
Review

Opportunities and Challenges of Using Artificial Intelligence in Predicting Clinical Outcomes and Length of Stay in Neonatal Intensive Care Units: Systematic Review

Samantha Tudor¹, MDH; Risha Bhatia², PhD; Michael Liem¹, PhD; Tafheem Ahmad Wani¹, PhD; James Boyd¹, PhD; Urooj Raza Khan¹, PhD

¹La Trobe University, Melbourne, Australia

²Monash University, Melbourne, Australia

Corresponding Author:

Samantha Tudor, MDH

La Trobe University

Plenty Rd Bundoora

Melbourne, 3086

Australia

Phone: 61 0422193167

Email: samantha_tudor@outlook.com

Abstract

Background: The use of artificial intelligence (AI) in health care has been steadily increasing for over 2 decades. Integrating AI into neonatal intensive care units (NICUs) has promise as it has the potential to reshape neonatal care and improve outcomes. However, challenges such as data quality, clinical interpretation, and ethical considerations may hinder AI's practical implementation in NICUs.

Objective: This study aims (1) to analyze the current AI research landscape for predicting clinical outcomes and length of stay in the NICU and (2) to explore the benefits and challenges of using AI in the NICU for these predictions.

Methods: A systematic review was conducted across 6 databases—PubMed, Embase, CINAHL, Cochrane Library, Informit, and La Trobe Library—to identify English-language peer-reviewed articles published between January 2017 and March 2023 that focused on the use of AI for predicting length of stay and clinical outcomes for NICU patients. Eligibility criteria excluded studies outside the NICU context or lacking predictive focus. Both prospective and retrospective designs were included. A thematic analysis of AI applications in NICUs from the articles identified was conducted.

Results: A total of 24 studies were included in the review, comprising 15 retrospective and 9 prospective designs. These studies primarily originated from the United States (13 studies), with others from Austria, Taiwan, and other countries. The studies evaluated AI applications in NICU settings to predict comorbidities (18/24), mortality (4/24), and length of stay (2/24). Sixteen studies were in the exploration stage, lacking cohesive AI strategies, while 8 demonstrated systematic exploration but no fully integrated solutions. The synthesis of results identified key applications of AI in NICU care, including data-driven insights and predictive models, advancements in medical imaging, improved risk stratification, and personalized neonatal care. AI showed promise in enhancing diagnostic accuracy and care planning, but significant challenges persist, such as data quality, model generalization, and ethical concerns. No studies reported a fully integrated AI ecosystem, highlighting the need for further research to bridge gaps and realize AI's transformative potential in neonatal care.

Conclusions: This review highlights the potential of AI in improving NICU care, particularly through predictive models, medical imaging, and personalized interventions. However, the evidence is limited by significant methodological variability, small sample sizes, risk of bias, and a lack of external validation in included studies. Many studies remain in exploratory phases without cohesive AI strategies or integration into clinical practice, limiting the practical applicability of findings. These results underscore the importance of addressing challenges such as data quality, model generalization, and ethical considerations to fully realize AI's potential in neonatal care. Future research should focus on robust validation, comprehensive implementation strategies, and ethical frameworks to ensure AI's effective and responsible integration into NICU settings.

(*J Med Internet Res* 2025;27:e63175) doi: [10.2196/63175](https://doi.org/10.2196/63175)

KEYWORDS

artificial intelligence; deep learning; machine learning; neonatal intensive care unit; NICU; predictive modeling; length of stay; clinical outcomes; systematic review

Introduction

Background

Neonatal intensive care units (NICUs) are specialized hospital units providing intensive medical care for critically ill newborns, particularly those born prematurely or with medical and surgical conditions, where close monitoring and specialized interventions and treatment are needed. One in 10 babies is born prematurely or sick, emphasizing the essential role of neonatal care in promoting their well-being and survival [1]. With an estimated 15 million premature births annually and around 1 million child deaths due to preterm complications each year [2], the significance of specialized care is evident. Surviving infants may encounter lifelong challenges, highlighting the need for effective neonatal care to ensure healthy development and improved long-term outcomes.

Within the complex ecosystem of the NICU, health care providers rely on the expertise of skilled clinicians and essential medical devices to deliver specialized care. The implementation of electronic medical records has made vast volumes of data more accessible in recent years, contributing to significant advances in medical research and technology, leading to improved survival rates for NICU patients [3]. However, these advancements come with increased demands on health care resources, including specialized staff, equipment, and facilities. Innovative approaches are essential to address these challenges and enhance overall health care efficacy.

One avenue for improvement is the prediction of the length of stay in the NICU. Accurate length of stay predictions are crucial for resource allocation, discharge planning, and optimizing care pathways [4]. Prolonged stays can have significant implications on neonatal development, parent-newborn interactions, and family well-being [5,6]. Additionally, predicting clinical outcomes is equally important, as it allows for monitoring essential developmental milestones, potential delays or improvements, and the lasting effects of medical interventions on NICU patients [7-12].

Health care providers face several challenges when predicting length of stay or clinical outcomes in the NICU setting. The vast amounts of data, including patient vitals, laboratory results, imaging data, and clinical notes, can be overwhelming to process manually [13]. Limited availability of human resources in high-stress NICU settings further compounds the challenge, where teams must balance heavy workloads and rapid decision-making [14].

Furthermore, the dynamic nature of neonatal care, marked by rapid changes in health status, adds another layer of complexity to accurate predictions. In this context, artificial intelligence (AI) technology emerges as a potential solution. AI refers to the capability of computer systems to carry out tasks that typically require human intelligence, including prediction, learning, and decision-making [15]. While traditional software

and other computer systems rely on predetermined rules and instructions to accomplish specific tasks, AI sets itself apart by its capacity to learn from data and continually enhance performance without explicit programming. This capability is primarily achieved through a branch of AI called machine learning [16].

Machine learning encompasses a diverse array of techniques, including supervised learning, where models are trained on labeled datasets (eg, historical patient outcomes); unsupervised learning, which aims to discover hidden structures and patterns within unlabeled data; and reinforcement learning, a paradigm focused on optimizing decision-making through iterative interactions with an environment and the receipt of rewards. Deep learning, a specialized form of machine learning, uses multilayered neural networks that mimic the human brain's processing abilities to uncover complex relationships in large datasets. In the context of this study, these techniques are especially valuable for predicting neonatal outcomes, such as length of stay or health conditions, using data from electronic medical records. By leveraging machine learning and deep learning, AI has the potential to transform decision-making in NICUs, providing more accurate, timely, and personalized insights to support neonatal care [12].

In health care, predictive analytics powered by AI has proven effective for disease diagnosis and risk stratification, offering insights from patient data, medical records, and imaging results [17,18]. However, AI integration in the NICU environment is not without its challenges. Concerns related to data privacy, algorithm bias, and the need for transparent and interpretable models raise ethical considerations [19,20]. Understanding these challenges is essential for the responsible and successful implementation of AI for predictive analytics in neonatal care.

Despite the growing body of research, existing reviews in this field [12,21,22] have primarily focused on technical performance metrics, such as model readiness and diagnostic accuracy, while overlooking critical themes such as stakeholder engagement, ethical considerations, and data integration. For example, Schouten et al [22] focus on model evaluation but overlook broader integration opportunities. Adegboro et al [21] highlight health impacts but neglect barriers like trust and collaboration, and McAdams et al [12] emphasize predictive performance without addressing implementation challenges such as data quality and explainability. To address these gaps, this study synthesizes recent research (2017-2023) through a thematic lens, providing actionable insights into opportunities such as predictive modeling and personalized neonatal care, as well as challenges such as ethical issues and stakeholder trust. A systematic review design was chosen to ensure a comprehensive and methodologically rigorous synthesis of available literature on AI applications in NICU settings. This approach allows for structured evidence appraisal and thematic synthesis across diverse study designs and outcomes, providing robust and

evidence-based recommendations to guide clinical, policy, and research decisions in this rapidly evolving field.

The importance of advancing research in this area becomes even more apparent when considering that other researchers have highlighted the lack of investigation into the practical implementation of AI in intensive care units. For instance, Vellido et al [23] lament the paucity of research on success factors for implementing machine learning in intensive care units, while Kim et al [24] point out that although challenges to implementing these technologies are acknowledged, they remain underexplored. This study seeks to address these gaps by providing a comprehensive and updated synthesis to guide future research and practical implementation in NICUs.

Research Focus and Aims

This review aimed to address the research question, “What are the opportunities and challenges in using AI in the NICU to predict length of stay and clinical outcomes?”

The objective was to analyze the current AI research landscape for predicting clinical outcomes and length of stay in the NICU, exploring the benefits and challenges of using AI in the NICU for these predictions. This will help identify gaps in AI applications for predicting clinical outcomes and length of stay in the NICU and offer recommendations for future research.

Methods

Methodology Overview

A systematic review methodology was applied to provide a robust and transparent synthesis of evidence related to AI applications in neonatal intensive care. This approach followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and involved systematic search, screening, data extraction, quality appraisal, and thematic analysis across multiple databases [25]. The review was tailored to the digital health context by focusing on studies with predictive outcomes in the NICU, aligning with current research priorities and ensuring methodological depth.

To enhance relevance and address recent advances, a targeted supplementary search of literature published post-March 2023 was conducted to identify key developments for discussion. These findings were integrated thematically into the Discussion section.

Database Search

English language peer-reviewed studies between January 2017 and March 2023 were searched in 6 databases (Embase, MEDLINE, CINAHL, Cochrane, Informit, and La Trobe Library). The search used relevant search strings and Medical Subject Headings (MeSH). The keywords included artificial intelligence, NICU, length of stay, and outcome, as detailed in [Table 1](#).

Table 1. Search method and number of results per database.

Database	Search method	Results, n
Ovid (MEDLINE and Embase)	(exp Artificial Intelligence/ OR AI OR “artificial intelligence” OR “machine learn*” OR “deep learn*” OR “neural network*”) AND (exp Intensive Care, Neonatal/ OR exp Infant, Premature/ OR exp Infant, Premature, Diseases/ OR exp Intensive Care Units, Neonatal/) AND (“length of stay” OR “LoS”).mp. OR (“outcome” OR “prognosis”).mp.)	249 (64 and 185)
CINAHL	((artificial intelligence OR ai OR a.i. OR “machine learning” OR “deep learning”) AND (nicu OR “neonatal intensive care unit” OR “special care” OR “baby unit” OR “newborn intensive care”)) AND (“length of stay” OR los OR “in-patient stay” OR “time in hospital” OR “time to discharge”) OR (outcomes OR prognosis) AND (English Language) AND (Published Date: 20170101-20230131)	21
Cochrane	(MH “Artificial Intelligence”) AND (MH “Intensive Care Units, Neonatal”)	1
Informit	“Artificial Intelligence” AND “Intensive Care Units, Neonatal”	0
La Trobe Library	((AI OR “artificial intelligence” OR “machine learn*” OR “deep learn*” OR “neural network*”) AND (NICU OR “neonatal intensive care unit*” OR “newborn intensive care unit*”) AND (“length of stay” OR “LoS” OR “outcome” OR “prognosis”)) AND (English language and last 5 years)	541

Eligibility Criteria

The inclusion criteria focused on NICU patients and the adoption of AI technology for predicting outcomes and length of stay. Outcomes were defined as measures related to the adoption of AI technology for predicting length of stay and clinical outcomes in NICU patients, including mortality, morbidity, growth and development, and other clinical parameters up to 5 years of age. This framework was consistently applied during the eligibility assessment to ensure a clear and systematic approach. Both retrospective and prospective studies were included for a comprehensive view. Retrospective studies offer historical insights through electronic health records, while prospective studies provide real-time data for dynamic observations. The

combination ensured a comprehensive evaluation, considering AI's opportunities and challenges in the NICU. Studies within the last 5 years, in English, and with full-text availability were included, aiming to understand AI's role in predicting NICU outcomes and length of stay. [Multimedia Appendix 1](#) details the reasons for exclusion of studies during the screening process.

Quality Appraisal

The quality appraisal process was conducted using a structured approach to ensure rigor and minimize bias. Multiple authors were involved in the selection and coding of studies to enhance reliability and reduce subjective influence. Initially, ST screened titles and abstracts for relevance, whereas URK and RB independently reviewed the screening, following which

discrepancies were resolved through discussion or consultation during regular meetings.

