



Published in final edited form as:

*J Perinatol.* 2024 January ; 44(1): 1–11. doi:10.1038/s41372-023-01848-5.

## Transforming neonatal care with artificial intelligence: challenges, ethical consideration, and opportunities

Brynne A. Sullivan<sup>1</sup>, Kristyn Beam<sup>2</sup>, Zachary A. Vesoulis<sup>3</sup>, Khyzer B. Aziz<sup>4</sup>, Ameena N. Husain<sup>5</sup>, Lindsey A. Knake<sup>6</sup>, Alvaro G. Moreira<sup>7</sup>, Thomas A. Hooven<sup>8</sup>, Elliott M. Weiss<sup>9,10</sup>, Nicholas R. Carr<sup>5</sup>, George T. El-Ferzli<sup>11</sup>, Ravi M. Patel<sup>12</sup>, Kelsey A. Simek<sup>5</sup>, Antonio J. Hernandez<sup>7</sup>, James S. Barry<sup>13</sup>, Ryan M. McAdams<sup>14</sup>

<sup>1</sup>Division of Neonatology, Department of Pediatrics, University of Virginia School of Medicine, Charlottesville, VA, USA.

<sup>2</sup>Department of Neonatology, Beth Israel Deaconess Medical Center, Boston, MA, USA.

<sup>3</sup>Division of Newborn Medicine, Department of Pediatrics, Washington University in St. Louis, St. Louis, MO, USA.

<sup>4</sup>Division of Neonatology, Department of Pediatrics, Johns Hopkins University, Baltimore, MD, USA.

<sup>5</sup>Division of Neonatology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA.

<sup>6</sup>Division of Neonatology, Department of Pediatrics, University of Iowa, Iowa City, IA, USA.

<sup>7</sup>Division of Neonatology, Department of Pediatrics, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.

<sup>8</sup>Division of Newborn Medicine, Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.

<sup>9</sup>Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, USA.

<sup>10</sup>Treuman Katz Center for Pediatric Bioethics and Palliative Care, Seattle Children's Research Institute, Seattle, WA, USA.

<sup>11</sup>Division of Neonatology, Department of Pediatrics, Ohio State University, Nationwide Children's Hospital, Columbus, OH, USA.

<sup>12</sup>Division of Neonatology, Department of Pediatrics, Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, GA, USA.

---

Reprints and permission information is available at <http://www.nature.com/reprints>

Correspondence and requests for materials should be addressed to Ryan M. McAdams. [mcadams@pediatrics.wisc.edu](mailto:mcadams@pediatrics.wisc.edu).

### AUTHOR CONTRIBUTIONS

RMM contributed to the conception and design. All authors contributed to the first draft. RMM, KB, ZAV, and BAS revised the article. ANH and KBA created the Figures. All authors contributed a final revision of the article, including providing intellectual content. All authors read and approved the final manuscript.

### COMPETING INTERESTS

The authors declare no competing interests.

<sup>13</sup>Division of Neonatology, Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO, USA.

<sup>14</sup>Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA.

## Abstract

Artificial intelligence (AI) offers tremendous potential to transform neonatology through improved diagnostics, personalized treatments, and earlier prevention of complications. However, there are many challenges to address before AI is ready for clinical practice. This review defines key AI concepts and discusses ethical considerations and implicit biases associated with AI. Next we will review literature examples of AI already being explored in neonatology research and we will suggest future potentials for AI work. Examples discussed in this article include predicting outcomes such as sepsis, optimizing oxygen therapy, and image analysis to detect brain injury and retinopathy of prematurity. Realizing AI's potential necessitates collaboration between diverse stakeholders across the entire process of incorporating AI tools in the NICU to address testability, usability, bias, and transparency. With multi-center and multi-disciplinary collaboration, AI holds tremendous potential to transform the future of neonatology.

---

## INTRODUCTION

Artificial intelligence (AI) is permeating society and will transform healthcare. As neonatologists, we must understand AI concepts to integrate these tools safely and effectively. We provide an overview of AI for neonatology and neonatal intensive care unit (NICU) care and thoughts about the future. First, we define critical concepts like machine learning (ML) and deep learning. We then discuss challenges and ethical considerations around developing and implementing AI in the NICU. Next, we review how AI may improve diagnosis, treatment, and prediction of outcomes in neonatology. Finally, we explore future directions for integrating AI into clinical care.

## DEFINITION OF AI, MACHINE LEARNING, AND DEEP LEARNING

AI encompasses a multitude of computer algorithms that can augment human abilities in problem-solving, classification, and prediction. While the term "AI" has gained popularity, the roots of AI trace back to the 1950s when engineers created intelligent machines to assist human tasks. Although not commonly referred to as AI, many clinicians routinely use such technology daily in the form of automated ECG interpretations.

ML, a subdomain of AI, utilizes algorithms that iteratively learn from patterns in large datasets, creating predictive models ranging from simple logistic regression to complex neural networks [1, 2]. The appeal of ML lies in its ability to uncover associations through guided (i.e., supervised ML) or unguided (i.e., unsupervised ML) discovery, potentially revealing important and new insights that are elusive to clinicians. Supervised learning involves teaching a computer pattern using labeled data, like showing a child pictures with names, while unsupervised learning allows the AI to find patterns without specific guidance, much like a child exploring and grouping toys on their own.

Deep learning, a subdomain of ML, uses artificial neural networks with several layers (“deep” structures) for more complex pattern recognition. These neural networks can be conceptualized as mimicking the structure and function of the human brain—where neurons are interconnected—to process and analyze complex data patterns and make decisions without explicit programming.

We will focus on specific considerations for developing, generating, and implementing AI models in neonatology. While many of the concepts and challenges discussed in this review and summarized in the figures apply to traditional statistical methods and epidemiological approaches, we concentrate on how they apply to the growing computational power of novel AI methods applied to large datasets. We will also discuss precautions, ethical considerations, and practical aspects of deploying these models in clinical care.

## OVERVIEW OF AI IN NEONATOLOGY

Although healthcare is generally slow to embrace technology, AI serves as an exception, showing marked advancements particularly in adult medicine, which outpaces the rapid growth in pediatrics and neonatology [3-5]. As of 2021, there were approximately 3000 publications about ML or AI in adult literature, compared to a mere 200 articles in neonatal literature, a noteworthy discrepancy considering the wealth of clinical data generated during a NICU hospitalization [4]. Nevertheless, neonatology represented the top specialty in a recent systematic review of pediatric ML models, accounting for 24% ( $n = 87$ ) of the 363 articles reviewed, followed by psychiatry at 20% ( $n = 73$ ) and neurology at 11% ( $n = 39$ ) [3]. Despite the increasing academic interest, implementation of AI in clinical settings remains limited.

Neonatology is uniquely positioned to generate impactful AI research due to the availability of complex, rich, and significant data volume generated over prolonged inpatient hospitalizations. There are opportunities to use existing stored and real-time data in applying AI algorithms to improve neonatal outcomes. For example, preterm birth is the leading cause of neonatal death in high-resource communities and its cause is multifactorial and variable. Neonatal–perinatal research has harnessed large, population-based datasets to predict preterm birth, which has begun to uncover patterns that identify targets for intervention [6-10]. However, using large electronic health record (EHR) data or insurance claims databases to train these models may include inherently biased data that generates predictions that are useless to clinicians or may have embedded implicit biases. Figure 1 summarizes the many challenges in the quality of data used for AI applications in neonatology. While data quality challenges any analysis, we hypothesize that such issues have a greater impact on analyses using AI methodology.

## APPLICATIONS OF AI IN NEONATOLOGY

Researchers have applied AI methods to many conditions or aspects of neonatology worthy of prediction. Here, we review prediction models applied to various outcomes of interest in neonatology with a focus on those translated to clinical care. Table 1 lists neonatal conditions that may potentially be addressed using AI technology.

## Sepsis

Diagnosing sepsis in premature infants is particularly challenging because its subtle and delayed manifestation can resemble those of immature organ systems [11]. Neonatal sepsis gradually becomes apparent as inflammation and organ dysfunction progress. Advanced analytics and predictive models can detect data patterns that alert the neonatal medical team of a subacute, potentially catastrophic illness. Given non-specific symptoms like apnea and respiratory distress and the infrequent occurrence of neonatal sepsis, clinicians face a dilemma: balancing early detection against the risk of over-treatment. Therefore, the urgency of designing sepsis alert systems is paramount, minimizing false alarms while empowering healthcare providers to interpret AI model outputs in the broader clinical context. This approach could guide decisions about antibiotic initiation and therapy duration, particularly among very low birthweight (VLBW) infants, where timely diagnosis and treatment significantly improve outcomes.

