能够辅助判断患者的血流动力学稳定性。此外，总胆红素与肌酐比值被用于同步评估肝肾功能损伤程度，该指标在严重感染性休克患者中可反映多器官功能障碍风险，为临床决策提供多维参考。

总的来说，临床专家评分特征由感染性休克指数，血尿素氮/肌酐比值，动脉血氧饱和度与吸入氧浓度比值，休克指数和总胆红素与肌酐比值共同整合。

本研究将提取的滑动窗口数值分析特征与临床专家评分特征整合到原始特征集中，构建完整的特征输入矩阵用于模型训练。

3.2. 网络模型

本研究提出了一种基于 LF-transformer 的改进模型架构，通过将原有模型中的多头注意力机制替换为算术块，增强了对表格型数据的特征表达能力。模型以 MIMIC-IV 数据集为基础，以 3.1 小节描述的方法提取的特征作为模型输入，输出则为脓毒症是否发生的二分类预测结果。

LF-transformer 的核心创新在于引入潜因子分解机制，分别对数据的行列特征进行提取。具体而言，通过 Column Transformer 和 Row Transformer 的矩阵乘法运算获取特征表示，随后利用 CLS Tokenizer 将数据转换为标准的 Transformer 输入格式，最终经由 CLS Query-wise Transformer 模块完成特征处理。

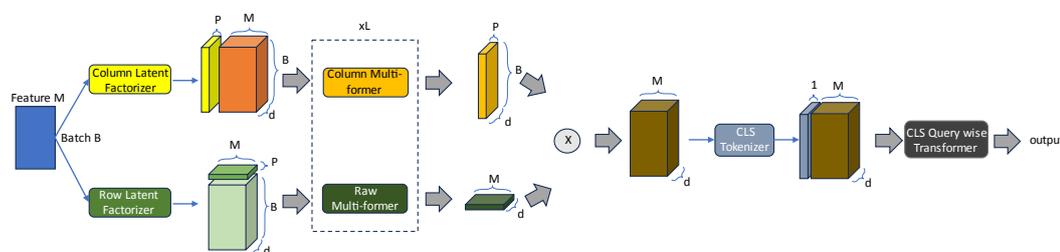


Figure 1. Model architecture

图 1. 模型整体框架图

为了进一步提升模型性能,本研究借鉴了 AMformer 中的算术块设计[9],通过引入乘法算术运算模块(Multi-former),有效地解决了表格数据中的稀疏交互问题。我们将这两种模型的优势特征相结合,构建了一个改进版的 LF-transformer 架构,以提高脓毒症早期预测的准确度。改进后的网络结构如图 1 所示。引入的乘法算术块如图 2 所示。

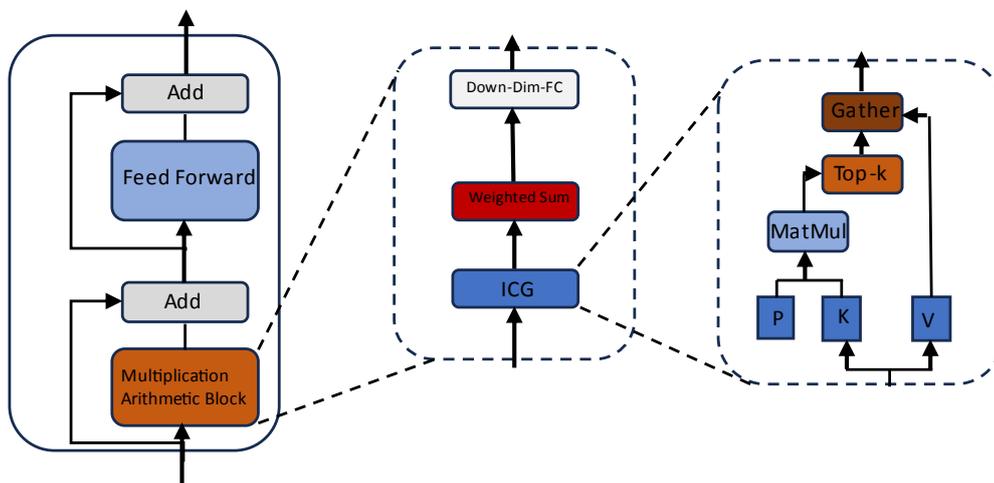


Figure 2. The overview of Multi-former

图 2. 乘法算术块示意图

经典 Attention 模块引入指数运算单元,用于捕获特征间的乘性交互关系,通过 Prompt tokens (P), 查询矩阵 $Q = XW^Q$, 键矩阵 $K = XW^K$, 值矩阵 $V = XW^V$, 可训练参数 $W^{Q/K/V} \in R^{d \times d}$, 构建特征间的交互表示。基于这些组件,经典 Attention 模块的加法流输出可表示为以下形式:

