POLITECNICO DI TORINO
Repository ISTITUZIONALE

A Pathway for Integrating Artificial Intelligence for Neonatal Neuromonitoring: From Code to Crib Side

Original
A Pathway for Integrating Artificial Intelligence for Neonatal Neuromonitoring: From Code to Crib Side / Kulakowski, Pawel; Alarcón, Ana; Skoric, Tamara; Seoni, Silvia; O'Sullivan, Mark; O'Toole, John M.. - In: THE JOURNAL OF PEDIATRICS. - ISSN 0022-3476. - 285:(2025). [10.1016/j.jpeds.2025.114708]

Availability:
This version is available at: 11583/3002240 since: 2025-07-30T12:36:00Z

Publisher:
Elsevier

Published
DOI:10.1016/j.jpeds.2025.114708

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

06 May 2026

A Pathway for Integrating Artificial Intelligence for Neonatal Neuromonitoring: From Code to Crib Side

Pawel Kulakowski, PhD[1], Ana Alarcón, MD, PhD[2,3], Tamara Skoric, PhD[4], Silvia Seoni, PhD[5], Mark O'Sullivan, PhD[6], and John M. O'Toole, PhD[7,8]

Effective brain monitoring is fundamental to providing optimal neurocritical care for neonates at risk of neurological injury. Different brain monitoring and neuroimaging modalities can provide complementary insights into neurological function and structure, which can aid in assessing injury, tracking development, and informing treatment decisions. Monitoring technologies include the electroencephalogram and near-infrared spectroscopy for brain activity and cerebral oxygenation, respectively, and imaging techniques such as magnetic resonance imaging (MRI) and cranial ultrasound for structural assessment.

These technologies produce complex data that can be difficult to interpret and often yield subjective conclusions, even among top experts. Indications of brain dysfunction or injury may be subtle and difficult to detect, and clear associations with the neurobiology are not always well established. Given these challenges, the field of neonatal neuromonitoring presents a compelling opportunity for the application of artificial intelligence (AI) to support analysis of these complex data. AI has the potential to support in the clinical decision process by uncovering new insights in the data, automating review tasks, and enabling scalable solutions across large, longitudinal, or multipatient neuromonitoring data sets.

In this Commentary, as a working group of experts from the European research project AI4NICU,[1] we focus on the case of AI applications for neonatal neuromonitoring. Monitoring brain function in neonates is a highly specialized area within pediatrics, distinguished by continuous and rapid clinical, scientific, and technological advances. Our working group consists of clinicians and biomedical engineers, with the aim of jointly mapped ethical or societal issues associated with the AI systems in the neonatal intensive care unit (NICU), defining the clinical importance of the performance of AI methods, and defining a unique assessment of AI performance in neonatal neuromonitoring. We believe AI tools have the potential to significantly improve clinical workflow in the NICU. Although the final diagnosis will always be made by a physician, AI clinical decision support tools could automate tedious and repetitive tasks, process data at a scale beyond human capacity and operate out-of-hours, and provide clinicians with relevant evidence to support informed decision-making.

In contrast to other reviews and commentaries on AI for neonatal care,[2-7] we present here a holistic approach to the development of AI tools for neonatal neuromonitoring, which is structured around 6 critical aspects that are potential bottlenecks for the tools' applicability, credibility, and efficacy (**Figure 1**). We begin with the issue of how patient data should be collected without raising concerns among patients, their parents, and health care professionals (section 2). We then comment on developing appropriate AI algorithms for the collected data and highlight some applications (section 3). We stress the importance of fluent cooperation between clinicians and engineers; with clinicians guiding application and purpose and engineers designing the algorithms, with continuous cooperation and communication throughout this iterative process. Later, we examine the potential for reluctance and distrust toward AI from both clinicians and patients' families, which may arise from either an AI algorithmic bias (section 4) or the inability to explain the AI predictions to evidence-driven medical staff (section 5).

Next, we describe regulatory frameworks and how they can impose a necessary compromise: although these regulations rightly prioritize patient safety and guard against the unchecked advancement of AI-enabled medical devices (AIMDs), they can also impose a significant financial and knowledge burden that may slow the pace of innovation and hinder development (section 6). Finally, we consider different aspects related to the deployment of the device: we highlight the importance of clinical validation studies, postmarket surveillance to ensure continued safety and

AI	Artificial intelligence
AIMD	AI-enabled medical device
FDA	Food and Drug Administration
HIE	hypoxic-ischemic encephalopathy
IMDRF	International Medical Device Regulators Forum
IT	Information technology
MDR	Medical Device Regulation
MRI	Magnetic resonance imaging
NICU	Neonatal intensive care unit
XAI	Explainable artificial intelligence

From the [1]Institute of Telecommunications, AGH University of Krakow, Krakow, Poland; [2]Department of Neonatology, Hospital Sant Joan de Déu and Hospital Clínic, BCNatal (Barcelona Centre for Maternal, Foetal and Neonatal Medicine), Institut de Recerca Sant Joan de Déu, Barcelona, Spain; [3]Faculty of Medicine and Health Sciences, Department of Surgery and Medical-Surgical Specialties, Universitat de Barcelona, Barcelona, Spain; [4]Faculty of Technical Science, University of Novi Sad, Novi Sad, Serbia; [5]Biolab, PoliToBIOMed Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy; [6]NeuroBell Ltd., Cork, Ireland; [7]CergenX Ltd, Dublin, Ireland; and [8]INFANT Research Centre, University College Cork, Cork, Ireland

0022-3476/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
https://doi.org/10.1016/j.jpeds.2025.114708