Progress has been made in predictive monitoring for late-onset sepsis (LOS) with heart rate characteristics (HRC) monitoring. Heart rate characteristics monitoring for LOS in VLBW infants is one of very few prediction models tested in a large, multicenter randomized clinical trial (RCT) [12]. The results of this RCT showed that display of the sepsis risk score, called the HRC index or HeRO score, reduced NICU mortality by 20% and death within 30 days of sepsis by 40% [12, 13]. While models that provide early warning of increased risk are invaluable tools, they should complement, rather than supplant, clinical judgment.

Methods to predict neonatal sepsis have included various approaches such as logistic regression, random forest, XGBoost, and neural networks, which span from basic statistics to more complex AI methods [14-16]. An early-onset sepsis calculator employs Bayesian statistics to adjust predicted probabilities based on clinical assessments. This prediction tool has been widely adopted due to its effectiveness in reducing antibiotic exposure [17-19]. Despite the varying results and adoption of early- and late-onset sepsis models, they provide promising foundations for refining the prediction and early detection of neonatal sepsis.

## Necrotizing enterocolitis

Necrotizing enterocolitis (NEC) is another potentially devastating complication of prematurity, affecting 2–7% of preterm infants born at <32 weeks' gestation [20]. The onset of NEC often exhibits non-specific clinical signs such as emesis, abdominal distention, and increased apnea with bradycardia and desaturation, resulting in a low positive predictive value for NEC.

The Bell's staging criteria outline that a definitive diagnosis of NEC may be contingent on radiographic findings of pneumatosis, portal venous gas, or pneumoperitoneum, or surgical exploration revealing necrotic bowel [21, 22]. However, many clinically diagnosed NEC cases do not exhibit these classical radiologic features [23], and the radiographic diagnosis of NEC suffers from high rates of false positivity. Therefore, developing AI to identify patterns in clinical data, physiology, and imaging is a promising alternative strategy for NEC diagnosis.

Recent studies have aimed to leverage various data sources and ML methods to enhance the current diagnostic strategy. Examples of ML methods applied to demographic and clinical data include two recent studies using prediction models to distinguish NEC with perforation from spontaneous intestinal perforation, which have different underlying pathophysiology but can present with similar clinical features [24, 25]. However, models using EHR data may be limited by including diagnostic criteria as input features.

Similar to sepsis risk prediction, studies incorporating physiologic data into ML models identify reduced heart rate variability [26, 27] and abnormal heart rate characteristics [28] as important predictors of imminent NEC diagnosis. Groups using -omics data in NEC ML prediction models have tested diverse modeling methods and novel biomarker predictors, including models using stool microbiota data [29], targeted stool sphingolipid metabolomics [30], and urinary biomarkers [31]. While many of these models demonstrate promising performance in retrospective cohorts, none have been implemented or tested prospectively.

Despite promising results of diverse AI methods providing more precise and timely NEC diagnoses than traditional methods, several concerns remain. Issues include inaccurate training dataset labels, heterogeneous clinical phenotypes in early-stage NEC, and impractical accessibility of necessary model inputs, such as stool samples and 'omics data. These issues have led to questions about the broad applicability of NEC prediction tools without external or prospective validation. These challenges must be addressed to successfully integrate these diagnostic technologies into neonatology practice.

### **Bronchopulmonary dysplasia**

Although neonatal care has advanced, allowing the smallest and most premature infants to survive, the rate of bronchopulmonary dysplasia (BPD) remains high, indicating that these advancements might be a contributing factor to its persistence [32]. Despite decades of research in the field, early BPD prediction and risk-based treatment strategies remain a challenge [33, 34]. The promise of AI and ML for BPD lies in the potential to predict long-term disease burden, guide treatment strategies, and discover novel risk factors [35].

Researchers have made significant strides in developing prediction tools for BPD using clinical data, physiologic data, imaging, and genomic biomarkers [36-42]. Both traditional statistical methods and AI and ML models have been employed to predict BPD, its severity, and potential outcomes. An early example of a BPD prediction model used logistic regression and multicenter data from the Neonatal Research Network [37]. This model was made available using an online calculator, allowing for widespread use in clinical practice. In a more recent example, perinatal variables, data on early-life respiratory support, and AI and ML methods were used to predict BPD-free survival with strong performance in a retrospective cohort [39]. This model was also made available online, which will allow for ongoing evaluation. We highlight these examples as tools available for evaluation and use at the bedside, where risk stratification could aid in decisions about treatment. These tools could enable targeted interventions for at-risk infants to improve respiratory outcomes.

Using genomic markers could provide an alternative strategy for precision targeting of BPD phenotypes [40]. Studies have combined AI and ML methods with targeted transcriptomics

[40] and exome sequencing [38] with high BPD prediction accuracy. Moreira, et al. combined ML with gene expression data to better understand mechanisms underpinning BPD and risk-stratify patients for developing BPD. Gene-centric ML models improved BPD prediction in the first postnatal week compared to a fixed clinical model using gestational age and birth weight [42]. Genomic and other ‘omics models for BPD prediction hold great promise but their inputs remains impractical for use in clinical care. Still, such prediction models add to what is known on signatures and phenotypes of BPD that will fuel future research and treatment developments.

### Retinopathy of prematurity

Retinopathy of prematurity (ROP), a potentially sight-threatening disorder, affects premature infants. Employing AI for the swift and accurate diagnosis of ROP could facilitate timely interventions and prevent persistent visual impairments. Conventionally, ROP diagnosis is conducted through manual examinations by expert ophthalmologists, a process that is labor-intensive, subjective, and susceptible to observer variability. In contrast, AI-based strategies have focused on perinatal and NICU data for early risk prediction and retinal image analysis for improving ROP screening efficiency and consistency.

Regarding the prediction of ROP, gestational age and birthweight are so profoundly linked to severe ROP that they offer little scope for improvement with additional physiologic data [43, 44]. Oxygenation measures, despite their association with ROP pathophysiology, do not contribute substantial predictive information for predicting severe ROP [45-47]. Postnatal weight gain data add to demographics in ROP prediction models [48] due to the role of insulin-like growth factor 1 in retinal vascular development. Several models incorporating postnatal growth data have been developed in large, multinational cohorts with adequate sensitivity and specificity for use as an adjunct to screening. However, weight gain due to pathologic conditions such as sepsis, hydrocephalus, or patent ductus arteriosus cause such models to fall short at an individual level.

AI analysis of retinal images, facilitated by the advent of retinal cameras, offers a method to improve ROP evaluation’s efficiency and inter-rater reliability [49, 50]. This approach has broadened the diagnostic potential beyond ROP, revealing connections with other systemic diseases like hypertension, renal failure, sleep apnea, and diabetes in adults [51]. Therefore, neonatal and infant retinal image analysis could be a diagnostic tool for disorders where vasculogenesis and angiogenesis may be compromised.

### Brain injury

Brain injury is a frequent and serious complication of prematurity, including hypoxic-ischemic encephalopathy (HIE) and preterm brain injury, such as intraventricular hemorrhage (IVH), white matter injury, and cerebellar hemorrhage. While neuroimaging, including cranial ultrasound (CUS) and brain MRI, has been the primary tool for prognosis, AI and ML innovations have attempted to augment standard imaging and functional brain assessment.

Neuroimaging interpretation is challenged by subjectivity, resulting in a lack of inter-rater reliability. Consequently, the AAP Choosing Wisely campaign advised against the

routine use of term-equivalent MRIs due to their low positive predictive value [52]. This recommendation has led to a risk-stratified imaging approach, where only infants with abnormal CUS or significant NICU complications undergo MRI [53]. While standardized MRI scoring systems have been developed for preterm and term HIE infants [54-57], their application is labor-intensive and necessitates a reader trained in the system. As a result, subjective interpretation of neonatal MRI remains the clinical standard.

