

·论著·人工智能应用于烧伤与创面诊疗的探索·

本文亮点:

- (1) 构建融合烧伤创面图像与临床量表特征的婴幼儿烧伤脓毒症风险预测模型。
- (2) 结合临床特征重要性与创面图像模型注意力可视化,提高所构建模型预测结果的可解释性。

Highlights:

- (1) This study constructed a risk prediction model for burn sepsis in infants and toddlers by integrating burn wound images with clinical scale features.
- (2) The interpretability of the constructed model's predictions was enhanced by combining the importance of clinical features with attention visualization of wound images model.



基于深度学习的多模态婴幼儿烧伤脓毒症风险预测模型构建和验证

沈宇禾¹ 章磊² 季易² 陈昇² 沈卫民²

¹南京医科大学生物医学工程与信息学院,南京 210029;²南京医科大学附属儿童医院烧伤整形外科,南京 210008

通信作者:沈卫民,Email:swmswmswm@sina.com

【摘要】 目的 构建并验证一种基于深度学习的多模态婴幼儿烧伤脓毒症风险预测模型(以下简称多模态预测模型)。**方法** 该研究为回顾性队列研究。2015年1月—2025年12月,南京医科大学附属儿童医院烧伤整形外科收治1709例符合入选标准的烧伤婴幼儿患者(以下简称患儿),烧伤总面积为1.0%~85.0%体表总面积。根据患儿的临床结局,将其分为脓毒症组[54例,其中男34例、女20例,年龄500.0(270.3,695.0)d]和非脓毒症组[1655例,其中男998例、女657例,年龄515.0(365.0,790.0)d]。统计患儿入院24h内的临床资料,并比较两组患儿烧伤总面积、Ⅲ度烧伤面积、合并吸入性损伤者比例、C反应蛋白(CRP)升高者比例、降钙素原升高者比例、凝血酶原时间(PT)、白蛋白水平、白细胞计数等。构建包含烧伤创面图像特征与配对的临床量表特征的多模态预测模型。将多模态预测模型与随机森林、XGBoost、TabTransformer、ResNet、Vision Transformer、Swin Transformer模型(6种对照模型)进行对比,评估其性能;采用五折交叉验证评估多模态预测模型的内部性能。采用基于梯度的特征重要性分析方法对多模态预测模型进行可解释性分析,量化临床特征的贡献。采用梯度加权类激活映射方法对多模态预测模型进行可视化分析。**结果** 与非脓毒症组相比,脓毒症组患儿烧伤总面积、Ⅲ度烧伤面积均明显更大(Z 值分别为8.266、7.945, $P<0.05$),合并吸入性损伤者比例、CRP升高者比例、降钙素原升高者比例和白细胞计数均明显升高(χ^2 值分别为4.914、7.803、8.859, $Z=1.969$, $P<0.05$),白蛋白水平明显下降($Z=-2.132$, $P<0.05$),PT明显延长($Z=-4.247$, $P<0.05$)。多模态预测模型在患儿烧伤脓毒症风险预测中的受试者操作特征曲线下面积为 0.931 ± 0.009 ,优于随机森林、XGBoost、TabTransformer、ResNet、Vision Transformer、Swin Transformer模型的 0.861 ± 0.031 、 0.862 ± 0.025 、 0.876 ± 0.019 、 0.897 ± 0.015 、 0.900 ± 0.015 、 0.916 ± 0.012 。同时,多模态预测模型在烧伤创面图像分

DOI: 10.3760/cma.j.cn501225-20260519-00200

收稿日期 2026-05-19

引用本文:沈宇禾,章磊,季易,等.基于深度学习的多模态婴幼儿烧伤脓毒症风险预测模型构建和验证[J].中华烧伤与创面修复杂志,2026,42(9):847-856. DOI: 10.3760/cma.j.cn501225-20260519-00200.
Shen YH,Zhang L, Ji Y, et al. Construction and validation of a multimodal risk prediction model for burn sepsis in infants and toddlers based on deep learning[J]. Chin J Burns Wounds, 2026, 42(9): 847-856. DOI: 10.3760/cma.j.cn501225-20260519-00200.



割中的戴斯相似系数优于 ResNet、Vision Transformer、Swin Transformer 模型。五折交叉验证显示,相较于 6 种对照模型,多模态预测模型整体呈现更优的预测性能。基于梯度的特征重要性分析显示,PT 的归一化重要性权重(1.000)最大,其次是舒张压(0.855)。多模态预测模型输出的创面区域与医师标注区域边界基本一致。 **结论** 构建的多模态预测模型可以准确评估患儿烧伤后脓毒症的发生风险,具有良好的区分度和临床实用性。

【关键词】 人工智能; 烧伤; 脓毒症; 深度学习; 婴幼儿; 预测模型; 多模态融合

基金项目:国家自然科学基金面上项目(62471240)

Construction and validation of a multimodal risk prediction model for burn sepsis in infants and toddlers based on deep learning

Shen Yuhe¹, Zhang Lei², Ji Yi², Chen Sheng², Shen Weimin²

¹School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 210029, China;

²Department of Burn and Plastic Surgery, Children's Hospital of Nanjing Medical University, Nanjing 210008, China