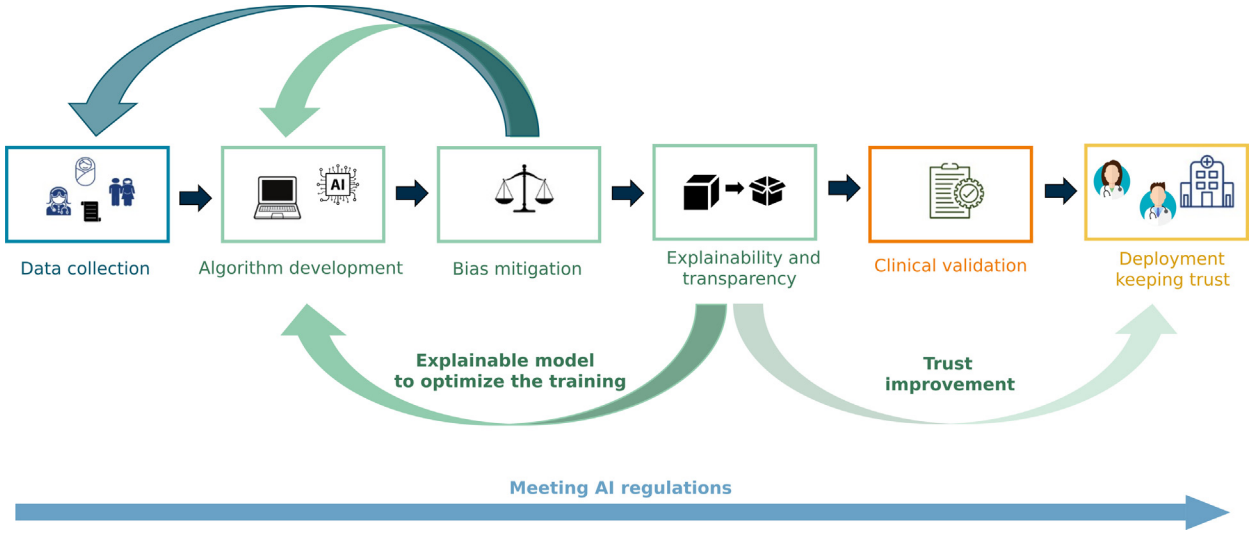

Figure 1. AI in neonatal neuromonitoring: from data collection to deployment.

identification of potential systematic misuse of the device, and the importance of including parent or /guardians in both the development and deployment stages of the device (section 7).

Data Collection and Curation

A key challenge in the development of AI in neonatal neuromonitoring is the need to use patient data for training and testing the models. Depending on the objectives (eg, enhancing diagnosis, outcome prediction, or personalized treatment) the types of data needed will vary and may include clinical information from electronic patient records, physiological data (such as EEG signals and cerebral oxygenation measurements), neuroimaging data, and more. The development, implementation, and reliance on AI systems in health care involve an increasing need to create and curate high-quality real-world patient data sets. Data collection for AI development in neonatal neuromonitoring requires a careful and ethical approach to ensure data accuracy and security while meeting the specific goals of improving neonatal neurological care through AI. In this process, the drive for innovation, care quality, and efficiency must be balanced with the patient's individual rights to privacy and confidentiality.[8]

Health care providers must implement organizational and technical measures at the earliest stages of design to ensure compliance with the principles of data protection "by design" and "by default".[9] AI projects in neonatal neuromonitoring often require oversight by the institutional review board or ethics committee during planning, as they usually involve research with human subject data and aim to develop medical devices or decision-support tools. In addition, the institutional data protection officer supervises the data journey to ensure compliance with privacy regulations and safeguard patient data throughout their lifecycle. For the use of patient data in developing AI technologies, informed consent can take various forms, including specific consent for a particular project, broad area-specific consent (eg, generic informed consent for the use of patient health data in neonatal research), or data anonymization with appropriate safeguards and ethics committee waivers when consent is not feasible. However, case studies exist in which data-driven technologies have used patient data without consent or research approval,[10] underscoring that the development of AI applications may require updates to current privacy laws and regulations. In any case, the AI community must comply with data protection regulations[9,11]; safeguarding privacy and security require deidentification of data and implementation of robust security measures, such as encryption and secure access controls. The challenge of anonymization lies in ensuring that it meets regulatory requirements while also retaining the necessary data of value to achieve the goals of AI technologies.[12] Anonymization tools like Nymiz (Nymiz Data Company), ARX (Helmholtz Center for Information Security – CISPA), and Amnesia (Athena Research Center) remove personally identifiable information using techniques such as differential privacy and pseudonymization. Equally important to privacy and security is ensuring data quality by avoiding socially constructed biases, inaccuracies, and errors.

Particularly sensitive is the repurposing of or granting third-party access to patient data for the testing and development of AI systems. In such cases, obtaining specific informed consent is required, ensuring that patient privacy is preserved, and maintaining data ownership within the hospital.

The stumbling blocks to multi-institutional data sharing primarily include legal and regulatory restrictions that can vary by country, data ownership rights, and concerns over

data misuse. Besides these policy-related barriers, there are technical challenges, such as insufficient information technology (IT) infrastructure in hospitals to ensure security during data sharing. Increasing awareness of the importance of data sharing and the potential benefits of AI in the NICU could spur initiatives aimed at harmonizing regulatory frameworks at the multi-institutional level and enhancing IT system standards within hospitals. Given that regulatory changes across countries can be protracted, a possible solution is the creation of local ethics-approved data biobanks operating within a single institution. Although this approach offers greater consistency by relying on less heterogeneous equipment, it also leads to smaller data sets with limited generalizability.