The included studies underwent quality assessment using modified WHO (World Health Organization) guidelines [26] and preferred QCC (Quality Checklist for Clinical Case Series) [27] due to varied study designs. [Multimedia Appendix 2 \[28-51\]](#) contains detailed ratings, evaluating bias and quality.

Data Extraction

Relevant data from the selected studies were systematically extracted and recorded in Microsoft Excel. A customized template, designed collaboratively by the research team specifically for this systematic review, was used for data extraction. The template included key features, such as study characteristics (background, methods, results, conclusions, and impact), study details (type of study, keywords, country, and number of participants), technology-related information (technology type and algorithm or model details), outcome (length of stay or outcome), evaluation measures, as well as insights on potential opportunities, challenges, discussed gaps, further research needs, and suggested improvement areas. The extracted data were organized and analyzed for synthesizing key findings and identifying and coding emerging themes relating to the research objectives using Thomas and Harden's method of thematic analysis [52]. Studies identified for full-text review were coded to identify recurring concepts, which were grouped into descriptive themes. Analytical themes, such as those related to challenges and opportunities, were developed to interpret the findings and provide actionable insights. This method ensured a structured and transparent synthesis of diverse qualitative data. The identification and development of themes were conducted iteratively, with refinements made after each round of discussions among the research team to achieve consensus and ensure accuracy. This collaborative process ensured a comprehensive and balanced synthesis of the evidence, leveraging diverse expertise to strengthen the review's findings.

Assessment of Maturity Levels in AI Adoption

The maturity levels of AI adoption in the included studies were assessed using a structured framework defining "exploring," "emerging/activating," or "integrated ecosystem" stages [53]. These stages were defined as follows:

Explorer (exploring): This stage reflects ad hoc efforts to leverage AI in health care, with no established benchmarks or strategic frameworks. Stakeholders, including policymakers and governments, may have begun exploring AI's potential but have not yet developed or drafted relevant policies or guidelines.

Emergent (emerging or activating): At this stage, efforts to use AI in health care are more systematic and linked to a national AI strategy with clearly defined priorities. Policies and guidelines supporting AI inclusion in public health are drafted, demonstrating progress toward structured integration.

Leader (integrated ecosystem): This stage represents the highest maturity level, where AI adoption is embedded in national strategies and aligned with public health goals. Policies and guidelines are implemented and continuously updated,

reflecting a well-established and operational AI ecosystem in health care.

These indicators encompassed several dimensions, such as the presence of national AI strategies, systematic exploration of AI applications, and the existence of policies or guidelines supporting AI integration. Careful consideration was given to the content, including mentions of AI strategies, the level of systematic exploration, and any references to comprehensive policies or frameworks. Each article was then categorized according to the maturity level that most accurately reflected its alignment with these dimensions. ST initially performed the categorization, which was then reviewed by URK and ML.

This methodical process ensured that the categorization of articles was based on discernible criteria, promoting consistency and reliability in assessing the state of AI adoption in neonatal care research. By applying these definitions and criteria, the assessment provided a clear understanding of the progress and gaps in AI maturity across the included studies.

Cohort Size Classification

Participant cohort sizes were categorized as small (fewer than 100 participants), medium (100 to 1000 participants), and large (more than 1000 participants) to standardize the classification of studies based on the scale of their population data.

Performance Measure Evaluation Categories

In evaluating the performance measures for the AI models used in the included studies, a systematic approach was used to rate their effectiveness based on the results obtained and the type of measure used, supported by statistical norms for machine learning [54]. The evaluation results were categorized as follows.

Area Under the Receiver Operating Characteristic Curve, Correlation, and Sensitivity and Specificity

Excellent (0.9-1): AI models within this category exhibited exceptional predictive accuracy and discrimination capabilities. They demonstrated a high degree of confidence in distinguishing between outcomes.

Good (0.8-0.9): Models in this range displayed strong predictive performance. They possessed a notable ability to differentiate between outcomes, indicating reliability in predictions.

Moderate (0.7-0.8): AI models within this range demonstrated moderate predictive performance. While reasonably discriminative, there was room for enhancement in their accuracy.

Normalized Average Root Mean Square Error

Excellent (0-lower values are better): AI models achieving lower values for normalized average root mean square error (RMSE) demonstrated superior predictive accuracy. Lower RMSE values indicated predictions that closely aligned with actual outcomes.

This systematic rating approach allows comprehensive evaluation of the performance of AI models in predicting clinical outcomes and length of stay in the NICU. It provided a structured framework for assessing the reliability and accuracy

of predictions based on the specific evaluation measures used in this study.

Reporting

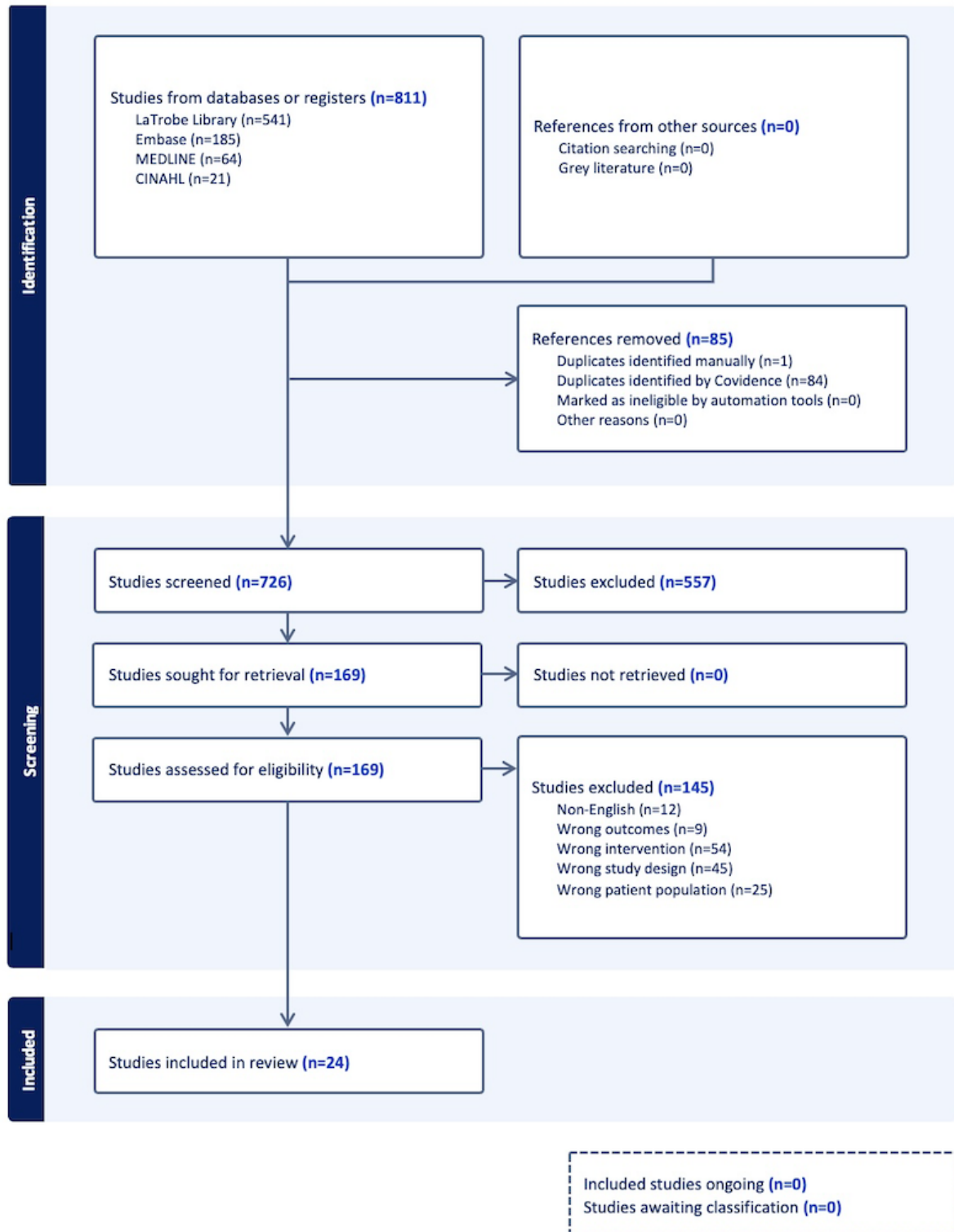
The PRISMA 2020 checklist was used to guide the reporting of this systematic review, ensuring a comprehensive and transparent presentation of the methodology and findings ([Multimedia Appendix 3](#)). Certain items (such as statistical synthesis or meta-analysis) were not applicable due to the heterogeneity of study designs and outcomes, but all relevant PRISMA elements were addressed in line with best practices for qualitative systematic reviews.

Results

Study Selection

A total of 811 studies were obtained and imported into the Covidence Systematic Review Software (Veritas Health Innovation). Among them, 85 duplicates were removed, resulting in 726 studies for screening. Out of the initial 726 studies, 169 were retained following a title and abstract review. The final selection comprised 24 articles deemed appropriate and included following a full review. A PRISMA flow diagram ([Figure 1](#)) illustrates the process of article selection and exclusion.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



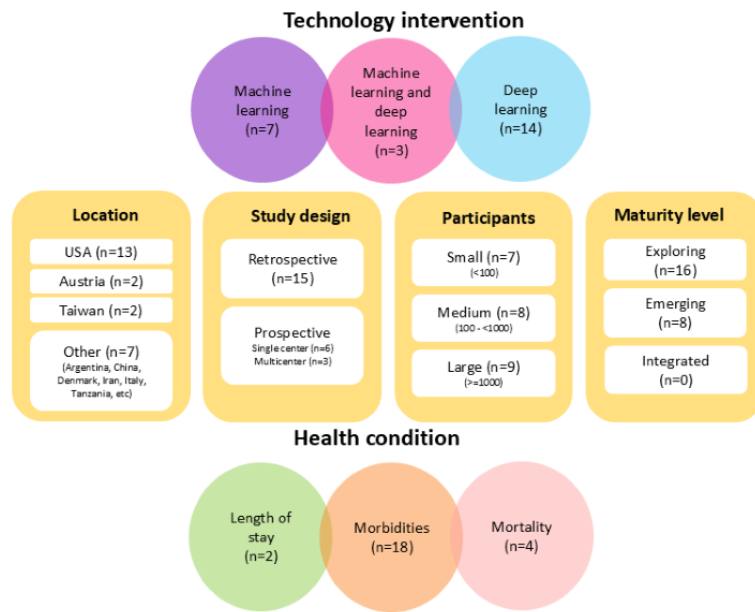
Study Characteristics

Key Features of Included Studies

Multimedia Appendix 4 [28-51] provides detailed characteristics of the included studies in a tabular format. The 24 included studies, as seen in Figure 2, include various technology interventions in NICU research: 7 using machine learning [28-34], 3 combining machine and deep learning [35-37], and 14 exclusively focusing on deep learning [38-51].

Geographically, 13 studies are from the United States [28,29,33,37,38,40,42,43,45,47-50], with others from Austria (2) [30,41], Taiwan (2) [44,46], and 6 in different countries including Argentina [32], China [36], Denmark [34], Iran [51], Italy [35], and Tanzania [31], along with 1 multinational dataset [39]. Study designs include 15 retrospective, 6 prospective single-center, and 3 multicenter studies. Participant cohorts varied, with 7 studies featuring small cohorts (fewer than 100), 8 with medium (100 to 1000), and 9 with large cohorts (1000+).

Figure 2. Study characteristics of AI predictions for NICU patients (n=24), including technology interventions, location, type of study design, number of participants, the AI maturity level, and the outcome predicted. AI: artificial intelligence; NICU: neonatal intensive care unit.



