

大数据赋能新生儿发病机制与智能预测研究进展

舒庆霞*, 滕 慧, 姚志军, 颜昌云, 杨惠质

湖南医药学院总医院, 湖南 怀化

收稿日期: 2026年8月4日; 录用日期: 2026年8月27日; 发布日期: 2026年9月7日

摘 要

新生儿疾病发病机制复杂, 传统研究模式受限于数据维度和分析方法, 难以实现精准的早期预警与个体化干预。大数据与人工智能技术的融合为这一领域带来了新的契机。本文简述了大数据技术在新生儿疾病发病机制解析及智能预测中的研究进展, 剖析现存瓶颈, 展望发展方向。

关键词

新生儿疾病, 大数据, 人工智能, 机器学习, 疾病预测, 发病机制

Research Progress on Neonatal Pathogenesis and Intelligent Prediction Empowered by Big Data

Qingxia Shu*, Hui Teng, Zhijun Yao, Changyun Yan, Huizhi Yang

General Hospital of Hunan University of Medicine, Huaihua Hunan

Received: August 4, 2026; accepted: August 27, 2026; published: September 7, 2026

Abstract

The pathogenesis of neonatal diseases is complex, and traditional research paradigms are constrained by data dimensionality and analytical methods, making it difficult to achieve precise early warning and individualized intervention. The integration of big data and artificial intelligence technologies has brought new opportunities to this field. This article briefly reviews the research progress of big data technologies in elucidating the pathogenesis of neonatal diseases and in intelligent

*通讯作者。

文章引用: 舒庆霞, 滕慧, 姚志军, 颜昌云, 杨惠质. 大数据赋能新生儿发病机制与智能预测研究进展[J]. 临床医学进展, 2026, 16(9): 676-681. DOI: 10.12677/acm.2026.1693094

prediction, analyzes the existing bottlenecks, and envisions future directions.

Keywords

Neonatal Diseases, Big Data, Artificial Intelligence, Machine Learning, Disease Prediction, Pathogenesis

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. 引言

新生儿期是儿童生长发育过程中最关键且最脆弱的阶段，也是各种严重疾病发生率和死亡率最高的时期之一。世界卫生组织数据显示，全球每年约有 244 万例新生儿死亡[1][2]，其中早产相关并发症、感染及围产期疾病仍是主要死亡原因。随着围产医学和新生儿重症监护(NICU)技术的持续发展，极低出生体重儿和超早产儿的存活率显著提升。然而，支气管肺发育不良(BPD)、坏死性小肠结肠炎(NEC)及新生儿呼吸窘迫综合征等严重疾病的发生率仍居高不下，是导致新生儿死亡、长期住院及远期神经发育障碍的重要原因。美国每年约有 18,000 例早产儿发生 BPD，其发病率并未随着 NICU 救治水平的提升而明显下降[3]；坏死性小肠结肠炎在极低出生体重儿中的发病率约为 5%~16%，具有起病隐匿、进展迅速、病死率高等特点，临床早期识别窗口极为有限[4]。传统依赖临床经验和单一指标的诊疗模式已难以满足精准预测、早期识别及个体化干预的需求。

近年来，大数据分析 with 人工智能技术的快速发展为新生儿疾病研究带来了新的机遇。依托机器学习、深度学习等算法，可从海量、多维、多模态数据中自动提取关键特征，识别变量间复杂的非线性关系，实现疾病风险预测、早期预警、精准风险分层及临床决策支持，有效提升模型的预测性能和临床应用价值[5]-[10]，推动新生儿医学从经验医学向精准医学、智慧医疗转型。本文简述了大数据技术在新生儿疾病发病机制解析及智能预测中的研究进展，剖析现存瓶颈，展望发展方向。

2. 多源大数据融合下新生儿疾病机制解析与智能预测应用

新生儿疾病发生发展是一个错综复杂、多因素交织的过程，涵盖遗传、环境、免疫、微生物、围产期因素等多方面的共同作用。传统研究多聚焦单一危险因素，难以识别变量间交互效应。多组学高通量数据、纵向临床队列数据、多模态影像与生理时序监测数据构成的多源大数据体系，结合机器学习、深度学习人工智能算法，可实现从分子发病机制、围产期致病交互规律到临床表型特征的全方位解析，同时同步落地疾病风险分层、实时预警等智能预测应用，实现机制研究与临床转化的深度融合。本文以数据类型为核心线索，整合疾病机制挖掘与智能预测模型应用，系统阐述大数据人工智能技术在新生儿重大疾病中的研究进展。