Technical limitations extend beyond IT infrastructure. Some of the equipment used in the NICU does not store data in readable formats, thereby limiting the potential for data to be used for research purposes, beyond simple data visualization for clinical staff. Another hurdle is the aggregation and linkage of diverse data streams, which are often stored separately and in different formats. These integration challenges can be mitigated by adopting standardized data schemes and developing multimodal data platforms with synchronized time-stamping, robust metadata structures, and usable interfaces.

Algorithm Development

AI is a general, nonspecific term that covers a range of statistical learning methods, many with little association to the broader field of machine intelligence. These methods are data-driven techniques that estimate model parameter values through a training process and can range in complexity from a simple linear regression model with a few parameters to a complex deep learning model with billions of parameters. In recent years, generative AI—such as the large language models behind chatbots—has grown into a multibillion dollar industry of geopolitical significance. Although it offers enormous potential for clinical applications, we are yet to see useful applications in neonatal neuromonitoring.[13,14] This commentary focuses instead on discriminative AI, a classification-type approach with a proven track record in this field. Beyond the usual guidelines that apply to developing AI models,[15] there are many challenges and potential pitfalls in developing these algorithms in the context of neonatal neuromonitoring applications.

First, in many instances, there is a lack of sufficient data, an obvious challenge for developing data-driven methods but also for testing the performance of the model. Data sets are often imbalanced with respect to the outcome of interest; for example, the prevalence of seizure is low in the neonatal population within a long-duration EEG recording. Second, data variability is high across neonates and therefore training and testing data sets should include large numbers of neonates without overlap in neonates between the training and testing data sets. Third, the type of AI model should be appropriate to the data set and application, for example, smaller models (less parameters) should be used for small data sets to avoid overfitting. Fourth, performance should be assessed using appropriate metrics. This is particularly important when the data set is imbalanced or when the ground truth is not clear. For example, if the ground truth is determined by the subjective interpretation of a signal, such as the EEG, then the model performance should be evaluated in relation to the inter-rater performance of the human reviewer.[16] Finally, the model should be validated on an independent held-out data set that was not used in the training process, and ideally this data set should be from a different center to the training data. Also, there should be no overlap in reviewers used in the training and validation sets. Although current discriminative AI models primarily rely on objective, quantifiable data, it is also important to incorporate qualitative insights from caregivers, nurses, and clinicians. Though harder to structure, these forms of subjective data could be captured through standardized assessments, scoring systems, or natural language processing techniques applied to clinical notes and parent-reported observations. This process will likely be better captured by generative AI models, a rapidly advancing area of research in health care.[13,14]

AI models for neonatal neuromonitoring can broadly fit into 3 categories: (1) predictive models that combine clinical variables to calculate the risk of brain dysfunction; (2) classification or segmentation models for neuroimaging data; and (3) classification models using physiological data. Here we present some examples in these areas to highlight the scope of the field but see elsewhere for a more comprehensive review.[2-6,17]

Similar to Apgar, Sarnat, or Clinical Risk Index for Babies scores, predictive AI models use readily available clinical variables to evaluate the clinical state of the neonate, for example, a machine-learning model that uses early clinical variables to predict hypoxic-ischemic encephalopathy (HIE).[18] Although AI models have a long history in neuroimaging, their application to neonatology remains relatively limited, likely due to the lack of neonatal data available. The AI methods, however, are mostly transferable. Examples using neonatal neuroimaging data include detecting intraventricular hemorrhage from cranial ultrasounds,[19] detecting acute bilirubin encephalopathy from multimodal MRI[20] and segmentation of the posterior limb of internal capsule in MRI of preterm neonates.[21] Physiological signals, such as cerebral oxygen saturation, can be difficult to interpret from visual inspection other than applying thresholds that indicate a preferred operating region. AI has the potential to extract more information than just applying thresholds. AI models of cerebral oxygenation recorded through near-infrared spectroscopy, for example, have been developed to detect intraventricular hemorrhage in preterm infants[22,23] and brain injury on MRI in term infants.[23]

Probably the most mature area in AI development for neonatal neuromonitoring is the application of AI methods to the EEG—in particular, EEG seizure detection.[16,24] Other AI applications in neonatal EEG include the prediction and

forecasting of seizures,[25,26] classification of sleep states,[27] estimation of brain maturation,[28-30] and grading of background activity.[31-33]

The EEG is a complex signal that is difficult to interpret visually and is subject to inter-rater variability. AI models have the potential to provide a more consistent and accurate interpretation of the EEG signal than a human observer.

Regulatory approval is required to deliver these models from the researchers' computer to the bedside (see Section 6 for more details). Ideally, this burden is passed to an existing commercial entity that has the financial resources and knowledge to negotiate the regulatory landscape. However, the reality is very few AI models make this transition. EEG seizure detection models are one of the few examples that have managed to make this transition: for example, the Nihon Kohden EEG system (Nihon Kohden Corporation)[34] the ANT Neuro Neo system[24] (ANT Neuro), and the Persyst-15 EEG-review software (Persyst Development) all include different regulatory approved seizure detection algorithms.