Corresponding author: Shen Weimin, Email: swmswmswm@sina.com

【Abstract】 **Objective** To construct and validate a multimodal risk prediction model (hereinafter referred to as the multimodal prediction model) for burn sepsis in infants and toddlers based on deep learning, thereby providing decision support for early clinical warning and precision intervention. **Methods** This retrospective cohort study included 1 709 burned infants and toddler patients with total burn area ranged from 1.0%–85.0% total body surface area who were admitted to the Department of Burn and Plastic Surgery of Children's Hospital of Nanjing Medical University and met the eligibility criteria from January 2015 to December 2025. According to clinical outcomes, infants and toddler patients were divided into sepsis group ($n=54$, 34 males and 20 females, aged 500.0 (270.3, 695.0) days) and non-sepsis group ($n=1\ 655$, 998 males and 657 females, aged 515.0 (365.0, 790.0) days). Clinical data collected within 24 h after admission were analyzed and compared between the two groups of infants and toddler patients, including the total burn area, full-thickness burn area, the proportion of infants and toddler patients with concomitant inhalation injury, elevated C-reactive protein (CRP), and elevated procalcitonin proportion, prothrombin time (PT), albumin, and white blood cell count. A multimodal prediction model was constructed by integrating wound image features and paired clinical scale features. Its performance was compared with that of random forest, XGBoost, TabTransformer, ResNet, Vision Transformer, and Swin Transformer models (six comparison models), and internal performance was evaluated using five-fold cross-validation. Gradient-based feature importance analysis was performed to interpret the multimodal prediction model and to quantify the contribution of clinical features. Gradient-weighted class activation mapping was applied for visualization of the multimodal prediction model. **Results** Compared with those in non-sepsis group, infants and toddler patients in sepsis group had significantly larger total burn area and full-thickness burn area (with Z values of 8.266 and 7.945, respectively, $P<0.05$). Additionally, the proportions of infants and toddler patients with concomitant inhalation injury, elevated CRP, and elevated procalcitonin, and white blood cell count were significantly higher (with χ^2 values of 4.914, 7.803, and 8.859, $Z=1.969$, respectively, $P<0.05$) while albumin levels were significantly lower ($Z=-2.132$, $P<0.05$), and prothrombin time (PT) was significantly prolonged ($Z=-4.247$, $P<0.05$). The multimodal prediction model achieved an area under the receiver operating characteristic curve of 0.931 ± 0.009 for predicting post-burn sepsis of infants and toddler patients, numerically higher than those of random forest, XGBoost, TabTransformer, ResNet, Vision Transformer, and Swin Transformer models (with values of 0.861 ± 0.031 , 0.862 ± 0.025 , 0.876 ± 0.019 , 0.897 ± 0.015 , 0.900 ± 0.015 , and 0.916 ± 0.012 , respectively). The Dice similarity coefficient of the multimodal prediction model for burn wound image segmentation was higher than that of the ResNet, Vision Transformer, and Swin Transformer models. Across five-fold cross-validation, the multimodal prediction model exhibited superior overall predictive performance compared to the six control models. Gradient-based feature importance analysis revealed that PT had the highest normalized importance weight (1.000), followed by diastolic blood pressure (0.855). The wound regions delineated by the multimodal prediction model were largely consistent with those manually annotated by physicians. **Conclusions** The constructed multimodal prediction model demonstrates good discriminative performance for assessing the risk of post-burn sepsis in infants and toddlers.

【Key words】 Artificial intelligence; Burns; Sepsis; Deep learning; Infants; Prediction model; Multimodal fusion

Fund program: General Program of National Natural Science Foundation of China (62471240)

烧伤是儿童常见意外伤害。婴幼儿皮肤屏障薄、免疫系统尚未成熟,烧伤后感染易扩散并进展为脓毒症。既往研究显示,烧伤婴幼儿患者脓毒症发生率为 6.4%,脓毒症病死率可达 53%,相关死亡占其烧伤死亡病例的 60% 以上^[1]。严重感染多发生于伤后早期,因此早期风险识别和干预尤为重要。

脓毒症的核心是感染所致宿主反应失调及器官功能障碍。烧伤后非感染性炎症与感染表现可能重叠,婴幼儿常缺乏典型症状、难以表达不适,早期识别尤其困难。血液标本微生物培养虽可提供病原学依据,但其结果获得通常需 24~72 h,敏感度还受抗菌药物影响,不能作为早期识别的唯一依据^[2-3]。传统预警评分等通用工具依赖有限临床变量,在烧伤人群中的受试者操作特征曲线下面积 (area under the receiver operating characteristic curve, AUROC) 为 0.70~0.79^[4],难以充分表征创面、生命体征和实验室指标的复杂交互。

已有机器学习研究多针对感染相关脓毒症、成人烧伤或年龄跨度较大的儿科人群^[4-7],数据来源、预处理与验证流程异质,且外部验证不足^[8-10]。针对 0~3 岁婴幼儿,仍缺少融合烧伤创面图像与临床特征的模型。本研究据此构建并内部验证基于深度学习的多模态婴幼儿烧伤脓毒症风险预测模型(以下简称多模态预测模型),利用婴幼儿患者入院早期信息评估烧伤后脓毒症风险,为临床预警和风险分层提供依据。

1 资料与方法

本回顾性队列研究遵循《赫尔辛基宣言》的基本原则,并获得南京医科大学附属儿童医院(以下简称本院)伦理委员会批准(批号:2024-03-047)。根据本院伦理委员会政策,本研究不干预婴幼儿患者治疗,豁免婴幼儿患者监护人的知情同意;所有研究数据均经匿名化处理。

1.1 入选标准

纳入标准:(1)年龄≤3 岁;(2)院前或院内明确诊断为烧伤,包括热液烫伤、火焰烧伤、化学烧伤和电烧伤;(3)伤后 48 h 内入院;(4)具备模型分析所需的基本临床资料。排除标准:(1)入院时已明确诊断为脓毒症或入院前已有明确血流感染证据;

(2)合并恶性肿瘤、严重先天性免疫缺陷疾病或自身免疫性疾病;(3)入院后 24 h 内死亡或主动出院;(4)关键预测变量缺失比例超过 10%。

本研究参照 2024 年儿童脓毒症和脓毒性休克国际共识提出的菲尼克斯脓毒症标准^[11-12]进行病例界定。依据该标准,儿童脓毒症定义为在疑似或确诊感染基础上出现危及生命的器官功能障碍,操作性判定以菲尼克斯脓毒症评分 (Phoenix sepsis score, PSS) ≥2 分为主要依据。该标准在中国儿科 ICU 患儿中的适用性,已通过本课题组前期基于中国儿科 ICU 队列的回顾性评价予以验证^[13]。本研究为回顾性研究,部分 PSS 相关指标在既往普通病房病历中并非系统记录,因此仅采用病历中可客观追溯且记录完整性较高的指标进行辅助判断,未对缺失项目进行主观推断或人为补分。脓毒症诊断由 2 名具有 10 年以上婴幼儿烧伤诊疗经验的主治医师独立判定,诊断不一致时由第三名主任医师裁定。

本研究纳入的指标均为婴幼儿患者入院 24 h 内的指标,其中涉及 PSS 相关指标的呼吸支持、胸部 X 线片记录、血小板计数、凝血酶原时间 (prothrombin time, PT)、纤维蛋白原、脉搏及收缩压/舒张压的记录完整性均为 100%,D-二聚体的记录完整性为 58.2% (628/1 079),侵入性机械通气、脉搏血氧饱和度/吸入氧浓度、平均动脉压、格拉斯哥昏迷量表评分及瞳孔反应/昏迷状态未进行系统统计。