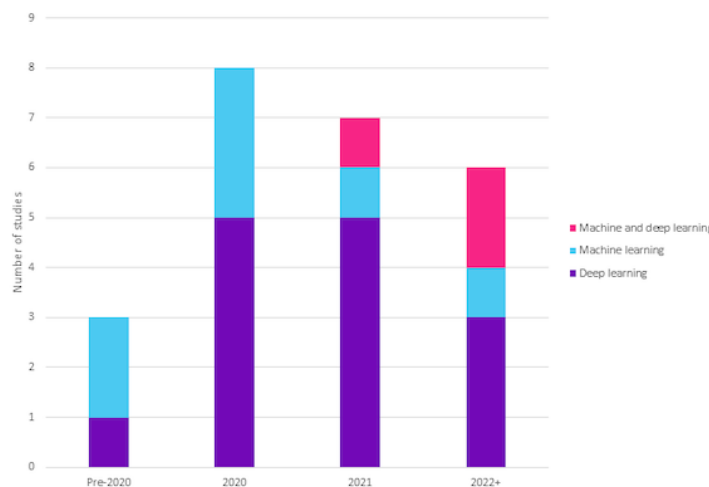
Of the 24 included studies, 16 fall under the “exploring” category, lacking a cohesive AI strategy [28,29,31,32,36,38-40,42,44-47,49-51]. Eight articles were classified as “emerging/activating,” showing systematic exploration but without a fully established AI ecosystem [30,33-35,37,41,43,48]. None were categorized under “integrated ecosystem,” indicating that a mature and fully established AI ecosystem in neonatal care remains a future aspiration.

The included 24 studies primarily analyze 3 key outcomes: length of stay (2 instances), morbidities (18 studies), and mortality (4 studies). These characteristics offer insights into the diverse range of studies in AI’s current NICU research.

Temporal Technological Trends

The 24 included studies depict temporal technology trends (Figure 3), revealing shifts in preference within neonatal care research. In 2020 and 2021, five studies emphasized deep learning [39-41,43-45,47,49-51], signifying its growing recognition in predicting outcomes. This trend continues in 2022, with 3 studies opting for deep learning [38,46,48]. Post-2021, studies integrate machine learning and deep learning [35-37] methods, reflecting evolving research methodologies and the importance of advanced AI, particularly deep learning, in neonatal care predictions.

Figure 3. Temporal technology trends of included studies (totaling 24) across different technologies—machine learning, deep learning, and a combination of both—based on their publication years.



Morbidities Studied

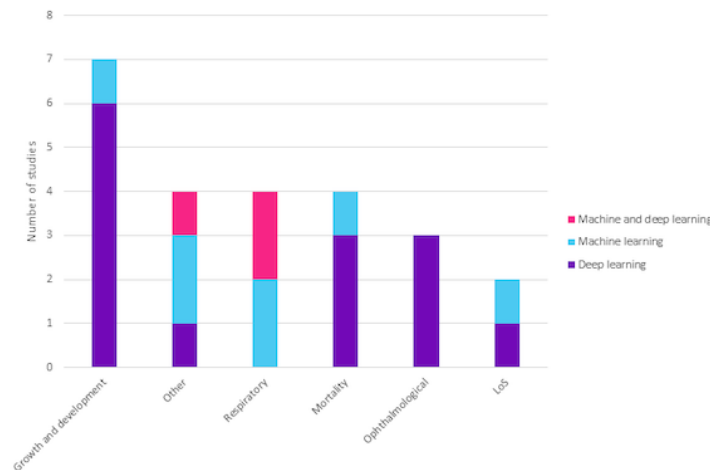
The 24 studies in this review outline AI technology’s applications in predicting specific NICU clinical outcomes (Figure 4), revealing prevalent trends. “Growth and

development” leads with 7 studies, mostly using deep learning [30,38,41-43,49,50], emphasizing AI’s role in assessing infant development. “Ophthalmological” outcomes are studied in 3 exclusively deep learning-driven articles [39,40,44].

“Respiratory” outcomes involve 4 studies, equally leveraging deep learning and machine learning studies [32,34-36], underscoring their relevance in neonatal respiratory health. The “other” category spans a variety of outcomes, with different AI techniques used. “Mortality” is studied by 4 articles, mostly using deep learning [31,45,48,51]. Lastly, “length of stay” involves 2 studies, evenly split between deep learning and

machine learning [28,46]. Overall, deep learning emerges as the predominant approach across categories, showcasing its diverse applicability in neonatal care. This comprehensive analysis accentuates AI’s complex role in enhancing infant health and well-being. [Multimedia Appendix 5](#) [28-51,55-66] summarizes clinical outcomes categories and descriptions of relevant studies.

Figure 4. Outcome type by technology type, illustrating the distribution of 24 included studies across different outcome types (growth and development, other, respiratory, mortality, ophthalmological, and LoS) and technology types (machine learning, both machine learning and deep learning, and deep learning). The chart provides insights into the focus areas of these studies and the technology used to address specific neonatal care outcomes. LoS: length of stay.



Predicting Outcomes Evaluation Measures

The 24 studies use diverse predictive model evaluation measures (Table 2) to assess machine or deep learning model accuracy in predicting outcomes or classifications in neonatal care. Figure 5 shows AI model performance across outcome categories, labeled “excellent,” “good,” “moderate,” or “not reported.” In the “excellent” category, deep learning excelled in “other,” “ophthalmological,” and “mortality” outcomes, while machine learning performed well in “respiratory” outcomes. “Good” models, primarily deep learning, demonstrated strength in “growth and development” and “mortality,” while machine

learning and deep learning showed promise in “other” and “respiratory” outcomes. “Moderate” performance was observed in deep learning for “growth and development” and “length of stay,” and in machine learning for “mortality.” Notably, 5 studies [32,35,40,41,49] did not report performance, including 2 that specified the measure used (eg, AUROC [area under the receiver operating characteristic curve] and error percentiles) but did not provide the results [35,41], and 3 that did not report either the measure or the results [32,40,49]. This figure highlights AI strengths in various outcomes, indicating research directions and areas needing improvement.

Table 2. Evaluation measures.

Evaluation measure	Number of studies	Study
Area under the receiver operating characteristic curve	15	[28,29,35-39,42,43,45-48,50,51]
Correlation (<i>r</i>)	1	[30]
Error percentiles	1	[41]
Normalized average root mean square error	1	[44]
Sensitivity and specificity	3	[31,33,34]
Not reported	3	[32,40,49]

Figure 5. Performance outcome by type of AI, displaying the performance assessment of AI models (machine learning, deep learning, and a combination of both) across different neonatal care outcome types (growth and development, other, respiratory, mortality, ophthalmological, and LoS). Performance is categorized into 4 levels: excellent, good, moderate, and NR. The chart provides insights into the quality of AI models in predicting specific neonatal care outcomes across different AI types and outcome categories. LoS: length of stay; NR: not reported.

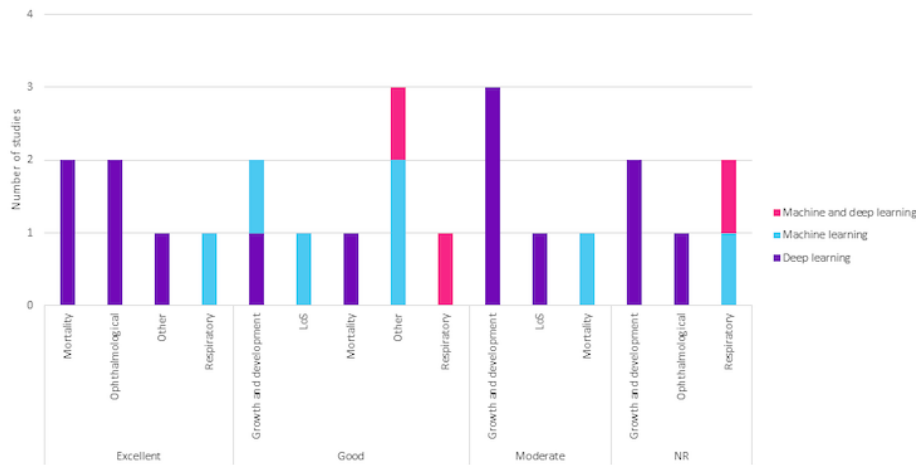


Table 2 demonstrates various evaluation measures for predicting neonatal outcomes. Among these, the most common metric is AUROC, used in 15 studies. Other measures such as correlation (*r*), error percentiles, and normalized average RMSE were less frequent, each found in a single study. Sensitivity and specificity were the focus of 3 studies. However, specific evaluation measures were not reported in 3 instances, showcasing diverse approaches to assessing AI models in neonatal care studies.

Performance by Outcome and Type of AI

Figure 5 provides an overview of the performance of AI models in predicting neonatal outcomes, categorized as “excellent,” “good,” “moderate,” or “not reported,” and further differentiated by the type of technology used. In the “excellent” category, deep learning models demonstrated exceptional predictive abilities across various outcomes, including “other,” “ophthalmological,” and “mortality,” while machine learning models were effective in predicting “respiratory” outcomes.

The “good” category saw deep learning models excelling in “growth and development” and “mortality,” with machine learning and deep learning models performing well in “other” and “respiratory” outcomes. Machine learning models also demonstrated proficiency in predicting “length of stay.” “Moderate” AI models, predominantly deep learning, showed moderate performance in “growth and development” and “length of stay,” while machine learning models displayed moderate performance in 1 “mortality” study.

Insights Into Potential Opportunities for AI in NICUs

Opportunity Themes for AI in NICU Research: Overview

Opportunities for AI applications in NICU have been themed into 5 categories (summarized in Table 3), from enhancing NICU care through data-driven insights and predictive models to using advancements in medical imaging, improving risk stratification, and personalizing neonatal care and interventions.

Table 3. Opportunity themes and findings.

Study	Application of AI ^a in the NICU ^b	Advancements in medical imaging	Data-driven insights and predictive models	Improving understanding and risk stratification	Personalized neonatal care and intervention
Iyer et al (2022) [30]	✓		✓		✓
He et al (2023) [36]	✓				
Ali et al (2022) [38]	✓				
Lin et al (2022) [46]	✓		✓		
Lee et al (2021) [45]	✓		✓		
Chen et al (2021) [39]	✓	✓	✓		
Gschwandtner et al (2020) [41]	✓				
He et al (2020) [43]	✓	✓			
Braun et al (2020) [28]	✓		✓	✓	✓
Choi et al (2020) [40]	✓	✓			
Saha et al (2020) [50]	✓	✓			✓
Huang et al (2020) [44]	✓	✓			
Hamilton et al (2020) [29]	✓		✓	✓	
He et al (2018) [42]	✓			✓	
Kausch et al (2022) [37]	✓				
Verder et al (2021) [34]	✓				
Patel et al (2022) [48]	✓		✓		
Ruixiang et al (2021) [49]	✓		✓		
Shalish et al (2017) [33]	✓		✓		
Sheikhtaheri et al (2021) [51]	✓				
Ofman et al (2019) [32]	✓		✓	✓	
Amodeo et al (2021) [35]	✓		✓	✓	✓
Kovacs et al (2021) [31]	✓				
Lure et al (2021) [47]	✓	✓	✓		✓

^aAI: artificial intelligence.

^bNICU: neonatal intensive care unit.

Advancements in Medical Imaging

Several studies within the selection highlight medical imaging advancements. Two studies focus on retinopathy of prematurity (ROP), showcasing the role of deep learning in standardizing disease assessment and predicting visual outcomes posttreatment [39,44]. Choi et al [40] also explored deep learning for ROP severity assessment, highlighting broader medical implications. He et al [43] introduced a multitask deep transfer learning model for early neurodevelopmental outcome prediction in preterm infants, emphasizing imaging's predictive role. Saha et al [50] used deep learning convolutional neural networks to forecast motor outcomes via brain diffusion MRI (magnetic resonance imaging) data. Lastly, Lure et al [47] used machine learning to distinguish neonatal conditions, enhancing clinical decision-making. These studies highlight medical imaging's

potential in early diagnosis and monitoring of neonatal conditions.

Data-Driven Insights and Predictive Models

Data-driven insights and predictive models in neonatal care have the potential to enhance outcomes [28-30,33,39,45,46,48,49]. They contribute to care planning by using machine learning to predict factors like length of stay, disease severity, and mortality risk, aiding health care providers [30,45,46]. Some explore risk assessment, detecting, and mitigating severe neonatal morbidities [28,29,50]. Two introduce adaptive machine learning algorithms for dynamic patient status adaptation [35,48]. Lastly, Ruixiang et al [49] highlight data-driven insights' transformative role in advancing neonatal research and medical care by improving predictions, care planning, and risk assessment.