2.1. 多组学大数据：分子机制解析

基因组、转录组、蛋白质组、代谢组、肠道微生物组等多组学数据，可从基因变异、分子表达、代谢紊乱、微生态失衡等微观层面揭示新生儿疾病核心发病机制，同时依托高维分子特征构建智能预测模型，实现疾病早期分子层面的风险筛查与亚型区分，突破传统临床表型预判的局限性。相较于传统单一指标检测，多组学大数据能够捕捉疾病发生前的隐匿分子异常，为新生儿支气管肺发育不良(BPD)、新生儿坏

死性小肠结肠炎(NEC)、早产儿败血症等重症的机制阐释与精准预测提供全新支撑。

在代谢组学方面,Chang 等人基于新生儿出生 48 小时内干血斑常规代谢筛查大数据,整合代谢指标构建新生儿代谢健康指数,突破传统单一筛查模式局限。该指数可精准实现早产儿 BPD、IVH、NEC 及早发性视网膜病变的多疾病风险分层评估,同时通过机器学习算法识别出最具预测价值的代谢物生物标志物,从代谢紊乱维度阐释了不同早产儿肺损伤、肠道损伤的异质性分子机制,解决了相同临床诊断下预后差异显著的机制难题[11]。山东第一医科大学附属省立医院依托中国北方新生儿协作网(SNN)大样本早产儿队列,在《Nature Communications》上发表研究成果。该研究系统梳理了早产儿住院期间常见的生长发育代谢等多种并发症的复杂共病模式,揭示了宫外生长发育迟缓(EUGR)合并肠外营养相关性胆汁淤积, EUGR 合并 BPD 等的共患模式,为早产儿复合重症的机制解析与风险预判提供了重要的本土数据支撑[12]。

在微生物组学研究方面,肠道微生态紊乱是早产儿 NEC、败血症等感染性、炎症性疾病的核心诱因。Karaca 等人通过 16S rRNA 宏基因组高通量测序技术,纳入大样本早产儿纵向菌群数据,系统对比 NEC 患儿与健康新生儿的肠道菌群定植特征,证实早产儿肠道菌群定植时序紊乱、有益菌丰度不足、条件致病菌过度增殖是 NEC 发生的核心分子机制,且在门、属、种多水平识别出 NEC 特征性菌群标志物[13]。

在基因组与转录组学层面,全基因组关联分析(GWAS)大数据挖掘发现,FGFR-3 和 FGFR-4 基因多态性变异与早产儿 BPD 肺间质纤维化、肺发育停滞密切相关[14];转录组学高通量测序识别出 IL-6、TNF- α 等炎症基因的异常持续表达模式,阐释了围产期炎症暴露诱发新生儿肺、脑多脏器损伤的核心机制[15]。

2.2. 纵向临床大数据:围产期致病交互机制挖掘与静态风险分层预测

围产期母婴暴露因素具有时序性、关联性、多交互性特征,传统横断面研究无法捕捉孕前、孕期、分娩期全程危险因素的动态作用规律。电子健康档案(EHR)积累的海量纵向临床大数据,完整覆盖母体基础状态、围产期并发症、分娩信息及新生儿出生后基线特征、诊疗轨迹,可系统解析“母-胎-婴”连续体的致病交互网络,同时依托基线静态数据构建智能预测模型,实现新生儿入院早期高危人群精准分层。

机制研究层面,传统研究多将早产、低出生体重作为新生儿重症的核心单一诱因,而纵向临床大数据的深度挖掘打破了这一认知局限。De Francesco 等人基于两个独立队列共 27,000 名母亲的 EHR 数据,关联超过 32,000 名新生儿,利用深度学习开展了从围产期早期至生后数月内多种不良结局的连续动态预测[16]。研究发现新生儿坏死性小肠结肠炎、支气管肺发育不良、动脉导管未闭以及脑瘫的风险并非仅由早产等新生儿自身因素驱动,而是与母体状态(如贫血)、产前特定药物暴露(包括吡哆美辛、西地那非)以及社会健康决定因素(如无家可归、监禁经历)存在显著关联;与此同时,新生儿脑室内出血的风险则与母体阿片类药物暴露、孕产妇肾功能衰竭及金黄色葡萄球菌定植状态等母源性因素密切相关。