1.2 临床资料及分组

2015 年 1 月—2025 年 12 月,本院烧伤整形外科收治 1 709 例符合入选标准的烧伤婴幼儿患者,其中男 1 256 例、女 453 例,年龄 0~1080.0 (576.0±255.0) d,烧伤总面积 1.0%~85.0% TBSA。根据烧伤婴幼儿患者的临床结局(依据为上述菲尼克斯脓毒症标准+医师核对),将其分为脓毒症组[54 例,其中男 34 例、女 20 例,年龄 500.0 (270.3, 695.0) d]和非脓毒症组[1 655 例,其中男 998 例、女 657 例,年龄 515.0 (365.0, 790.0) d]。两组患者性别、年龄比较,差异均无统计学意义($\chi^2=0.998$, $Z=-1.954$, P 值分别为 0.051、0.541)。

1.3 数据收集与烧伤创面图像标注

收集以下信息。(1)人口学特征:年龄、性别、体

重、民族;(2)烧伤特征:烧伤总面积、Ⅲ度烧伤面积、烧伤深度(采用三度四分法分级)、致伤因素(热液、火焰、化学物质、电)、烧伤部位(头面颈、躯干及四肢),以及烧伤创面图像;(3)入院临床状况:烧伤至入院时间、院前处理情况(换药或输液),以及入院时体温、心率、呼吸频率、血压和胸部X线片记录;(4)实验室指标:入院后首次血常规(白细胞计数、中性粒细胞绝对值、淋巴细胞绝对值、血小板计数)、C反应蛋白(C-reactive protein, CRP)、降钙素原、白蛋白、血红蛋白、凝血功能指标(PT、纤维蛋白原和D-二聚体)、血乳酸及肝肾功能指标(白蛋白、前白蛋白、视黄醇结合蛋白等)水平及入院时创面标本和发热后血液标本微生物培养结果;(5)器官功能及治疗相关信息:有无少尿现象、是否合并吸入性损伤、是否发生休克及其类型、是否行机械通气、是否使用血管活性药物;(6)住院结局资料:出院转归、院内死亡情况及住院天数。比较两组婴幼儿患者烧伤总面积、Ⅲ度烧伤面积、合并吸入性损伤者比例、CRP升高者比例、降钙素原升高者比例、PT、热液烫伤者比例、白蛋白水平、白细胞计数等。

本研究共纳入 1 709 例婴幼儿患者对应的 1 709 张烧伤创面图像。创面局部感染的视觉征象主要包括创面渗出物增多或伴脓性分泌物、创面颜色异常、创周红肿加深、坏死组织或焦痂增多、创面边缘炎症反应明显及肉芽组织生长不良等。烧伤创面图像标注采用ITK-SNAP软件(美国宾夕法尼亚大学图像计算与科学实验室)完成。由 2 名主治医师独立对主要烧伤创面区域进行人工勾画,并生成创面区域掩膜;对创面边界或疑似局部感染视觉征象存在分歧时,由 2 名医师讨论后达成一致,形成最终标注结果。烧伤创面图像标注人员与脓毒症诊断判定人员相互独立,以减少主观判断偏倚。

1.4 基于深度学习的多模态预测模型的构建

1.4.1 模型框架 构建多模态预测模型框架,该框架整合烧伤婴幼儿患者的烧伤创面图像特征与配对的临床量表特征。其中,临床量表特征包含患者的以下指标:生理指标,如心率、呼吸频率、体温、血压等高采样频率指标;实验室指标,如降钙素原、白细胞计数和CRP等,用以共同构成感染风险的生物化学维度的判定依据;临床文本特征,包括医师的日常病程记录、护理评估笔记及影像学报告中提取的结构化信息。基于此,构建Transformer的多尺度混合深度神经网络模型,用于融合婴幼儿烧

伤创面图像与临床风险因素特征,预测烧伤后脓毒症发生风险。该模型框架由 2 个子网络组成。

子网络 1 采用Swin Transformer U型分割结构^[14]提取烧伤创面图像特征并完成语义分割。依次通过 4 个编码阶段预处理图像:第 1 个阶段进行图块嵌入并接入 2 个 Swin Transformer 模块,后续 3 个阶段通过图块合并降低分辨率增加通道数,即 F_{E1} 、 F_{E2} 、 F_{E3} 、 F_E 通道数依次为 96、192、384 和 768。解码器包含 3 个阶段,通过图块扩展和跳跃连接逐级恢复空间分辨率,输出创面分割图及多尺度特征 $\{F_E, F_{D1}, F_{D2}, F_{D3}\}$ 。在创面分割监督下,子网络 1 同时学习创面区域与感染相关视觉表征,供后续多模态融合。

子网络 2 用于提取多维临床特征,涵盖患者基本信息、生命体征、生物化学检验、烧伤信息及合并症等病历资料的原始数据,并由医师整理为表格化数据。预处理阶段首先依据医学合理范围及原始病历记录对异常值进行核查,若判断为明显录入错误且无法进一步确认,则作为缺失值处理。对于最终纳入分析的病例,其连续数值特征采用 K 近邻插补进行缺失值填补,以利用相似病例间的信息降低简单均值或中位数填补可能带来的偏倚;类别特征采用训练集中对应变量的众数进行填补。随后将临床量表特征划分为类别特征和连续数值特征,并对类别特征进行编码,以便后续网络正确学习其临床语义。采用TabTransformer结构^[15]分别处理两类特征:类别特征通过多头注意力结构提取语义关联,连续数值特征经层归一化处理,随后将两者叠加,获得完整的临床烧伤-脓毒症语义,用 F_{TAB} 表示。

将子网络 1 输出的多尺度图像特征块 $\{F_E, F_{D1}, F_{D2}, F_{D3}\}$ 与子网络 2 输出的多维临床特征 F_{TAB} 在通道维度进行早期拼接融合,形成 $F = \{F_{TAB}, F_E, F_{D1}, F_{D2}, F_{D3}\}$,经线性层压缩至 256 维,再通过全连接层预测婴幼儿烧伤患者发生脓毒症的概率。

多模态预测模型采用戴斯损失与二元交叉熵的混合损失函数进行训练,其定义如下: $L =$

$$0.5L_{Dice} + 0.5L_{BCE} \quad \text{其中, } L_{Dice} = 1 - \frac{2 \sum_i (p_i \times g_i)}{\sum_i p_i + \sum_i g_i},$$

针对创面分割任务, p_i 表示模型对第 i 个像素的预测概率, g_i 为对应的真实创面金标准; $L_{BCE} = -[\omega_{pos} \times y \times \ln(\hat{y}) + \omega_{neg} \times (1 - y) \times \ln(1 - \hat{y})]$, y 为真实脓毒症标签, $\hat{y}$ 为模型预测的脓毒症概率, ω_{pos} 与 ω_{neg} 分别为正样本(脓毒症)与负样本(非脓毒症)的权重,

