

机器学习在新生儿坏死性小肠结肠炎手术指征预测中的应用研究进展

申桂源, 薛江*

山东大学齐鲁第二医院儿童医学中心, 山东 济南

收稿日期: 2026年7月28日; 录用日期: 2026年8月22日; 发布日期: 2026年9月1日

摘要

新生儿坏死性小肠结肠炎是早产儿重症监护中的急危重症, 其中约20%~40%的患儿需接受手术治疗, 但手术时机选择困难, 目前临床决策依赖腹部X线、超声等影像学动态评估与经验性判断, 缺乏统一量化预测工具。近年来, 机器学习技术在NEC手术指征预测中得到广泛应用。本文系统综述机器学习在该领域的研究进展, 分析对比现有模型的构建策略、算法特点与预测效能, 总结其优势与不足, 并探讨临床转化面临的挑战与未来方向, 以期机器学习在NEC手术预测中的进一步发展提供参考。

关键词

新生儿, 坏死性小肠结肠炎, 机器学习, 手术指征, 预测模型

Research Progress on the Application of Machine Learning in Predicting Surgical Indications for Necrotizing Enterocolitis in Neonates

Guiyuan Shen, Jiang Xue*

Children's Medical Center, The Second Qilu Hospital of Shandong University, Jinan Shandong

Received: July 28, 2026; accepted: August 22, 2026; published: September 1, 2026

*通讯作者。

文章引用: 申桂源, 薛江. 机器学习在新生儿坏死性小肠结肠炎手术指征预测中的应用研究进展[J]. 临床医学进展, 2026, 16(9): 59-67. DOI: 10.12677/acm.2026.1693020

Abstract

Necrotizing enterocolitis (NEC) is an acute and critical disease in preterm infants under neonatal intensive care. Approximately 20%~40% of affected infants require surgical treatment, yet the selection of optimal surgical timing poses great challenges. At present, clinical decision-making depends on serial imaging evaluations including abdominal radiography and ultrasonography as well as empirical judgment, and standardized quantitative prediction tools are lacking. In recent years, machine learning techniques have been widely adopted for predicting surgical indications of NEC. This article systematically reviews research advances in machine learning in this field, analyzes and compares construction strategies, algorithm characteristics and predictive performance of existing models, summarizes their strengths and limitations, and discusses challenges and future directions for clinical translation, so as to provide a reference for further development of machine learning in surgical prediction of NEC.

Keywords

Neonates, Necrotizing Enterocolitis, Machine Learning, Surgical Indications, Prediction Model

Copyright © 2026 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. 前言

新生儿坏死性小肠结肠炎(necrotizing enterocolitis, NEC)是新生儿重症监护室中最为常见的胃肠道急症之一,严重威胁早产儿,尤其是极低出生体重儿的生命健康[1]。随着围产医学不断进步,早产儿和极低出生体重儿的存活率上升,NEC的发病率也随之提高,约20%~40%的Bell分期III期患儿需接受外科手术干预[2]。且NEC病情进展迅速,数小时至数天内即可由肠壁水肿发展为全层坏死穿孔,留给临床决策的时间窗十分有限[3]。然而,现阶段NEC手术决策仍主要依托主管医师的临床经验,通过观察患儿临床表现及结合腹部X线等影像学检查综合评估手术时机。但约25%~50%合并肠穿孔的NEC患儿影像学并无游离气体征象[4],同时,不同医师对腹部体征、影像学改变的判读存在差异。由此可见,依靠传统手段评估NEC手术时机存在明显局限性。

近年来,机器学习及人工智能NEC辅助诊疗中得到广泛应用[5][6],在NEC手术指征预测领域也取得了显著进展。早期研究主要围绕围产期因素和静态临床参数构建传统预测模型,初步验证了手术型NEC风险分层的可行性。然而,仅凭围产期因素的预测效能有限,提示需引入更多动态临床指标加以改进。在此基础上,研究者将视野拓展至影像组学与多模态数据融合,进一步构建了多模态人工智能系统,整合腹部X线图像与临床参数,结合放射组学特征预测手术时机,实现信息互补,显著提升预测效能。与此同时引入时序预测模型,推动了预测模式从“静态判别”向“动态预警”的转变。上述研究表明机器学习从单模态静态建模到多模态动态预警逐步演进,预测效能逐步提升,但当前研究仍以单中心回顾性分析为主,模型的可解释性与泛化能力尚待进一步验证。本文系统综述机器学习在NEC手术指征预测中的研究进展,总结现有模型的构建策略与算法特点,分析当前面临的挑战,并对未来临床转化方向进行展望,以期为NEC外科决策的精准化和智能化提供参考。

