Applications of Artificial Intelligence in Neonatology

Roberto Chioma †, Annamaria Sbordone †, Maria Letizia Patti, Alessandro Perri, Giovanni Vento and Stefano Nobile *

Department of Life Sciences and Public Health, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, 00168 Rome, Italy
* Correspondence: stefano.nobile@policlinicogemelli.it
† These authors contributed equally to this work.

Abstract: The development of artificial intelligence methods has impacted therapeutics, personalized diagnostics, drug discovery, and medical imaging. Although, in many situations, AI clinical decision-support tools may seem superior to rule-based tools, their use may result in additional challenges. Examples include the paucity of large datasets and the presence of unbalanced data (i.e., due to the low occurrence of adverse outcomes), as often seen in neonatal medicine. The most recent and impactful applications of AI in neonatal medicine are discussed in this review, highlighting future research directions relating to the neonatal population. Current AI applications tested in neonatology include tools for vital signs monitoring, disease prediction (respiratory distress syndrome, bronchopulmonary dysplasia, apnea of prematurity) and risk stratification (retinopathy of prematurity, intestinal perforation, jaundice), neurological diagnostic and prognostic support (electroencephalograms, sleep stage classification, neuroimaging), and novel image recognition technologies, which are particularly useful for prompt recognition of infections. To have these kinds of tools helping neonatologists in daily clinical practice could be something extremely revolutionary in the next future. On the other hand, it is important to recognize the limitations of AI to ensure the proper use of this technology.

Keywords: infant; machine learning; predictive models; clinical algorithm

1. Introduction

The term “artificial intelligence” (AI) refers to the general ability of computing algorithms to emulate human decision-making, while machine learning (ML) is a subdivision of AI that includes techniques that enable machines to learn from data without explicit programming [1]. Over recent decades, both have deeply influenced personalized diagnostics and therapeutics, drug discovery, and medical imaging [2]. Potentially, these approaches could significantly improve our understanding of disease and therapeutic efficacy in both children and infants [3]. Algorithms are employed in ML for the classification of data and to make predictions. ML encompasses three broad subgroups: supervised ML, unsupervised ML, and neural networks. Supervised ML employs labeled datasets to train an algorithm to perform data classification; in particular, for prognostic applications. Unsupervised ML is generally used as a tool to explore data and requires less human input. Neural networks take inspiration from biological neural networks, can “learn” from missed predictions, and are useful to process imaging data. To date, supervised ML has mostly been used in pediatrics and neonatology to design predictive models. For example, applying readily understandable methods, algorithms designed to identify patients at risk of pneumonia, urinary tract infection, bacterial meningitis, intra-abdominal injury, or clinically relevant traumatic brain injury have been obtained [4].

Although in many situations AI clinical decision-support tools may seem superior to rule-based tools, their use may result in additional challenges. These include the need for vast datasets, the challenges of generalizability, the presence of unbalanced data (i.e., due
Neonatal critical care units usually produce consistent amounts of generally under-utilized data that can be computed and analyzed with AI. In this review, we present and discuss the most recent and impactful applications of AI in neonatology and highlight the future directions for research relating to this population of patients.

2. Applications of Artificial Intelligence

2.1. AI and Neuromonitoring

The revolution in AI—and, particularly, the subcategory of ML—has influenced research on the neuromonitoring of critically sick neonates over the last couple of decades. With the advancement in processing power and data storage, AI now enables computer systems to scrutinize and analyze enormous volumes of information, deciphering illness patterns. Clinical neuromonitoring can generate big data, particularly in terms of electroencephalogram (EEG) data [5]. An EEG is a non-invasive recording of cerebral cortex electrical activity that permits real-time assessment of cortical background function, which has prognostic utility for neonates. As EEGs also permit the differentiation of epileptic seizures from nonepileptic events and can detect nonconvulsive seizures, they are also the reference tool for diagnosing seizures [6]. Continuous EEG (cEEG) recordings maximize the diagnostic utility of this tool, as they are more likely to record electrographic events, but also its prognostic significance, permitting the assessment of the evolution in background activity over time [7,8]. In any case, interpreting the plethora of information generated requires expertise and is resource-consuming, aspects that limit the diffusion of cEEG monitoring.