临床预测应用层面,依托纵向临床基线数据构建的静态机器学习模型,是 NICU 入院早期风险分层的核心工具,可快速识别高危新生儿,指导早期精准干预。Shu 等人基于 3341 例极低出生体重儿回顾性队列,整合 47 项母体围产期、新生儿出生早期基线变量,采用随机森林、XGBoost、LightGBM 等 7 种主流机器学习算法,构建新生儿死亡、BPD、NEC、败血症、IVH 多结局预测模型[17]。研究证实树模型集成算法对各类重症的区分效能显著优于传统逻辑回归统计模型,所有结局的精确率-召回率曲线下面积均显著高于疾病基线患病率,可精准完成极低出生体重儿入院早期风险分层[17]。相较于传统模型,该机器学习模型可自动挖掘多变量非线性交互效应,精准量化母体高危因素、新生儿基线指标与疾病结局的关联强度,解决了传统模型变量筛选片面、预测精度低的问题。

2.3. 多模态影像与生理时序大数据:临床表型机制解析与动态预警预测

新生儿重症存在显著临床异质性,相同诊断下患儿脏器损伤程度、远期预后差异极大,传统静态诊

断模式无法区分疾病内表型、捕捉隐匿性损伤。多模态影像大数据可揭示脏器结构微损伤机制，持续生理监测时序数据可反映功能动态异常规律，二者融合结合时序机器学习算法，既可阐释疾病异质性表型的核心机制，又能实现新生儿病情实时动态预警，破解传统诊疗滞后、预判不准的痛点。

影像组学大数据聚焦脏器结构微表型机制与远期预后预测。新生儿脑损伤、BPD、持续性肺动脉高压等重症的隐匿性结构损伤无法通过常规影像肉眼识别，是导致预后异质性的核心原因。Wagner 等人基于极早产儿生后早期头颅 MRI 影像大数据，提取高维影像组学空间特征，通过机器学习模型筛选出脑白质微结构损伤、灰质发育异常的特异性特征图谱，证实这些肉眼不可见的微结构损伤是极早产儿远期神经运动功能障碍的核心机制，且影像组学特征对远期不良结局的预测性能显著优于传统临床评分与影像指标[18]。首都医科大学附属北京儿童医院最新研究构建婴幼儿脑发育规范化轨迹模型，基于临床二维磁共振影像(2D MRI)和影像组学特征实现个体脑发育水平的定量评估，精准区分生理性发育滞后与病理性脑损伤，进一步完善了新生儿脑损伤的表型分型机制与预后预测体系[19]。

生理时序信号大数据聚焦功能动态紊乱机制与实时病情预警。NICU 新生儿需持续监测心率、呼吸、血氧饱和度、脑电、脑氧等高频生理信号，海量时序数据中蕴含疾病早期隐匿功能异常规律。Steiner 等人联合振幅整合脑电图、NIRS 脑氧监测多模态时序数据，构建深度学习融合模型，精准捕捉缺氧缺血性脑病新生儿早期脑功能电生理波动、脑氧代谢紊乱的动态特征，阐释了缺氧、炎症双重损伤路径下的差异化神经损伤亚型机制，该模型对远期不良神经结局的预测 AUC 可达 0.96~0.99，实现了早期隐匿损伤的精准识别[20]。

在动态预警应用方面，时序机器学习模型突破了传统静态评估、固定阈值判断的局限，可实现新生儿疾病全程实时风险追踪。Bennis 等人纳入 513 例早产儿临床数据以及呼吸与氧合时序数据，构建 BPD 动态预测模型，AUROC 达 0.83，显著优于仅依托静态临床数据的预测模型，可精准捕捉生后一周内 BPD 的早期功能异常信号，实现病情渐进性进展的动态预警[21]。最新研究进一步优化多模态时序融合模型，整合心率变异性、呼吸节律、血氧波动、脑电信号多维时序特征，构建新生儿重症早期预警体系(EWS)，可提前 24~72 小时预判败血症、NEC、重度脑损伤的发病风险，有效弥补了 NICU 传统被动诊疗的短板[22]。

