








## REVIEW ARTICLE

# Artificial intelligence in the diagnosis of necrotising enterocolitis in newborns

Arkadiusz Sitek<sup>1,2</sup><sup>✉</sup>, Joanna Seliga-Siwecka<sup>3</sup><sup>✉</sup>, Szymon Płotka<sup>1,4</sup>, Michał K. Grzeszczyk<sup>1</sup>, Szymon Seliga<sup>1</sup>, Krzysztof Włodarczyk<sup>5</sup> and Renata Bokinić<sup>3</sup>

© The Author(s), under exclusive licence to the International Pediatric Research Foundation, Inc 2022

Necrotising enterocolitis (NEC) is one of the most common diseases in neonates and predominantly affects premature or very-low-birth-weight infants. Diagnosis is difficult and needed in hours since the first symptom onset for the best therapeutic effects. Artificial intelligence (AI) may play a significant role in NEC diagnosis. A literature search on the use of AI in the diagnosis of NEC was performed. Four databases (PubMed, Embase, arXiv, and IEEE Xplore) were searched with the appropriate MeSH terms. The search yielded 118 publications that were reduced to 8 after screening and checking for eligibility. Of the eight, five used classic machine learning (ML), and three were on the topic of deep ML. Most publications showed promising results. However, no publications with evident clinical benefits were found. Datasets used for training and testing AI systems were small and typically came from a single institution. The potential of AI to improve the diagnosis of NEC is evident. The body of literature on this topic is scarce, and more research in this area is needed, especially with a focus on clinical utility. Cross-institutional data for the training and testing of AI algorithms are required to make progress in this area.

*Pediatric Research* (2023) 93:376–381; <https://doi.org/10.1038/s41390-022-02322-2>

**IMPACT:**

- Only a few publications on the use of AI in NEC diagnosis are available although they offer some evidence that AI may be helpful in NEC diagnosis.
- AI requires large, multicentre, and multimodal datasets of high quality for model training and testing. Published results in the literature are based on data from single institutions and, as such, have limited generalisability.
- Large multicentre studies evaluating broad datasets are needed to evaluate the true potential of AI in diagnosing NEC in a clinical setting.

**INTRODUCTION**

Necrotising enterocolitis (NEC) is one of the most serious conditions in newborns, affecting up to 10% of very-low-birth-weight infants. In the most premature population, mortality rates can increase to as high as 60%.<sup>1</sup>

The precise aetiology remains unclear. However, fundamental risk factors, such as prematurity, enteral feeding, intestinal colonisation, and bowel ischaemia, have been well established.<sup>2</sup> NEC mainly affects the distal ileum and colon. Clinical symptoms include feeding intolerance, abdominal distension, and bloody stools. In the most severe cases, patients may present with abdominal wall erythema, apnoeic spells, lethargy, and septic shock.<sup>3</sup> The suspected diagnosis is confirmed with typical findings on abdominal radiography (AR), including pneumatosis intestinalis (PI), portal vein gas (PVG), and in extreme cases, pneumoperitoneum.<sup>4</sup>

In 1978, Bell developed a staging system for NEC, which was modified by Walsh and Kliegmann in the mid-1980s.<sup>5</sup> A combination of clinical symptoms and AR findings allows for the

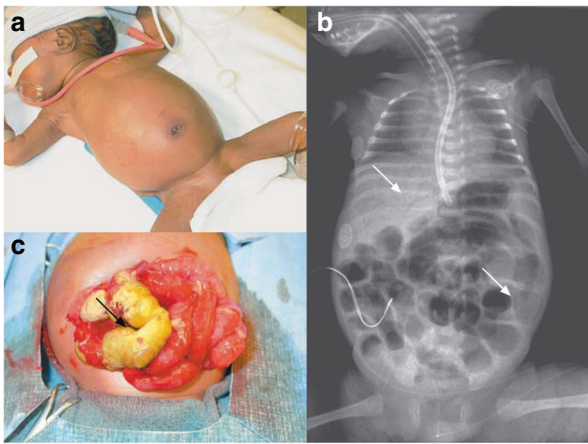
grading of interventions and standardised treatment. This widely adopted classification is based on plain radiology findings, despite the first reports on ultrasound (US) use for NEC diagnosis being published before Walsh and Kliegmann<sup>5</sup> in the early 1980s.<sup>6,7</sup> Abdominal ultrasonography can depict PI, PVG, and pneumoperitoneum (in some cases, the head of the AR); however, it also provides other crucial information, such as bowel wall viability (thickness or thinning) and free abdominal fluid. These additional findings are helpful for the diagnosis and management of NEC.