A number of ML algorithms based on EEG background activity have been proposed with various aims. Much research has been undertaken in the area of automated EEG interpretation regarding the early severity grading of perinatal hypoxic-ischemic encephalopathy (HIE) in order to provide diagnostic decision support [9–11]. Recently, advanced techniques for signal processing and ML, such as convolutional neural network structures able to self-extract convolutional features from raw EEGs [12], have demonstrated strong performance metrics for classifying HIE severity grades [13]. Pavel et al. also investigated the potential role of AI in predictive modeling for electrographic seizures in neonates with HIE [14], aiming to identify infants at the highest risk of subsequent seizures early. ML models were developed separately and in combination for qualitative EEG (assessed by a neurophysiologist) and quantitative EEG (automatically generated by weighting the quantity of discontinuous activity) features, as well as for clinical features. Interestingly, the automated quantitative EEG analysis performed as well as the analysis executed by an expert neurophysiologist, augmenting the predictive value of these models by adding clinical information. These reports emphasize the potential of using ML to process the background EEG activity of neonates with HIE.

AI has also been extensively investigated in sleep studies of term and preterm infants, using EEG data to automatically classify neonatal sleep stages [15,16]. Sleep studies are particularly relevant in the management of sick infants due to their prognostic value, as the expression of alternating vigilance states (i.e., sleep–wake cycling) is a positive prognostic factor and a global marker of neurological well-being [17,18]. Detection of sleep states could also enhance neuroprotective care, as less interruption during deep sleep can be a therapeutic approach. In any case, the research on AI-driven tools evaluating neonatal cortical background activity is still in the preclinical stage.
The detection of paroxysmal events is another key purpose of EEG monitoring. The neonatal population is at the highest risk of seizures compared with any other time in life, and seizures are the most frequent neurological emergency in this age group [19]. The incidence of seizures is estimated at approximately 8% in neonates and has a parabolic relation to gestational age, with seizures being more frequent in infants born at less than 30 and more than 36 weeks of gestation [19]. Early recognition of seizures is crucial as they are oftentimes a manifestation of an underlying pathological condition, such as HIE, meningitis, or stroke [20,21], and because recent evidence suggests that early diagnosis and treatment of seizures improve the response to medication [22]. Furthermore, it is widely established that impaired long-term neurodevelopmental outcomes are associated with an increased seizure burden [23–25]. The recurrence of seizures itself appears to bring additional neurodevelopmental repercussions, independently of the underlying etiology [26]. On the other hand, treatment for nonepileptic events exposes neonates to unneeded harmful medications [27]. The diagnosis of seizures is particularly difficult in the neonatal population as the events are usually electrographic only, their manifestation can be masked by medication, and, even when present, they can be challenging to differentiate from normal neonatal movements [28]. EEG monitoring is thus crucial to identify neonatal seizures, and cEEG monitoring is the recommended standard of care for this aim [29].

According to current guidelines, 24 h cEEG monitoring should be performed for high-risk neonates to screen for seizures, and 24 h of seizure freedom are recommended if seizing activity is detected. Among infants with HIE undergoing therapeutic hypothermia, up to 72 h of EEG recording should be considered, as rewarming is the most likely phase for delayed seizures [29]. In any case, cEEG monitoring is not available in many NICUs due to resource restrictions. The interpretation of data generated with cEEGs is difficult and time-consuming and requires the availability of experienced neurophysiologists on a 24/7 basis. These aspects limit the availability of this crucial neuromonitoring tool, as only half of NICUs report the use of cEEGs, indicating an implementation issue [30]. Amplitude-integrated EEG (aEEG) monitoring has been adopted as a surrogate for cEEG monitoring, gaining universal approval and use in NICUs worldwide. An aEEG is a simplified bedside tool that displays one—or, more commonly, two—channels of EEG data quantitatively converted via smoothing, rectification, and filtering. Then, the cortical electrical activity is time-compressed and presented in a semi-logarithmic chart [31]. The role of aEEG monitoring in the assessment of background cortical activity is well-established [32] and, compared with cEEG monitoring, faster and more straightforward to apply and review. In any case, its sensitivity, specificity, and interobserver agreement are critically suboptimal in detecting seizures [33,34], and experts suggest a complementary role, rather than replacing cEEG monitoring, for it [7,35].