2. 机器学习在 NEC 手术指征预测中的建模方法

2.1. 数据来源与特征筛选

目前, NEC 手术指征预测模型的输入特征可归纳为临床特征、影像特征和时序特征三大类。

临床特征是最基础的数据来源, 主要包括围产期因素(胎龄、出生体重、出生窒息史等)、生命体征(心率、呼吸、血压、体温等)、实验室指标[7](血常规、C 反应蛋白、血小板计数、血气分析、乳酸等)以及喂养耐受情况(腹胀、胃潴留、便血等腹部体征)。Kim 等[8]利用韩国极低出生体重儿全国队列数据, 证实围产期因素在手术型 NEC 预测中具有重要价值。Cho 等[9]进一步发现, 出生后早期的心率变异性异常是 NEC 进展为手术型的重要预警信号。

影像特征的提取是近年来的研究热点[10]。传统方法依赖放射科医师对腹部 X 线平片的主观判读, 而影像组学技术可从平片中定量提取大量高维影像特征(如纹理特征、形状特征、强度特征等), 将影像信息转化为可计算的数据。Li 等[11]的研究表明, 基于腹部 X 线片的影像组学特征联合临床信息可有效预测 NEC 手术需求。此外, 卷积神经网络(convolutional neural network, CNN)可实现端到端的特征自动提取, 无需人工定义特征, 直接从原始图像中学习手术指征相关的潜在病灶模式[12]。

时序特征的引入是机器学习相较于传统方法的显著优势。NEC 的病程呈动态演变, 单一时间点的指标难以完整反映疾病状态。长短期记忆网络(long short-term memory, LSTM) [13]等循环神经网络能够处理不等长的时间序列数据, 捕捉生命体征和实验室指标的动态变化趋势, 实现超前 1~2 d 的手术需求预测。这一能力使模型从“静态判断”升级为“动态预警”, 更契合临床实际需求。

2.2. 常用机器学习算法及其特点

根据算法架构和学习范式, 目前应用于 NEC 手术指征预测的机器学习算法可分为传统机器学习算法和深度学习算法两大类。

传统机器学习算法中, 逻辑回归(logistic regression, LR)是最基础的分类模型, 因其良好的可解释性而被广泛用作基准模型。多项研究表明, LR 在预测 NEC 手术时机中的表现并不逊于更复杂的模型[14]。随机森林(random forest, RF) [15]作为集成学习代表, 通过构建多棵决策树并投票输出结果, 具有抗过拟合、能处理高维数据、可评估特征重要性等优势, 在多项研究中表现优异。支持向量机(support vector machine, SVM)擅长处理小样本和高维数据, 通过核函数将数据映射到高维空间实现非线性分类[16]。Sun 等[17]采用 SVM 结合迁移学习策略, 在无气腹表现的 NEC 早产儿中实现了 AUC 达 0.937 的预测效能。梯度提升算法(如 XGBoost、LightGBM、CatBoost)通过串行迭代地拟合前一轮的残差, 具有预测精度高、训练速度快、内置正则化防过拟合等优点[18]。Jin 等[19]利用 CatBoost 构建的预测模型内部验证 AUC 达 0.905。不同算法的适用场景存在差异: LR 适合作为基线且对可解释性要求较高的场景; RF 和梯度提升算法适合特征维度高、变量间存在复杂交互的数据集; SVM 在小样本高维数据中具有一定优势。

深度学习算法在 NEC 手术指征预测中的应用尚处于起步阶段, 但展现出独特优势。CNN 在影像特征自动提取方面具有天然优势。多中心研究显示, 基于 Efficientnet-b0 的 CNN 模型对手术型 NEC 的预测 AUC 达 0.837 [12]。LSTM 专为处理时间序列数据而设计, 能够记忆长期依赖关系, 适用于基于连续生命体征监测数据的前瞻性手术需求预测, 已有研究报道其预测精确率达 0.913 [13]。多模态融合模型是近年来的前沿方向, 通过融合影像数据与临床数据、实验室数据, 实现信息互补。深度学习的局限在于对数据量要求高、计算资源消耗大且可解释性较差, 但这些不足正随着技术发展逐步改善。