通常根据训练集中两类样本数量的反比进行设定,以缓解数据不平衡对模型训练的影响。见图 1。

1.4.2 模型训练和参数调整 模型训练采用自适应矩估计优化器,初始学习率为 0.000 1,权重衰减系数为 0.001。学习率调度采用平台期降低学习率策略并监控验证集的 AUROC,当 AUROC 连续 5 个训练轮次未提高时,采用学习率×衰减因子×0.5 进行自适应下调。训练批次为 8,最多训练轮次为 100,模型在 PyTorch 2.0 框架下完成训练。

1.5 对比模型与消融模型

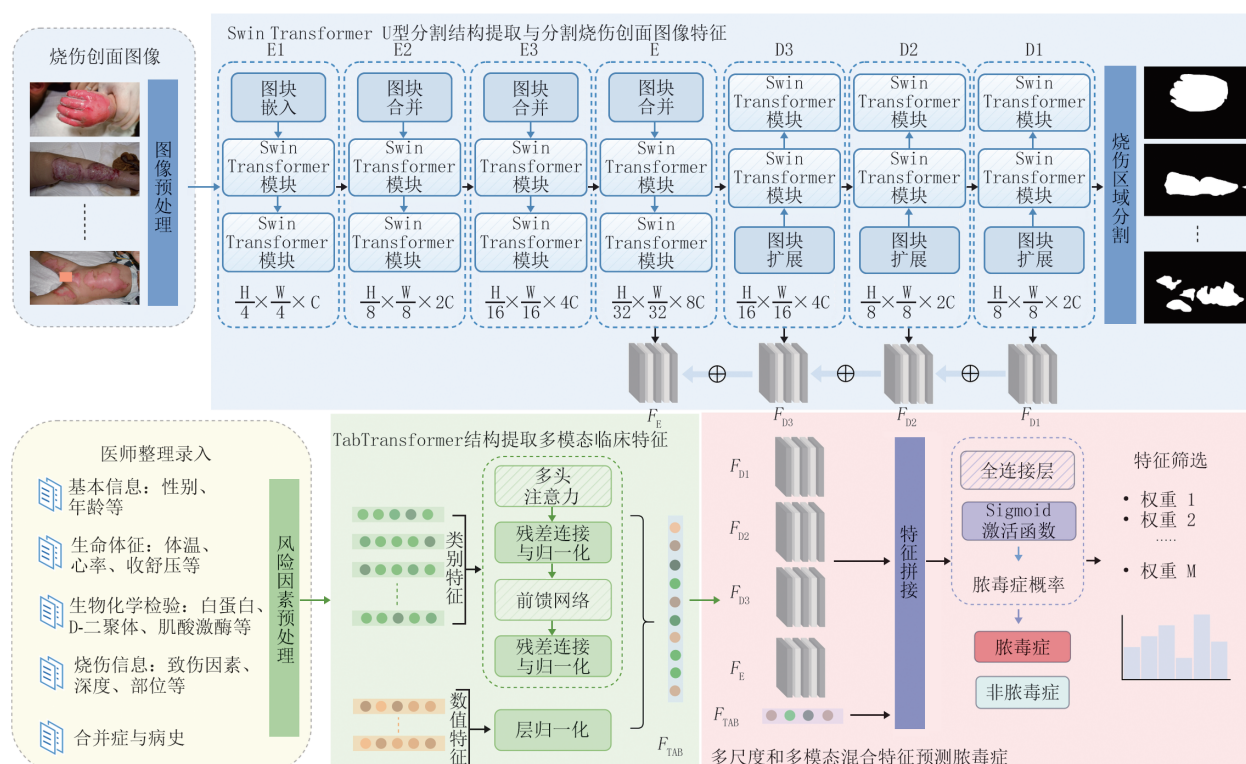
通过设置多种烧伤临床量表模型和创面图像模型,评估多模态预测模型的性能。

临床量表模型仅输入表格化临床数据,包括类别特征与连续数值特征。作为对照模型的 3 种临床量表模型如下。(1)随机森林模型:采用随机搜索交叉验证进行超参数调优,搜索范围包括决策树数量、最大深度、最小叶节点样本数、特征采样比例及类别权重等参数,最终模型启用类别平衡权重,以缓解脓毒症组与非脓毒症组样本之间的类别不平衡。(2)XGBoost 模型:采用随机搜索交叉验证进行超参数调优,搜索范围包括弱学习器数量、学习率、

最大树深度、子采样比例、列采样比例、L1/L2 正则化系数及正负样本权重。目标函数设为二元逻辑回归,评估指标包括对数损失和 AUROC。(3)TabTransformer 模型:采用与多模态预测模型子网络 2 相同的结构配置,利用多头注意力机制提取类别特征间的语义交互关系,连续数值特征经层归一化处理后与类别特征融合,最终通过全连接层输出脓毒症发生概率^[15]。

创面图像模型仅输入预处理后的烧伤创面图像,同时完成烧伤创面分割并输出脓毒症发生概率。各模型采用相同的病例级划分和评价流程,并在训练数据内独立优化。作为对照模型的 3 种创面图像模型如下。(1)ResNet 模型:采用 ResNet-34 编码器及 U 型解码器,分类与分割任务共享编码器权重^[16];(2)Vision Transformer 模型:采用 Vision Transformer 编码器提取全局特征,用解码器重构分割结果,并分类标记预测风险^[17];(3)Swin Transformer 模型:Swin Transformer 编码器结合 U 型解码器,利用多尺度特征完成分类与分割^[14]。

为评价多模态融合策略的增量贡献,本研究在保持 TabTransformer+Swin Transformer 双分支框架



注:模型框架由烧伤创面图像分支(Swin Transformer U)、临床量表分支(TabTransformer)及多尺度特征融合与分类模块组成;烧伤创面图像分支同时完成创面分割,融合特征用于输出脓毒症风险概率

图 1 人工智能多模态婴幼儿烧伤脓毒症风险预测模型框架

Figure 1 Framework of the artificial intelligence-based multimodal risk prediction model for burn sepsis in infants and toddlers

及分类头结构一致的前提下设置 3 组消融实验: (1) 仅输入临床表格数据, 图像分支置空或不参与特征融合; (2) 仅输入烧伤创面图像数据, 表格分支置空或不参与特征融合; (3) 同时输入临床表格数据与烧伤创面图像数据, 并进行特征拼接融合。