Hence, given the key role but the limited accessibility of cEEGs, significant research effort has been invested over recent decades into the development of automatic seizure detection algorithms (SDAs) [36]. After the pioneering work of Liu et al. [37], who proposed a computerized detection system for neonatal seizures back in 1992, several different approaches have been reported, improved, and validated. While the initial SDAs originated from modifications of algorithms designed for adults [37,38], showing performances far from acceptable for clinical purposes, the progressive complexity of the algorithmic solutions and the techniques for signal processing led to reliable SDAs capable of analyzing channels of raw EEG data in real time and continuously, as shown in Figure 1 [39,40].
After repeated training and offline analysis [11,39], the SDA developed by a group from Cork was tested in a multicenter clinical investigation by assessing the accuracy of the recognition of electrographic seizures with and without the algorithm as a cot-side support tool for the clinical team. Although the predefined target of enhancing the detection of individual infants in NICUs in which no EEG expertise is readily available, suggesting that the advantage offered by the SDA could be greater if tested with teams less experienced in bedside EEG interpretation. Although further research is warranted to demonstrate improvements in diagnostic performance, ML techniques can be safely applied in clinical care, and it

![Seizure probability output](image-url)

An important theoretical limitation of these systems is that the definition of ground truth is not absolute in the detection of seizures. An SDA can only perform as well as the neurophysiologist who performed the annotation of the training data. The insightful report by Stevenson et al. employed the gold standard for EEG monitoring—multichannel cEEG monitoring—to determine the interobserver agreement regarding neonatal seizure identification among three international experts, which was reported to be 83% [41]. These results can be understood as the upper limit for the performance of any machine detecting seizures.

In recent years, the first important steps in the transition of SDAs to the cot-side have been made. As part of the monitoring framework used to conduct an antiseizure medication trial, a 2019 study introduced an SDA in the clinical setting [42]. The real-time review systematically included the algorithm, sending direct email alerts to study neurologists if they chose to activate this function. Although all the study neurologists found the SDA useful, a high false detection rate was reported, and only two out of nine used the instant messaging alert function, others finding false-positive alerts to be too disruptive. Eventually, the first randomized clinical trial evaluating the impact of ML on the recognition of neonatal seizures in real time in an NICU was published in 2020 [43]. After repeated training and offline analysis [11,39], the SDA developed by a group from Cork was tested in a multicenter clinical investigation by assessing the accuracy of the recognition of electrographic seizures with and without the algorithm as a cot-side support tool for the clinical team. Although the predefined target of enhancing the detection of individual infants with seizures was not met, the algorithm significantly improved the recognition of seizure hours. These results were more pronounced on weekends as compared to weekdays (45.3% in the non-algorithm group vs. 66.0% in the algorithm group). The authors speculated that the latter finding may have more reflected the situation in NICUs in which no EEG expertise is readily available, suggesting that the advantage offered by the SDA could be greater if tested with teams less experienced in bedside EEG interpretation. Although further research is warranted to demonstrate improvements in diagnostic performance, ML techniques can be safely applied in clinical care, and it

Figure 1. An EEG viewer incorporating an SDA. In the upper panel the output of the SDA is displayed as a graph of seizure probability. The trace changes to red and an annotation is made when a seizure is identified. The interface also displays the cEEG and aEEG (source: [39]).
appears likely that they will become an integral part of the neurological evaluation of vulnerable neonates.