The Risk of AI Bias

AI bias has already been identified as an issue in medical applications,[35-37] including applications in neonatal monitoring, such as early detection of HIE. External validation of AI algorithms intended for EEG grading in a HIE cohort on "unseen" data sets confirmed the presence of model bias.[32] Data set bias can inadvertently occur in medical research due to small and skewed databases for training and testing AI algorithms, which are not representative of the general population. For example, the dominant presence of a specific group of infants with similar gestational or postnatal age, or place of birth (low- or middle- or high-income country), can skew AI algorithms toward a particular group of patients. Furthermore, AI algorithms intended for the classification or prediction of outcome of rare neurological disorders are limited by small data sets, for example, HIE, with a prevalence of approximately 1.5-3 per 1000 live births in high-income countries and 3-20 per 1000 live births worldwide.[38] AI models for seizure prediction or grading EEG in HIE[25,39-41] use data set sizes ranging from 47 to 181 infants, while data sets <200 would be considered small for AI training. A lack of open data sets (limited to 2 data sets for EEG, for example,[22,42] is also a constraining factor for external validation of AI models. There is also substantial diversity in the equipment used between countries. Multichannel continuous electroencephalography is mainly used in highly specialized centers, following American Clinical Neurophysiology Society guidelines, while the prevalence of amplitude-integrated EEG or single-channel EEG is still high in clinical practice.[43,44]

The Food and Drug Administration (FDA) identified the crucial importance of bias mitigation in AI-based software for medical devices.[45] The FDA intends to support piloting real-world AI performance monitoring algorithms and developing a methodology to identify and eliminate bias.[46]

The efforts of the scientific community are currently oriented in 3 directions: (a) encouraging clinical research centers to collect and curate data sets, (b) sharing data sets across centers or countries (where possible within the constraints of national or regional guidelines and law), and (c) supporting open-source code of AI-based software and trained models. There is a growing trend of publishing open-source AI models through open platforms (such as Zenodo and GitHub),[25,47] which allows for fair comparison of AI algorithms' performances on their own data sets. An alternative to sharing the trained models in the form of computer code is to provide an online service to run the models to make it easier for nonexperts. Recently, the Baba Center (New Children's Hospital, Helsinki University Hospital) developed a cloud platform called BabaCloud, which hosts different algorithms for neonatal EEG. This platform supports researchers and nontechnical users in analyzing their own data with AI models by uploading anonymized EEG data, which is deleted after processing.[31] Federated learning is another potential solution for preserving data privacy in sensitive areas such as neonatal monitoring and improving AI model performances by joint training of multiple data sets across different geographies. The basic idea of federated learning is to enable sharing of the training model parameters across centers instead of sharing the data directly between stakeholders. The AI model parameters are updated via training on local data sets and then are aggregated to a central model. Limiting factors include a slow training process for model development, as the parameters are updated over an internet connection. This can make it difficult to iterate over different model designs during the development process, a typical procedure for an experimental science of AI model building. To date, we are not aware of a successful example in this area of neuromonitoring technologies for neonates. Other limitations include the requirement of developing and applying appropriate data harmonization procedures across data sets before training, although this can be overcome by defining a clear guide to data collection and processing in the first instance. Despite these limitations, federated learning does have the potential to develop AI models on much larger and more diverse data, and with that, the potential to reduce bias.

Algorithm Transparency and Explainability

Despite the potential benefits of AI in neonatal neuromonitoring, a major barrier to its clinical adoption lies in the lack of transparency and explainability of these algorithms.[48] AI systems, particularly deep learning models, are considered "black boxes" because they generate predictions without explaining how decisions are made. This stands in stark contrast to traditional statistical methods, such as logistic regression, which are inherently interpretable but often lack the predictive power of complex AI models. In neonatal neuromonitoring for brain injury, where early interventions can alter lifelong outcomes, clinicians may hesitate to act on AI-based alerts—such as those predicting seizure onset or

abnormal EEG patterns—without a clear understanding of how those predictions are made. Moreover, this lack of transparency also presents other challenges, such as meeting regulatory standards and safely integrating AI into clinical workflows, which complicate its widespread adoption in health care.

To address these concerns, explainable AI (XAI) techniques have been developed, offering clinicians the ability to understand the rationale behind AI decisions, thereby improving diagnostic accuracy and reducing the risk of errors.[49] XAI methods help clinicians identify which clinical features contribute most to AI predictions, enabling critical