$$O^A = \text{soft max} \left(\frac{QK^T}{\sqrt{d}} \right) V \in R^{N \times d} \quad (1)$$

针对表格型数据中特征交互的稀疏性特点,传统软注意力机制在构建特征间关联时存在效率问题——具有显著相关性的特征对在位置上往往并不相邻。为此,本研究引入了乘法算术块(见图 2)。其中,ICG (Interaction Candidate Generator)模块作为候选交互信息生成器,该机制通过在 QW^T 的每一行中选择 top-k 个最大值,并将其他元素通过大负常数掩蔽处理,使其在经过 softmax 后趋近于 0。该模块采用稀疏特征交互策略,每个特征仅与其最相关的 k 个特征进行交互,其中 k 为远小于输入特征维度 N 的固定常数。这种设计在保持特征交互稀疏性的同时显著降低了计算复杂度。随后,通过全连接层将特征映射至原始维度的 50%,实现特征降维。

本次模型训练采用的损失函数为加权交叉熵损失函数,它的表达式为:

$$\text{Loss} = -[w \cdot y \log(p) + (1-y) \log(1-p)] \quad (2)$$

加权交叉熵损失函数通过量化预测概率分布与真实标签分布之间的差异来指导模型训练。在二分类问题中,真实标签 y 取值为 $\{0, 1\}$, 而 p 代表模型对正类的预测概率。该损失函数的优化目标是 minimized 预测分布与真实分布之间的信息差异,从而提升模型的分类性能。为了减少样本不均衡对模型训练的影响,加入了权重 $W \in R^+ > 1$ 用于增强模型训练对阳性样本的关注度。交叉熵损失函数通常与 softmax 激活函数配合使用,其中 softmax 函数负责将模型的原始输出映射到 $(0, 1)$ 区间上的概率分布,随后基于此概率分布计算交叉熵损失。

由于网络模型输出的结果为分布于(0, 1)区间上的数值, 因此还需要选取合适的阈值作类别划分。本研究采用了约登指数寻找合适阈值, 具体步骤如下: (1) 以特定步长遍历从 0 到 1 的候选阈值点。(2) 根据阈值将预测概率转换为二分类结果, 计算该阈值下的敏感度(真阳性率)以及特异度(1 - 假阳性率)并计算约登指数 = 敏感度 + 特异度 - 1; (3) 选取使得约登指数最大的阈值最为选取阈值。

4. 模型训练与评估

4.1. 训练环境

本文的实验训练环境配置如表 2 所示。

Table 2. Experimental environment

表 2. 实验环境

硬件环境		软件环境	
CPU	Intel 12400F	操作系统	Windows11
GPU	NVIDIA RTX4060	Cuda 版本	12.1
内存	32G	Pytorch 版本	2.0

4.2. 模型超参数设置

在模型训练之前, 我们对关键超参数进行了系统性配置以优化模型性能。数据集划分方面, 采用 8:2 的比例将数据集分为训练集和验证集, 以评估模型的泛化能力。为确保实验结果的可复现性, 我们统一设定随机种子为 42。在优化策略上, 采用 Adam 优化器并设置初始学习率为 $5e-4$, 同时实施学习率衰减策略, 当验证集损失在连续 5 个 epoch 内没有改善时, 将学习率降低至原来的 0.1 倍。批处理大小(batch size)设置为 64, 在计算资源效率和训练稳定性之间取得平衡。模型训练轮次(epochs)设定为 150 轮, 并引入早停机制, 当验证集性能在连续 10 个 epoch 内无显著提升时终止训练, 以防止过拟合。这些超参数的选择基于网格搜索结果, 在多次实验验证后确定。