Recently, AI has also been employed in the field of neonatal neuroimaging and developmental follow-up. The internal capsule is a subcortical white matter structure composed of afferent and efferent fiber tracts connecting the cerebral hemispheres to other subcortical structures, the brainstem, and the spinal cord. The posterior limb of the internal capsule (PLIC) is one of the structures in which myelination commences in the neonatal brain. Importantly, the correct and timely maturation of the PLIC, particularly in terms of myelination, is crucial to neurological outcomes in term and preterm neonates. Abnormal or asymmetrical signals detected through magnetic resonance imaging (MRI) in the PLIC have been associated with perinatal complications, development of hemiplegia, and poor neurodevelopmental outcomes [44]. These findings are especially relevant in neonatal care, as the rate of neurodevelopmental abnormalities among infants with perinatal problems remains high despite advances in clinical practice [45]. In particular, the prevalence of cerebral palsy, a severe condition carrying important life-long burdens and associated with an abnormal PLIC MRI signal, has been increasing over recent decades, being reported as well above 2.0 per 1000 live births [46], with a significant increase with decreasing gestational age [47]. A recent study proposed an ML pipeline for the automated segmentation and quantification of the PLIC in premature infants undergoing MRI [48]. The authors showed good accuracy for the AI method compared to expert analysis, and successfully applied their ML pipeline to data from a large dataset, proving generalizability. Although the proposed method seems particularly promising, assessment of its translation into clinical practice is warranted.

Valavani et al. used ML to predict language abilities at 2 years of corrected age in premature infants by analyzing MRI features and perinatal clinical information [49]. In their study, patterns of delayed myelination and some clinical features (i.e., twin status, sex, antenatal steroid administration, breastfeeding) could predict language delay. The authors concluded that ML might support services and facilitate targeted early interventions, potentially improving the life quality of children born preterm. Other authors have proposed a self-training deep neural network model using MRI data (brain functional connectome) obtained in very preterm infants at term-equivalent age to predict neurodevelopmental deficit at 2 years of age [50]. They applied a self-training technique with a small cohort, obtaining good predictive accuracy; the results will need to be replicated in large cohorts. Balta et al. recently provided new insight into the automated analysis of infants’ general movements [51], an important screening examination for childhood neuromotor disorders. The assessment of infants’ spontaneous movements is particularly informative, as some patterns have been identified to reliably predict cerebral palsy at the age of 3 to 5 months (i.e., the absence of fidgety movements or the presence of predominantly cramped synchronized movements). The authors demonstrated the feasibility of analyzing infants’ general movements with an automated method they developed involving processing videos obtained with an inexpensive RGB-D camera at their homes. The proposed protocol seems promising, as it makes it possible to perform this screening assessment in a familiar environment, improving infants’ compliance and providing an evaluation of the true neurodevelopmental performance, which is sometimes challenging in the clinical environment. Furthermore, this approach could be particularly helpful to overcome logistic barriers in families with reduced access to health services; for example, those living in rural and remote areas. However, despite demonstrating feasibility, the authors could not conduct meaningful statistical analyses due to the small sample size. Research on the validation of the developed protocol and its application in clinical practice with a larger number of infants is warranted.
2.2. AI and the Respiratory System

Respiratory distress syndrome (RDS) is the most common respiratory consequence of prematurity. It is caused by primary pulmonary surfactant deficiency and affects more than 80% of infants born before 29 weeks of gestational age [52,53]. Bronchopulmonary dysplasia (BPD) is a chronic respiratory condition experienced by neonates born at the early stages of pulmonary development [54] that leads to poor long-term respiratory, cardiovascular, and neurodevelopmental outcomes. Its prevalence is around 45% in infants born <28 weeks of gestational age and it is associated with longer hospitalization and a higher risk of comorbidities, such as poor somatic growth and neurodevelopmental impairment [55,56]. BPD also carries substantial social and economic costs [57,58]. The risk of developing BPD is higher in subgroups of infants who experience worse RDS and a worse early respiratory course [59]. The diagnosis of RDS is clinical and is not based on a single specific diagnostic test. As previously mentioned, RDS is caused by lung immaturity and the derived deficit of surfactant in the alveoli, leading to high surface tension and the collapse of respiratory units. The cornerstone of the management of patients with RDS is administrating exogenous surfactant, aiming to enhance pulmonary compliance and minimize exposure to mechanical ventilation and the derived ventilator-induced lung injury [60]. In any case, prophylactic therapy with surfactant is not a cost-effective strategy [53], and unnecessary or inappropriate administration can lead to potentially dangerous side effects (i.e., airway lesions, pneumothorax, pulmonary hemorrhage). Some authors have investigated the role of biomarkers of RDS, such as the lecithin/sphingomyelin ratio (L/S ratio) from gastric aspirates or amniotic fluid samples. The L/S ratio can give information about endogenous pulmonary surfactant production, which is decreased in preterm infants with RDS. Unfortunately, the calculation of the L/S ratio is still confined to research and not available at the bedside due to technical complexities. To overcome these limitations and provide a bedside point-of-care test method useful for clinicians, Ahmed et al. tested a machine-learning algorithm also useful for the study of other biological samples that are interpretable in the mid-infrared spectrum [61]. Their findings are particularly insightful as, after clinical validation, the employment of ML-guided devices able to quantify biomarkers of RDS in real time might potentially guide interventions in preterm infants with respiratory symptoms. In any case, considerable concerns about signal-to-noise and signal-to-background limitations remain, and the applicability of this new technology at the bedside does not appear forthcoming.