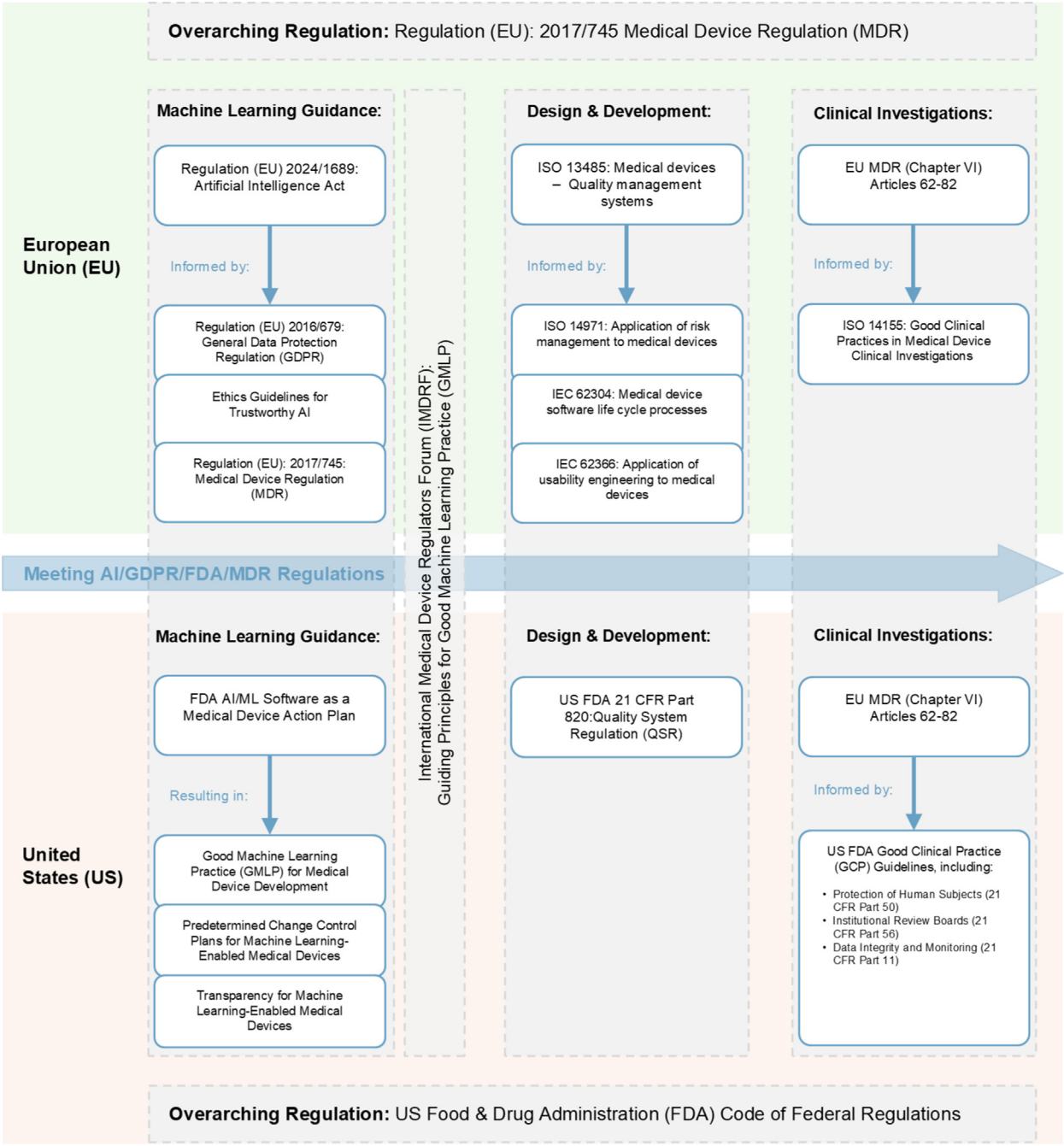

Figure 2. Regulations and standards involved in the development and deployment of an AI-enabled medical device for both the US and European Union (EU). Conformity Européenne - CE mark (EU) and FDA clearance (US) are commonly accepted by several other jurisdictions, with either no additional filing, or abbreviated filings with the individual countries.

evaluation in sensitive domains such as brain monitoring. Although not directly focused on brain injury, Marvin et al[50] developed an interpretable AI approach to predict NICU admissions, identifying preterm birth history and infant weight as key factors. Similarly, Juraev et al used XAI not only to highlight the most influential features but also to describe the rules the algorithm followed to predict mortality in the NICU.[51] Such transparency allows domain experts to verify AI-generated predictions.

XAI is also valuable in analyzing unstructured data like images[52] and signal analysis.[53] For example, Ornek et al combined an AI model with XAI to classify neonatal thermograms as healthy or pathological.[53] The XAI revealed which parts of the image were most important for the AI's predictions, ensuring that the model focused on relevant features rather than unrelated regions, such as the background. By revealing the regions of interest that influenced the model's predictions, clinicians could verify that the AI system focused on medically relevant areas, improving both trust and diagnostic accuracy.

Currently, XAI is used mainly during the development phases of AI models to improve performance and build confidence during validation.[54] However, future research should focus on integrating XAI into real-time clinical workflows, allowing clinicians to benefit from explainable insights during patient care. A promising direction for research is the development of XAI-driven indicators, such as reliability scores, that correlate a model's reasoning with its confidence in predictions.[55] These indicators would not only serve as performance metrics during model validation but could also accompany predictions in real time. Providing clinicians with a reliability score would offer an easily interpretable measure of the model's confidence in its predictions. This would not only enhance trust in AI systems but also allow clinicians to make more informed and confident decisions, ensuring AI's successful integration into real-time clinical workflows for neonatal neurological care.

By addressing the challenges of model explanation and transparency, XAI can play a pivotal role in building trust, meeting regulatory requirements, and ensuring AI systems are safely and effectively integrated into neonatal neuromonitoring.

AI Regulations

Due to the recent influx of AIMDs on the market, regulators including the FDA in the US and the European Union via the Medical Device Regulation (MDR) have focused on introducing AI-specific guidance to ensure the safety and efficacy of AIMDs. These AI-specific guidances are generally aligned with the overarching territory's regulations, as seen in Figure 2.

According to the FDA database for AIMDs, as of 25th June 2024, there are over 950 AIMDs cleared for market in the US. Approximately 82% of these products are in the radiology domain, which suggests that they are clinical decision support-type systems or products.[56] The FDA does not currently have any specific regulation or guidance released for AIMDs. However, 2 draft guidances from the FDA, one focusing on "AI in Drug and Biological Products" and the other on "AIMD Lifecycle Management"[57,58] should help inform AIMD development.

In Europe, the MDR, which was introduced in 2018 to supersede the Medical Device Directive, provides stricter requirements for all medical devices. The MDR specifically included new medical device classifications for Software as a Medical Device (SaMD), including AI-enabled devices. In addition, in Europe the AI Act will require medical device manufacturers (including manufacturers of SaMD and AIMDs) to be dually compliant with the MDR and AI Act. The majority of these rules will become effective in August 2026.[59]

What do these regulations mean for the field of AIMD research and development? They introduce a number of requirements and considerations, some of which are aligned with the International Medical Device Regulators Forum (IMDRF) guiding principles for Good Machine Learning Practice for medical devices, and act as good guiding frameworks for researchers to build trustworthy AI products for medical applications, including:

1. Risk-based frameworks, such as IMDRF Risk Categorization or the FDA SaMD Framework, must be used to determine the risk-profile of the product.
2. Independence of train and test data sets, as well as ensuring that the data sets and study populations used to validate the AIMDs are representative of the intended patient population.
3. Transparency and oversight, with a focus on defining a clear and well-understood intended use or purpose throughout the total product life cycle.
4. Continuous monitoring and postmarket surveillance to maintain or ensure safety and effectiveness through real-world evidence.