Improving Understanding and Risk Stratification

The theme presents a cohesive effort toward understanding neonatal conditions and individual risk stratification through data-driven approaches. Three articles focus on machine learning–driven risk assessment for conditions like bronchopulmonary dysplasia (BPD), in-hospital length of stay, and mortality risk [42,45,46]. Leveraging extensive datasets, these studies offer critical insights into associated risk factors, enabling early intervention. Additionally, 3 articles delve into NICU usage trends [28], severe morbidity risk [29], and chronic lung disease determinants in very low birth weight infants [32]. Two studies emphasize dynamic machine learning algorithms for mortality prediction and pulmonary hypertension (PH) [35,48]. Together, these studies deepen our understanding of neonatal conditions, reshaping disease paradigms and paving the way for more effective interventions.

Personalized Neonatal Care and Intervention

Numerous studies contribute to the theme of “personalized neonatal care and intervention,” extending the horizons of

individualized care for neonates. These studies explore personalized care plans and interventions. They investigate trends in NICU usage [28], assess risk factors for severe neonatal morbidity [29], predict motor outcomes [50], and differentiate between critical neonatal conditions [47]. By leveraging predictive models and continuous monitoring, these studies seek to tailor care strategies to the specific needs of each neonate. This data-driven approach holds the potential to revolutionize neonatal care by optimizing interventions and leading to better clinical outcomes.

Insights Into Potential Challenges for AI in NICUs

Challenge Themes Overview

Literature indicated some concerns that span multiple facets of AI application in NICU, ranging from data quality and quantity issues to clinical interpretability, model generalization, clinical and diagnostic variability, and ethical and regulatory considerations. Table 4 provides a condensed overview of the key thematic challenges encountered across the various included studies.

Table 4. Challenge themes and findings.

Study	Data quality and quantity challenges	Clinical interpretability and usability	Model generalization and validation	Clinical and diagnostic variability	Ethical and regulatory challenges
Iyer et al (2022) [30]		✓		✓	
He et al (2023) [36]	✓		✓		
Ali et al (2022) [38]	✓				
Lin et al (2022) [46]					
Lee et al (2021) [45]	✓			✓	
Chen et al (2021) [39]		✓	✓		✓
Gschwandtner et al (2020) [41]		✓	✓	✓	
He et al (2020) [43]		✓			
Braun et al (2020) [28]		✓			
Choi et al (2020) [40]	✓		✓	✓	
Saha et al (2020) [50]	✓		✓	✓	
Huang et al (2020) [44]	✓				
Hamilton et al (2020) [29]					
He et al (2018) [42]	✓				
Kausch et al (2022) [37]		✓			
Verder et al (2021) [34]	✓		✓	✓	
Patel et al (2022) [48]					✓
Ruixiang et al (2021) [49]	✓				
Shalish et al (2017) [33]		✓		✓	
Sheikhtaheri et al (2021) [51]	✓		✓	✓	
Ofman et al (2019) [32]			✓	✓	✓
Amodeo et al (2021) [35]			✓	✓	
Kovacs et al (2021) [31]		✓			✓
Lure et al (2021) [47]					

Data Quality and Quantity

Several studies face recurring challenges with health care data in the NICU, including issues related to quality, quantity, and availability [34,36,38,40,42,44,45,49-51]. Addressed by multiple articles, these challenges span data quality's influence on predictive models, the need for larger datasets, variability due to data quality, and concerns about both data quality and quantity. The authors highlight these challenges in developing effective machine learning solutions for neonatal care. Overcoming these hurdles is crucial for successful machine learning implementation, ultimately enhancing health care outcomes for neonates.

Clinical Interpretability and Usability

Challenges in clinical interpretability and usability arise as machine-generated insights need integration into clinical

practice. Studies emphasize the need for standardized measures for neurodevelopmental assessment, ethical data sharing, complex functional connectivity estimation, user-friendly tools in low-resource settings, and distinguishing specific illnesses [28,30,31,33,37,39,41,44]. Overcoming these challenges is crucial for meaningful AI integration in the NICU and its adoption by health care professionals, and improving health care outcomes for these patients.

Model Generalization and Validation

The challenge of ensuring AI model reliability across diverse populations in neonatal care is evident. He et al [36] advocate for external validation of AI models for BPD severity prediction across various populations. Chen et al [39] stress the need for adaptable and generalizable models, particularly for diverse patient demographics and different camera systems in ROP. Similarly, Choi et al [40] focus on creating versatile deep

learning scales for ROP applicable across various centers. Saha et al [50] emphasize the importance of generalizing predictions for diverse neonatal populations in motor outcome prediction. Verder et al [34] highlight the necessity of generalizing and validating models for accurate predictions in diverse clinical scenarios for BPD. Sheikhtaheri et al [51] discuss the validation of predictive models for neonatal deaths in NICUs, recognizing the need for external validation. Ofman et al [32] note the challenge of studying disease determinants across multiple centers in chronic lung disease. Amodeo et al [35] explore the complexities of predicting PH outcomes in a diverse patient population.

Clinical and Diagnostic Variability

Handling clinical and diagnostic variability is a critical challenge in the NICU, with particular significance in fields such as ophthalmology and critical care, where even subtle variations can have a profound impact on outcomes [30,37,40,45,47,50]. Iyer et al [30] highlight the need for standardized, objective, and scalable measures for neurodevelopmental assessment, contrasting with current subjective and nonscalable methods. Similarly, Lee et al [45] note the necessity for a larger training dataset and consideration of site-specific differences to improve model performance. In ophthalmology, Choi et al [40] highlight the diagnostic variability and subjective quantification of severity levels, hindering the interpretation of clinical trial data. Saha et al [50] raise concerns about the limited sample size and heterogeneity of brain injuries, which may lead to overfitting and poor prediction performance. Furthermore, Shalish et al [33] point out practice variability and uncertainty in defining extubation failure, complicating clinical decision-making.

Addressing imbalanced data and the need for standardized physiological testing are highlighted by Ofman et al [32] and Sheikhtaheri et al [51], respectively. Moreover, Amodeo et al [35] underscore the challenges in measuring lung vascularization accurately, given the techniques' reproducibility issues and the impact of hemodynamic changes during neonatal transition. These challenges collectively illustrate the complexity and variability inherent in neonatal care, underscoring the importance of developing robust AI models and data-driven solutions to enhance precision and reliability in clinical decision-making.

Ethical and Regulatory Challenges

The integration of machine learning in neonatal care introduces ethical and regulatory complexities. Chen et al [39] emphasize the ethical challenge of data sharing, weighing the practicality against privacy concerns in multi-institutional datasets. The study highlights regulatory hurdles in validating models across diverse populations and camera systems. Patel et al [48] note the regulatory risks of overfitting single-site datasets and ethical concerns around addressing missing data. They also note the need to evaluate model performance with evolving clinical cohorts, which has regulatory implications. Ofman et al [32] underline ethical and regulatory issues linked to nonstandardized physiologic testing in NICUs, impacting research and clinical practices. Kovacs et al [31] note the ethical concern of deploying tools in resource-limited settings, balancing computational limitations with clinicians' interpretability. These articles

emphasize the intricate ethical and regulatory challenges in AI integration in neonatal care, requiring careful navigation for ethical standards and regulatory compliance.

Discussion

Key Findings

The objective of this systematic review was to analyze the current AI research landscape for predicting clinical outcomes and length of stay in the NICU, exploring the benefits and challenges of using AI in the NICU for these predictions. The surge in studies between 2020 and 2021 reflects a growing recognition of AI's potential to reshape neonatal care, possibly accelerated by the COVID-19 pandemic's impact on health care systems and driving innovative solutions to mitigate disruptions. The sustained research momentum, including studies in 2022, signifies an enduring commitment to exploring AI's role in the NICU.

This study contributes to the growing field of AI in NICUs by addressing gaps left by previous reviews and advancing the discourse on critical factors for successful AI adoption. Unlike earlier reviews, which predominantly focused on technical aspects and model performance, this study emphasizes emerging themes, including the importance of explainability, multidisciplinary collaboration, and stakeholder engagement. For instance, while [22] and [12] underscore technical barriers and model readiness, they lack focus on real-time data integration and actionable pathways for clinical implementation. This study builds on their findings by highlighting the transformative potential of AI in enabling predictive modeling, personalized care, and multidimensional data synthesis, essential for advancing NICU practices. Furthermore, this review synthesizes research through a broader thematic lens and incorporates the most recent evidence from 2017 to 2023. The systematic review design ensures a methodologically rigorous and transparent synthesis of the latest research, enabling relevance and depth in addressing practical challenges such as ethical concerns, stakeholder trust, and system integration. By using structured quality appraisal and thematic analysis across diverse study types, this review offers actionable insights that are not only academically robust but also critically important for guiding health care professionals, policymakers, and technologists in implementing AI solutions tailored to the unique complexities of NICU environments.

In terms of technologies used, deep learning emerges as a dominant technology across various clinical outcomes, highlighting its effectiveness in automatically identifying relevant features from raw data, especially in medical imaging and clinical analysis. However, the field's maturity in AI integration within neonatal care remains in the early exploration stage, lacking a mature and cohesive ecosystem.

The growing presence of various national guidelines and policies highlights the progress of AI implementation within health care [67-69]. However, there is a need to prioritize the development of a national AI strategy tailored for neonatal care due to the unique challenges of this environment. Specific considerations for neonatal care include the need for specialized models that

account for the unique physiology and comorbidities of premature infants, as well as the heightened need of ensuring patient safety and data privacy in a patient population that have continuous and complete medical records from birth.

The study reveals substantial opportunities for the use of AI in the NICU to predict length of stay and patient outcomes. These cover diverse areas: advancements in medical imaging, data-driven insights and predictive models, improved understanding and risk stratification, and personalized neonatal care and intervention.

The first opportunity focuses on AI's potential to transform neonatal care in the NICU by leveraging machine learning to perform predictive analytics to help inform the clinician during the clinical decision-making process, aiding in earlier interventions. However, there are research gaps in translating AI's theoretical effectiveness into real-world NICU settings [3-6,70-74]. This brings model generalization and validation, the third challenge theme, which highlights the importance of ensuring AI models perform reliably across diverse patient populations. External validation across different populations is necessary for reliability and adaptability. The unique patient populations and scenarios in the NICU demand tailored approaches for model generalization and validation, vital for trustworthy predictions.

The second opportunity highlights the use of AI in NICU medical imaging. Six studies primarily use deep learning to analyze medical images, notably for conditions such as ROP. AI-driven imaging enhances accuracy, facilitating early intervention and personalized care plans. These findings not only affirm anticipated AI benefits in medical imaging but also underline its relevance in the NICU, particularly focusing on neonatal conditions. AI-powered imaging is key for precise diagnosis, standardized disease assessment, and predicting critical outcomes, aligning closely with literature expectations [15,75,76]. It enables tailored, timely care plans for individual neonatal needs, a significant advantage in this setting. The challenge here is handling clinical and diagnostic variability in these settings, where even subtle variations can have an impact on outcomes. Addressing this challenge involves the development of models and algorithms that can adapt to diverse clinical contexts, accounting for nuances that might otherwise be overlooked. It calls for an ongoing effort to refine models and diagnostic tools to accommodate the inherent variability in neonatal care, ultimately improving the precision and reliability of clinical decision-making.

The third and fourth opportunities highlight the transformative power of data-driven insights and predictive models in the NICU. Thirteen studies within these themes leverage AI to translate health care data into actionable insights, aiding decision-making. They encompass predictive modeling, spanning length of stay, disease severity, and mortality risk. These studies deepen the comprehension of neonatal conditions and their determinants, aligning with the literature's focus on advancing scientific understanding and medical care in the NICU [44,76-81]. By using data-driven approaches, they pave the way for redefining disease paradigms, ultimately enhancing interventions and care strategies. The challenge here is data

quality and quantity, as it involves acquiring high-quality, sufficient data for AI model training. Data incompleteness and inconsistency, common issues in health care data, align with existing literature. Overcoming these challenges requires advanced algorithms and robust data management strategies within the NICU [80].