Towards improving imaging interpretation, researchers have developed AI-based analytic pipelines to perform automated brain segmentation and volume measurements. These methods use image analysis to replace painstaking manual segmentation to label regional boundaries with minimal human intervention, facilitating research to uncover associations between regional brain volumes and neurologic outcomes [58-60]. Studies that utilize ML to identify features associated with long-term neurodevelopmental outcomes are underway [61], signifying a promising direction for ML in neonatal brain injury research.

Recent advancements for brain injury also include AI-enabled seizure detection algorithms designed specifically for neonates [62, 63]. Mathieson et al., described one of the earliest neonatal AI seizure detection algorithms, based on a support vector machine model [62]. This strategy was able to recognize neonatal seizure with 75% accuracy, far exceeding clinical detection of seizures, while still short of trained neurologists. A subsequent iteration of the model by the same research group utilized a real-time display of seizure probability based on an AI-algorithm (called Algorithm for Neonatal Seizure Recognition or ANSer) compared to no algorithm display. The algorithm group had a greater percentage of correctly identified seizures (66 vs 45%) and provided a highly accessible alert to bedside providers without requiring specific expertise in reading the EEG. Other ML models predict seizures associated with HIE using the combination of EEG and clinical data [64-68]. Automated algorithms such as these could be valuable in the care of infants with HIE in hospitals without a pediatric neurology specialist, but are not yet widely used or available in the clinical setting. Potential benefits include earlier seizure treatment, reduced seizure burden, and improved long-term outcomes [69, 70]. See Table 1 for additional opportunities to apply AI and ML technology to neonatal neurocritical care.

### Optimization of oxygen therapy

Oxygen therapy, a necessity in managing lung disease in preterm infants, has transformed neonatology and substantially reduced mortality. However, oxygen misuse has been linked to ROP and blindness due to excessive oxygen saturation [71]. Current evidence, including large clinical trials and meta-analyses, supports maintaining pulse oximetry (SpO<sub>2</sub>) goal oxygen saturations of 91–95% to minimize the risk of mortality and NEC, although this higher target (compared to 85–89%) may increase the risk of ROP and BPD [72]. However, this generalized approach doesn't consider individual risk differences and inaccuracies in pulse oximeter readings, particularly among Black neonatal and pediatric populations [73, 74].

In the NICU, oxygen therapy is commonly adjusted by manually modifying the fraction of inspired oxygen (FiO<sub>2</sub>), which can be inconsistent, tedious, and triggers an excess of alarms. Automated oxygen titration systems based on proportional-integral-derivative algorithms

have been developed and implemented in many European NICUs. Still, their adoption has been limited due to their inability to handle the non-linearity and individual variability of oxygen response. This includes inconsistent changes in SpO<sub>2</sub> with equal changes in FiO<sub>2</sub> [75-77].

ML can navigate this non-linear SpO<sub>2</sub>/FiO<sub>2</sub> relationship and quickly adapt to individual differences. In a recent review, four different ML strategies (decision tree, nearest neighbor, Bayes, and support vector machine) were evaluated to improve detection of true hypoxemia on pulse oximetry, achieving a remarkable specificity of more than 99% across all 4 models, and the greatest sensitivity of 87% in the decision tree model [78]. This marks a significant step towards reducing false alarms and system lability using only a software overlay on top of existing hardware, an economical and practical strategy. Optimal ML-based oxygen titration models could allow accurate, individualized, and efficient neonatal oxygen therapy, but widespread clinical implementation remains elusive.

## **CHALLENGES IN AND SOLUTIONS TO IMPLEMENTING AI IN NEONATOLOGY**

### **Data quality and availability**

The high volume of data generated by NICU patients makes neonatology a logical field to apply AI technologies. However, the utility and accuracy of these technologies fundamentally rely on the data quality used for model development [79]. NICU data have diverse sources and complex structures that pose specific challenges for AI applications. Hurdles in each stage of data processing are highlighted in Fig. 1, including the lack of data standardization across centers and the prevalence of missing data. Addressing data quality issues is essential for ensuring AI model reliability and sharing data to facilitate the utilization of large, multicenter, representative datasets needed for high-quality AI development. An interdisciplinary approach can be instrumental in overcoming these challenges. Involving clinicians early in development enriches the process with medical expertise. Even those not trained in advanced data analytics can provide valuable insights into data provenance, quality, and context. A collaborative, multidisciplinary approach can help create more robust, reliable AI models that effectively transform NICU care.

More investment is needed in creating and curating datasets, and historical data should not be considered the default ground truth. Therefore, datasets may need expiration dates if not updated. We also need to create registries for AI tools whereby tools designed in different institutions using different populations could be captured as metadata and used to help determine when the underlying algorithms may be suitable for a particular patient. To build such a registry, technology and policies should be developed to enable it to be used as a part of an ecosystem [80].

### **Transparency and interpretability**

The current state of transparency and interpretability in neonatal AI is unsatisfactory [81, 82]. Best practices, standards, guidelines, and tools are needed to improve the current state of transparency and interpretability in neonatal AI [83-85]. In practice, AI and ML



models often require healthcare providers to reconcile precise quantitative results with limited knowledge of how it arrived at the prediction. This “black box” issue of transparency makes many in healthcare uncomfortable with predictive models. Despite the remarkable ability of these models to sift through vast datasets, identify patterns, and make decisions, understanding the computation behind these machine-learned decisions remains complex, especially when predictions do not align with human logic. At present, there is a lack of consensus on how to critically appraise the development and performance of AI healthcare models and tools [86, 87].

Explainability in AI refers to the degree to which a human can understand the decision-making process of a model. In many instances, there is a desire that decisions made by AI systems need to be explained and justified, especially when those decisions may indicate treatment plans with significant risk have significant impacts [82]. Furthermore, without fully explainable models, providers may not be able to choose the correct intervention (e.g., model predicts sepsis but not antibiotic choice). Methods such as feature importance (e.g., Shapley Additive explanations) are useful for large predictive models, and saliency maps are valuable tools in image recognition. However, despite their utility, these techniques present challenges, which include substantial computational costs and potential for misinterpretations, particularly in the context of image analysis [83].

Responsible deployment of AI-based clinical decision support is a shared task; providers must gain a working knowledge of the complexities of AI, how models and features are selected, and the potential limitations of these predictive models [84, 85, 88] and developers must continue to push boundaries to generate new healthcare-specific technical advances to demystify AI’s decision-making processes without compromising its efficacy.

Beyond the initial inception of an AI model, it is equally essential to conduct thorough inspections or audits of performance to ensure trust in using AI throughout its entire lifecycle, as populations, treatments, and disease incidence evolves over years. This process can be achieved by establishing a team of experts from various fields who work with key stakeholders to define the boundaries and context for the assessment [89].

Additionally, experimental evaluation of AI-models should be considered using RCTs, which not only consider the performance of the model (e.g., area under the receiver operating characteristic (AUROC)) but also the impact such a model may have on clinical practice and outcomes. For example, a model may have a high AUROC in predicting sepsis, but when testing empirically may not improve sepsis-related outcomes such as mortality. Despite many published sepsis prediction models, the HeRO trial is the only RCT performed to evaluate a sepsis risk model in the clinical setting [12]. While the RCT remains the gold standard for testing the impact of a novel technology or device, the number of infants needed to show clinically significant impact from a prediction tool, in many cases, is not ethically or financially feasible.

It is also essential to ensure that multidisciplinary teams who evaluate the application are selected to represent all stakeholders, including patients, legal perspectives, and scientific and technological expertise. Gathering consensus among domain experts is critical, and

clear communication and documentation of any disagreements or tradeoffs are necessary for transparency and clarity for future users.

### **Ethical considerations and minimizing bias**

Figure 2 summarizes considerations for ensuring ethical AI/ML in neonatology. Bias, equity, and fairness are essential considerations in AI development. Systematic gaps in healthcare can result from device or test performance discrepancies, unequal access to resources, and both implicit and explicit racism [74, 90-92]. Racial and ethnic disparities in neonatal outcomes, especially mortality, remain despite narrowing gaps [93, 94]. Such biases become integrated into the clinical data used for AI and ML models, potentially compromising data availability, quality, and accuracy. ML algorithms can unintentionally propagate these biases, posing concerns for privacy and safety. Developers must be aware of and mitigate these biases when using clinical data to train AI and ML algorithms.