1.6 模型验证

采用五折交叉验证评估多模态预测模型的内部性能。以婴幼儿患者为数据划分单位, 按照脓毒症结局进行分层, 将全部病例划分为 5 个互斥子集。每轮以其中 4 个子集作为训练集, 剩余 1 个子集作为验证集, 重复 5 轮, 使每例婴幼儿患者均且仅在 1 个验证折中接受模型评价。在数据划分过程中, 兼顾不同入院年份病例的分布, 确保各折均包含 2015—2025 年各年的病例, 以减少病例集中于某一特定年份或治疗阶段造成的时间分布偏倚。

绘制受试者操作特征曲线, 评估多模态预测模型相较于其他 6 种模型的预测性能。考虑到本研究阳性事件比例较低, 同时绘制精确率-召回率曲线, 以更全面评价类别不平衡条件下多模态预测模型对阳性病例的识别能力。

1.7 模型的可解释性分析

分别针对多模态预测模型中烧伤创面图像与临床特征进行可解释性分析。

采用基于梯度的特征重要性分析方法对多模态预测模型的临床特征 (如 PT、舒张压、白蛋白水平、CRP、降钙素原、中性粒细胞计数、血糖水平、脉搏、纤维蛋白原、白细胞计数等) 进行可解释性分析。该方法利用模型输出对输入特征的梯度信息, 量化模型脓毒症预测输出对各临床特征变化的局部敏感性。其中, 类别特征通过其嵌入向量的梯度范数衡量贡献, 连续数值特征则基于标准化后数值输入的梯度评估重要性, 所有特征的重要性得分经绝对值归一化映射至 0~1 区间。

采用梯度加权类激活映射方法, 对多模态预测模型处理烧伤创面图像时的高响应区域进行可视化分析。该方法通过计算脓毒症类别预测得分对图像编码器目标特征层输出的梯度, 评估各空间位置对模型预测结果的贡献。随后计算各通道权重并加权融合, 生成空间热图, 用于创面区域的可视化与脓毒症风险预测。

1.8 模型预测性能评价与统计学处理

模型的预测采用 AUROC 进行评价。同步计算准确率、精确率、召回率、特异度、F1 分数以及模型

创面区域分割的戴斯相似系数 (Dice similarity coefficient, DSC)。对于五折交叉验证产生的配对性能指标, 由于每个模型仅产生 5 个验证折的配对结果, 所以 P 值最小为 0.062 5。该 P 值主要作为补充性统计信息, 性能差异需结合均值、标准差及跨验证折方向一致性共同解释。

采用 SPSS 26.0 及 Python 3.11 软件进行数据分析。对于连续变量数据, 符合正态时, 以 $\bar{x} \pm s$ 表示, 模型间五折交叉验证性能指标的配对比较采用 Wilcoxon 符号秩检验; 不符合正态分布时, 以 $M(Q_1, Q_3)$ 表示, 组间比较采用 Mann-Whitney U 检验。分类变量数据以频数 (百分比) 表示, 组间比较采用 χ^2 检验或 Fisher 精确概率法。所有检验为双侧检验, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 两组婴幼儿患者临床资料

与非脓毒症组相比, 脓毒症组婴幼儿患者烧伤总面积、Ⅲ度烧伤面积均明显更大 ($P < 0.05$), 合并吸入性损伤者比例、CRP 升高者比例、降钙素原升高者比例和白细胞计数均明显升高 ($P < 0.05$), 白蛋白水平明显下降 ($P < 0.05$), PT 明显延长 ($P < 0.05$)。见表 1。

2.2 多模态预测模型的预测性能

多模态预测模型的准确率、精确率、召回率、特异度、F1 分数、AUROC 均优于随机森林、XGBoost、TabTransformer、ResNet、Vision Transformer、Swin Transformer 模型; 同时, 多模态预测模型在创面区域分割中的 DSC 优于 ResNet、Vision Transformer、Swin Transformer 模型。见表 2。

消融实验结果显示, 多模态预测模型的准确率、精确率、召回率、特异度、F1 分数、AUROC 均优于临床量表模型和创面图像模型, DSC 高于创面图像模型, 见表 3。

五折交叉验证显示, 多模态预测模型的准确率、精确率、召回率、特异度、F1 分数、AUROC 优于随机森林模型 (P 值均为 0.062 5)、XGBoost 模型 (P 值分别为 0.062 5、0.062 5、0.875 0、0.062 5、0.062 5、0.062 5)、TabTransformer 模型 (P 值均为 0.062 5)、ResNet 模型 (P 值均为 0.062 5)、Vision Transformer 模型 (P 值分别为 0.125 0、0.062 5、0.187 5、0.062 5、0.062 5、0.062 5)、Swin Transformer 模型 (P 值分别为 0.125 0、0.125 0、0.187 5、0.062 5、0.062 5、

表1 两组烧伤婴幼儿患者临床资料比较

Table 1 Comparison of clinical data between infants and toddlers with burns in sepsis group and non-sepsis group						
组别	例数	烧伤总面积[%TBSA, $M(Q_1, Q_3)$]	Ⅲ度烧伤面积[% TBSA, $M(Q_1, Q_3)$]	合并吸入性损伤[例(%)]	热液烫伤[例(%)]	白细胞计数[$\times 10^9/L$, $M(Q_1, Q_3)$]
脓毒症组	54	40.0(20.0, 60.0)	45.0(30.0, 64.0)	11(20.4)	50(92.6)	13.62(11.27, 17.85)
非脓毒症组	1 655	10.0(6.0, 15.0)	20.0(8.0, 35.0)	82(4.9)	1 511(91.3)	13.41(10.03, 17.50)
统计量值		$Z=8.266$	$Z=7.945$	$\chi^2=4.914$	$\chi^2=1.080$	$Z=1.969$
P值		<0.001	<0.001	<0.001	0.280	0.049

组别	例数	CRP升高[例(%)]	降钙素原升高[例(%)]	白蛋白水平[g/L, $M(Q_1, Q_3)$]	PT[s, $M(Q_1, Q_3)$]
脓毒症组	54	54(100)	53(98.1)	39.25(35.40, 41.85)	12.30(11.40, 13.10)
非脓毒症组	1 655	763(46.1)	631(38.1)	39.90(36.90, 42.60)	11.50(10.70, 11.90)
统计量值		$\chi^2=7.803$	$\chi^2=8.859$	$Z=-2.132$	$Z=-4.247$
P值		<0.001	<0.001	0.033	<0.001

注:TBSA为体表总面积,CRP为C反应蛋白,PT为凝血酶原时间;CRP和降钙素原截断值分别为10 mg/L、0.5 ng/mL^[18-19];所有指标均为入院24 h内数据

表2 多模态婴幼儿烧伤脓毒症风险预测模型与6种对照模型的性能比较($\bar{x} \pm s$,样本数为1 709)