The final opportunity, personalized neonatal care and intervention, signals a significant shift in NICU health care. Studies focusing on tailoring care plans and interventions for individual neonatal needs explore personalized care plans, risk assessment, predicting motor outcomes, and distinguishing critical neonatal conditions. These efforts, using predictive models and continuous monitoring, aim to optimize interventions. Personalized care plans carry substantial implications, potentially reducing unnecessary treatments, cutting health care costs, and improving clinical outcomes. These align with literature stressing the importance of individualized neonatal care [71,82]. This is closely related to challenges of clinical interpretability and usability, as well as managing clinical and diagnostic variability. Clinical interpretability and usability focus on the transparency and practicality of AI-driven models within the NICU. Health care providers in the NICU demand accurate predictions and a clear understanding of AI outputs.

Recent Advances in AI Research in NICU Settings

Building on the findings of this review, more recent AI research in NICU settings (post-2023) demonstrates continued advancement across key clinical domains, particularly neurological development, ophthalmological conditions, and respiratory outcomes. A growing body of literature emphasizes the integration of longitudinal and multisource datasets to strengthen predictive modeling. For instance, studies focused on neurodevelopmental outcomes have begun incorporating nationwide longitudinal clinical and sociodemographic data to more accurately identify risks of cognitive delay [83,84]. In the respiratory domain, pilot studies have explored the early detection of BPD by analyzing volatile organic compounds in exhaled breath, identifying novel biomarkers to support early diagnosis and intervention [85]. Similarly, machine learning techniques have been applied to uncover predictive clinical variables spanning the perinatal and neonatal periods, enhancing early risk stratification for respiratory morbidity [86]. Recent developments also highlight a shift toward multifactorial models that enable personalized care planning. Notably, the Mendelian Phenotype Search Engine (MPSE) uses natural language processing and machine learning to automatically identify high-risk neonates likely to benefit from whole genome sequencing, improving diagnostic yield and enabling timely interventions within the first 48 hours of NICU admission [87]. Parallel advancements in ophthalmology include AI-assisted diagnosis for ROP, with recent studies validating model outputs through clinician review to enhance diagnostic accuracy and trust [88,89]. These post-2023 developments align with our review findings and reinforce the trajectory toward integrated, real-time, and patient-centered AI solutions in neonatal care.

Regardless of any application, ethical and regulatory considerations and challenges remain in all applications. It

emphasizes the complex landscape requiring careful navigation. Key ethical concerns include data sharing dilemmas and safeguarding patient data privacy and security. Regulatory hurdles, especially in validating AI models across diverse populations and settings, require adaptable frameworks. Only 4 studies directly mention these ethical and regulatory challenges, suggesting a potential research gap in the NICU field of AI, requiring more exploration and consideration. Ensuring responsible AI adoption is pivotal for equity in health care access and maintaining high standards of care.

Recent studies confirm these challenges and suggest strategies to address them. For instance, learning models, enabling collaborative training across institutions without sharing raw data, have emerged as a practical solution to mitigate privacy risks while enhancing model robustness [21,90]. Large-scale neonatal research networks, such as the Vermont Oxford Network, have emerged as examples that offer collaborative frameworks for aggregating high-quality data to overcome issues of data scarcity and variability [12]. Challenges like clinical usability and model generalization can be addressed through explainable artificial intelligence (XAI), which enhances transparency and builds clinician trust [22]. This can be achieved by transparent data practices and clear disclosures about imbalanced datasets, which are essential for meeting the needs of underserved populations [90]. Additionally, integrating ethics review boards early in the AI development process can guide data governance and ensure compliance with privacy laws [12]. Addressing ethical challenges also requires diverse teams of researchers, engineers, informaticists, and neonatologists to tackle these problems using equity-focused frameworks, particularly for addressing historically understudied challenges in neonatal care. Multicenter validation and standardized data collection are also important to ensure model reliability and address variability across clinical environments [12]. Lastly, centralized AI platforms could be used to streamline model integration into workflows, support continuous updates, and address performance degradation over time [22]. Regular audits, along with clear documentation of model development and trade-offs, further ensure responsible implementation [90]. These strategies provide actionable pathways for advancing robust, ethical, and equitable AI adoption in NICUs, paving the way for impactful neonatal care.

Future directions for NICU opportunities present key research areas. First, integrating AI with medical imaging, especially for conditions like ROP, demands refined technology, larger datasets, and real-world validation. These challenges pertaining to data integration have also been highlighted in previous reviews [12,22]. Second, enhancing data-driven insights and predictive models requires broader clinical scenarios and improved data quality via collaborative NICU efforts. Building on recommendations in prior reviews to explore clinician trust in AI systems, this study highlights the need for actionable strategies to enhance transparency and multidisciplinary collaboration [21]. Understanding neonatal conditions better, exploring diverse risk factors, and fostering multidisciplinary collaborations are key priorities. Practical AI application in NICUs, optimizing resource allocation and care, requires real-world implementation and thorough assessment. Lastly,

refining scalable personalized care plans and interventions, while ethically considering AI personalization, remains essential for expanding AI's role and improving neonatal outcomes.

Conclusions

Opportunities and Challenges

This systematic review has highlighted substantial opportunities for AI in the NICU. Advancements in medical imaging, combined with AI, have the potential to improve diagnostic accuracy and enable early intervention in this field. Furthermore, data-driven insights and predictive models offer the opportunity to enhance clinical decision-making and deepen our understanding of neonatal conditions. Additionally, personalized neonatal care could optimize health care delivery for individual neonates.

Simultaneously, the study has revealed crucial challenges in integrating AI in the NICU. Issues related to data quality emphasize the need for robust data management strategies. Ensuring clinical interpretability and usability is essential to ensure AI tools align smoothly with clinical workflows, especially in the high-stress NICU environment. Moreover, achieving model generalization and validation across diverse patient populations and addressing clinical and diagnostic variability are essential considerations. Ethical and regulatory challenges, including data privacy, security, and model validation, underscore the importance of responsible AI adoption in the NICU.

These findings align with existing literature, revealing the unique complexities of the NICU context. The identified research gaps include the need for practical AI implementations within the NICU, considering resource constraints and clinical requirements. Additionally, there is a need for the development of ethical frameworks and regulatory guidelines tailored specifically to the NICU environment. Future research should focus on practical implementation, ethical frameworks, and regulatory guidelines to realize AI's potential in the NICU.

This research holds potential for neonatal care, benefiting a range of stakeholders. Health care professionals, including neonatologists, nurses, and other clinicians, can gain directly from the insights provided in this study. Health care institutions that manage NICUs also stand to benefit, as the research highlights the importance of practical AI implementation within the NICU, enabling them to optimize resource allocation and improve patient care. This research serves as an important resource for advancing AI technology in neonatal care, fostering a future of improved health care delivery and enhanced well-being for NICU patients and their families.

Limitations and Future Directions

The exploration of the potential of AI in the NICU to predict clinical outcomes and length of stay comes with a recognition of strengths and limitations that shape the scope and methodology of this study. The search strategy in this systematic review was designed to be comprehensive, incorporating a range of keywords and subject headings to capture relevant literature. The search terms were carefully chosen in collaboration with experienced research librarians to ensure inclusivity; however,

the complexity of the AI field and evolving terminologies might have introduced limitations. It is possible that studies using emerging AI-related terminology were not captured. Furthermore, the inclusion criteria focused on academic research articles, which inadvertently excluded insights from private organizations and the grey literature. This limitation could result in missing valuable perspectives and data that could enhance understanding of AI applications in the NICU. Moreover, the decision to restrict the study to English full-text articles introduced a language bias, primarily focusing on studies from English-speaking countries. As a result, research conducted in other languages may have been excluded, limiting the diversity and global representation in this review. Additionally, this study's eligibility criteria specifically focused on interventions emphasizing predictive aspects of AI, which could have led to an incomplete portrayal of AI's multifaceted role in the NICU. Nonpredictive aspects of AI, while relevant, were not the central focus of this systematic review.

Despite these potential limitations, this study provides valuable insights into the challenges and opportunities in the emerging field of AI predictions in the NICU. The decision to focus on academic research articles ensured a rigorous examination of peer-reviewed literature, lending credibility to the findings. While recognizing these limitations, the study serves as a foundation for further exploration and analysis in this dynamic and evolving field. Our review considered studies published until March 2023, providing a snapshot of advancements in AI applications in NICUs during that period. Post the review period, new developments include the use of ensemble machine learning models for predicting NICU length of stay with high accuracy through hybrid approaches such as Classifier Fusion-LoS, Gradient Boosting, CatBoost, and Recurrent Neural Networks, which can result in better resource management and quality of

care [91,92]. Additionally, researchers have developed advanced AI models, which can enable clinicians to interpret predictions for sepsis more effectively and accurately [93,94]. Furthermore, new research combining proteomics with explainable machine learning methods has facilitated the identification of biomarkers for critical conditions, such as posthemorrhagic ventricular dilation, offering enhanced early intervention capabilities in neonates [95]. While these advancements build upon the opportunities identified in our study, particularly in the domains of explainability, real-time applications, and personalized neonatal care, the studies also highlight the existing challenges including the lack of high-quality and comprehensive datasets and the need for clinician training to effectively use these AI tools [91-94].

The systematic review process has exposed several research gaps. First, there is a noticeable gap concerning the maturity level of AI implementation in NICUs. While existing studies have laid the foundation, further research is needed to extend AI maturity to the next stages of emerging and integration. Second, the scarcity of studies addressing ethical and regulatory aspects of AI in NICUs is evident. These areas, such as liability in cases of adverse outcomes resulting from AI predictions, warrant a comprehensive investigation. Furthermore, the absence of studies reporting unfavorable or null results suggests potential reporting bias within the field. Finally, collaborative partnerships and patient and family engagement are instrumental in advancing AI research in NICUs.

Overall, these future directions and improvements collectively contribute to the ongoing evolution of AI research in NICUs, fostering a more comprehensive understanding of the opportunities and challenges while refining the methodology for more effective and robust reviews in the future.

Acknowledgments

The authors would like to acknowledge the support and contributions of La Trobe University in supporting the study by providing access to resources and databases used in this study. We are especially grateful to the library team for their invaluable insights and expertise in conducting the literature review.

Data Availability

All data supporting the findings of this study are provided in the multimedia appendices included with this manuscript. Additional information or clarification can be requested from the corresponding author.

Authors' Contributions

Conceptualization: ST
Data curation: ST
Formal analysis: ST, RB, TAW, JB, URK
Investigation: ST, TAW
Methodology: TAW, URK
Project administration: ML, TAW, URK
Resources: ML, JB
Supervision: RB, ML, JB, URK
Validation: RB, ML, URK
Visualization: URK
Writing – original draft: ST
Writing – review & editing: ST, RB, ML, TAW, JB, URK

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full-text excluded studies—reasons.

[\[DOCX File , 38 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Quality criteria checklist.

[\[DOCX File , 20 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

PRISMA 2020 checklist.

[\[DOCX File , 43 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Included studies.

[\[DOCX File , 21 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Clinical outcomes description.