To address this, researchers are exploring less biased data sources, like multi-omics and physiologic data [15, 95-100]. Assessing a model's performance end-to-end, from development to clinical deployment, can also be beneficial. To address bias and fairness during development, we can increase the model's capacity, include data from a minority group of patients, or adjust resource allocation.

Testability, usability, and safety of AI technology are important considerations for ethical AI in neonatology (Fig. 2). Additionally, we must avoid creating a generation of physicians with automation bias [101]. An over-reliance on automated systems, or "automation complacency," can lead to diminished awareness of potential errors, reduced vigilance, and an oversight of ethical considerations. Striking a balance between human judgment and automated decision-making is crucial. This involves regular evaluation of AI systems, continuous education of algorithms and those using the algorithms, and a comprehensive understanding of the model's limitations. Responsibility for these tasks typically falls on AI developers, data scientists, biostatisticians, and engineers but should be broadened to include those in the healthcare field.

Creating health equity requires deliberate effort, particularly when addressing the impact of structural and historical bias. The process involves the entire pipeline lifecycle, from data collection to deployment of results, and requires careful consideration of models and methods [89, 102]. Measuring the downstream impact of AI predictions and implementing policies that promote equitable outcomes are key aspects of this effort [103]. Ethical considerations should be integrated into the primary design process rather than being an afterthought [62-64]. Data governance is essential to detect and address systemic, computational, and human biases [65-67].

As AI systems continue to advance, the involvement of neonatologists, hospital leadership, and medical ethicists is vital to bridge the gap between technological complexities and user understanding. Efforts must be made to integrate clinical teams with data scientists, fostering communication, understanding, and partnerships that align with a shared vision to best benefit infants and their families.

Trust is paramount for successful AI implementation and requires input from clinical users, data scientist, and patients early in development. Transparency in model development should include data definitions, inclusion and exclusion decisions, and cultural and patient preferences. Making implicit decisions explicit is essential, and stakeholders, including patients, should be engaged in the co-design process from the beginning. For example, social scientists trained in the social determinants of health can provide a critical perspective [68, 69, 104].

Table 2 demonstrates principles for ethically developing and implementing AI in the NICU. AI's potential in healthcare is intrinsically linked to ongoing efforts by global and federal regulators. As AI continues to reshape patient care, these bodies must define healthcare teams' responsibilities, ensuring AI's ethical and efficient integration in neonatology.

### **Regulatory aspects of AI**

Regulatory bodies such as the US Food and Drug Administration (FDA) and the European Union's AI Act have established initial frameworks for addressing aspects of medical AI applications, such as lifecycle regulation, algorithmic bias, and user transparency [105, 106]. However, AI's dynamic and multifaceted nature in medicine has stimulated ongoing debates regarding its regulation, specifically whether AI should be regulated as a device or a system to support patient care. The FDA's Artificial Intelligence and Machine Learning Software as a Medical Device Action Plan [107] outlines government involvement to ensure the accuracy and reliability of AI tools. With its rapid expansion and potential, the global digital health market requires careful navigation, underpinned by strong ethical guidelines adopted by regulatory bodies to protect stakeholders.

### **Future directions and opportunities**

Rapid advancements in technology, data storage, and connectivity present opportunities to incorporate AI tools to create 'smart' NICUs. Table 3 summarizes emerging and existing opportunities for AI applications in neonatology. While we reviewed many promising AI and ML models, few AI tools have been approved and implemented for use in the NICU. As of Fall 2023, the FDA had approved around 692 AI-enabled devices or "software as a medical device," primarily in radiology, with a mere eight receiving reimbursement codes from the Centers for Medicare and Medicaid Services. No such codes exist for neonatal or pediatric care, which highlights a major financial barrier to adoption.

In our group opinion, the future of AI research in neonatology lies in multicenter collaboration, data sharing, transparency, and ethical consideration toward developing methods and systems that result in rigorously tested AI tools. A shift in focus from model development to technology implementation will require multi-disciplinary input for clinical integration, regulatory aspects, and ongoing evaluation. Even with successful development, testing, and implementation of accurate, unbiased models, the future of AI-driven NICU care is uncertain. To give an example, the HRC index was developed to predict imminent late-onset sepsis risk in very low birth weight infants. The model was externally validated, tested in a 9-NICU randomized controlled trial, shown to improve mortality, commercialized as the HeRO system, and FDA 501k cleared for implementation. Despite this success,

most NICUs have not implemented the HeRO system and those with the technology likely experience variable user engagement. For AI technology to impact patient outcomes in the future, we have considerable work to do in all stages of development, but a major roadblock could be getting clinicians to embrace its potential.

To leverage the power of AI to enhance the quality and precision of clinical care for neonates, we founded a group NeomIND-AI (Neonatal Machine learning, INnovations, Development, and Artificial Intelligence). Our goal is to create a future where neonatal care is more personalized, efficient, and effective, and where every child has the best possible start in life (<https://neomindai.com>).

## CONCLUSION

AI holds tremendous potential to transform neonatology through enhanced diagnostics, individualized treatments, and proactive prevention of complications. However, ethical challenges and biases must be addressed diligently before fully integrating these emerging technologies into clinical practice. By upholding principles of transparency, accountability, and human-centered values, and overcoming barriers to implementation and adoption, we can harness AI to create a brighter future for neonatal medicine.

## REFERENCES

1. Rowe M. An introduction to machine learning for clinicians. *Acad Med.* 2019;94:1433–6. [PubMed: 31094727]
2. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018;319:1317–8. [PubMed: 29532063]
3. Hoodbhoy Z, Masroor Jeelani S, Aziz A, Habib MI, Iqbal B, Akmal W, et al. Machine learning for child and adolescent health: A systematic review. *Pediatrics.* (2021) Jan;147.
4. 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) Apr 18;101346.
5. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med.* 2023;388:1201–8. [PubMed: 36988595]
6. Sun Q, Zou X, Yan Y, Zhang H, Wang S, Gao Y, et al. Machine learning-based prediction model of preterm birth using electronic health record. *J Health Eng.* 2022;2022:9635526.
7. Abraham A, Le B, Kosti I, Straub P, Velez-Edwards DR, Davis LK, et al. Dense phenotyping from electronic health records enables machine learning-based prediction of preterm birth. *BMC Med.* 2022;20:333. [PubMed: 36167547]
8. Weber A, Darmstadt GL, Gruber S, Foeller ME, Carmichael SL, Stevenson DK, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. *Ann Epidemiol.* 2018;28:783–9.e1. [PubMed: 30236415]
9. Gao C, Osmundson S, Velez Edwards DR, Jackson GP, Malin BA, Chen Y. Deep learning predicts extreme preterm birth from electronic health records. *J Biomed Inf.* 2019;100:103334.
10. Jehan F, Sazawal S, Baqui AH, Nisar MI, Dhingra U, Khanam R, et al. Multiomics characterization of preterm birth in low- and middle-income countries. *JAMA Netw Open.* 2020;3:e2029655. [PubMed: 33337494]
11. Wynn JL, Polin RA. Progress in the management of neonatal sepsis: the importance of a consensus definition. *Pediatr Res.* 2018;83:13–5. [PubMed: 29019470]
12. Moorman JR, Carlo WA, Kattwinkel J, Schelonka RL, Porcelli PJ, Navarrete CT, et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. *J Pediatr.* 2011;159:900–6.e1. [PubMed: 21864846]