2.3. 模型评价体系

模型评价需从多个维度综合衡量。受试者操作特征曲线下面积(area under the curve, AUC)是最核心的

评估指标, 反映模型区分手术型与非手术型 NEC 的能力。灵敏度与特异度分别反映模型识别手术病例和排除非手术病例的能力, 在临床场景中需根据误判代价权衡二者的优先级。校准曲线用于评估模型预测概率与实际发生率之间的一致性, 校准良好的模型才能为临床决策提供可靠的概率参考。决策曲线分析 (decision curve analysis, DCA) 通过计算不同阈值下的净临床获益来评价模型的临床实用性, 而非单纯追求统计指标的优化。此外, SHAP 等可解释性分析方法可量化各特征对预测结果的贡献, 为临床医生理解模型决策提供依据[20]。

3. 机器学习在 NEC 手术时机预测中的应用进展

3.1. 基于临床与实验室数据的预测模型

3.1.1. 聚焦临床参数的建模

Kim [8] 等利用韩国新生儿网络 (Korean Neonatal Network, KNN) 全国队列数据, 纳入 16,385 例极低出生体重儿 (其中 NEC 患儿 1085 例, 手术型 654 例、非手术型 431 例), 基于出生后 1 周内获取的 38 个围产期变量 (包含母体、产前及产后因素), 构建了预测手术型 NEC 的集成学习模型, AUC 为 0.721。该研究的核心价值在于验证了仅凭围产期因素进行早期风险分层的可行性——这些数据在出生后 1 周内即可获取, 无需等待 NEC 临床症状出现。然而, AUC 仅 0.721 的预测效能表明, 单纯依赖围产期因素尚不足以实现精准预测, 后续研究需引入更多动态临床指标加以改进。

3.1.2. 聚焦动态时序的建模

与静态模型不同, Cui [13] 等采用长短期记忆网络 (LSTM) 处理 NEC 患儿的时间序列数据, 实现超前预测手术需求。该研究纳入 791 例 NEC 患儿 (手术组 257 例、内科组 534 例), 选取 35 个临床特征, 使用 focal loss 作为损失函数以缓解类别不平衡问题。模型在提前 1 天预测手术需求时精确率达 0.913 ± 0.034 , 提前 2 天预测时精确率为 0.905 ± 0.036 。该研究的创新之处在于将 NEC 手术预测从“静态判别”推进至“动态预警”——模型能够利用每日更新的临床数据持续评估手术风险, 更契合 NEC 病情快速演变的临床实际。

3.2. 基于影像组学与多模态数据的预测模型

3.2.1. 基于影像组学特征的建模

Li 等[11] 聚焦于腹部 X 线平片未显示气腹的 NEC 患儿, 评估影像组学对手术决策的预测价值。该研究纳入 171 例 NEC 新生儿, 分别构建基于临床特征、影像组学特征及二者联合的预测模型。结果显示, XGBoost 联合模型在训练集 AUC 达 0.988, 测试集 AUC 达 0.959。这一发现使得在传统影像学判读“阴性”的病例中仍能捕捉到与手术需求相关的潜在信息, 对于临床手术决策具有重要意义。

3.2.2. 多模态融合模型

Gao [18] 等构建了包含特征工程、机器学习和深度学习组件的多模态 AI 系统, 基于 1823 例 NEC 患儿 (1992 张腹部 X 线图像和临床参数), 利用 LightGBM 分类器结合放射组学特征预测手术需求。系统总体准确度达 0.96, 手术预测 AUC 达 0.941。该研究识别出心率、血红蛋白、C 反应蛋白、胎龄、肠鸣音减弱、腹部压痛等关键预测因素。Qi 等[21] 则采用改进的联合非负矩阵分解算法融合临床与影像多模态数据, 通过超图约束挖掘两类数据特征间的高阶关联, 并使用邻接矩阵正则化防止过拟合, 最终确认 14 个临床特征。采用随机森林、支持向量机和逻辑回归三种分类器进行二元分类, 随机森林分类的 AUC 达 0.8。

多模态模型的性能优势突出, 影像与临床数据的融合能够实现信息互补, 弥补单一模态的信息不足。

3.3. 不同模型的效能比较

现有模型的预测效能存在较大差异, AUC 范围介于 0.721 至 0.964 之间。Kim 等基于围产期因素的集成模型 AUC 为 0.721; Qi 等的多模态融合模型结合随机森林分类 AUC 为 0.8; Efficientnet-b0 纯影像模型对手术型 NEC 的 AUC 为 0.837; CatBoost 模型内部验证 AUC 达 0.905; 随机森林在 Li 等的研究中 AUC 达 0.922; Gao 等的多模态 LightGBM 模型 AUC 达 0.941; 最新 Dual Swin Transformer 多模态模型 AUC 达 0.920。这一梯度清晰地表明: 特征维度越丰富、模态融合越充分, 模型的预测效能通常越高。