4.3. 模型评估

为了全面评估模型性能, 本研究采用了一系列互补的评价指标。其中, AUROC (Area Under the Receiver Operating Characteristic curve)通过计算 ROC 曲线下的面积来评估模型的性能, 该指标对类别不平衡不敏感, 取值范围为[0, 1], 其中 1 表示完美分类, 0.5 表示随机分类。

在二分类评估中, 我们还采用了以下指标:

敏感度(Sensitivity), 也称召回率(Recall), 用于评估模型识别正类样本的能力:

$$Sen = TP / (TP + FN) \quad (3)$$

特异度(Specificity), 用于评估模型识别负类样本的能力:

$$Spec = TN / (TN + FP) \quad (4)$$

其中, TP (True Positive)表示正确识别的正类样本数, TN (True Negative)表示正确识别的负类样本数, FP (False Positive)表示误判为正类的负类样本数, FN (False Negative)表示误判为负类的正类样本数。这些指标共同构成了一个完整的评估体系, 能够从不同角度反映模型的预测性能。我们将基于这些指标, 将本研究提出的模型与现有研究成果进行系统性对比, 以验证模型的优越性。

在评估过程中, 我们特别关注模型在不同临床场景下的表现。例如, 在早期预警系统中, 较高的敏

敏感度往往比特异度更为重要, 因为漏诊(假阴性)可能带来更严重的后果。相反, 在某些筛查场景中, 可能需要更高的特异度以避免过度医疗。因此, 这些指标的选择和权衡也反映了模型在实际临床应用中的适用性考量。

我们同时使用了多种经典模型(包括经典的 LF-Transformer、长短期记忆网络(LSTM)、XGboost), 在 MIMIC-IV 数据集上进行验证, 各个模型的 ROC 曲线见图 3。