3. 人工智能在 NICU 临床应用中面临的挑战和发展展望

尽管人工智能在新生儿疾病机制解析、风险分层与实时预警中展现出良好应用价值，但目前在 NICU 临床落地仍存在诸多特异性瓶颈。NICU 高频监护产生海量生理时序数据，但新生儿躁动、体位变动、设备干扰等问题导致数据噪声大、有效信号占比低、缺失值多，难以精准区分生理性波动与病理性异常，直接降低模型预测稳定性与准确性；同时早产儿处于快速动态发育阶段，生理特征与疾病规律随胎龄、日龄持续演变，现有人工智能模型多基于单中心回顾性静态数据训练，缺乏多中心前瞻性外部验证，泛化能力与时间时效性不足，难以适配新生儿全病程动态诊疗需求；此外，NICU 研究涉及母婴双联敏感数据，存在双重隐私泄露风险，而目前尚无针对新生儿人工智能诊疗的专属伦理监管规范，数据共享脱敏、算法公平性管控体系不完善，易出现小众胎龄群体的算法歧视问题；且深度学习模型普遍存在可解释性差的“黑箱”缺陷，模型决策逻辑不透明，叠加各医疗中心数据标准不统一、多源数据异构性强等问题，极大限制了人工智能模型的临床认可度与规模化推广应用。

针对上述 NICU 场景下的应用短板，未来可从数据优化、模型迭代、伦理合规、技术转化多维度推进人工智能技术的临床落地与升级。后续需建立 NICU 多中心统一的数据采集、质控与脱敏标准，研发新生儿专属数据降噪与修复算法，规范多源异构数据格式，从源头提升建模数据质量；依托增量学习、迁移学习技术构建动态自适应智能模型，适配早产儿不同发育阶段的生理特征与疾病演变规律，解决传

统模型时效性弱、泛化性差的问题,实现全病程精准风险预警;基于联邦学习、隐私计算等技术开展跨中心协同建模,在保障母婴敏感数据安全、实现数据不出院的前提下构建大样本训练数据集,同时完善新生儿人工智能临床应用专属伦理规范,明确数据权属、知情同意与算法公平评价标准,筑牢技术合规发展底线;最后大力发展可解释人工智能技术,明晰模型决策机制与核心致病关联特征,结合多模态融合算法与大语言模型,搭建覆盖风险筛查、实时预警、辅助诊断、个体化干预的全流程 NICU 智能诊疗体系,推动新生儿医学向精准化、智能化、个体化方向高质量发展。

致 谢

感谢吉首大学科研项目(Jdzd24003)的支持。

基金项目

吉首大学科研项目(编号 Jdzd24003)。

参考文献

- [1] WHO (2020) The Top 10 Causes of Fudan University Death. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [2] WHO (2021) World Patient Safety Day 2021. <https://www.who.int/news-room/events/detail/2021/09/17/default-calendar/world-patient-safety-day-2021>
- [3] McOmber, B.G., Moreira, A.G., Kirkman, K., Acosta, S., Rusin, C. and Shivanna, B. (2024) Predictive Analytics in Bronchopulmonary Dysplasia: Past, Present, and Future. *Frontiers in Pediatrics*, **12**, Article 1483940. <https://doi.org/10.3389/fped.2024.1483940>
- [4] 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.
- [5] Hulsén, T., Jamuar, S.S., Moody, A.R., Karnes, J.H., Varga, O., Hedensted, S., *et al.* (2019) From Big Data to Precision Medicine. *Frontiers in Medicine*, **6**, Article 34. <https://doi.org/10.3389/fmed.2019.00034>
- [6] Nielsen, A.N. and Smyser, C.D. (2025) Machine Learning for Outcome Prediction in Preterm Infants: Opportunities and Challenges. *Pediatric Research*, **98**, 759-761. <https://doi.org/10.1038/s41390-025-04131-9>
- [7] Pablo, R.J., Roberto, D., Victor, S., Isabel, G., Paul, C. and Elizabeth, O. (2022) Big Data in the Healthcare System: A Synergy with Artificial Intelligence and Blockchain Technology. *Journal of Integrative Bioinformatics*, **19**, Article ID: 20200035. <https://doi.org/10.1515/jib-2020-0035>
- [8] Shilo, S., Rossman, H. and Segal, E. (2020) Axes of a Revolution: Challenges and Promises of Big Data in Healthcare. *Nature Medicine*, **26**, 29-38. <https://doi.org/10.1038/s41591-019-0727-5>
- [9] Teji, J.S., Jain, S., Gupta, S.K. and Suri, J.S. (2022) NeoAI 1.0: Machine Learning-Based Paradigm for Prediction of Neonatal and Infant Risk of Death. *Computers in Biology and Medicine*, **147**, Article ID: 105639. <https://doi.org/10.1016/j.compbiomed.2022.105639>
- [10] Zhang, Z., Xiao, Q. and Luo, J. (2023) Infant Death Prediction Using Machine Learning: A Population-Based Retrospective Study. *Computers in Biology and Medicine*, **165**, Article ID: 107423. <https://doi.org/10.1016/j.compbiomed.2023.107423>
- [11] Chang, A.L., Reiss, J.D., Culos, A., Becker, M., Mayo, J.A., Marić, I., *et al.* (2026) Quantitative Assessment of Neonatal Health Using Dried Blood Spot Metabolite Profiles and Deep Learning. *Science Translational Medicine*, **18**, eadv4942. <https://doi.org/10.1126/scitranslmed.adv4942>
- [12] Yang, C., Shi, R., Yuan, J., Ding, Y., Geng, H., Li, H., *et al.* (2026) Hypertensive Disorders of Pregnancy Contribute to Developmental and Metabolic Multimorbidity in Preterm Infants. *Nature Communications*, **17**, Article No. 8855. <https://doi.org/10.1038/s41467-026-75869-x>
- [13] Karaca, C. and Takcı, H.A.M. (2025) Role of Gut Microbiome in Developing Necrotizing Enterocolitis. *Folia Microbiologica*, **70**, 197-204. <https://doi.org/10.1007/s12223-024-01217-5>
- [14] Abdellatif, M.A.K., Eyada, E., Rabie, W., Abdelaziz, A. and Shahin, W. (2022) Genetic and Biochemical Predictors of Neonatal Bronchopulmonary Dysplasia. *Journal of Pediatric Genetics*, **11**, 173-178. <https://doi.org/10.1055/s-0040-1721740>
- [15] Ahmed, A.E., Ibrahim, F.E., Linga, B.S.G., Alqudah, A.M. and Al-Dewik, N. (2025) Comparative Analysis of Preterm