13. Fairchild KD, Schelonka RL, Kaufman DA, Carlo WA, Kattwinkel J, Porcelli PJ, et al. Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial. *Pediatr Res*. 2013;74:570–5. [PubMed: 23942558]
14. Peng Z, Varisco G, Liang R-H, Kommers D, Cottaar W, Andriessen P, et al. DeepLOS: deep learning for late-onset sepsis prediction in preterm infants using heart rate variability. *Smart Health*. 2022;26:100335.
15. 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*. 2023;93:1913–21. [PubMed: 36593281]
16. Cabrera-Quiros L, Kommers D, Wolvers MK, Oosterwijk L, Arents N, van der Sluijs-Bens J, et al. Prediction of late-onset sepsis in preterm infants using monitoring signals and machine learning. *Crit Care Explor*. 2021;3:e0302. [PubMed: 33532727]
17. Puopolo KM, Draper D, Wi S, Newman TB, Zupancic J, Lieberman E, et al. Estimating the probability of neonatal early-onset infection on the basis of maternal risk factors. *Pediatrics*. 2011;128:e1155–63. [PubMed: 22025590]
18. Escobar GJ, Puopolo KM, Wi S, Turk BJ, Kuzniewicz MW, Walsh EM, et al. Stratification of risk of early-onset sepsis in newborns 34 weeks' gestation. *Pediatrics*. 2014;133:30–6. [PubMed: 24366992]
19. Kuzniewicz MW, Puopolo KM, Fischer A, Walsh EM, Li S, Newman TB, et al. A quantitative, risk-based approach to the management of neonatal early-onset sepsis. *JAMA Pediatr*. 2017;171:365–71. [PubMed: 28241253]
20. Battersby C, Santhalingam T, Costeloe K, Modi N. Incidence of neonatal necrotising enterocolitis in high-income countries: a systematic review. *Arch Dis Child Fetal Neonatal Ed*. 2018;103:F182–9. [PubMed: 29317459]
21. Walsh MC, Kliegman RM. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr Clin North Am*. 1986;33:179–201. [PubMed: 3081865]
22. Bell MJ, Ternberg JL, Feigin RD, Keating JP, Marshall R, Barton L, et al. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. *Ann Surg*. 1978;187:1–7. [PubMed: 413500]
23. Battersby C, Longford N, Costeloe K, Modi N, UK Neonatal Collaborative Necrotising Enterocolitis Study Group. Development of a gestational age-specific case definition for neonatal necrotizing enterocolitis. *JAMA Pediatr*. 2017;171:256–63. [PubMed: 28046187]
24. 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;56:1703–10. [PubMed: 33342603]
25. Son J, Kim D, Na JY, Jung D, Ahn J-H, Kim TH, et al. Development of artificial neural networks for early prediction of intestinal perforation in preterm infants. *Sci Rep*. 2022;12:12112. [PubMed: 35840701]
26. Meister AL, Gardner FC, Browning KN, Travagli RA, Palmer C, Doheny KK. Vagal tone and proinflammatory cytokines predict feeding intolerance and necrotizing enterocolitis risk. *Adv Neonatal Care*. 2021;21:452–61. [PubMed: 34847103]
27. Doheny KK, Palmer C, Browning KN, Jairath P, Liao D, He F, et al. Diminished vagal tone is a predictive biomarker of necrotizing enterocolitis-risk in preterm infants. *Neurogastroenterol Motil*. 2014;26:832–40. [PubMed: 24720579]
28. Stone ML, Tatum PM, Weitkamp JH, Mukherjee AB, Attridge J, McGahren ED, et al. Abnormal heart rate characteristics before clinical diagnosis of necrotizing enterocolitis. *J Perinatol*. 2013;33:847–50. [PubMed: 23722974]
29. Lin YC, Salieb-Aouissi A, Hooven TA. Interpretable prediction of necrotizing enterocolitis from machine learning analysis of premature infant stool microbiota. *BMC Bioinforma*. 2022;23:104.
30. Rusconi B, Jiang X, Sidhu R, Ory DS, Warner BB, Tarr PI. Gut sphingolipid composition as a prelude to necrotizing enterocolitis. *Sci Rep*. 2018;8:10984. [PubMed: 30030452]
31. Sylvester KG, Ling XB, Liu GY, Kastenber ZJ, Ji J, Hu Z, et al. A novel urine peptide biomarker-based algorithm for the prognosis of necrotising enterocolitis in human infants. *Gut*. 2014;63:1284–92. [PubMed: 24048736]

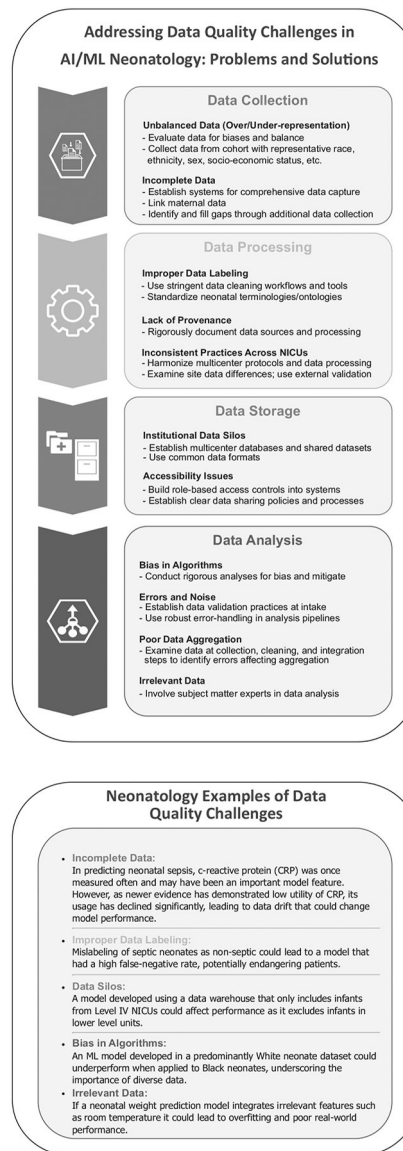
32. Stoll BJ, Hansen NI, Bell EF, Walsh MC, Carlo WA, Shankaran S, et al. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *JAMA*. 2015;314:1039–51. [PubMed: 26348753]
33. Williams E, Greenough A. Advances in treating bronchopulmonary dysplasia. *Expert Rev Respir Med*. 2019;13:727–35. [PubMed: 31244360]
34. Principi N, Di Pietro GM, Esposito S. Bronchopulmonary dysplasia: clinical aspects and preventive and therapeutic strategies. *J Transl Med*. 2018;16:36. [PubMed: 29463286]
35. He W, Zhang L, Feng R, Fang W-H, Cao Y, Sun S-Q, et al. Risk factors and machine learning prediction models for bronchopulmonary dysplasia severity in the Chinese population. *World J Pediatr*. 2023;19:568–76. [PubMed: 36357648]
36. Valenzuela-Stutman D, Marshall G, Tapia JL, Mariani G, Bancalari A, Gonzalez Á, et al. Bronchopulmonary dysplasia: risk prediction models for very-low- birthweight infants. *J Perinatol*. 2019;39:1275–81. [PubMed: 31337853]
37. Laughon MM, Langer JC, Bose CL, Smith PB, Ambalavanan N, Kennedy KA, et al. Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants. *Am J Respir Crit Care Med*. 2011;183:1715–22. [PubMed: 21471086]
38. Dai D, Chen H, Dong X, Chen J, Mei M, Lu Y, et al. Bronchopulmonary dysplasia predicted by developing a machine learning model of genetic and clinical information. *Front Genet*. 2021;12:689071. [PubMed: 34276789]
39. Leigh RM, Pham A, Rao SS, Vora FM, Hou G, Kent C, et al. Machine learning for prediction of bronchopulmonary dysplasia-free survival among very preterm infants. *BMC Pediatr*. 2022;22:542. [PubMed: 36100848]
40. Jia M, Li J, Zhang J, Wei N, Yin Y, Chen H, et al. Identification and validation of cuproptosis related genes and signature markers in bronchopulmonary dysplasia disease using bioinformatics analysis and machine learning. *BMC Med Inf Decis Mak*. 2023;23:69.
41. Fairchild KD, Nagraj VP, Sullivan BA, Moorman JR, Lake DE. Oxygen desaturations in the early neonatal period predict development of bronchopulmonary dysplasia. *Pediatr Res*. 2019;85:987–93. [PubMed: 30374050]
42. Moreira A, Tovar M, Smith AM, Lee GC, Meunier JA, Cheema Z, et al. Development of a peripheral blood transcriptomic gene signature to predict bronchopulmonary dysplasia. *Am J Physiol Lung Cell Mol Physiol*. 2023;324:L76–87. [PubMed: 36472344]
43. Sullivan BA, McClure C, Hicks J, Lake DE, Moorman JR, Fairchild KD. Early heart rate characteristics predict death and morbidities in preterm infants. *J Pediatr*. 2016;174:57–62. [PubMed: 27113378]
44. Sullivan BA, Wallman-Stokes A, Isler J, Sahni R, Moorman JR, Fairchild KD, et al. Early pulse oximetry data improves prediction of death and adverse outcomes in a two-center cohort of very low birth weight infants. *Am J Perinatol*. 2018;35:1331–8. [PubMed: 29807371]
45. Hartnett ME. Pathophysiology of retinopathy of prematurity. *Annu Rev Vis Sci*. 2023;9:39–70. [PubMed: 37164029]
46. Di Fiore JM, Kaffashi F, Loparo K, Sattar A, Schluchter M, Foglyano R, et al. The relationship between patterns of intermittent hypoxia and retinopathy of prematurity in preterm infants. *Pediatr Res*. 2012;72:606–12. [PubMed: 23037873]
47. Di Fiore JM, Bloom JN, Orge F, Schutt A, Schluchter M, Cheruvu VK, et al. A higher incidence of intermittent hypoxemic episodes is associated with severe retinopathy of prematurity. *J Pediatr*. 2010;157:69–73. [PubMed: 20304417]
48. Athikarisamy S, Desai S, Patole S, Rao S, Simmer K, Lam GC. The use of postnatal weight gain algorithms to predict severe or type I retinopathy of prematurity: a systematic review and meta-analysis. *JAMA Netw Open*. 2021;4:e2135879. [PubMed: 34812847]
49. Wang J, Ji J, Zhang M, Lin J-W, Zhang G, Gong W, et al. Automated explainable multidimensional deep learning platform of retinal images for retinopathy of prematurity screening. *JAMA Netw Open*. 2021;4:e218758. [PubMed: 33950206]
50. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136:803–10. [PubMed: 29801159]