A specific bottleneck for AI development in the neonatal community, that is, impacted by Good Machine Learning Practice and IMDRF guiding principles is the data imbalance and small sample sizes that are common in neonatal anomalies, such as HIE. This introduces challenges in conducting suitably powered prospective clinical trials to validate the performance of AIMDs with real-world evidence and drives the need for accessible, retrospective data sets with sufficient power to prove safety and efficacy in a representative intended population. However, there is a risk of bias when using retrospective data that needs to be considered prior to accessing data for testing. Relatedly, the FDA has specific requirements and guidelines for the use of "out-of-US" data in regulatory submissions, which aim to ensure the data meet their standards and are relevant to the US population.

System Design and Deployment

Engaging with stakeholders, including neonatologists, nurses, and other health care professionals, is essential to ensure that AI systems are codesigned to meet clinical needs

and integrate seamlessly into existing workflows. Regular feedback from end users helps identify areas for improvement or areas of concern. Communication is crucial, and sharing findings and advancements with both the medical community and the public promotes access and transparency and helps build trust.

Another important step is to engage parents and patient advocacy groups to align AI solutions with the priorities of those directly impacted by neonatal brain injury and its long-term consequences. This user-centric approach not only enhances the relevance and effectiveness of AI systems but also addresses ethical and practical concerns, resulting in solutions that are well-informed and widely accepted. It is crucial to keep parents involved after implementation to ensure that their experiences and feedback continue to shape improvements and maintain trust.

AI medical devices should be monitored over longer periods of time than the minimum regulatory postmarket requirements. Safety, clinical utility, and efficacy should be reassessed in the context of potential shifts in clinical practice. Carefully designed randomized clinical trials would help reinforce the utility of the AI device. The AI models themselves should be expected to improve over time, by training with more data, increased data diversity, better methods, and more computation—a trend that has accelerated in the last decade and is likely to continue. This may require recertification with regulatory bodies or could be developed through a regulatory framework that allows for this type of iterative improvement (for example, the FDA's Predetermined Change Control Plan). This process safeguards against the risks of being locked into outdated or obsolete AI tools.

Conclusions

The pathway from development of academic AI models to clinically validated tools for neonatal neurocritical care will be both lengthy and complex. The process requires: (1) sufficiently large and diverse data sets to represent the expected population heterogeneity; (2) an awareness of bias and its potential negative impact; (3) transparency in model development, and, in particular, explainability of models where possible, to build trust and confidence; (4) adherence to AI MDRs throughout the life cycle of the product; and (5) clinical evaluation through rigorous validation and postmarket surveillance. We should not be discouraged by this complicated process. Recent advances in AI make this an ideal time to advance our efforts. The potential of AI to assist in improving neonatal outcomes for this vulnerable population is likely to be worth the effort. ■

CRediT authorship contribution statement

Pawel Kulakowski: Writing – review & editing, Writing – original draft, Project administration, Conceptualization. **Ana Alarcón:** Writing – review & editing, Writing – original draft. **Tamara Skoric:** Writing – review & editing, Writing – original draft. **Silvia Seoni:** Writing – review & editing, Writing – original draft, Methodology. **Mark O'Sullivan:** Writing – review & editing, Writing – original draft, Methodology. **John M. O'Toole:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of Competing Interest

This work was supported by the EU COST Action AI4NICU (Maximizing impact of multidisciplinary research in early diagnosis of neonatal brain injury) under Grant CA20124.

Submitted for publication Feb 13, 2025; last revision received May 28, 2025; accepted Jun 21, 2025.

Reprint requests: Pawel Kulakowski, PhD, Institute of Telecommunications, AGH University of Krakow, Czarnowiejska 74, Krakow 30-054, Poland. E-mail: kulakowski@agh.edu.pl

References

1. The website of the COST Action CA20124 AI4NICU "Maximising impact of multidisciplinary research in early diagnosis of neonatal brain injury". Accessed July 10, 2025. https://ai4nicu.org/
2. Chioma R, Sbordone A, Patti ML, Perri A, Vento G, Nobile S. Applications of artificial intelligence in neonatology. Appl Sci 2023;13:3211. https://doi.org/10.3390/APP13053211
3. Keles E, Bagci U. The past, current, and future of neonatal intensive care units with artificial intelligence: a systematic review. NPJ Digit Med 2023;6:220. https://doi.org/10.1038/s41746-023-00941-5
4. Beam K, Sharma P, Levy P, Beam AL. Artificial intelligence in the neonatal intensive care unit: the time is now. J Perinatol 2023;44:131-5. https://doi.org/10.1038/s41372-023-01719-z
5. Husain A, Knake L, Sullivan B, Barry J, Beam K, Holmes E, et al. AI models in clinical neonatology: a review of modeling approaches and a consensus proposal for standardized reporting of model performance. Pediatr Res 2024;2024:1-11. https://doi.org/10.1038/s41390-024-03774-4
6. Rallis D, Baltogianni M, Kapetaniou K, Giapros V. Current applications of artificial intelligence in the neonatal intensive care unit. BioMedInformatics 2024;4:1225-48. https://doi.org/10.3390/BIOMEDINFORMATICS4020067
7. Indrio F, Pettoello-Mantovani M, Giardino I, Masciari E. The role of artificial intelligence in pediatrics from treating illnesses to managing children's overall well-being. J Pediatr 2024;275:114291. https://doi.org/10.1016/J.JPEDS.2024.114291
8. Council of Europe. Report on the impact of artificial intelligence on the doctor-patient relationship. Council of Europe. 2023. Accessed July 10, 2025. https://www.coe.int/en/web/bioethics/report-impact-of-ai-on-the-doctor-patient-relationship
9. General Data Protection Regulation (GDPR). Legal text. Accessed July 10, 2025. https://gdpr-info.eu/
10. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7:351-67. https://doi.org/10.1007/S12553-017-0179-1
11. U.S. Department of Health and Human Services. HIPAA for professionals. U.S. Department of Health and Human Services. 2024. Accessed July 10, 2025. https://www.hhs.gov/hipaa/index.html
12. Digital Europe, Healthcare AI Data Ethics 2030 vision. Digital Europ. 2019. Accessed July 10, 2025. https://digital-europe-website-v1.s3.fr-par.scw.cloud/uploads/2019/02/Healthcare-AI-Data-Ethics-2030-vision.pdf
13. McDuff D, Schaekermann M, Tu T, Palepu A, Wang A, Garrison J, et al. Towards accurate differential diagnosis with large language models. Nature 2025;2025:451-7. https://doi.org/10.1038/S41586-025-08869-4
14. Tu T, Schaekermann M, Palepu A, Saab K, Freyberg J, Tanno R, et al. Towards conversational diagnostic artificial intelligence. Nature 2025;642:442-50. https://doi.org/10.1038/S41586-025-08866-7