- and Full-Term Births Using Transcriptomic Profiling. <https://doi.org/10.57945/manara.32051301>
- [16] De Francesco, D., Reiss, J.D., Roger, J., Tang, A.S., Chang, A.L., Becker, M., *et al.* (2023) Data-Driven Longitudinal Characterization of Neonatal Health and Morbidity. *Science Translational Medicine*, **15**, eadc9854. <https://doi.org/10.1126/scitranslmed.adc9854>
- [17] Shu, C., Zebda, R., Espinosa, C., Reiss, J., Debuyserie, A., Reber, K., *et al.* (2025) Early Prediction of Mortality and Morbidities in VLBW Preterm Neonates Using Machine Learning. *Pediatric Research*, **97**, 2056-2064. <https://doi.org/10.1038/s41390-024-03604-7>
- [18] Wagner, M.W., So, D., Guo, T., Erdman, L., Sheng, M., Ufkes, S., *et al.* (2022) MRI Based Radiomics Enhances Prediction of Neurodevelopmental Outcome in Very Preterm Neonates. *Scientific Reports*, **12**, Article No. 11872. <https://doi.org/10.1038/s41598-022-16066-w>
- [19] Luo, H., Li, H., Xiong, S., Wang, R., Li, Z., Wang, J., *et al.* (2026) Clinical 2D MRI Radiomics for Infant Brain Age and Trajectory Modeling. *iScience*, **29**, Article ID: 116623. <https://doi.org/10.1016/j.isci.2026.116623>
- [20] Steiner, M., Urlesberger, B., Giordano, V., Kasprian, G., Glatter, S., Oberleitner-Leeb, C., *et al.* (2022) Outcome Prediction in Neonatal Hypoxic-Ischaemic Encephalopathy Using Neurophysiology and Neuroimaging. *Neonatology*, **119**, 483-493. <https://doi.org/10.1159/000524751>
- [21] Bennis, F.C., Onland, W., van der Vorst, J.P., Hoogendoorn, M., Hutten, G.J., van Kaam, A.H., *et al.* (2026) Prediction of Bronchopulmonary Dysplasia Seven Days after Birth Using Respiratory and Oxygenation Timeseries with Machine Learning. *Pediatric Research*. <https://doi.org/10.1038/s41390-026-05301-z>
- [22] Huo, H., Lu, Y., Zhou, J., Zhang, L., Pei, J. and Hu, C. (2026) High-Throughput Analysis of Multimodal Monitoring Data: The Role of Machine Learning in Early Warning Systems for High-Risk Neonates. *BMJ Paediatrics Open*, **10**, e004705. <https://doi.org/10.1136/bmjpo-2026-004705>