51. Zekavat SM, Raghu VK, Trinder M, Ye Y, Koyama S, Honigberg MC, et al. Deep learning of the retina enables phenome- and genome-wide analyses of the microvasculature. *Circulation*. 2022;145:134–50. [PubMed: 34743558]
52. Ho T, Dukhovny D, Zupancic JAF, Goldmann DA, Horbar JD, Pursley DM. Choosing wisely in newborn medicine: five opportunities to increase value. *Pediatrics*. 2015;136:e482–9. [PubMed: 26195536]
53. Mohammad K, Scott JN, Leijser LM, Zein H, Afifi J, Piedboeuf B, et al. Consensus approach for standardizing the screening and classification of preterm brain injury diagnosed with cranial ultrasound: a canadian perspective. *Front Pediatr*. 2021;9:618236. [PubMed: 33763394]
54. Kidokoro H, Neil JJ, Inder TE, New MR. imaging assessment tool to define brain abnormalities in very preterm infants at term. *AJNR Am J Neuroradiol*. 2013;34:2208–14. [PubMed: 23620070]
55. Barkovich AJ, Hajnal BL, Vigneron D, Sola A, Partridge JC, Allen F, et al. Prediction of neuromotor outcome in perinatal asphyxia: evaluation of MR scoring systems. *AJNR Am J Neuroradiol*. 1998;19:143–9. [PubMed: 9432172]
56. Trivedi SB, Vesoulis ZA, Rao R, Liao SM, Shimony JS, McKinsty RC, et al. A validated clinical MRI injury scoring system in neonatal hypoxic-ischemic encephalopathy. *Pediatr Radio*. 2017;47:1491–9.
57. Weeke LC, Groenendaal F, Mudigonda K, Blennow M, Lequin MH, Meiners LC, et al. A novel magnetic resonance imaging score predicts neurodevelopmental outcome after perinatal asphyxia and therapeutic hypothermia. *J Pediatr*. 2018;192:33–40.e2. [PubMed: 29246356]
58. Gruber N, Galijasevic M, Regodic M, Grams AE, Siedentopf C, Steiger R, et al. A deep learning pipeline for the automated segmentation of posterior limb of internal capsule in preterm neonates. *Artif Intell Med*. 2022;132:102384. [PubMed: 36207089]
59. Richter L, Fetit AE. Accurate segmentation of neonatal brain MRI with deep learning. *Front Neuroinformatics* 2022;16:1006532.
60. Shen DD, Bao SL, Wang Y, Chen YC, Zhang YC, Li XC, et al. An automatic and accurate deep learning-based neuroimaging pipeline for the neonatal brain. *Pediatr Radio*. 2023;53:1685–97.
61. Weiss RJ, Bates SV, Song Y, Zhang Y, Herzberg EM, Chen Y-C, et al. Mining multisite clinical data to develop machine learning MRI biomarkers: application to neonatal hypoxic ischemic encephalopathy. *J Transl Med*. 2019;17:385. [PubMed: 31752923]
62. Mathieson SR, Stevenson NJ, Low E, Marnane WP, Rennie JM, Temko A, et al. Validation of an automated seizure detection algorithm for term neonates. *Clin Neurophysiol*. 2016;127:156–68. [PubMed: 26055336]
63. Pavel AM, Rennie JM, de Vries LS, Blennow M, Foran A, Shah DK, et al. A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. *Lancet Child Adolesc Health*. 2020;4:740–9. [PubMed: 32861271]
64. Raurale SA, Boylan GB, Mathieson SR, Marnane WP, Lightbody G, O’Toole JM. Grading hypoxic-ischemic encephalopathy in neonatal EEG with convolutional neural networks and quadratic time-frequency distributions. *J Neural Eng*. 2021;18:046007.
65. Pavel AM, O’Toole JM, Proietti J, Livingstone V, Mitra S, Marnane WP, et al. Machine learning for the early prediction of infants with electrographic seizures in neonatal hypoxic-ischemic encephalopathy. *Epilepsia* 2023;64:456–68. [PubMed: 36398397]
66. Pliego J. Surgical correction of mitral valve stenosis under direct vision using extracorporeal circulation. *Gac Med Mex*. 1964;94:423–33. [PubMed: 4891440]
67. O’Shea A, Lightbody G, Boylan G, Temko A. Neonatal seizure detection from raw multichannel EEG using a fully convolutional architecture. *Neural Netw*. 2020;123:12–25. [PubMed: 31821947]
68. Raeisi K, Khazaei M, Croce P, Tamburro G, Comani S, Zappasodi F. A graph convolutional neural network for the automated detection of seizures in the neonatal EEG. *Comput Methods Prog Biomed*. 2022;222:106950.
69. Srinivasakumar P, Zempel J, Trivedi S, Wallendorf M, Rao R, Smith B, et al. Treating EEG seizures in hypoxic ischemic encephalopathy: a randomized controlled trial. *Pediatrics* 2015;136:e1302–9. [PubMed: 26482675]