15. Chen PHC, Liu Y, Peng L. How to develop machine learning models for healthcare. Nat Mater 2019;18:410-4. https://doi.org/10.1038/S41563-019-0345-0
16. Hogan R, Mathieson SR, Luca A, Ventura S, Griffin S, Boylan GB, et al. Scaling convolutional neural networks achieves expert level seizure detection in neonatal EEG. NPJ Digit Med 2025;8:17. https://doi.org/10.1038/s41746-024-01416-x
17. Sakore SS, Devi S, Mahapure P, Kamble M, Jadhav P. Artificial intelligence applications in neonatal critical care: a scoping review. J Clin Neonatol 2024;13:102-9. https://doi.org/10.4103/JCN.JCN_13_24
18. Murray AL, O'Boyle DS, Walsh BH, Murray DM. Validation of a machine learning algorithm for identifying infants at risk of hypoxic ischaemic encephalopathy in a large unseen data set. Arch Dis Child Fetal Neonatal Ed 2025;110:279-84. https://doi.org/10.1136/ARCHDISCHILD-2024-327366
19. Ahmad T, Guida A, Stewart S, Barrett N, Jiang X, Vincer M, et al. Can deep learning classify cerebral ultrasound images for the detection of brain injury in very preterm infants? Eur Radiol 2024;35:1948-58. https://doi.org/10.1007/S00330-024-11028-4
20. Zhang H, Xia S. Enhancing acute bilirubin encephalopathy diagnosis with multi-modal MRI: a deep learning approach. Appl Sci 2024;14:2464. https://doi.org/10.3390/APP14062464
21. 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. https://doi.org/10.1016/J.ARTMED.2022.102384
22. O'Toole JM, Mathieson SR, Raurale SA, Magarelli F, Marnane WP, Lightbody G, et al. Neonatal EEG graded for severity of background abnormalities in hypoxic-ischaemic encephalopathy. Sci Data 2023;10:129. https://doi.org/10.1038/s41597-023-02002-8
23. Ashoori M, O'Toole JM, Garvey AA, O'Halloran KD, Walsh B, Moore M, et al. Machine learning models of cerebral oxygenation (rcSO2) for brain injury detection in neonates with hypoxic-ischaemic encephalopathy. J Physiol 2024;602:6347-60. https://doi.org/10.1113/JP287001
24. Tapani KT, Nevalainen P, Vanhatalo S, Stevenson NJ. Validating an SVM-based neonatal seizure detection algorithm for generalizability, non-inferiority and clinical efficacy. Comput Biol Med 2022;145:105399. https://doi.org/10.1016/J.COMPBIOMED.2022.105399
25. 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. https://doi.org/10.1111/EPI.17468
26. Bernardo D, Kim J, Cornet MC, Numis AL, Scheffler A, Rao VR, et al. Machine learning for forecasting initial seizure onset in neonatal hypoxic–ischemic encephalopathy. Epilepsia 2024;66:89-103. https://doi.org/10.1111/EPI.18163
27. Montazeri Moghadam S, Nevalainen P, Stevenson NJ, Vanhatalo S. Sleep State Trend (SST), a bedside measure of neonatal sleep state fluctuations based on single EEG channels. Clin Neurophysiol 2022;143:75-83. https://doi.org/10.1016/J.CLINPH.2022.08.022
28. Pavlidis E, O'Toole JM, Pisani F, Boylan GB, Stevenson NJ. EEG maturational age estimation: a comparison of visual and automated interpretation of the EEG in preterm infants. J Clin Med 2025;14:3528. https://doi.org/10.3390/JCM14103528
29. O'Toole JM, Boylan GB, Vanhatalo S, Stevenson NJ. Estimating functional brain maturity in very and extremely preterm neonates using automated analysis of the electroencephalogram. Clin Neurophysiol 2016;127:2910-8. https://doi.org/10.1016/J.CLINPH.2016.02.024
30. Stevenson NJ, Oberdorfer L, Koolen N, O'Toole JM, Werther T, Klebermass-Schrehof K, et al. Functional maturation in preterm infants measured by serial recording of cortical activity. Sci Rep 2017;7:12969. https://doi.org/10.1038/s41598-017-13537-3
31. Moghadam SM, Airaksinen M, Nevalainen P, Marchi V, Hellström-Westas L, Stevenson NJ, et al. An automated bedside measure for monitoring neonatal cortical activity: a supervised deep learning-based electroencephalogram classifier with external cohort validation. Lancet Digit Health 2022;4:e884-92. https://doi.org/10.1016/S2589-7500(22)00196-0
32. 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. https://doi.org/10.1088/1741-2552/ABE8AE
33. O'Sullivan ME, Lightbody G, Mathieson SR, Marnane WP, Boylan GB, O'Toole JM. Development of an EEG artefact detection algorithm and its application in grading neonatal hypoxic-ischemic encephalopathy. Expert Syst Appl 2023;213:118917. https://doi.org/10.1016/J.ESWA.2022.118917
34. 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. https://doi.org/10.1016/S2352-4642(20)30239-X
35. Michelson KN, Klugman CM, Kho AN, Gerke S. Ethical considerations related to using machine learning-based prediction of mortality in the pediatric intensive care unit. J Pediatr 2022;247:125-8. https://doi.org/10.1016/J.JPEDS.2021.12.069
36. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. NPJ Digit Med 2023;6:113. https://doi.org/10.1038/s41746-023-00858-z
37. Ott MA. Bias in, bias out: ethical considerations for the application of machine learning in pediatrics. J Pediatr 2022;247:124. https://doi.org/10.1016/J.JPEDS.2022.01.035
38. Kurinczuk JJ, White-Koning M, Badawi N. Epidemiology of neonatal encephalopathy and hypoxic-ischaemic encephalopathy. Early Hum Dev 2010;86:329-38. https://doi.org/10.1016/J.EARLHUMDEV.2010.05.010
39. Yu S, Marnane WP, Boylan GB, Lightbody G. Neonatal hypoxic-ischemic encephalopathy grading from multi-channel EEG time-series data using a fully convolutional neural network. Technologies 2023;11:151. https://doi.org/10.3390/TECHNOLOGIES11060151
40. Stevenson NJ, Korotchikova I, Temko A, Lightbody G, Marnane WP, Boylan GB. An automated system for grading EEG abnormality in term neonates with hypoxic-ischaemic encephalopathy. Ann Biomed Eng 2013;41:775-85. https://doi.org/10.1007/S10439-012-0710-5
41. Ahmed R, Temko A, Marnane W, Lightbody G, Boylan G. Grading hypoxic–ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine. Clin Neurophysiol 2016;127:297-309. https://doi.org/10.1016/J.CLINPH.2015.05.024
42. Stevenson NJ, Tapani K, Lauronen L, Vanhatalo S. A dataset of neonatal EEG recordings with seizure annotations. Sci Data 2019;6:190039. https://doi.org/10.1038/SDATA.2019.39
43. Shellhaas RA, Chang T, Tsuchida T, Scher MS, Riviello JJ, Abend NS, et al. The American clinical neurophysiology society's guideline on continuous electroencephalography monitoring in neonates. J Clin Neurophysiol 2011;28:611-7. https://doi.org/10.1097/WNP.0B013E31823E96D7
44. Sandoval Karamian AG, Wusthoff CJ. Current and future uses of continuous EEG in the NICU. Front Pediatr 2021;9:768670. https://doi.org/10.3389/FPED.2021.768670
45. Artificial intelligence and machine learning in software as a medical device | FDA. Accessed July 10, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
46. Software as a medical device (SaMD) action plan. 2021. Accessed July 10, 2025. www.fda.gov
47. O'Toole JM, Boylan GB, Lloyd RO, Goulding RM, Vanhatalo S, Stevenson NJ. Detecting bursts in the EEG of very and extremely premature infants using a multi-feature approach. Med Eng Phys 2017;45:42-50. https://doi.org/10.1016/J.MEDENGPHY.2017.04.003
48. Varghese J. Artificial intelligence in medicine: chances and challenges for wide clinical adoption. Visc Med 2020;36:443-9. https://doi.org/10.1159/000511930
49. Reddy S. Explainability and artificial intelligence in medicine. Lancet Digit Health 2022;4:e214-5. https://doi.org/10.1016/S2589-7500(22)00029-2