Table 2 Performance comparison between the multimodal risk prediction model for burn sepsis in infants and toddlers and six control models							
模型类别	准确率(%)	精确率(%)	召回率(%)	特异度(%)	F1分数(%)	AUROC	DSC(%)
多模态婴幼儿烧伤脓毒症风险预测模型	99.24±0.34	85.1±5.4	90±6	99.52±0.16	88±5	0.931±0.009	95.1±1.3
随机森林模型	97.66±0.29	57.3±6.1	66±8	98.62±0.16	61±3	0.861±0.031	—
XGBoost模型	98.19±0.49	47.6±15.4	88±14	98.33±0.39	61±17	0.862±0.025	—
TabTransformer模型	97.54±0.65	62.9±1.6	62±12	98.79±0.07	62±7	0.876±0.019	—
ResNet模型	97.54±0.39	68.4±6.1	60±7	98.97±0.17	64±6	0.897±0.015	91.8±1.2
Vision Transformer模型	98.60±0.38	75.8±5.6	80±11	99.21±0.16	77±6	0.900±0.015	93.2±1.3
Swin Transformer模型	98.60±0.52	75.8±5.6	81±14	99.21±0.16	78±7	0.916±0.012	93.4±1.5

注:AUROC为受试者操作特征曲线下面积,DSC为戴斯相似系数;“—”表示无此统计量值

表3 多模态婴幼儿烧伤脓毒症风险预测模型与两种不同输入模态模型的消融实验结果比较($\bar{x} \pm s$,样本数为1 709)

Table 3 Comparison of ablation experiments between the multimodal risk prediction model for burn sepsis in infants and toddlers and two different input modalities models							
模型类别	准确率(%)	精确率(%)	召回率(%)	特异度(%)	F1分数(%)	AUROC	DSC(%)
多模态婴幼儿烧伤脓毒症风险预测模型	99.24±0.34	85.1±5.4	90±6	99.52±0.16	88±5	0.931±0.009	95.1±1.3
临床量表模型	96.82±0.71	61.4±3.6	61±12	98.21±0.09	61±8	0.844±0.014	—
创面图像模型	97.90±0.66	73.4±5.6	80±14	98.64±0.14	77±8	0.901±0.012	92.2±1.6

注:多模态婴幼儿烧伤脓毒症风险预测模型融合了临床表格数据和烧伤创面图像;AUROC为受试者操作特征曲线下面积,DSC为戴斯相似系数;“—”表示无此统计量值

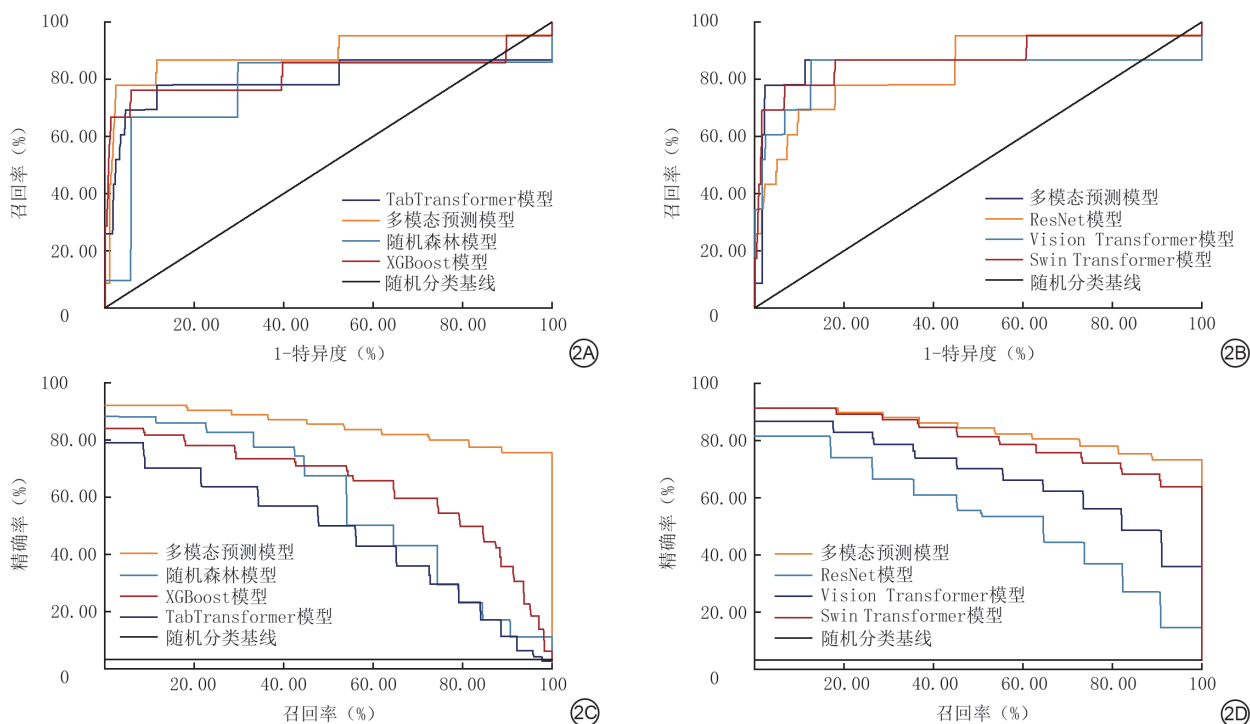
0.062 5);同时,多模态预测模型在创面区域分割中的DSC高于ResNet、Vision Transformer、Swin Transformer模型(P 值均为0.062 5)。即,多模态预测模型较对照模型整体呈现更优的预测性能。

多模态预测模型的平均AUROC和平均精确率分别为0.931、0.898,均高于随机森林模型的0.861与0.672、XGBoost模型的0.862与0.711、TabTransformer模型的0.876与0.705及ResNet模

型的0.897与0.568、Vision Transformer模型的0.900与0.720和Swin Transformer模型的0.916与0.875模型,见图2。

2.3 多模态预测模型的可解释性分析

基于梯度的特征重要性分析显示,PT的归一化重要性权重最大,为1.000,随后依次为舒张压的0.855、白蛋白水平的0.551、CRP的0.543、降钙素原的0.517、中性粒细胞绝对值的0.392、血糖水平



注:多模态预测模型为多模态婴幼儿烧伤脓毒症风险预测模型的简称;临床量表模型包括随机森林、XGBoost、TabTransformer模型,创面图像模型包括ResNet、Vision Transformer和Swin Transformer模型