影响预测效能的关键因素包括以下四个方面。样本量方面, Kim [8]等基于 16,385 例的全国队列数据虽规模最大, 但因仅使用静态围产期因素而效能有限; Gao [18]等纳入 1823 例患儿, 为多模态模型提供了充足的训练基础; 而部分小样本研究(如 Qi 等仅 45 例)的泛化能力受限。数据质量方面, 多中心数据显著优于单中心数据[22], 外部验证是评估模型泛化能力的关键环节。特征选择策略方面, LASSO 回归和 Boruta 算法等[23]规范化特征筛选方法有助于减少过拟合、提升模型可解释性。类别不平衡处理方面, NEC 手术病例相对少见, Cui 等[13]采用 focal loss 作为损失函数有效缓解了这一问题; 此外, 过采样(如 SMOTE)和欠采样也是常用策略。

综上, 当前研究呈现从单模态向多模态融合、从静态预测向动态时序预警演进的趋势。但现阶段相关研究多为单中心回顾性研究, 模型的泛化能力和可解释性仍有欠缺, 后续可开展大样本多中心前瞻性队列研究进一步深入探索。

4. 机器学习模型的临床转化: 挑战与展望

4.1. 技术层面

4.1.1. EHR 数据的基础设施适配

实际电子健康档案(EHR)数据的复杂性是临床转化的主要技术障碍。与实验室中使用的静态、清洗后的数据集不同, 临床 EHR 数据具有高度异质性, 且不同医疗机构的新生儿体征记录、实验室检验指标、影像学报告、生命监护数据的采集频次、字段定义、量化标准尚未统一, 如腹胀、喂养不耐受等 NEC 核心临床指征仍依赖医护主观记录, 缺乏标准化量化数值, 导致模型特征变量无法与 EHR 实时数据精准匹配, 极易出现数据错配、特征缺失、维度冗余等问题。构建可复用、标准化的数据基础设施是模型得以稳定运行的基本前提, 也是建模本身面临的显著挑战。若能依托企业级数据仓库, 将分散于各医疗系统的原始临床数据按通用数据模型进行标准化映射, 形成涵盖患者信息、就诊记录、诊断、用药及实验室结果等核心内容的分析型数据集, 并实现定期更新, 则有望满足临床决策对数据时效性的要求, 为模型实时预测奠定数据基础。然而, 上述基础设施的建设涉及大量数据清洗、字段映射与缺失值处理工作, 其实施复杂度与维护成本显著高于实验室研究阶段, 是模型从回顾性验证走向前瞻性部署的首要技术瓶颈。

4.1.2. 实时数据流的处理瓶颈

在实时数据流搭建层面, NEC 病情进展快, 手术时间窗短, 要求手术指征预测模型具备实时动态预警能力, 而当前临床数据传输模式无法满足模型实时运算需求。一方面, 临床监护数据、检验数据、影像数据存在录入、审核、归档时间差, 无法实现秒级、分钟级同步推送, 导致模型输入数据滞后于患儿真实病情, 难以捕捉 NEC 病情快速恶化的关键窗口期, 无法精准预判紧急手术指征。另一方面, 现有机机器学习模型多为静态批量预测模式, 基于固定时间节点的数据集完成运算, 缺乏针对新生儿动态时序数据的流式运算架构, 无法实现数据实时更新、模型动态迭代预测。对于 NICU 而言, 如何在有限的 IT 基础设施下支撑实时特征提取、模型推理与结果推送, 是亟待解决的技术难题。

4.2. workflow层面

4.2.1. 人机交互的认知负荷

在人机交互模式方面, 现有 NEC 手术指征预测模型多为“黑箱式”输出模式, 仅输出手术风险结果或预测概率值, 缺乏完整的决策依据、特征权重分析与病情演变趋势解读, 可解释性差。对于 NEC 这类高风险新生儿重症, 临床医护人员不仅需要风险预测结果, 更需要明确模型判定手术指征的核心依据, 如是否基于肠壁积气、血气恶化、腹胀进行性加重等关键指标, 以此结合临床经验综合判断。决策逻辑不透明, 这是临床医生难以信任并采纳 AI 工具的核心原因。