[\[DOCX File , 20 KB-Multimedia Appendix 5\]](#)

References

1. WHO recommendations for care of the preterm or low-birth-weight infant. World Health Organization. Nov 15, 2022. URL: <https://www.who.int/publications/i/item/9789240058262> [accessed 2025-09-05]
2. Ohuma EO, Moller A, Bradley E, Chakwera S, Hussain-Alkhateeb L, Lewin A, et al. National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. *Lancet*. 2023;402(10409):1261-1271. [doi: [10.1016/s0140-6736\(23\)00878-4](https://doi.org/10.1016/s0140-6736(23)00878-4)]
3. Cheong JLY, Olsen JE, Huang L, Dalziel KM, Boland RA, Burnett AC, et al. Members of the Victorian Infant Collaborative Study Group. Changing consumption of resources for respiratory support and short-term outcomes in four consecutive geographical cohorts of infants born extremely preterm over 25 years since the early 1990s. *BMJ Open*. 2020;10(9):e037507. [FREE Full text] [doi: [10.1136/bmjopen-2020-037507](https://doi.org/10.1136/bmjopen-2020-037507)] [Medline: [32912950](https://pubmed.ncbi.nlm.nih.gov/32912950/)]
4. Gruenberg DA, Shelton W, Rose SL, Rutter AE, Socaris S, McGee G. Factors influencing length of stay in the intensive care unit. *Am J Crit Care*. 2006;15(5):502-509. [doi: [10.4037/ajcc2006.15.5.502](https://doi.org/10.4037/ajcc2006.15.5.502)]
5. Araki S, Saito T, Ichikawa S, Saito K, Takada T, Noguchi S, et al. [Family-centered care in neonatal intensive care units: combining intensive care and family support]. *J UOEH*. 2017;39(3):235-240. [doi: [10.7888/juoeh.39.235](https://doi.org/10.7888/juoeh.39.235)] [Medline: [28904275](https://pubmed.ncbi.nlm.nih.gov/28904275/)]
6. Santos J, Pearce SE, Stroustrup A. Impact of hospital-based environmental exposures on neurodevelopmental outcomes of preterm infants. *Curr Opin Pediatr*. 2015;27(2):254-260. [doi: [10.1097/mop.0000000000000190](https://doi.org/10.1097/mop.0000000000000190)]
7. Avila-Alvarez A, Solar Boga A, Bermúdez-Hormigo C, Fuentes Carballal J. Extrauterine growth restriction among neonates with a birthweight less than 1500 grams. *An Pediatr (Engl Ed)*. 2018;89(6):325-332. [doi: [10.1016/j.anpede.2018.02.004](https://doi.org/10.1016/j.anpede.2018.02.004)]
8. Bahado-Singh RO, Sonek J, McKenna D, Cool D, Aydas B, Turkoglu O, et al. Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix. *Ultrasound Obstet Gynecol*. 2019;54(1):110-118. [FREE Full text] [doi: [10.1002/uog.20168](https://doi.org/10.1002/uog.20168)] [Medline: [30381856](https://pubmed.ncbi.nlm.nih.gov/30381856/)]
9. Hsu J, Chang Y, Cheng H, Yang C, Lin C, Chu S, et al. Machine learning approaches to predict in-hospital mortality among neonates with clinically suspected sepsis in the neonatal intensive care unit. *J Pers Med*. 2021;11(8):695. [FREE Full text] [doi: [10.3390/jpm11080695](https://doi.org/10.3390/jpm11080695)] [Medline: [34442338](https://pubmed.ncbi.nlm.nih.gov/34442338/)]
10. Irles C, González-Pérez G, Carrera Muiños S, Michel Macias C, Sánchez Gómez C, Martínez-Zepeda A, et al. Estimation of neonatal intestinal perforation associated with necrotizing enterocolitis by machine learning reveals new key factors. *Int J Environ Res Public Health*. 2018;15(11):2509. [FREE Full text] [doi: [10.3390/ijerph15112509](https://doi.org/10.3390/ijerph15112509)] [Medline: [30423965](https://pubmed.ncbi.nlm.nih.gov/30423965/)]
11. Juraev F, El-Sappagh S, Abdukhamidov E, Ali F, Abuhmed T. Multilayer dynamic ensemble model for intensive care unit mortality prediction of neonate patients. *J Biomed Inform*. 2022;135:104216. [FREE Full text] [doi: [10.1016/j.jbi.2022.104216](https://doi.org/10.1016/j.jbi.2022.104216)] [Medline: [36208833](https://pubmed.ncbi.nlm.nih.gov/36208833/)]

12. McAdams RM, Kaur R, Sun Y, Bindra H, Cho SJ, Singh H. Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: a systematic review. *J Perinatol*. 2022;42(12):1561-1575. [doi: [10.1038/s41372-022-01392-8](https://doi.org/10.1038/s41372-022-01392-8)] [Medline: [35562414](https://pubmed.ncbi.nlm.nih.gov/35562414/)]
13. Klock P, Erdmann AL. [Caring for newborns in a NICU: dealing with the fragility of living/surviving in the light of complexity]. *Rev Esc Enferm USP*. 2012;46(1):45-51. [FREE Full text] [doi: [10.1590/s0080-62342012000100006](https://doi.org/10.1590/s0080-62342012000100006)] [Medline: [22441264](https://pubmed.ncbi.nlm.nih.gov/22441264/)]
14. Samra HA, McGrath JM, Rollins W. Patient safety in the NICU: a comprehensive review. *J Perinat Neonatal Nurs*. 2011;25(2):123-132. [doi: [10.1097/jpn.0b013e31821693b2](https://doi.org/10.1097/jpn.0b013e31821693b2)]
15. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge. MIT Press; 2016.
16. Van Zuylen H. Difference between artificial intelligence and traditional methods. *Artificial Intelligence Applications to Critical Transportation Issues [Monograph]*. Nov 2012:3-5. [doi: [10.17226/22690](https://doi.org/10.17226/22690)]
17. Szolovits P, Patil RS, Schwartz WB. Artificial intelligence in medical diagnosis. *Ann Intern Med*. 1988;108(1):80-87. [doi: [10.7326/0003-4819-108-1-80](https://doi.org/10.7326/0003-4819-108-1-80)] [Medline: [3276267](https://pubmed.ncbi.nlm.nih.gov/3276267/)]
18. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2(10):719-731. [FREE Full text] [doi: [10.1038/s41551-018-0305-z](https://doi.org/10.1038/s41551-018-0305-z)] [Medline: [31015651](https://pubmed.ncbi.nlm.nih.gov/31015651/)]
19. Han JH, Yoon SJ, Lee HS, Park G, Lim J, Shin JE, et al. Application of machine learning approaches to predict postnatal growth failure in very low birth weight infants. *Yonsei Med J*. 2022;63(7):640-647. [FREE Full text] [doi: [10.3349/ymj.2022.63.7.640](https://doi.org/10.3349/ymj.2022.63.7.640)] [Medline: [35748075](https://pubmed.ncbi.nlm.nih.gov/35748075/)]
20. Panesar A. Ethics of intelligence. In: *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. Berkeley, CA. Apress; 2019:207-254.
21. Adegboro CO, Choudhury A, Asan O, Kelly MM. Artificial intelligence to improve health outcomes in the NICU and PICU: a systematic review. *Hosp Pediatr*. 2022;12(1):93-110. [doi: [10.1542/hpeds.2021-006094](https://doi.org/10.1542/hpeds.2021-006094)] [Medline: [34890453](https://pubmed.ncbi.nlm.nih.gov/34890453/)]
22. Schouten JS, Kalden MACM, van Twist E, Reiss IKM, Gommers DAMPJ, van Genderen ME, et al. From bytes to bedside: a systematic review on the use and readiness of artificial intelligence in the neonatal and pediatric intensive care unit. *Intensive Care Med*. 2024;50(11):1767-1777. [doi: [10.1007/s00134-024-07629-8](https://doi.org/10.1007/s00134-024-07629-8)] [Medline: [39264415](https://pubmed.ncbi.nlm.nih.gov/39264415/)]
23. Vellido A, Ribas V, Morales C, Ruiz Sanmartín A, Ruiz Rodríguez JC. Machine learning in critical care: state-of-the-art and a sepsis case study. *Biomed Eng Online*. 2018;17(Suppl 1):135. [FREE Full text] [doi: [10.1186/s12938-018-0569-2](https://doi.org/10.1186/s12938-018-0569-2)] [Medline: [30458795](https://pubmed.ncbi.nlm.nih.gov/30458795/)]
24. Kim SY, Kim S, Cho J, Kim YS, Sol IS, Sung Y, et al. A deep learning model for real-time mortality prediction in critically ill children. *Crit Care*. 2019;23(1):279. [FREE Full text] [doi: [10.1186/s13054-019-2561-z](https://doi.org/10.1186/s13054-019-2561-z)] [Medline: [31412949](https://pubmed.ncbi.nlm.nih.gov/31412949/)]
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. [FREE Full text] [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
26. Tricco AC, Langlois EV, Straus SE. Rapid reviews to strengthen health policy and systems: a practical guide. World Health Organization. Aug 10, 2017. URL: <https://wkc.who.int/resources/publications/i/item/2017-08-10-rapid-reviews-to-strengthen-health-policy-and-systems-a-practical-guide> [accessed 2025-09-02]
27. Duval D, Pearce-Smith N, Palmer JC, Sarfo-Annin JK, Rudd P, Clark R. Critical appraisal in rapid systematic reviews of COVID-19 studies: implementation of the quality criteria checklist (QCC). *Syst Rev*. 2023;12(1):55. [FREE Full text] [doi: [10.1186/s13643-023-02219-z](https://doi.org/10.1186/s13643-023-02219-z)] [Medline: [36973811](https://pubmed.ncbi.nlm.nih.gov/36973811/)]
28. Braun D, Braun E, Chiu V, Burgos AE, Gupta M, Volodarskiy M, et al. Trends in neonatal intensive care unit utilization in a large integrated health care system. *JAMA Netw Open*. 2020;3(6):e205239. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.5239](https://doi.org/10.1001/jamanetworkopen.2020.5239)] [Medline: [32556257](https://pubmed.ncbi.nlm.nih.gov/32556257/)]
29. Hamilton EF, Dyachenko A, Ciampi A, Maurel K, Warrick PA, Garite TJ. Estimating risk of severe neonatal morbidity in preterm births under 32 weeks of gestation. *J Matern Fetal Neonatal Med*. 2020;33(1):73-80. [doi: [10.1080/14767058.2018.1487395](https://doi.org/10.1080/14767058.2018.1487395)] [Medline: [29886760](https://pubmed.ncbi.nlm.nih.gov/29886760/)]
30. Iyer KK, Leitner U, Giordano V, Roberts JA, Vanhatalo S, Klebermass-Schrehof K, et al. Bedside tracking of functional autonomic age in preterm infants. *Pediatr Res*. 2023;94(1):206-212. [doi: [10.1038/s41390-022-02376-2](https://doi.org/10.1038/s41390-022-02376-2)] [Medline: [36376508](https://pubmed.ncbi.nlm.nih.gov/36376508/)]
31. Kovacs D, Msanga DR, Mshana SE, Bilal M, Oravcova K, Matthews L. Developing practical clinical tools for predicting neonatal mortality at a neonatal intensive care unit in Tanzania. *BMC Pediatr*. 2021;21(1):537. [FREE Full text] [doi: [10.1186/s12887-021-03012-4](https://doi.org/10.1186/s12887-021-03012-4)] [Medline: [34852794](https://pubmed.ncbi.nlm.nih.gov/34852794/)]
32. Ofman G, Caballero MT, Alvarez Paggi D, Marzec J, Nowogrodzki F, Cho H, et al. The discovery BPD (D-BPD) program: study protocol of a prospective translational multicenter collaborative study to investigate determinants of chronic lung disease in very low birth weight infants. *BMC Pediatr*. 2019;19(1):227. [FREE Full text] [doi: [10.1186/s12887-019-1610-8](https://doi.org/10.1186/s12887-019-1610-8)] [Medline: [31279333](https://pubmed.ncbi.nlm.nih.gov/31279333/)]
33. Shalish W, Kanbar LJ, Rao S, Robles-Rubio CA, Kovacs L, Chawla S, et al. Prediction of extubation readiness in extremely preterm infants by the automated analysis of cardiorespiratory behavior: study protocol. *BMC Pediatr*. 2017;17(1):167. [FREE Full text] [doi: [10.1186/s12887-017-0911-z](https://doi.org/10.1186/s12887-017-0911-z)] [Medline: [28716018](https://pubmed.ncbi.nlm.nih.gov/28716018/)]