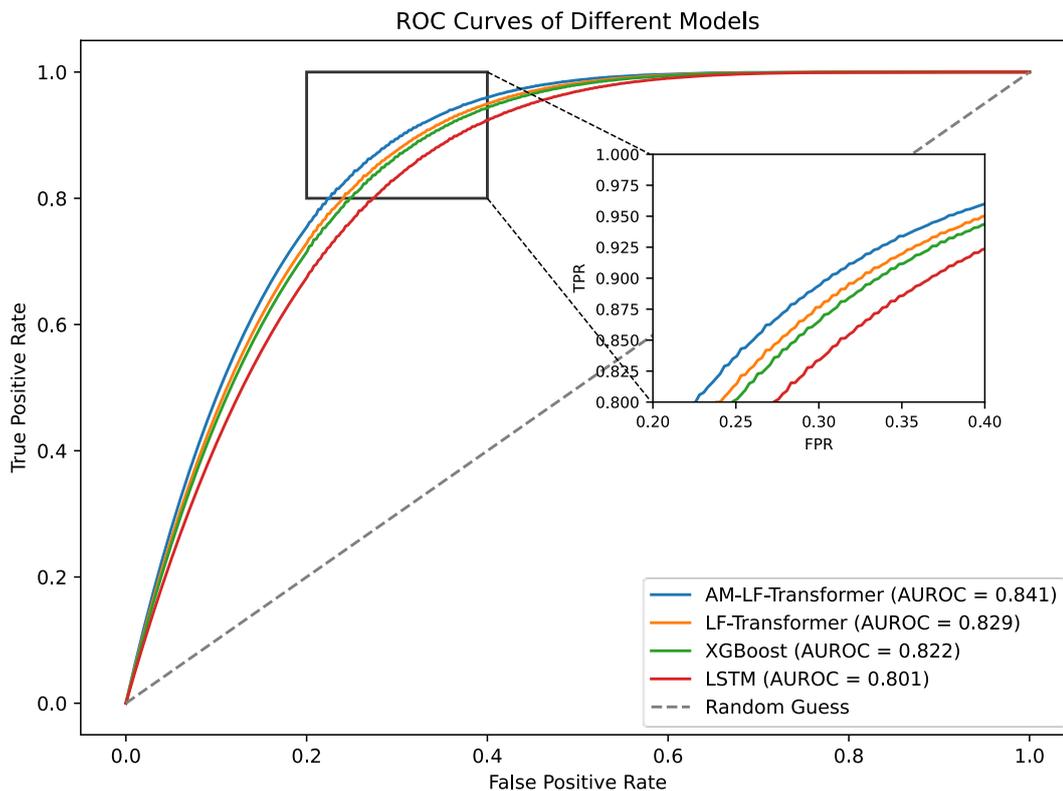


Figure 3. ROC curves of different models

图 3. 各模型评估 ROC 曲线

实验结果表明, 改进后的 LF-transformer 模型在各项评价指标上均展现出优越性。在 AUROC 评估中, 改进模型取得了 0.841 的最优表现, 相比原始 LF-Transformer (0.829)、XGBoost (0.822) 和 LSTM (0.801) 模型均有显著提升。这一结果验证了引入 AMformer 模块对增强模型在脓毒症早期预测任务上的有效性。针对于其他评价指标, 所取得的结果见表 3。

Table 3. Evaluation metrics of different models

表 3. 各模型各个评估指标

	AUROC	敏感度	特异度
改进 LF-Transformer	0.841	0.759	0.763
LF-Transformer	0.829	0.746	0.748
XGBoost	0.822	0.732	0.738
LSTM	0.801	0.712	0.715

这些评价指标的综合比较充分说明, 通过引入乘法算术块结构, 改进型 LF-transformer 在特征交互建

模方面获得了显著提升，不仅提高了模型对脓毒症病例的识别能力(敏感度)，同时也保持了较高的特异度，从而在整体预测性能上实现了突破。这对于临床实践中早期识别高风险患者具有一定意义。

最后，采用 SHAP 算法对改进后的 LF-transformer 模型进行可解释性分析。通过 SHAP 算法计算所有特征在模型预测过程中的重要性，并为每个样本计算 SHAP 值，绘制特征权重图以反映特征与预测结果之间的关系。根据变量的重要性值对各特征进行显著性排序，结果如图 4 所示。分析结果表明，特征如 FiO_2 、Temperature、WBC、GCS、BUN/CR 和 $\text{SpO}_2_{\text{min}}$ 对模型判断患者是否为脓毒症的影响较为明显。