图2 多模态预测模型与6种对照模型预测性能的比较。2A.临床量表模型与多模态预测模型的受试者操作特征曲线;2B.创面图像模型与多模态预测模型的受试者操作特征曲线;2C.临床量表模型与多模态预测模型的精确率-召回率曲线;2D.创面图像模型与多模态预测模型的精确率-召回率曲线

Figure 2 Comparison of predictive performance between the multimodal prediction model and six control models

的0.349、脉搏的0.332、纤维蛋白原的0.322、白细胞计数的0.284。

可视化分析结果显示,多模态预测模型输出的创面区域与医师标注区域边界基本一致,DSC为95.1%。梯度加权类激活映射热图的高响应区域主要分布于创面及渗出、坏死等视觉表现较明显的区域。见图3。

3 讨论

本研究基于1709例烧伤婴幼儿患者的临床资料数据,构建了多模态预测模型。该模型预测脓毒症的AUROC为 0.931 ± 0.009 ,其准确率、精确率、召回率、特异度和F1分数均高于随机森林、XGBoost、TabTransformer模型及3种创面图像模型。在消融实验中,所构建模型的AUROC高于临床量表模型或创面图像模型,提示临床病理生理信息与创面表征具有互补性;此外,所构建模型在烧伤创面图像分割中的DSC为 $(95.1 \pm 1.3)\%$,高于ResNet、Vision Transformer和Swin Transformer模型。需要强调的是,五折交叉验证的模型间比较,差异均无统计学

意义(最小 P 值为0.0625),因此结果应解释为内部验证中的数值优势,尚不能认定所构建模型确定优于所有对照模型。既往儿科脓毒症预测、数据驱动表型和早期预警研究已提示机器学习的应用潜力^[20-26],本研究进一步补充了0~3岁烧伤婴幼儿患者的多模态证据。

本组烧伤婴幼儿患者的脓毒症发生率为3.16%(54/1709),与既往儿科及烧伤相关报道^[27-29]存在一定差异,可能与病例构成、诊断标准和中心实践不同有关。与非脓毒症组比较,脓毒症组婴幼儿患者烧伤总面积和Ⅲ度烧伤面积更大,合并吸入性损伤者、CRP升高者及降钙素原升高者的比例更高,同时白蛋白水平降低,PT延长。白蛋白水平降低可能同时反映炎症反应、营养状态和毛细血管通透性变化。可解释性分析中,PT的重要性权重最大,随后依次为舒张压、白蛋白水平、CRP、降钙素原和中性粒细胞绝对值等。这些结果与烧伤或脓毒症风险模型及生物标志物研究的总体方向一致^[30],也说明单一烧伤面积或单个炎症指标难以覆盖全部风险信息。



图3 采用多模态婴幼儿烧伤脓毒症风险预测模型对2例婴幼儿烧伤患者进行的创面分割与注意力可视化。3A.入院时婴幼儿患者的烧伤创面原始图像;3B.高年资医师标注的烧伤创面区域金标准;3C.模型的创面区域分割结果;3D.梯度加权类激活映射热图;3E、3F、3G、3H.分别为与图3A、3B、3C、3D对应的另一例婴幼儿患者图像

Figure 3 Wound segmentation and attention visualization in two burned infants and toddlers using the multimodal risk prediction model for burn sepsis in infants and toddlers

临床特征重要性与创面可视化多模态预测模型结果提供了互补解释。临床分支显示凝血、循环、营养和炎症相关变量共同参与风险判别;创面分支的高响应区域主要位于创面及渗出、坏死等视觉表现明显的部位,所构建模型对烧伤创面的分割边界与医师标注基本一致。创面图像分支同时获得较高DSC,也支持可视化位置与分割任务的一致性。该结果有助于核查多模态预测模型是否利用了与烧伤创面相关的信息,并提示图像表征可补充结构化临床数据。基于梯度的特征重要性和创面注意力热图反映了多模态预测模型敏感区域,虽然不能替代病因判断,但可辅助治疗方案的制订。

本模型仅使用入院24 h内可获得的临床量表特征和烧伤创面图像,适合定位为血液标本微生物培养结果明确前的辅助风险分层工具,可为急诊分诊和早期监测提供参考。五折交叉验证及消融实验结果支持临床指标特征和烧伤创面图像的互补价值,也为后续嵌入电子病历的自动化预警提供基础。既往研究已探索深度学习脓毒症模型的临床应用^[31];在此基础上,本研究验证了所构建模型在风险识别、特征互补和结果解释方面的应用潜力,为后续嵌入电子病历、开展前瞻性验证并推动临床转化奠定了较为扎实的基础。

本研究仍有以下局限。首先,单中心回顾性设

计与仅54例阳性患者限制了所构建模型的稳定性与外部适用性,不同地区病例构成和临床实践差异可能造成选择偏倚,需在多中心前瞻性队列中验证。其次,病例跨越11年,其间救治流程、抗菌药物使用策略和重症支持水平可能变化;尽管各折尽量覆盖不同年份病例,仍未行严格的时间亚组分析和时间外验证,不能完全排除时间相关混杂因素。再次,该模型仅纳入烧伤婴幼儿患者入院24 h内静态信息,未反映后续体温、心率、呼吸频率和血压等动态轨迹。最后,既往普通病房病历中部分PSS相关指标记录不完整,可能影响病例判定的完整性。后续应扩大样本量,统一变量及临床表现采集标准,开展外部与多时间验证,并在确认泛化性能后评估多模态预测模型对临床决策和患儿结局的影响。

利益冲突 所有作者声明不存在利益冲突

作者贡献声明 沈宇禾:模型构建、统计分析与论文撰写;章磊、李易、陈昇:收集并整理数据;沈卫民:论文修改、研究指导

参考文献

- [1] 杨宗城,汪仕良,周一平.实用烧伤外科手册[M].北京:人民军医出版社,2001.
- [2] 张小娇,韩素鸽,张茜.《新生儿败血症》摘译[J].中华新生儿科杂志,2026,41(2):127-128. DOI:10.3760/cma.j.cn101451-20250508-00161.
- [3] 王小红,宋洲洋,朱欢,新生儿医院获得性败血症危险因素分析及预测模型建立[J].中华医院感染学杂志,2026,36(2):232-237. DOI:10.11816/cn.ni.2026-258570.