可解释性人工智能(Explainable AI, XAI)的引入为此提供了解决方案。SHAP (SHapley Additive exPlanation)是目前应用最广泛的模型解释方法之一[20]。其核心思想源于合作博弈论, 通过计算每个特征对预测结果的贡献, 量化各变量在模型决策中的重要性并进行特征重要性排序。在 NEC 手术指征预测中, SHAP 分析可识别出对手术决策影响最大的关键预测因子, 为临床医生提供“为什么”的答案——当模型提示某例患儿需要手术时, 医生可同时看到该预测所依赖的关键因素(如“CRP 较前日升高 50%”或“腹部 X 线片特定影像组学特征超出阈值”), 从而独立判断模型推理是否合理。这种透明化机制有助于建立临床对 AI 工具的信任, 并为模型的持续优化提供反馈依据, 需要进一步加强完善。

4.2.2. 警报疲劳与临床脱敏

NICU 环境中已有大量监护仪警报, 任何新增的预测性警报都可能加剧临床医生的认知负担。若模型产生大量假阳性警报(即预测需手术但实际保守治疗成功), 医护人员可能逐渐对系统产生“脱敏”, 出现预警忽视、延迟处置等问题, 这反而削弱了系统的临床价值。同时, 模型缺乏分层预警与智能降噪机制, 未结合患儿实际病情动态变化进行差异化阈值调整, 难以区分一过性指标异常与真性病情恶化所致的手术风险升高, 进一步加剧无效警报输出。此外, 临床缺乏标准化的预警处置流程, 未明确不同等级警报的响应时限与复核标准, 导致处置随意性强, 易引发过度诊疗或处置滞后等风险。

4.3. 管理层面

4.3.1. 模型性能监控与 MLOps 体系建设

临床数据分布、诊疗流程、疾病发病特征会随时间、地域、诊疗技术发展发生动态变化, 易引发模型性能漂移, 而现有落地体系缺乏常态化、自动化的模型性能监控机制, 因此模型性能可能因数据漂移或概念漂移而逐渐衰退。这就要求建立标准化的 MLOps (Machine Learning Operations)体系, 涵盖持续监控、定期评估与有序更新的闭环流程。对于 NEC 预测模型, 关键性能指标(敏感性、特异性、AUC、校准度)应设定容忍阈值和预警机制, 一旦超出预设范围即触发临床审查。

4.3.2. 版本迭代的监管合规

现有模型多为单中心、小样本的回顾性分析, 缺乏外部验证, 未建立基于多中心临床真实数据的持续优化机制, 使得模型本身泛化能力薄弱, 难以适配不同医疗机构的诊疗场景与患儿人群特征。另一方面, 行业尚未形成 NEC 专科预测模型的迭代规范, 缺乏数据更新、算法优化、版本更新、前后版本性能对比验证的标准化流程, 模型迭代无统一标准、无质量管控、无效果核验, 极易出现迭代后模型性能不稳定、适配性下降等问题。同时, 模型研发团队与临床运维团队权责分离, 临床应用中发现的预测缺陷、适配问题无法及时反馈至研发端, 导致模型长期停留在初始版本, 无法适配临床诊疗动态发展需求。

4.3.3. 法律责任界定的模糊地带

AI 辅助诊疗模型的临床应用仍处于法律监管空白阶段, 当前法律框架下, 传统产品责任、医疗过失

和职业疏忽等原则难以直接适用于 AI 辅助决策场景。首先, 模型预测结果仅为临床辅助决策依据, 不具备诊疗法律效力, 但当模型出现漏判、误判, 导致手术时机延误、过度手术、患儿不良预后等医疗纠纷时, 无法明确责任主体, 难以界定责任归属, 区分研发团队算法缺陷、临床数据质量问题、医护人员决策失误、运维管理漏洞等不同环节的责任边界。其次, 模型训练数据、迭代数据涉及大量新生儿隐私医疗数据, 数据采集、脱敏、使用、共享环节的合规标准尚不统一, 存在数据泄露、隐私侵权的法律风险。此外, 目前尚无针对新生儿专科 AI 诊疗模型的准入标准、质控规范与监管体系, 模型研发、落地、运维、迭代全流程缺乏统一监管, 临床应用的合规性、安全性、有效性无法得到制度化保障, 极大限制了模型的规模化临床推广。