Raimondi et al. focused on lung ultrasound, a technique able to quantify pulmonary aeration, and its ability to correlate with respiratory status in sick neonates, applying AI-assisted analysis [62]. They built a dataset of scans (600 ultrasound frames) to perform a texture and ML analysis with the aim of correlating the ultrasound and the mean grayscale intensity with the oxygenation status. To do so, they enrolled a cohort of 75 neonates with various degrees of respiratory distress of diverse origin. Despite a significant linear correlation between the grayscale ML analysis and blood gases indexes being found, the relatively small sample size, the inhomogeneous etiology of the respiratory distress, and the variable postnatal age warrant further research on this topic with larger datasets derived from more uniform cohorts. The representative lung ultrasound findings and the calculated intensity histograms are shown in Figure 2.

A particular topic of research interest in pediatric respiratory care is risk stratification for BPD, aiming to identify infants that could benefit from preventive interventions, such as corticosteroids, or treatment for a patent ductus arteriosus using a standardized risk estimation tool. The US National Institute of Child Health and Human Development has recently endorsed a prognostic tool, the BPD Outcome Estimator [63], which is widely employed to guide steroid administration and family counseling. In any case, the estimator is only applicable to neonates of White, Black, or Hispanic descent. Patel et al. recently designed an ML algorithm for respiratory outcome prediction for extremely premature infants of Asian heritage, aiming to fill this gap in knowledge [64]. As a demonstration of the model, a Web app was developed and is currently consultable online. In any
Apnea of prematurity (AOP) is another common issue detected in premature infants admitted to NICUs. It is correlated with the immaturity of the respiratory control system in these fragile patients and can be triggered by several conditions, including infections, gastroesophageal reflux, and seizures [68]. AOP is diagnosed in more than 50% of preterm babies and is almost universal in infants with extremely low birth weight [69]. It is associated with brain damage and poor neurodevelopmental outcomes and thus needs prompt recognition and treatment. AOP can be categorized as central (lack of diaphragmatic contraction due to the cessation of respiratory drive), obstructive (due to airway obstruction), or mixed apnea (characterized by both central and obstructive features). As a standard of care, bedside monitors are set to generate an alarm in cases of a temporary drop in the
amplitude of the thoracic impedance, as it reflects a drop in thoracic motion and, indirectly, in respiratory effort [70]. This strategy permits the identification of central apneas with suboptimal accuracy, as reported in clinical studies, showing a high rate of false positive episodes [71]. Varisco and colleagues developed and validated an enhanced AOP detection model based on ML to automatically recognize true apnea by analyzing data from the electrocardiographic monitoring of infants [72]. The authors found that AOP often occurred in clusters and that breathing patterns were more disturbed in patients with more frequent central AOP and concluded that AI leads to improved apnea detection compared to traditional methods with fewer false alarms. Considering how alarm fatigue is becoming a common issue in modern NICUs, exposing sick infants to the risk of missed alarms and potentially disrupting consequences, AI may radically change daily clinical practice. In any case, the limited sample size (20 preterm infants overall) constitutes an important flaw in the proposed method, as does the lack of external or prospective validation.