70. Glass HC, Glidden D, Jeremy RJ, Barkovich AJ, Ferriero DM, Miller SP. Clinical neonatal seizures are independently associated with outcome in infants at risk for hypoxic-ischemic brain injury. *J Pediatr*. 2009;155:318–23. [PubMed: 19540512]
71. Hartnett ME, Lane RH. Effects of oxygen on the development and severity of retinopathy of prematurity. *J AAPOS*. 2013;17:229–34. [PubMed: 23791404]
72. Askie LM, Darlow BA, Finer N, Schmidt B, Stenson B, Tarnow-Mordi W, et al. Association between oxygen saturation targeting and death or disability in extremely preterm infants in the neonatal oxygenation prospective meta-analysis collaboration. *JAMA*. 2018;319:2190–201. [PubMed: 29872859]
73. Ruppel H, Makeneni S, Faerber JA, Lane-Fall MB, Foglia EE, O’Byrne ML, et al. Evaluating the accuracy of pulse oximetry in children according to race. *JAMA Pediatr*. 2023;177:540–3. [PubMed: 36939727]
74. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. *J Perinatol*. 2022;42:79–85. [PubMed: 34642469]
75. Alvarez D, Hornero R, Abásolo D, del Campo F, Zamarrón C. Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection. *Physiol Meas*. 2006;27:399–412. [PubMed: 16537981]
76. Lu X, Jiang L, Chen T, Wang Y, Zhang B, Hong Y, et al. Continuously available ratio of SpO<sub>2</sub>/FiO<sub>2</sub> serves as a noninvasive prognostic marker for intensive care patients with COVID-19. *Respir Res*. 2020;21:194. [PubMed: 32698822]
77. Sadeghi Fathabadi O, Gale TJ, Lim K, Salmon BP, Dawson JA, Wheeler KI, et al. Characterisation of the oxygenation response to inspired oxygen adjustments in preterm infants. *Neonatology* 2016;109:37–43. [PubMed: 26554825]
78. Ostojic D, Guglielmini S, Moser V, Fauchère JC, Bucher HU, Bassler D, et al. Reducing false alarm rates in neonatal intensive care: a new machine learning approach. *Adv Exp Med Biol*. 2020;1232:285–90. [PubMed: 31893422]
79. Kristiansen TB, Kristensen K, Uffelmann J, Brandslund I. Erroneous data: the Achilles’ heel of AI and personalized medicine. *Front Digit Health*. 2022;4:862095. [PubMed: 35937419]
80. Elmore JG, Lee CI. Data quality, data sharing, and moving artificial intelligence forward. *JAMA Netw Open*. 2021;4:e2119345. [PubMed: 34398208]
81. Miller DD. The medical AI insurgency: what physicians must know about data to practice with intelligent machines. *npj Digital Med*. 2019;2:62.
82. Scott IA, Carter SM, Coiera E. Exploring stakeholder attitudes towards AI in clinical practice. *BMJ Health Care Inform*. 2021;28:e100450.
83. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;3:e745–50. [PubMed: 34711379]
84. Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *npj Digital Med*. 2020;3:53.
85. TerKonda SP, Fish EM. Artificial intelligence viewed through the lens of state regulation. *Intell Based Med*. 2023;7:100088.
86. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Progn Res*. 2018;2:11. [PubMed: 31093561]
87. Pencina MJ, Goldstein BA, D’Agostino RB. Prediction models - development, evaluation, and clinical application. *N Engl J Med*. 2020;382:1583–6. [PubMed: 32320568]
88. Parikh RB, Helmchen LA. Paying for artificial intelligence in medicine. *npj Digital Med*. 2022;5:63.
89. Chen MM, Golding LP, Nicola GN. Who will pay for AI? *Radiol. Artif Intell* 2021;3:e210030. [PubMed: 34142090]
90. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. *N. Engl J Med* 2020;383:2477–8. [PubMed: 33326721]
91. Williams DR, Rucker TD. Understanding and addressing racial disparities in health care. *Health Care Financ Rev*. 2000;21:75–90. [PubMed: 11481746]



92. Martin AE, D'Agostino JA, Passarella M, Lorch SA. Racial differences in parental satisfaction with neonatal intensive care unit nursing care. *J Perinatol*. 2016;36:1001–7. [PubMed: 27583386]
93. Sullivan BA, Doshi A, Chernyavskiy P, Husain A, Binai A, Sahni R, et al. Neighborhood deprivation and association with neonatal intensive care unit mortality and morbidity for extremely premature infants. *JAMA Netw Open*. 2023;6:e2311761. [PubMed: 37166800]
94. Travers CP, Carlo WA, McDonald SA, Das A, Ambalavanan N, Bell EF, et al. Racial/ethnic disparities among extremely preterm infants in the united states from 2002 to 2016. *JAMA Netw Open*. 2020;3:e206757. [PubMed: 32520359]
95. Esteban-Escañó J, Castán B, Castán S, Chóliz-Ezquerro M, Asensio C, Laliena AR, et al. Machine learning algorithm to predict acidemia using electronic fetal monitoring recording parameters. *Entropy*. 2021;24:68. [PubMed: 35052094]
96. Goldstein B, Fiser DH, Kelly MM, Mickelsen D, Ruttimann U, Pollack MM. Decomplexification in critical illness and injury: relationship between heart rate variability, severity of illness, and outcome. *CritCare Med*. 1998;26:352–7.
97. Ellenby MS, McNames J, Lai S, McDonald BA, Krieger D, Sciabassi RJ, et al. Uncoupling and recoupling of autonomic regulation of the heart beat in pediatric septic shock. *Shock* 2001;16:274–7. [PubMed: 11580109]
98. Badke CM, Marsillio LE, Weese-Mayer DE, Sanchez-Pinto LN. Autonomic nervous system dysfunction in pediatric sepsis. *Front Pediatr*. 2018;6:280. [PubMed: 30356758]
99. Papaioannou VE, Maglaveras N, Houvarda I, Antoniadou E, Vretzakis G. Investigation of altered heart rate variability, nonlinear properties of heart rate signals, and organ dysfunction longitudinally over time in intensive care unit patients. *J Crit Care*. 2006;21:95–103. [PubMed: 16616632]
100. Griffin MP, Lake DE, Bissonette EA, Harrell FE, O'Shea TM, Moorman JR. Heart rate characteristics: novel physiometers to predict neonatal infection and death. *Pediatrics* 2005;116:1070–4. [PubMed: 16263991]
101. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inf Assoc*. 2012;19:121–7.
102. Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, et al. Recommendations for the standardization and interpretation of the electrocardiogram: part I: The electrocardiogram and its technology: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. *Circulation* 2007;115:1306–24. [PubMed: 17322457]
103. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2021;18:465–78. [PubMed: 33526938]
104. Wysocki O, Davies JK, Vigo M, Armstrong AC, Landers D, Lee R, et al. Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI-driven clinical decision-making. *Artif Intell*. 2023;316:103839.
105. Global digital health market forecast 2025 | Statista [Internet]. 2023. Available from: <https://www.statista.com/statistics/1092869/global-digital-health-market-size-forecast/>
106. Torous J, Stern AD, Bourgeois FT. Regulatory considerations to keep pace with innovation in digital health products. *npj Digital Med*. 2022;5:121.
107. Artificial Intelligence and Machine Learning in Software as a Medical Device | FDA [Internet]. 2023. Available from: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>



**Fig. 1. Overview and neonatology specific examples of a systematic data quality framework.** This flowchart depicts the key phases in an end-to-end data quality process. It begins with initial data acquisition in the Data Collection phase from sources such as electronic medical records, medical devices, research databases, and literature reviews. The next Data Processing stage involves activities such as data inspection, anomaly detection, cleaning, transformation, and integration to curate the dataset. The processed quality data is then stored and managed in the Data Storage phase. Subsequently in the Data Analysis phase, statistical analyses and visualizations are performed to derive insights and identify data quality issues to refine the overall collection and processing. Effective data governance and metadata management are critical throughout each phase to ensure accuracy, transparency, and reproducibility. The systematic workflow promotes high quality data essential for robust analytics and decision making in healthcare applications. We hypothesize that the importance of addressing these data quality challenges is amplified when using ML methods

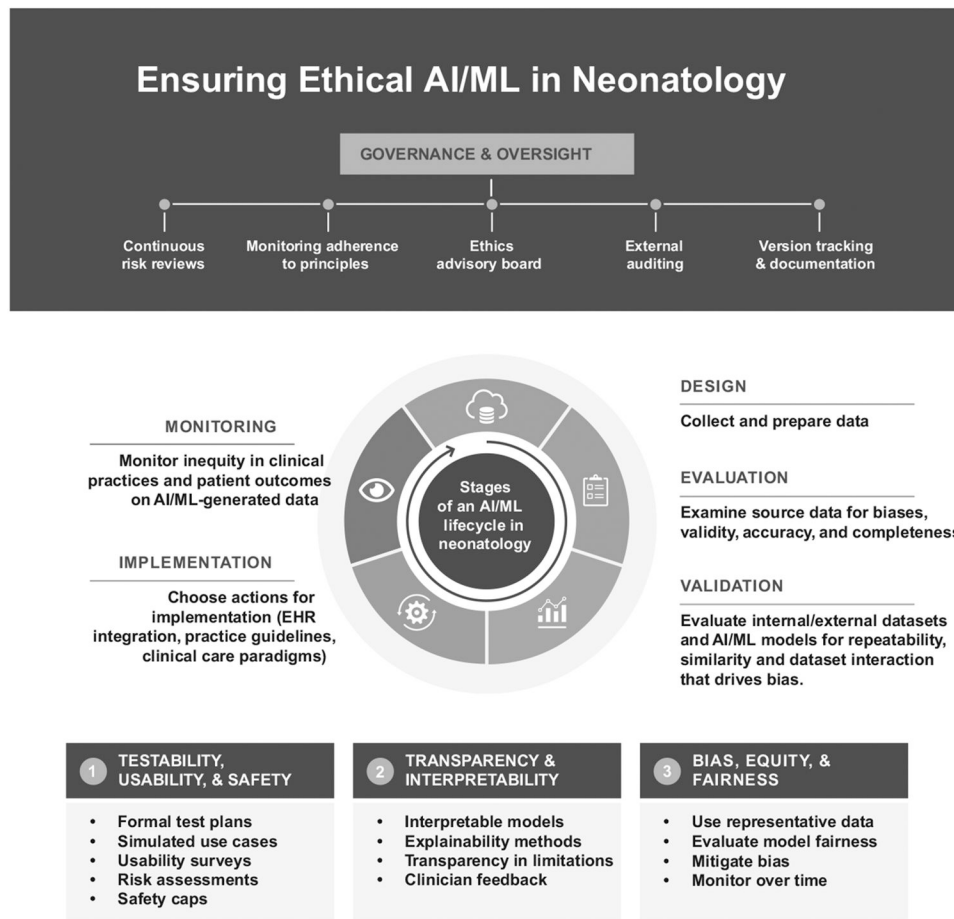
over traditional statistical methods, but this hypothesis has not been tested. Thus, we highlight these challenges in the context of ML for neonatal care, however, most can be applied to any analysis in any population.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 2. Stages of the AI/ML development lifecycle in neonatology, highlighting important ethical considerations.**