5. 机器学习在 NEC 预测中的数据科学挑战

5.1. “手术指征”标签的主观性与不确定性

对于 NEC 当前尚无统一的手术指征共识, 决策常依赖临床医生的主观判断与经验积累。绝对指征仅包括肠穿孔(气腹征), 但该征象仅在一部分穿孔患儿中出现, 且穿孔出现时往往意味着病情已进入晚期, 错过了最佳手术时机; 相对指征如门静脉积气、肠壁积气、CRP 及 PCT 升高、血小板减少等, 其特异性与主观性差异较大, 导致决策争议普遍存在。临床医生普遍认为决策困难源于“手术指征缺乏客观标准”, 对于不良结局、同事批评及实施不必要手术的“恐惧”, 这些心理因素都可以影响决策过程[24]。除此之外, 标签噪声对模型的影响不容低估, 不同医师对同一患儿的“应否手术”判断可能不一致, 同一医师在不同时间点的判断亦可能漂移。以这类主观标签训练的模型, 本质上是学习“主流临床实践”, 而非学习“最佳决策”, 其预测边界因而具有内在模糊性。

5.2. 类别不平衡

在整体 NICU 住院人群中, NEC 总体发病率仅 1%~3%, 手术阳性样本占比不足 0.5%, 阴性保守病例占据绝对数据优势。极端失衡的数据分布会引导模型训练偏向整体准确率优化, 优先学习轻症与阴性病例特征, 最终导致模型对高危手术患儿的识别灵敏度、阳性预测值显著降低, 存在漏诊风险。尽管这一不平衡程度远低于某些罕见病场景, 但仍需关注其对模型训练的影响。现有研究已对此有所回应。Cui 等采用 Focal Loss 作为损失函数, 通过降低易分类样本的权重、聚焦于难分类样本, 有效缓解了类别不平衡对模型性能的负面影响。然而, 外部验证的缺乏仍是该领域的重要缺口——目前绝大多数模型仅在单中心数据上验证, 其在不同 NICU 人群中的泛化能力尚待检验。

5.3. 时序数据的不规则采样与缺失

NEC 病情演变是一个动态过程, 手术决策的高度依赖生命体征、血气、血常规、腹部体征等指标的动态时序变化。然而, 临床数据、实验室检查和影像学检查均由临床医生根据病情变化灵活安排, 病情加重时检查频率升高, 好转时则减少。这种“采样与病情相关”的特性, 使时序数据呈现高度不规则性。

Cui 等的数据报告直观反映了这一现象, 当以 1 天为采样间隔时, 35 个特征的平均缺失率高达 72.01%; 延长至 3 天间隔时, 缺失率仍达 62.01%。为应对缺失问题, 该研究采用了前向填充与后向填充策略, 离散特征以-1 填补, 连续特征以组内均值填补。但简单填充方法可能引入偏倚——均值填充会低估数据变异性, 前向填充在长间隔下可能将遥远过去的数值误作当前状态的合理近似。更先进的时序建模方法已在其他医学领域探索, 如结合时间衰减机制的 GRU-D [25]、基于 Neural ODE 的连续患者状态模型[26], 以及通过元素级注意力机制同时处理稀疏测量、缺失模式与时间不规则性的 Transformer 变体[27], 但这些方法尚未在 NEC 手术预测中得到充分应用。

6. 小结

本文系统梳理了机器学习在 NEC 手术指征预测领域的研究进展与转化困境。机器学习在 NEC 手术指征预测领域已取得显著进展。现有研究表明, 机器学习模型能够有效整合临床参数、影像特征和时序数据, 其预测效能已初步显示出优于传统方法的潜力。然而, 该领域整体仍处于从“概念验证”向“临床转化”过渡的早期阶段。

当前的核心瓶颈集中在数据质量与模型泛化两个方面[7], 绝大多数研究为单中心、小样本的回顾性分析, 普遍存在样本类别不平衡、手术标签主观模糊、时序临床数据不规则缺失等数据科学缺陷, 缺乏多中心外部验证, 可解释性不足导致临床信任度低。同时, 模型落地还面临 EHR 系统适配不足、人机协同模式不完善、警报疲劳突出、运维迭代体系缺失及法律权责界定模糊等临床转化难题。机器学习在 NEC 手术指征预测中的定位应当是辅助决策工具, 而非替代临床判断, 旨在为医生提供一个客观、量化的风险评估依据和早期预警参考, 辅助其在有限的决策窗口期内做出更及时、精准的判断。相信在未来通过建立高质量的多中心数据库、推进前瞻性临床验证, 可以不断提升模型质量, 从而为改善 NEC 患儿的长期预后及生存质量提供助力。