- [4] Drysch M, Reinkemeier F, Puszcz F, et al. Streamlined machine learning model for early sepsis risk prediction in burn patients[J]. NPJ Digit Med, 2025, 8(1): 621. DOI: 10.1038/s41746-025-02078-z.
- [5] Yong L, Zhenzhou L. Deep learning based prediction of in-hospital mortality for sepsis[J]. Sci Rep, 2024, 14(1): 372. DOI:10.1038/s41598-023-49890-9.
- [6] Chaganti S, Singh V, Gent AE, et al. Evaluating the impact of common clinical confounders on performance of deep-learning-based sepsis risk assessment[J]. Front Artif Intell, 2025, 8:1452471. DOI: 10.3389/frai.2025.1452471.
- [7] Har-Shai L, Gershov S, Lagziel T, et al. Machine learning-based predictive model for fever and adverse clinical events in hospitalized pediatric burn patients[J]. J Burn Care Res, 2026, 47(5): 1467-1473. DOI: 10.1093/jbcr/irag074.
- [8] Moor M, Rieck B, Horn M, et al. Early prediction of sepsis in the ICU using machine learning: a systematic review[J]. Front Med (Lausanne), 2021, 8: 607952. DOI: 10.3389/fmed.2021.607952.
- [9] Deng HF, Sun MW, Wang Y, et al. Evaluating machine learning models for sepsis prediction: a systematic review of methodologies[J]. iScience, 2022, 25(1): 103651. DOI: 10.1016/j.isci.2021.103651.
- [10] Kainth D, Prakash S, Sankar MJ. Diagnostic performance of machine learning-based models in neonatal sepsis: a systematic review[J]. Pediatr Infect Dis J, 2024, 43(9): 889-901. DOI: 10.1097/INF.0000000000004409.
- [11] Schlapbach LJ, Watson RS, Sorce LR, et al. International consensus criteria for pediatric sepsis and septic shock[J]. JAMA, 2024, 331(8):665-674. DOI: 10.1001/jama.2024.0179.
- [12] Menon K, Schlapbach LJ, Akech S, et al. Criteria for pediatric sepsis-a systematic review and meta-analysis by the pediatric sepsis definition taskforce[J]. Crit Care Med, 2022, 50(1):21-36. DOI: 10.1097/CCM.0000000000005294.
- [13] Fan J, Shen H, Zhu L, et al. The Phoenix sepsis score criteria in critically ill children: evaluation using a retrospective, single-center PICU cohort in China, 2019-2024[J]. Pediatr Crit Care Med, 2025, 26(12): e1476-e1484. DOI: 10.1097/PCC.0000000000003833.
- [14] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 10012-10022. DOI: 10.1109/ICCV48922.2021.00986.
- [15] Huang X, Khetan A, Cvitkovic M, et al. TabTransformer: tabular data modeling using contextual embeddings[C]//Proceedings of the First Workshop on Weakly Supervised Learning (WeaSuL), ICLR 2021. Virtual: WeaSuL, 2021.
- [16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778. DOI:10.1109/CVPR.2016.90.
- [17] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]//International Conference on Learning Representations. Virtual Event, Austria: OpenReview. net, 2021.
- [18] 刘政枝, 王少华, 苏立贤, 等. 探讨降钙素原及超敏 C-反应蛋白在早产儿医院感染败血症早期诊断意义[J]. 中国实验诊断学, 2022, 26(3):397-400. DOI: 10.3969/j.issn.1007-4287.2022.03.023.
- [19] 王胜云, 陈德昌. 降钙素原和 C-反应蛋白与脓毒症患者病情严重程度评分的相关性研究及其对预后的评估价值[J]. 中华危重病急救医学, 2015, 27(2):97-101. DOI: 10.3760/cma.j.issn.2095-4352.2015.02.00.
- [20] Alpern ER, Scott HF, Balamuth F, et al. Derivation and validation of predictive models for early pediatric sepsis[J]. JAMA Pediatr, 2026, 179(12): 1318-1325. DOI: 10.1001/jamapediatrics.2025.3892.
- [21] Luo W, Xiong L, Wang J, et al. Development and performance evaluation of a clinical prediction model for sepsis risk in burn patients[J]. Medicine (Baltimore), 2024, 103(48): e40709. DOI: 10.1097/MD.00000000000040709.
- [22] Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric sequential organ failure assessment score and evaluation of the Sepsis-3 definitions in critically ill children [J]. JAMA Pediatr, 2017, 171(10): e172352. DOI: 10.1001/jamapediatrics.2017.2352.
- [23] Azizi S, Hoveidamanesh S, Bagheri T, et al. Modern machine learning techniques used in prediction of sepsis and bloodstream infection in burn patients: a systematic review [J]. Burns, 2026, 52(5):107965. DOI: 10.1016/j.burns.2026.107965.
- [24] Qin Y, Kernan KF, Fan Z, et al. Machine learning derivation of four computable 24-h pediatric sepsis phenotypes to facilitate enrollment in early personalized anti-inflammatory clinical trials[J]. Crit Care, 2022, 26(1): 128. DOI: 10.1186/s13054-022-03977-3.
- [25] Shin Y, Cho KJ, Lee Y, et al. Multicenter validation of a deep-learning-based pediatric early-warning system for prediction of deterioration events[J]. Acute Crit Care, 2022, 37(4):654-666. DOI: 10.4266/acc.2022.00976.
- [26] Shi X, Wang X, Yang H, et al. Accurate prediction of sepsis from pediatric emergency department to PICU using a machine-learning model[J]. Front Pediatr, 2025, 13: 1610187. DOI: 10.3389/fped.2025.1610187.
- [27] Bhargava V, George L, Malloy M, et al. The role of a single dose of vancomycin in reducing clinical sepsis in premature infants prior to removal of peripherally inserted central catheter: a retrospective study[J]. Am J Perinatol, 2018, 35(10):990-993. DOI: 10.1055/s-0038-1632391.
- [28] Watson RS, Carrol ED, Carter MJ, et al. The burden and contemporary epidemiology of sepsis in children[J]. Lancet Child Adolesc Health, 2024, 8(9): 670-681. DOI: 10.1016/S2352-4642(24)00140-8.
- [29] Jiang Y, Wang W, Xu R, et al. Construction and efficacy evaluation of a model for early diagnosis of pediatric sepsis based on LASSO-logistic regression[J]. Front Pediatr, 2025, 13:1624278. DOI: 10.3389/fped.2025.1624278.
- [30] Tsurumi A, Flaherty PJ, Que YA, et al. A preventive tool for predicting bloodstream infections in children with burns[J]. Shock, 2023, 59(3):393-399. DOI: 10.1097/SHK.0000000000002075.
- [31] Boussina A, Shashikumar SP, Malhotra A, et al. Impact of a deep learning sepsis prediction model on quality of care and survival[J]. NPJ Digit Med, 2024, 7(1):14. DOI: 10.1038/s41746-023-00986-6.