34. Verder H, Heiring C, Ramanathan R, Scoutaris N, Verder P, Jessen TE, et al. Bronchopulmonary dysplasia predicted at birth by artificial intelligence. *Acta Paediatr.* 2021;110(2):503-509. [FREE Full text] [doi: [10.1111/apa.15438](https://doi.org/10.1111/apa.15438)] [Medline: [32569404](https://pubmed.ncbi.nlm.nih.gov/32569404/)]
35. Amodeo I, De Nunzio G, Raffaelli G, Borzani I, Griggio A, Conte L, et al. A machine and deep learning approach to predict pulmonary hypertension in newborns with congenital diaphragmatic hernia (CLANNISH): protocol for a retrospective study. *PLoS One.* 2021;16(11):e0259724. [FREE Full text] [doi: [10.1371/journal.pone.0259724](https://doi.org/10.1371/journal.pone.0259724)] [Medline: [34752491](https://pubmed.ncbi.nlm.nih.gov/34752491/)]
36. He W, Zhang L, Feng R, Fang W, Cao Y, Sun S, et al. Risk factors and machine learning prediction models for bronchopulmonary dysplasia severity in the Chinese population. *World J Pediatr.* 2023;19(6):568-576. [FREE Full text] [doi: [10.1007/s12519-022-00635-0](https://doi.org/10.1007/s12519-022-00635-0)] [Medline: [36357648](https://pubmed.ncbi.nlm.nih.gov/36357648/)]
37. Kausch SL, Brandberg JG, Qiu J, Panda A, Binai A, Isler J, et al. Cardiorespiratory signature of neonatal sepsis: development and validation of prediction models in 3 NICUs. *Pediatr Res.* 2022;93:1913-1921. [doi: [10.1101/2022.09.28.22280469](https://doi.org/10.1101/2022.09.28.22280469)]
38. Ali R, Li H, Dillman JR, Altaye M, Wang H, Parikh NA, et al. A self-training deep neural network for early prediction of cognitive deficits in very preterm infants using brain functional connectome data. *Pediatr Radiol.* 2022;52(11):2227-2240. [FREE Full text] [doi: [10.1007/s00247-022-05510-8](https://doi.org/10.1007/s00247-022-05510-8)] [Medline: [36131030](https://pubmed.ncbi.nlm.nih.gov/36131030/)]
39. Chen JS, Coyner AS, Ostmo S, Sonmez K, Bajimaya S, Pradhan E, et al. Deep learning for the diagnosis of stage in retinopathy of prematurity: accuracy and generalizability across populations and cameras. *Ophthalmol Retina.* 2021;5(10):1027-1035. [FREE Full text] [doi: [10.1016/j.oret.2020.12.013](https://doi.org/10.1016/j.oret.2020.12.013)] [Medline: [33561545](https://pubmed.ncbi.nlm.nih.gov/33561545/)]
40. Choi RY, Brown JM, Kalpathy-Cramer J, Chan RVP, Ostmo S, Chiang MF, et al. ImagingInformatics in Retinopathy of Prematurity Consortium. Variability in plus disease identified using a deep learning-based retinopathy of prematurity severity scale. *Ophthalmol Retina.* 2020;4(10):1016-1021. [FREE Full text] [doi: [10.1016/j.oret.2020.04.022](https://doi.org/10.1016/j.oret.2020.04.022)] [Medline: [32380115](https://pubmed.ncbi.nlm.nih.gov/32380115/)]
41. Gschwandtner L, Hartmann M, Oberdorfer L, Fürbass F, Klebermaß-Schrehof K, Werther T, et al. Deep learning for estimation of functional brain maturation from EEG of premature neonates. 2020. Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; July 20-24, 2020; Montreal, QC, Canada. [doi: [10.1109/embc44109.2020.9175380](https://doi.org/10.1109/embc44109.2020.9175380)]
42. He L, Li H, Holland SK, Yuan W, Altaye M, Parikh NA. Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework. *Neuroimage Clin.* 2018;18:290-297. [FREE Full text] [doi: [10.1016/j.nicl.2018.01.032](https://doi.org/10.1016/j.nicl.2018.01.032)] [Medline: [29876249](https://pubmed.ncbi.nlm.nih.gov/29876249/)]
43. He L, Li H, Wang J, Chen M, Gozdas E, Dillman JR, et al. A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Sci Rep.* 2020;10(1):15072. [FREE Full text] [doi: [10.1038/s41598-020-71914-x](https://doi.org/10.1038/s41598-020-71914-x)] [Medline: [32934282](https://pubmed.ncbi.nlm.nih.gov/32934282/)]
44. Huang C, Kuo R, Li C, Ting DS, Kang EY, Lai C, et al. Prediction of visual outcomes by an artificial neural network following intravitreal injection and laser therapy for retinopathy of prematurity. *Br J Ophthalmol.* 2020;104(9):1277-1282. [doi: [10.1136/bjophthalmol-2019-314860](https://doi.org/10.1136/bjophthalmol-2019-314860)] [Medline: [31744795](https://pubmed.ncbi.nlm.nih.gov/31744795/)]
45. Lee J, Cai J, Li F, Vesoulis ZA. Predicting mortality risk for preterm infants using random forest. *Sci Rep.* 2021;11(1):7308. [FREE Full text] [doi: [10.1038/s41598-021-86748-4](https://doi.org/10.1038/s41598-021-86748-4)] [Medline: [33790395](https://pubmed.ncbi.nlm.nih.gov/33790395/)]
46. Lin W, Wu T, Chen Y, Chang Y, Lin C, Lin Y. Predicting in-hospital length of stay for very-low-birth-weight preterm infants using machine learning techniques. *J Formos Med Assoc.* 2022;121(6):1141-1148. [FREE Full text] [doi: [10.1016/j.jfma.2021.09.018](https://doi.org/10.1016/j.jfma.2021.09.018)] [Medline: [34629242](https://pubmed.ncbi.nlm.nih.gov/34629242/)]
47. Lure AC, Du X, Black EW, Irons R, Lemas DJ, Taylor JA, et al. Using machine learning analysis to assist in differentiating between necrotizing enterocolitis and spontaneous intestinal perforation: a novel predictive analytic tool. *J Pediatr Surg.* 2021;56(10):1703-1710. [doi: [10.1016/j.jpedsurg.2020.11.008](https://doi.org/10.1016/j.jpedsurg.2020.11.008)] [Medline: [33342603](https://pubmed.ncbi.nlm.nih.gov/33342603/)]
48. Patel AK, Trujillo-Rivera E, Morizono H, Pollack MM. The criticality index-mortality: a dynamic machine learning prediction algorithm for mortality prediction in children cared for in an ICU. *Front Pediatr.* 2022;10:1023539. [FREE Full text] [doi: [10.3389/fped.2022.1023539](https://doi.org/10.3389/fped.2022.1023539)] [Medline: [36533242](https://pubmed.ncbi.nlm.nih.gov/36533242/)]
49. Ruixiang L, Mingrong Y, Li C, Rongxiu Z. Early physical linear growth of small-for-gestational-age infants based on computer analysis method. *J Healthc Eng.* 2021:7227928. [FREE Full text] [doi: [10.1155/2021/7227928](https://doi.org/10.1155/2021/7227928)] [Medline: [34697568](https://pubmed.ncbi.nlm.nih.gov/34697568/)]
50. Saha S, Pagnozzi A, Bourgeat P, George JM, Bradford D, Colditz PB, et al. Predicting motor outcome in preterm infants from very early brain diffusion MRI using a deep learning convolutional neural network (CNN) model. *Neuroimage.* 2020;215:116807. [FREE Full text] [doi: [10.1016/j.neuroimage.2020.116807](https://doi.org/10.1016/j.neuroimage.2020.116807)] [Medline: [32278897](https://pubmed.ncbi.nlm.nih.gov/32278897/)]
51. Sheikhtaheri A, Zarkesh MR, Moradi R, Kermani F. Prediction of neonatal deaths in NICUs: development and validation of machine learning models. *BMC Med Inform Decis Mak.* 2021;21(1):131. [FREE Full text] [doi: [10.1186/s12911-021-01497-8](https://doi.org/10.1186/s12911-021-01497-8)] [Medline: [33874944](https://pubmed.ncbi.nlm.nih.gov/33874944/)]
52. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol.* 2008;8(1):45. [FREE Full text] [doi: [10.1186/1471-2288-8-45](https://doi.org/10.1186/1471-2288-8-45)] [Medline: [18616818](https://pubmed.ncbi.nlm.nih.gov/18616818/)]
53. Reimagining global health through artificial intelligence: the roadmap to AI maturity. Broadband Commission. Sep 2020. URL: https://broadbandcommission.org/wp-content/uploads/2021/02/WGAIinHealth_Report2020.pdf [accessed 2025-09-05]
54. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning.* New York: Springer; 2013.

55. Peyton C, Einspieler C. General movements: a behavioral biomarker of later motor and cognitive dysfunction in NICU graduates. *Pediatr Ann*. 2018;47(4):e159-e164. [doi: [10.3928/19382359-20180325-01](https://doi.org/10.3928/19382359-20180325-01)] [Medline: [29668025](https://pubmed.ncbi.nlm.nih.gov/29668025/)]
56. WHO child growth standards: growth velocity based on weight, length and head circumference: methods and development. World Health Organization. Nov 12, 2009. URL: <https://www.who.int/publications/i/item/9789241547635> [accessed 2025-09-05]
57. Lenke MC. Motor outcomes in premature infants. *Newborn Infant Nurs Rev*. 2003;3(3):104-109. [doi: [10.1016/s1527-3369\(03\)00032-1](https://doi.org/10.1016/s1527-3369(03)00032-1)]
58. Kim SJ, Port AD, Swan R, Campbell JP, Chan RP, Chiang MF. Retinopathy of prematurity: a review of risk factors and their clinical significance. *Surv Ophthalmol*. 2018;63(5):618-637. [FREE Full text] [doi: [10.1016/j.survophthal.2018.04.002](https://doi.org/10.1016/j.survophthal.2018.04.002)] [Medline: [29679617](https://pubmed.ncbi.nlm.nih.gov/29679617/)]
59. Smith J, Pieper CH, Maree D, Gie RP. Compliance of the respiratory system as a predictor for successful extubation in very-low-birth-weight infants recovering from respiratory distress syndrome. *S Afr Med J*. 1999;89(10):1097-1102. [Medline: [10582068](https://pubmed.ncbi.nlm.nih.gov/10582068/)]
60. Perlman JM, Volpe JJ. Prevention of neonatal intraventricular hemorrhage. *Clin Neuropharmacol*. 1987;10(2):126-142. [doi: [10.1097/00002826-198704000-00003](https://doi.org/10.1097/00002826-198704000-00003)] [Medline: [3332606](https://pubmed.ncbi.nlm.nih.gov/3332606/)]
61. Volpe JJ. Neurobiology of periventricular leukomalacia in the premature infant. *Pediatr Res*. 2001;50(5):553-562. [doi: [10.1203/00006450-200111000-00003](https://doi.org/10.1203/00006450-200111000-00003)] [Medline: [11641446](https://pubmed.ncbi.nlm.nih.gov/11641446/)]
62. Caplan MS. Neonatal necrotizing enterocolitis: introduction. *Semin Perinatol*. 2008;32(2):69. [doi: [10.1053/j.semperi.2008.02.001](https://doi.org/10.1053/j.semperi.2008.02.001)] [Medline: [18346529](https://pubmed.ncbi.nlm.nih.gov/18346529/)]
63. Neu J, Modi N, Caplan M. Necrotizing enterocolitis comes in different forms: historical perspectives and defining the disease. *Semin Fetal Neonatal Med*. 2018;23(6):370-373. [doi: [10.1016/j.siny.2018.07.004](https://doi.org/10.1016/j.siny.2018.07.004)] [Medline: [30100524](https://pubmed.ncbi.nlm.nih.gov/30100524/)]
64. Pace E, Yanowitz T. Infections in the NICU: neonatal sepsis. *Semin Pediatr Surg*. 2022;31(4):151200. [FREE Full text] [doi: [10.1016/j.sempedsurg.2022.151200](https://doi.org/10.1016/j.sempedsurg.2022.151200)] [Medline: [36038218](https://pubmed.ncbi.nlm.nih.gov/36038218/)]
65. Thébaud B. Bronchopulmonary dysplasia. *Nat Rev Dis Primers*. 2019;5(1):77. [doi: [10.1038/s41572-019-0134-8](https://doi.org/10.1038/s41572-019-0134-8)] [Medline: [31727885](https://pubmed.ncbi.nlm.nih.gov/31727885/)]
66. Abman SH. Pulmonary hypertension: the hidden danger for newborns. *Neonatology*. 2021;118(2):211-217. [FREE Full text] [doi: [10.1159/000516107](https://doi.org/10.1159/000516107)] [Medline: [33951650](https://pubmed.ncbi.nlm.nih.gov/33951650/)]
67. Artificial intelligence in healthcare. Australian Medical Association. Aug 8, 2023. URL: <https://www.ama.com.au/articles/artificial-intelligence-healthcare> [accessed 2025-09-05]
68. Harnessing artificial intelligence for health. World Health Organization. URL: <https://www.who.int/teams/digital-health-and-innovation/harnessing-artificial-intelligence-for-health> [accessed 2025-09-05]
69. A national policy roadmap for artificial intelligence in healthcare. Australian Alliance for Artificial Intelligence in Healthcare. URL: https://www.mq.edu.au/_data/assets/pdf_file/0005/1281758/AAAiH_NationalAgendaRoadmap_20231122.pdf [accessed 2025-09-05]
70. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. *Healthc Manage Forum*. 2020;33(1):10-18. [FREE Full text] [doi: [10.1177/0840470419873123](https://doi.org/10.1177/0840470419873123)] [Medline: [31550922](https://pubmed.ncbi.nlm.nih.gov/31550922/)]
71. Kwok TC, Henry C, Saffaran S, Meeus M, Bates D, Van Laere D, et al. Application and potential of artificial intelligence in neonatal medicine. *Semin Fetal Neonatal Med*. 2022;27(5):101346. [FREE Full text] [doi: [10.1016/j.siny.2022.101346](https://doi.org/10.1016/j.siny.2022.101346)] [Medline: [35473694](https://pubmed.ncbi.nlm.nih.gov/35473694/)]
72. Hintz SR, Bann CM, Ambalavanan N, Cotten CM, Das A, Higgins RD, et al. Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Predicting time to hospital discharge for extremely preterm infants. *Pediatrics*. 2010;125(1):e146-e154. [FREE Full text] [doi: [10.1542/peds.2009-0810](https://doi.org/10.1542/peds.2009-0810)] [Medline: [20008430](https://pubmed.ncbi.nlm.nih.gov/20008430/)]
73. Snowdon A. Digital health: a framework for healthcare transformation. HIMSS. 2020. URL: <https://www.himss.org/sites/hde/files/media/file/2022/12/21/dhi-white-paper.pdf> [accessed 2025-09-05]
74. White RD, Smith JA, Shepley MM. Recommended standards for newborn ICU design, eighth edition. *J Perinatol*. 2013;33(Suppl 1):S2-S16. [doi: [10.1038/jp.2013.10](https://doi.org/10.1038/jp.2013.10)] [Medline: [23536026](https://pubmed.ncbi.nlm.nih.gov/23536026/)]
75. Kim JH, Sampath V, Canvasser J. Challenges in diagnosing necrotizing enterocolitis. *Pediatr Res*. 2020;88(Suppl 1):16-20. [doi: [10.1038/s41390-020-1090-4](https://doi.org/10.1038/s41390-020-1090-4)] [Medline: [32855507](https://pubmed.ncbi.nlm.nih.gov/32855507/)]
76. Leon C, Carrault G, Pladys P, Beuchee A. Early detection of late onset sepsis in premature infants using visibility graph analysis of heart rate variability. *IEEE J Biomed Health Inform*. 2021;25(4):1006-1017. [doi: [10.1109/jbhi.2020.3021662](https://doi.org/10.1109/jbhi.2020.3021662)]
77. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94-98. [FREE Full text] [doi: [10.7861/futurehosp.6-2-94](https://doi.org/10.7861/futurehosp.6-2-94)] [Medline: [31363513](https://pubmed.ncbi.nlm.nih.gov/31363513/)]
78. Hartley C. Toward personalized medicine for pharmacological interventions in neonates using vital signs. *Paediatr Neonatal Pain*. 2021;3(4):147-155. [FREE Full text] [doi: [10.1002/pne2.12065](https://doi.org/10.1002/pne2.12065)] [Medline: [35372840](https://pubmed.ncbi.nlm.nih.gov/35372840/)]
79. Kwok TNC, Henry C, Saffaran S, Meeus M, Bates D, Van Laere D, et al. Application and potential of artificial intelligence in neonatal medicine. *Semin Fetal Neonatal Med*. 2022;27(5):101346. [FREE Full text] [doi: [10.1016/j.siny.2022.101346](https://doi.org/10.1016/j.siny.2022.101346)] [Medline: [35473694](https://pubmed.ncbi.nlm.nih.gov/35473694/)]
80. Lu Z, Qian P, Bi D, Ye Z, He X, Zhao Y, et al. Application of AI and IoT in clinical medicine: summary and challenges. *Curr Med Sci*. 2021;41(6):1134-1150. [doi: [10.1007/s11596-021-2486-z](https://doi.org/10.1007/s11596-021-2486-z)] [Medline: [34939144](https://pubmed.ncbi.nlm.nih.gov/34939144/)]