参考文献

- [1] Jiang, S., Yan, W., Li, S., Zhang, L., Zhang, Y., Shah, P.S., *et al.* (2020) Mortality and Morbidity in Infants < 34 Weeks' Gestation in 25 NICUs in China: A Prospective Cohort Study. *Frontiers in Pediatrics*, **8**, Article ID: 33. <https://doi.org/10.3389/fped.2020.00033>
- [2] 杜京斌, 黄金狮. 新生儿坏死性小肠结肠炎的外科手术治疗现状[J]. 中国实用儿科杂志, 2024, 39(5): 334-338.
- [3] Sitek, A., Seliga-Siwecka, J., Plotka, S., Grzeszczyk, M.K., Seliga, S., Włodarczyk, K., *et al.* (2023) Artificial Intelligence in the Diagnosis of Necrotising Enterocolitis in Newborns. *Pediatric Research*, **93**, 376-381. <https://doi.org/10.1038/s41390-022-02322-2>
- [4] Maheshwari, A., Singh, R., Mishra, V., Cuna, A., Schwartz, D.M. and Chan, S. (2022) Imaging for Diagnosis and Assessment of Necrotizing Enterocolitis. *Newborn*, **1**, 182-189. <https://doi.org/10.5005/jp-journals-11002-0002>
- [5] 姜齐, 张莉. 人工智能在新生儿坏死性小肠结肠炎辅助诊疗中的研究进展[J]. 中国当代儿科杂志, 2025, 27(10): 1281-1285.
- [6] McElroy, S.J. and Lueschow, S.R. (2023) State of the Art Review on Machine Learning and Artificial Intelligence in the Study of Neonatal Necrotizing Enterocolitis. *Frontiers in Pediatrics*, **11**, Article ID: 1182597. <https://doi.org/10.3389/fped.2023.1182597>
- [7] Duci, M., Verlatto, G., Moschino, L., Uccheddu, F. and Fascetti-Leon, F. (2025) Advances in Artificial Intelligence and Machine Learning for Precision Medicine in Necrotizing Enterocolitis and Neonatal Sepsis: A State-of-the-Art Review. *Children*, **12**, Article 498. <https://doi.org/10.3390/children12040498>
- [8] Kim, S.H., Oh, Y.J., Son, J., Jung, D., Kim, D., Ryu, S.R., *et al.* (2024) Machine Learning-Based Analysis for Prediction of Surgical Necrotizing Enterocolitis in Very Low Birth Weight Infants Using Perinatal Factors: A Nationwide Cohort Study. *European Journal of Pediatrics*, **183**, 2743-2751. <https://doi.org/10.1007/s00431-024-05505-7>
- [9] Cho, H., Lee, E.H., Lee, K. and Heo, J.S. (2022) Machine Learning-Based Risk Factor Analysis of Necrotizing Enterocolitis in Very Low Birth Weight Infants. *Scientific Reports*, **12**, Article No. 21407. <https://doi.org/10.1038/s41598-022-25746-6>
- [10] Hu, M., Ju, X., Zhang, X., Li, X., Zhang, Z., Chen, Z., *et al.* (2025) New Biomarkers to Predict the Need for Surgery of Necrotizing Enterocolitis: A Study Based on Abdominal X-Ray Radiomics and Machine Learning. *BioMedical Engineering Online*, **24**, Article No. 144. <https://doi.org/10.1186/s12938-025-01466-6>
- [11] Li, Y., Wu, K., Yang, H., Wang, J., Chen, Q., Ding, X., *et al.* (2024) Surgical Prediction of Neonatal Necrotizing Enterocolitis Based on Radiomics and Clinical Information. *Abdominal Radiology*, **49**, 1020-1030. <https://doi.org/10.1007/s00261-023-04157-9>
- [12] Weller, J.H., Scheese, D., Tragesser, C., Yi, P.H., Alaish, S.M. and Hackam, D.J. (2024) Artificial Intelligence Vs. Doctors: Diagnosing Necrotizing Enterocolitis on Abdominal Radiographs. *Journal of Pediatric Surgery*, **59**, Article 161592. <https://doi.org/10.1016/j.jpedsurg.2024.06.001>
- [13] Cui, C., Qiu, L., Li, L., Chen, F., Liu, X., Sun, H., *et al.* (2024) A Time Series Algorithm to Predict Surgery in Neonatal