2.3. **AI and the Eye**

Retinopathy of prematurity (ROP), one of the most frequent causes of childhood blindness, is a multifactorial disease related to risk factors such low birth weight (BW), low gestational age (GA), oxygen exposure, infections, and blood transfusions [73]. To detect and possibly treat this condition in a timely manner, it is important to start the screening at the right time and to continue with regular monitoring and evaluations. Current ROP screening involves a first eye examination and several follow-up examinations performed longitudinally during hospitalization and after discharge if needed. The examination can be quite stressful for the baby and laborious for the ophthalmologists. Moreover, it is not always possible to follow a correct ROP screening program in remote areas [74]. Over the last few years, applications of AI in the diagnosis of this condition have been evaluated: in a recent large, international, multicenter study, a deep-learning method was trained to detect one of the features of the affected retina and to give an automated score, showing good results [75]. This important report provided a proof of concept, demonstrating that a deep-learning comprehensive screening platform could potentially improve objective ROP diagnosis and screening accessibility. Another set of authors have recently developed a deep-learning approach for the prediction of ROP and its severity in a large multicenter study [74]. The model they trained is based on retinal images from the first ROP screening and clinical data for the patient available before or at the time of the first ROP screening, making it possible to include retinal status in the predictive model. The artificial intelligence’s training was based on three schemes (majority voting, one-vote veto, and image-level methods) and allowed it to generate the predictive probability of retinal photographs coupled with clinical risk factors. The deep learning-based system had similar accuracy compared to a commonly used ROP score algorithm (the ROP score) and could even have better efficiency in the detection and interpretation of pathological features compared to the classical medical evaluation. Despite the promising results of the study, larger external validation with other prospective multicenter datasets is warranted. Potentially, more sophisticated ML algorithms could provide more important prognostic details about the condition (i.e., precise staging, zone, plus disease).

2.4. **AI and Vital Signs Monitoring**

The monitoring of vital signs in newborns, such as respiratory and heart rates and body temperature, is one of the key elements of modern neonatal care. For example, an imbalance in vital signs is one of the early signs of infectious diseases, which represent one of the main causes of mortality and morbidity worldwide. To monitor these parameters in newborns, sensors are placed in direct contact with the skin, potentially causing skin injury, stress, and discomfort for babies. Non-contact measurement methods (i.e., camera recordings, laser Doppler vibrometry, and others) have recently been developed to continuously record vital parameters and avoid such complications in fragile patients [76,77]. Advanced algorithms are needed to work on the images deriving from automatic camera-based
measurements of vital signs. The tracking of regions of interest (ROIs) and signal fusion for vital signs extraction are important steps in the performance of camera-based temperature monitoring. Deep learning-based approaches may lead to robust, real-time assessment of vital parameters, such as body temperature alterations, as suggested by Lyra et al. [78]. In their work, clinical data for neonates were used to train a model based on deep learning in order to achieve real-time extraction of surface temperature. The analysis of the results was challenging for several reasons (e.g., the parameters and data related to the changes in the positions of infants’ bodies during the recording). However, the authors showed the feasibility of low-cost, embedded GPUs for real-time temperature monitoring of neonates, and further studies are expected to expand this method in clinical practice.

2.5. AI and the Gastrointestinal System

Intestinal perforation (IP) in preterm infants is a life-threatening condition. It can be spontaneous or secondary to necrotizing enterocolitis and requires urgent surgical intervention. Even if risk factors for IP have been elucidated, preventive measures and prediction models are not easily developed because the pathogenesis of this condition in preterm babies is still not fully understood. ML can be helpful, allowing a meticulous assessment of complex conditions starting from large datasets. Given the deep clinical impact of such conditions on outcomes such as death and quality of life, it is undeniable that having an AI tool helping and supporting clinicians in management could be extremely helpful. A recent study developed an AI algorithm trained on a large national dataset relating to the clinical features of infants [79]. The proposed algorithm predicted and characterized IP preterm infants with excellent accuracy, outperforming all other classic ML algorithms. The authors speculated that the assessment of different clinical data, such as vital signs, radiologic findings, biomarkers, and laboratory results, might enable the development of a more accurate ML model, eventually making it possible to achieve early prediction of IP and improved outcomes for preterm babies. In the nutrition field, a recent study by Han et al. focused on the possibility of using AI to predict postnatal growth failure and thus plan timely interventions. ML models were developed employing different techniques with a large dataset (approximately 8000 very low birth weight infants from various Chinese NICUs), and their feasibility and good predictive performance were demonstrated [80]. In any case, despite the vast size of the dataset, it did not include important information regarding enteral and parenteral nutrition, limiting the study results.