81. Shaheen MY. Applications of artificial intelligence (AI) in healthcare: a review. ScienceOpen Preprints. [doi: [10.14293/s2199-1006.1.sor.ppvry8k.v1](https://doi.org/10.14293/s2199-1006.1.sor.ppvry8k.v1)]
82. Rochefort CM, Rathwell BA, Clarke SP. Rationing of nursing care interventions and its association with nurse-reported outcomes in the neonatal intensive care unit: a cross-sectional survey. BMC Nurs. 2016;15(1):46. [FREE Full text] [doi: [10.1186/s12912-016-0169-z](https://doi.org/10.1186/s12912-016-0169-z)] [Medline: [27489507](https://pubmed.ncbi.nlm.nih.gov/27489507/)]
83. Bowe AK, Lightbody G, Staines A, Murray DM, Norman M. Prediction of 2-year cognitive outcomes in very preterm infants using machine learning methods. JAMA Netw Open. 2023;6(12):e2349111. [FREE Full text] [doi: [10.1001/jamanetworkopen.2023.49111](https://doi.org/10.1001/jamanetworkopen.2023.49111)] [Medline: [38147334](https://pubmed.ncbi.nlm.nih.gov/38147334/)]
84. Chung HW, Chen J, Chen H, Ko F, Ho S, Taiwan Premature Infant Follow-Up Network. Developing a practical neurodevelopmental prediction model for targeting high-risk very preterm infants during visit after NICU: a retrospective national longitudinal cohort study. BMC Med. 2024;22(1):68. [FREE Full text] [doi: [10.1186/s12916-024-03286-2](https://doi.org/10.1186/s12916-024-03286-2)] [Medline: [38360711](https://pubmed.ncbi.nlm.nih.gov/38360711/)]
85. Tenero L, Piazza M, Sandri M, Ferrante G, Giacomello E, Ficial B, et al. Early diagnosis of bronchopulmonary dysplasia with e-nose: a pilot study in preterm infants. Sensors (Basel). 2024;24(19):6282. [FREE Full text] [doi: [10.3390/s24196282](https://doi.org/10.3390/s24196282)] [Medline: [39409322](https://pubmed.ncbi.nlm.nih.gov/39409322/)]
86. Montagna S, Magno D, Ferretti S, Stelluti M, Gona A, Dionisi C, et al. Combining artificial intelligence and conventional statistics to predict bronchopulmonary dysplasia in very preterm infants using routinely collected clinical variables. Pediatr Pulmonol. 2024;59(12):3400-3409. [doi: [10.1002/ppul.27216](https://doi.org/10.1002/ppul.27216)] [Medline: [39150150](https://pubmed.ncbi.nlm.nih.gov/39150150/)]
87. Peterson B, Juarez EF, Moore B, Hernandez EJ, Frise E, Li J, et al. MPSE identifies newborns for whole genome sequencing within 48 h of NICU admission. NPJ Genom Med. 2025;10(1):47. [FREE Full text] [doi: [10.1038/s41525-025-00506-3](https://doi.org/10.1038/s41525-025-00506-3)] [Medline: [40506467](https://pubmed.ncbi.nlm.nih.gov/40506467/)]
88. Chen S, Zhao X, Wu Z, Cao K, Zhang Y, Tan T, et al. Multi-risk factors joint prediction model for risk prediction of retinopathy of prematurity. EPMA J. 2024;15(2):261-274. [FREE Full text] [doi: [10.1007/s13167-024-00363-7](https://doi.org/10.1007/s13167-024-00363-7)] [Medline: [38841619](https://pubmed.ncbi.nlm.nih.gov/38841619/)]
89. Coyner AS, Young BK, Ostmo SR, Grigorian F, Ells A, Hubbard B, et al. Use of an artificial intelligence-generated vascular severity score improved plus disease diagnosis in retinopathy of prematurity. Ophthalmology. 2024;131(11):1290-1296. [doi: [10.1016/j.ophtha.2024.06.006](https://doi.org/10.1016/j.ophtha.2024.06.006)] [Medline: [38866367](https://pubmed.ncbi.nlm.nih.gov/38866367/)]
90. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in healthcare. Annu Rev Biomed Data Sci. 2021;4(1):123-144. [FREE Full text] [doi: [10.1146/annurev-biodatasci-092820-114757](https://doi.org/10.1146/annurev-biodatasci-092820-114757)] [Medline: [34396058](https://pubmed.ncbi.nlm.nih.gov/34396058/)]
91. Erdogan Yildirim A, Canayaz M. Machine learning-based prediction of length of stay (LoS) in the neonatal intensive care unit using ensemble methods. Neural Comput Appl. 2024;36(23):14433-14448. [doi: [10.1007/s00521-024-09831-7](https://doi.org/10.1007/s00521-024-09831-7)]
92. Ganatra HA, Latifi SQ, Baloglu O. Pediatric intensive care unit length of stay prediction by machine learning. Bioengineering (Basel). 2024;11(10):962. [FREE Full text] [doi: [10.3390/bioengineering11100962](https://doi.org/10.3390/bioengineering11100962)] [Medline: [39451338](https://pubmed.ncbi.nlm.nih.gov/39451338/)]
93. Iqbal F, Chandra P, Lewis L, Acharya D, Purkayastha J, Shenoy P, et al. Application of artificial intelligence to predict the sepsis in neonates admitted in neonatal intensive care unit. J Neonatal Nurs. 2024;30(2):141-147. [FREE Full text] [doi: [10.1016/j.jnn.2023.07.016](https://doi.org/10.1016/j.jnn.2023.07.016)]
94. Yang M, Peng Z, van Pul C, Andriessen P, Dong K, Silvertand D, et al. Continuous prediction and clinical alarm management of late-onset sepsis in preterm infants using vital signs from a patient monitor. Comput Methods Programs Biomed. 2024;255:108335. [FREE Full text] [doi: [10.1016/j.cmpb.2024.108335](https://doi.org/10.1016/j.cmpb.2024.108335)] [Medline: [39047574](https://pubmed.ncbi.nlm.nih.gov/39047574/)]
95. Vignolle GA, Bauerstätter P, Schönthaler S, Nöhammer C, Olischar M, Berger A, et al. Predicting outcomes of preterm neonates post intraventricular hemorrhage. Int J Mol Sci. 2024;25(19):10304. [FREE Full text] [doi: [10.3390/ijms251910304](https://doi.org/10.3390/ijms251910304)] [Medline: [39408633](https://pubmed.ncbi.nlm.nih.gov/39408633/)]

Abbreviations

- AI:** artificial intelligence
- AUROC:** area under the receiver operating characteristic curve
- BPD:** bronchopulmonary dysplasia
- MeSH:** Medical Subject Headings
- MPSE:** Mendelian Phenotype Search Engine
- MRI:** magnetic resonance imaging
- NICU:** neonatal intensive care unit
- PH:** pulmonary hypertension
- PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- QCC:** Quality Checklist for Clinical Case Series
- RMSE:** root mean square error
- ROP:** retinopathy of prematurity
- WHO:** World Health Organization
- XAI:** explainable artificial intelligence

Edited by A Schwartz; submitted 14.Jun.2024; peer-reviewed by MAA Bayoumi, E Wachira, J Schouten; comments to author 03.Dec.2024; revised version received 26.Jan.2025; accepted 17.Aug.2025; published 03.Oct.2025

Please cite as:

Tudor S, Bhatia R, Liem M, Wani TA, Boyd J, Raza Khan U

Opportunities and Challenges of Using Artificial Intelligence in Predicting Clinical Outcomes and Length of Stay in Neonatal Intensive Care Units: Systematic Review

J Med Internet Res 2025;27:e63175

URL: <https://www.jmir.org/2025/1/e63175>

doi: [10.2196/63175](https://doi.org/10.2196/63175)

PMID:

©Samantha Tudor, Risha Bhatia, Michael Liem, Tafheem Ahmad Wani, James Boyd, Urooj Raza Khan. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 03.Oct.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.