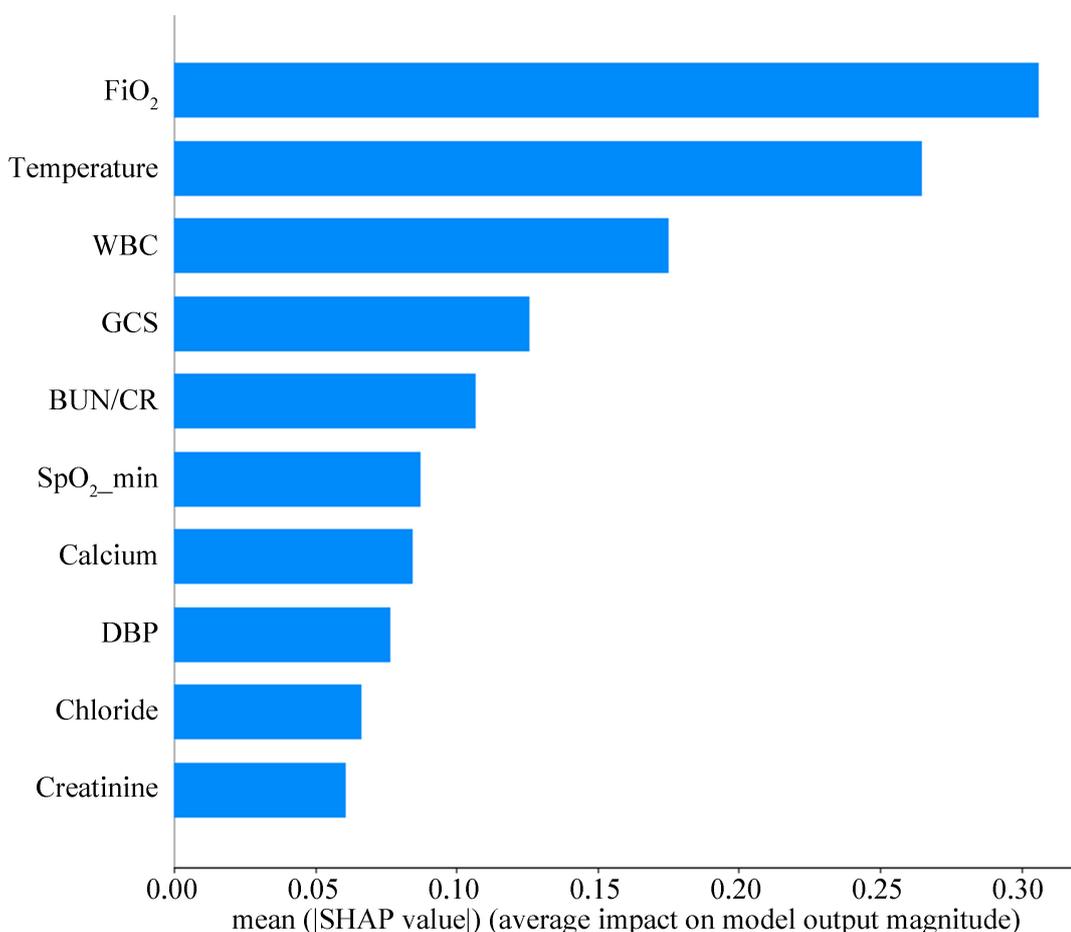


Figure 4. Feature importance analysis of the improved LF-transformer model (SHAP values)

图 4. 改进后的 LF-transformer 模型特征重要性分析(SHAP 值)

5. 结论

本研究提出的改进 LF-Transformer 模型通过动态 top-k 特征选择机制(ICG 模块)与乘法算术块的特征交互增强方法，有效解决了传统机器学习方法在稀疏表格数据中动态特征交互捕捉不足的问题。基于 MIMIC-IV 数据集的实验验证表明，该模型在脓毒症早期预测任务中达到 AUROC 0.841，特异性 0.763，敏感度 0.759，性能显著优于传统方法。研究成果为 ICU 脓毒症智能预警系统的开发提供了有效的算法支持。

参考文献

- [1] Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M., *et al.* (2016) The Third

International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, **315**, 801.

<https://doi.org/10.1001/jama.2016.0287>

- [2] World Health Organization (2017) Improving the Prevention, Diagnosis, and Clinical Management of Sepsis. https://apps.who.int/gb/ebwha/pdf_files/eb140/b140_12-en.pdf
- [3] Hu, C., Li, L., Huang, W., Wu, T., Xu, Q., Liu, J., *et al.* (2022) Interpretable Machine Learning for Early Prediction of Prognosis in Sepsis: A Discovery and Validation Study. *Infectious Diseases and Therapy*, **11**, 1117-1132. <https://doi.org/10.1007/s40121-022-00628-6>
- [4] Srimedha, B.C., Naveen Raj, R. and Mayya, V. (2022) A Comprehensive Machine Learning Based Pipeline for an Accurate Early Prediction of Sepsis in ICU. *IEEE Access*, **10**, 105120-105132. <https://doi.org/10.1109/access.2022.3210575>
- [5] Al-Mualemi, B.Y. and Lu, L. (2021) A Deep Learning-Based Sepsis Estimation Scheme. *IEEE Access*, **9**, 5442-5452. <https://doi.org/10.1109/access.2020.3043732>
- [6] Zhou, A., Beyah, R. and Kamaleswaran, R. (2022) OnAI-Comp: An Online AI Experts Competing Framework for Early Sepsis Detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **19**, 3595-3603. <https://doi.org/10.1109/tcbb.2021.3122405>
- [7] Dai, H., Hwang, H. and Tseng, V.S. (2023) Poems: Policy Network-Based Early Warning Monitoring System for Sepsis in Intensive Care Units. *IEEE Journal of Biomedical and Health Informatics*, **27**, 3610-3621. <https://doi.org/10.1109/jbhi.2023.3272486>
- [8] Na, K., Lee, J. and Kim, E. (2024) LF-Transformer: Latent Factorizer Transformer for Tabular Learning. *IEEE Access*, **12**, 10690-10698. <https://doi.org/10.1109/access.2024.3354972>
- [9] Cheng, Y., Hu, R., Ying, H., Shi, X., Wu, J. and Lin, W. (2024) Arithmetic Feature Interaction Is Necessary for Deep Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 11516-11524. <https://doi.org/10.1609/aaai.v38i10.29033>