50. Marvin G, Alam MGR. Explainable feature learning for predicting neonatal intensive care unit (NICU) admissions. In: Proceedings of 2021 IEEE international conference on biomedical engineering, computer and information technology for health, BECITHCON 2021, Dhaka, Bangladesh; 2021.
51. 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. https://doi.org/10.1016/j.jbi.2022.104216
52. Raab D, Theissler A, Spiliopoulou M. XAI4EEG: spectral and spatio-temporal explanation of deep learning-based seizure detection in EEG time series. Neural Comput Appl 2023;35:10051-68. https://doi.org/10.1007/s00521-022-07809-x
53. Ornek AH, Ceylan M. Explainable artificial intelligence (XAI): classification of medical thermal images of neonates using class activation maps. Traitement Du Signal 2021;38:1271-9. https://doi.org/10.18280/ts.380502
54. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). Comput Methods Programs Biomed 2022;226:107161. https://doi.org/10.1016/j.cmpb.2022.107161
55. Seoni S, Molinari F, Rajendra Acharya U, Lih OS, Barua PD, García S, et al. Application of spatial uncertainty predictor in CNN-BiLSTM model using coronary artery disease ECG signals. Inf Sci 2024;665:120383. https://doi.org/10.1016/j.ins.2024.120383
56. Artificial intelligence and machine learning (AI/ML)-Enabled medical devices | FDA. Accessed July 10, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices
57. Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products | FDA. Accessed May 23, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
58. Artificial Intelligence-Enabled Device Software Functions. Lifecycle management and marketing submission recommendations | FDA. Accessed May 23, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing
59. Regulation - EU - 2024/1689 - EN - EUR-Lex. Accessed July 10, 2025. https://eur-lex.europa.eu/eli/reg/2024/1689/oj