- Necrotizing Enterocolitis. *BMC Medical Informatics and Decision Making*, **24**, Article No. 304. <https://doi.org/10.1186/s12911-024-02695-w>
- [14] Li, X., Zhang, L., Gao, H., Wang, Y., Huang, F., Li, D., *et al.* (2025) The Prediction Models for the Optimal Timing of Surgical Intervention for Necrotizing Enterocolitis: Nomogram vs. Five Machine Learning Models. *Pediatric Surgery International*, **41**, Article No. 260. <https://doi.org/10.1007/s00383-025-06163-y>
 - [15] Lure, A.C., Du, X., Black, E.W., Irons, R., Lemas, D.J., Taylor, J.A., *et al.* (2021) Using Machine Learning Analysis to Assist in Differentiating between Necrotizing Enterocolitis and Spontaneous Intestinal Perforation: A Novel Predictive Analytic Tool. *Journal of Pediatric Surgery*, **56**, 1703-1710. <https://doi.org/10.1016/j.jpedsurg.2020.11.008>
 - [16] Verhoeven, R., Kupers, T., Brunsch, C.L., Hulscher, J.B.F. and Kooi, E.M.W. (2024) Using Vital Signs for the Early Prediction of Necrotizing Enterocolitis in Preterm Neonates with Machine Learning. *Children*, **11**, Article 1452. <https://doi.org/10.3390/children11121452>
 - [17] Sun, D., Xie, C., Zhao, Y., Liao, J., Zhang, Y., Hua, K., *et al.* (2026) Transfer Learning Prediction of Surgical Necrotizing Enterocolitis in Preterm Infants without Pneumoperitoneum on Abdominal X-Ray. *Translational Pediatrics*, **15**, 68-68. <https://doi.org/10.21037/tp-2025-1-867>
 - [18] Gao, W., Pei, Y., Liang, H., Lv, J., Chen, J. and Zhong, W. (2021) Multimodal AI System for the Rapid Diagnosis and Surgical Prediction of Necrotizing Enterocolitis. *IEEE Access*, **9**, 51050-51064. <https://doi.org/10.1109/access.2021.3069191>
 - [19] Jin, X., Sun, W., Li, Y., Su, Y., Xu, L. and Zhu, X. (2025) Use of CatBoost Algorithm to Identify the Need for Surgery in Infants with Necrotizing Enterocolitis. *Frontiers in Pediatrics*, **13**, Article ID: 1465278. <https://doi.org/10.3389/fped.2025.1465278>
 - [20] Lundberg, S.M. and Lee, S. (2017) A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 4768-4777.
 - [21] Qi, G., Huang, S., Lai, D., Li, J., Zhao, Y., Shen, C., *et al.* (2022) An Improved Joint Non-Negative Matrix Factorization for Identifying Surgical Treatment Timing of Neonatal Necrotizing Enterocolitis. *Bosnian Journal of Basic Medical Sciences*, **22**, 972-981. <https://doi.org/10.17305/bjbms.2022.7046>
 - [22] Li, Y., Wu, K., Mou, Y., Zhong, C., Zhang, G., Tan, R., *et al.* (2025) Artificial Intelligence Aids Doctors in Diagnosing Necrotizing Enterocolitis and Predicting Surgery Using Abdominal Radiographs: A Multicenter Study. *Quantitative Imaging in Medicine and Surgery*, **15**, 10722-10737. <https://doi.org/10.21037/qims-2024-2867>
 - [23] Li, Y., Zhang, P., Wu, J., Wang, Y., Chen, Y., Sheng, S., *et al.* (2025) Machine Learning-Based Prediction of Surgical Intervention in Preterm Infants with Necrotizing Enterocolitis: A Retrospective Cohort Study. *Children*, **13**, Article 21. <https://doi.org/10.3390/children13010021>
 - [24] Bethell, G.S., Hall, N.J., Battersby, C., Knight, M. and Darlington, A. (2025) Surgeons and Neonatologists Views about Surgical Decision-Making in Necrotising Enterocolitis. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, **110**, 595-601. <https://doi.org/10.1136/archdischild-2025-328480>
 - [25] Che, Z., Purushotham, S., Cho, K., Sontag, D. and Liu, Y. (2018) Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports*, **8**, Article No. 6085. <https://doi.org/10.1038/s41598-018-24271-9>
 - [26] Chauhan, V.K., Thakur, A., O'Donoghue, O., Rohanian, O., Molaei, S. and Clifton, D.A. (2024) Continuous Patient State Attention Model for Addressing Irregularity in Electronic Health Records. *BMC Medical Informatics and Decision Making*, **24**, Article No. 117. <https://doi.org/10.1186/s12911-024-02514-2>
 - [27] Zeng, Z., Lin, M., Liu, Y., Yao, S. and Gong, X. (2025) HET-VQVAE: A Novel Encoder-Decoder Framework for Irregularly-Sampled Multivariate Time Series of Patients. *Journal of Big Data*, **12**, Article No. 247. <https://doi.org/10.1186/s40537-025-01303-6>