The development of AI/ML models in healthcare follows a cyclical workflow that includes design, implementation, testing, deployment, monitoring, and retraining. Oversight by diverse stakeholders, nationally and locally, is critical for governance. This includes continuous risk reviews, monitoring adherence to principles, eliciting input through advisory boards, and enabling external auditing and clinician feedback. At each stage, unique ethical challenges arise that must be proactively addressed. In the design phase, considerations such as testability, usability, safety, bias, fairness, transparency, and interpretability should be prioritized from the start. Representative and unbiased data collection is crucial during implementation, along with privacy and security protections. Throughout testing, model performance and safety should be rigorously evaluated across diverse groups. Monitoring performance post-deployment enables continuous improvement through retraining. Overall, ethics should not be an afterthought but instead integrated into every step. The figure emphasizes that thoughtful design and testing parameters, mitigating bias and lack of equity, and ensuring comprehensibility for clinicians can promote better, more ethical AI in neonatology. However, this requires extensive collaboration between computer scientists, clinicians, and ethics experts across the entire development lifecycle.



Neonatal conditions with opportunities for improved management with AI and selected challenges.


Table 1.

Condition	Importance	Challenges with Current Diagnosis	Opportunities for AI Improvements	Challenges in Implementing AI
Bronchopulmonary dysplasia (BPD)	Leading cause of morbidity in premature infants; impacts lung development	Diagnosis delayed, based on clinical and radiographic criteria at 36 weeks PMA	Earlier prediction or risk stratification may enable targeted early intervention	Heterogeneous phenotypes require large, diverse cohorts; integration into complex clinical workflows
Intraventricular hemorrhage (IVH)	Bleeding in the brain may cause neurologic impairment	Cranial ultrasounds have low sensitivity in detecting mild IVH; subject to interpreter variability	AI-assisted ultrasound analysis could improve IVH detection and prognostic capabilities	Need for explainable AI to build provider trust; costs of technology adoption
Late-onset Sepsis (LOS)	Systemic inflammatory condition; significant cause of morbidity in premature infants	Diagnosis relies on nonspecific clinical signs and blood culture results	AI models can identify patterns in clinical and physiologic data that predict or detect LOS	High false positive rate; diagnostic criteria for LOS varies in training data
Necrotizing enterocolitis (NEC)	Intestinal and systemic inflammatory condition; significant cause of morbidity in premature infants	Diagnosis relies on nonspecific clinical signs and radiographic finding with low specificity	AI models can identify patterns in clinical and physiologic data that predict or detect NEC	Difficult to obtain reliable labeled training data, avoiding over-diagnosis
Retinopathy of prematurity (ROP)	Abnormal retinal vascular proliferation; Leading cause of childhood blindness	Specialist exam required; subject to interobserver variability	AI analysis of retinal images can improve efficiency & accuracy of ROP eye exams	Need for standardized high-quality image inputs; medicolegal implications of AI diagnosis
Neonatal seizures	Seizures indicate neurological impairment; require rapid treatment but subtle manifestations	Specialist training required for EEG interpretation; lack of 24/7 EEG review	AI-assisted seizure detection on continuous EEG may enable faster diagnosis and treatment	Requires continuous EEG monitoring, avoiding over-treatment

**Table 2.**

Framework for ethically developing and deploying AI in the NICU.

Stage	Guiding Values	Key Actions and Considerations
 Design	Beneficence	<ul style="list-style-type: none"> <li>• Collaborate with neonatal networks</li> <li>• Prioritize patient benefit</li> <li>• Use neonatal-specific datasets</li> </ul>
	Fidelity	<ul style="list-style-type: none"> <li>• Establish protocols for regular model updates</li> <li>• Track performance over time</li> <li>• Set expiration dates for model re-evaluation</li> </ul>
	Fairness	<ul style="list-style-type: none"> <li>• Review data collection methods for potential biases</li> <li>• Assess fairness and validity across population</li> <li>• Simulate model on diverse datasets</li> <li>• Solicit feedback from diverse stakeholders</li> </ul>
 Evaluation	Safety Accountability Transparency	<ul style="list-style-type: none"> <li>• Emphasize prototyping and simulation testing</li> </ul>
		<ul style="list-style-type: none"> <li>• Proceed with caution before NICU integration</li> </ul>
		<ul style="list-style-type: none"> <li>• Solicit feedback from diverse stakeholders</li> </ul>
		<ul style="list-style-type: none"> <li>• Use multi-center and external cohorts to ensure generalizability</li> </ul>
 Validation	Autonomy Justice Education	<ul style="list-style-type: none"> <li>• Involve all stakeholders in decision-making</li> </ul>
		<ul style="list-style-type: none"> <li>• Ensure equal access across diverse settings</li> </ul>
		<ul style="list-style-type: none"> <li>• Comprehensive training programs for all users</li> </ul>
		<ul style="list-style-type: none"> <li>• Plan for re-calibration and updated implementation</li> </ul>
 Implementation		

Stage	Guiding Values	Key Actions and Considerations	
	Justice	<ul style="list-style-type: none"><li>• Develop methods to detect biases</li><li>• Ensure fairness across neonatal populations</li></ul>	
	Accountability	<ul style="list-style-type: none"><li>• Define roles and responsibilities for AI management</li><li>• Regular performance assessments</li></ul>	
	Vigilance	<ul style="list-style-type: none"><li>• Implement consistent monitoring post-deployment</li><li>• Continuous algorithm vigilance in NICU</li></ul>	

**Table 3.**

Selected examples of emerging or existing AI/ML applications in neonatology.

Technology/Application	Explanation	Device/Data Source	Pros	Cons
Automated seizure detection	AI algorithms analyze EEG data to detect seizures	Continuous EEG waveforms	High accuracy in some reports, faster treatment	Performance variable; requires large datasets
ROP risk	Prediction models use clinical data to risk stratify for ROP screening	EHR demographics, growth, clinical variables	Higher sensitivity than GA/BW screening criteria	Limited accuracy gains
ROP severity	AI image analysis of retinal images to classify severity	RetCam images	Less inter-rater variability, efficient diagnosis	Requires specialized equipment
Surfactant dosing optimization	Suggests personalized doses based on lung mechanics	Airway flow and pressure data from sensors	Precision dosing; reduce lung injury	Requires specialized equipment; unproven benefit
Automated echocardiogram analysis	AI extracts key cardiac echo features	Stored or real-time echocardiogram images	Efficient diagnosis; Standardizes measures	Large datasets required; Qualitative review still needed
Video/audio analytics	Analyzes video and audio to detect various clinical targets	Video camera and microphone with AI	Unobtrusive monitoring; Identify subtle changes	Privacy concerns; Large datasets needed
Extubation failure	ML models that analyze clinical data and continuous physiologic data	EHR, bedside monitor, and ventilator data	Reduce extubation failure; Assistance in use of corticosteroids	Need for continuous physiology data storage
Neonatal sepsis monitoring	Displays ML model output as risk of imminent sepsis diagnosis	Continuous physiologic monitoring data	Earlier diagnosis and treatment, reduced mortality	High false positive rate, fear of over-treatment