2.6. AI and Neonatal Jaundice

A very interesting application of AI for the future is in jaundice diagnosis. A Saudi Arabian group explored the possibility of using deep learning and ML models for the diagnosis of neonatal jaundice using a dataset comprising images from a smartphone camera. The study population included late preterm and term newborns. In this study, the authors trained a neural network to recognize jaundice from the information obtained from the images of the skin and the eyes [81]. Guardalia et al. studied the clinical data for a large population of newborns with the intention of building a risk stratification tool for neonatal jaundice using an ML approach and without using bilirubin measurements. Their model showed good accuracy in the risk stratification of neonatal jaundice [82].

3. Discussion

Artificial intelligence has great potential to improve our understanding of disease and to enhance therapeutic effectiveness—in neonatal medicine as well as elsewhere—as reported in recent studies. We presented the most relevant studies in this field and found that the most important applications of AI involved neuromonitoring, assessment, and prediction of significant respiratory diseases and prediction of retinopathy of prematurity, neonatal sepsis, jaundice, and intestinal perforation. These conditions represent serious complications of prematurity and may lead to permanent consequences and even death. The development of point-of-care tools is another promising approach, enabling clinicians
to predict and effectively treat some conditions, such as RDS. In any case, there are some ethical challenges in the application of AI to neonatal clinical care. Permission to participate is typically provided by parents acting as legally authorized representatives since neonates are not capable of consenting. Considering the need for proxy consent, newborn infants need additional protection as research participants. Furthermore, parents of sick or premature infants have to face numerous difficulties in the early weeks and months of life of their babies. Mental, physical, and emotional fatigue may impair the capacity of parents to provide informed consent. Hence, particular attention has been given to the respect for rigorous ethical standards in perinatal research [83]. There are also several ethical and jurisdictional pitfalls regarding the application of ML in clinical decision-making. For example, if the validated algorithm were to make a mistake, who should be considered accountable for it? Other limitations of AI include the need for large datasets to “train” the models, the presence of unbalanced data, issues of generalizability, a lack of evidence-based care for some neonatal conditions, maturational variation, and costs. Further important issues to be faced when applying AI to neonatal care are the need for proper performance assessment, the need for external validation to enhance generalizability, and the interpretability of the model. The results of the studies presented in our manuscript are promising. Moreover, the majority of the diseases we discussed represent serious neonatal conditions and may lead to permanent consequences and even death. The prevalence of the diseases discussed in this narrative review, along with estimates of related direct and indirect costs, have been previously reported [84–86] and, in our opinion, justify including the assessment of AI on the research agenda.

4. Conclusions

There is a growing interest in the medical field regarding the modern applications of artificial intelligence in neonatal care. In this narrative review, we presented some of the most recent and significant findings related to the topic in order to provide an overview of the potential of AI applications in neonatology.

Author Contributions: Conceptualization, S.N.; methodology, R.C., A.S. and S.N.; writing—original draft preparation, R.C., A.S. and S.N.; writing—review and editing, S.N., R.C., A.S., M.L.P., A.P. and G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the nature of the manuscript.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Acknowledgments: We thank Geraldine B. Boylan, Sean Mathieson, and Francesco Raimondi for providing us with insightful images.

Conflicts of Interest: The authors declare no conflict of interest.

References


64. Patel, M.; Sandhu, J.; Chou, F.S. Developing a machine learning-based tool to extend the usability of the NICHD BPD Outcome Estimator to the Asian population. PLoS ONE 2022, 17, e0272709. [CrossRef] [PubMed]

65. Xing, W.; He, W.; Li, X.; Chen, J.; Cao, Y.; Zhou, W.; Shen, Q.; Zhang, X.; Ta, D. Early severity prediction of BPD from chest X-ray images using deep learning: A study at the 28th day of oxygen inhalation. Comput. Methods Programs Biomed. 2022, 221, 106869. [CrossRef] [PubMed]


81. Althnian, A.; Almanea, N.; Aloboud, N. Neonatal Jaundice Diagnosis Using a Smartphone Camera Based on Eye, Skin, and Fused Features with Transfer Learning. *Sensors* 2021, 21, 7038. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.