Despite several studies, NEC remains a conundrum. Timely diagnosis and treatment implementation remain a major challenge for neonatologists worldwide. Despite numerous efforts, morbidity rates have not improved over time. This is mainly due to the fact that early symptoms of NEC are often non-specific and difficult to distinguish from benign illnesses, such as apnoea of prematurity or feeding intolerance. Moreover, NEC is a rare disease; hence, it is difficult to gain professional experience and develop expertise in its timely diagnosis and treatment. The rarity of NEC also hinders the creation of effective clinical protocols

<sup>1</sup>Sano Centre for Computational Medicine, Cracow, Poland. <sup>2</sup>Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. <sup>3</sup>Department of Neonatology and Neonatal Intensive Care, Medical University of Warsaw, Warsaw, Poland. <sup>4</sup>Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands. <sup>5</sup>Main Library, Medical University of Warsaw, Warsaw, Poland. ✉email: [asitek@mgh.harvard.edu](mailto:asitek@mgh.harvard.edu); [joanna.seliga@wum.edu.pl](mailto:joanna.seliga@wum.edu.pl)

Received: 22 April 2022 Accepted: 3 September 2022

Published online: 4 October 2022



**Fig. 1** Clinical and radiographic features of necrotising enterocolitis. Panel **a** shows an infant with a shiny, distended abdomen with periumbilical erythema. (Photograph courtesy of Dr David Kays, Department of Pediatric Surgery, University of Florida.) In the radiograph shown in panel **b**, the upper arrow points to portal air, and the lower arrow points to a ring of intramural gas, which is indicative of pneumatosis intestinalis. (Radiograph courtesy of Dr Jonathan Williams, Department of Pediatric Pathology, University of Florida.) In Panel **c**, the arrow points to an area of necrotic bowel in a patient with necrotising enterocolitis. (Photograph courtesy of Dr David Kays, Department of Pediatric Surgery, University of Florida.) Figure reused Neu and Walker<sup>3</sup> with permission from the *New England Journal of Medicine*.

owing to the lack of high-quality data. Electronic health records are often incomplete and lack crucial information, such as the outcomes of AR or US studies. Moreover, data collection procedures vary from country to country, increasing the difficulty of creating and benchmarking NEC diagnosis protocols.

Artificial intelligence (AI) is a broad discipline with the main purpose of providing machines the ability to perform actions that require domain knowledge and intelligence (driving a car, conducting surgery, patient diagnosis, etc.). Over the last decades, we have witnessed rapid progress in machine learning (ML), a subdomain of AI that has also made big strides in healthcare.<sup>8</sup> ML is a data-driven method that can assist in decision making in healthcare. Data-driven methods are constructed solely based on retrospective data with no expert input. Despite several reported successes with retrospective datasets, AI methods have yet to have an impact on clinical practice<sup>8</sup> and experience commercial success.<sup>9</sup>

ML is divided into three categories: unsupervised, supervised, and reinforcement learning. Supervised learning is used when the ground truth (GT) is available (e.g., whether the newborn has NEC or not).<sup>10</sup> The supervised ML model predicts the GT from the input data. In data science, the input variables are often referred to as features or explanatory variables. Among the most popular supervised ML models are support vector machines (SVMs), decision trees (DTs), linear regressions (LRs), and naive Bayes (NB). Unsupervised learning is often used when there is no available GT or when the goal is to find new patterns in the data. Typical tasks in unsupervised learning are clustering (grouping data samples) and anomaly detection (e.g., detecting arrhythmia in the ECG signal). Reinforcement learning is devoted to training an agent's (a car, AI-controlled computer game player, or an autonomous robotic surgeon) behaviour to achieve the maximum reward (e.g., successful surgery).<sup>11</sup> This type of ML is frequently used in robotics.

The greatest limitation of AI systems is that they adequately generalise their behaviour to an unknown environment (e.g., diagnosing a patient with previously unseen symptoms). With the

rise of computational capabilities, such as those of artificial neural networks (ANNs), this issue has been partially overcome. Deep ANNs have opened the path to a new area of ML: deep learning (DL). Deep refers to the fact that, in deep networks, information is processed by several layers of ANN (hundreds or even thousands), whereas in shallow networks, there are only a few of these. DL has quickly become the go-to tool for several tasks owing to its robustness and multiple developed DL architectures of ANN suited for different problems. For example, the early architectures of deep convolutional neural networks (CNNs) were able to improve image recognition tasks by 20% (e.g., if provided a picture of a horse, the algorithm classifies the picture as showing a horse) compared with standard computer vision algorithms.<sup>12</sup> Currently, state-of-the-art CNNs surpass human performance in image recognition tasks. Furthermore, deep ANNs completely transform the abilities of computer programmes to understand spoken or written languages.<sup>13</sup>

This narrative review aimed to summarise the currently available literature on the use of AI in diagnosing NEC, highlight open issues, and identify future directions for implementing AI in clinical neonatal practice (Fig. 1).

## METHODS

Our study was a narrative review based on a systematic search strategy to gather facts from the available literature. This is in contrast to a classic systematic review, which is designed to provide an answer to a defined empirical question. Our aim was to gather data on the use of AI in NEC diagnosis. Reports were included or excluded if they met our inclusion or exclusion criteria, respectively (Table 1).

PubMed, Embase, arXiv, and IEEE Xplore databases were searched for this narrative review. In consultation with a research librarian (K.W.), a standardised search strategy was employed using a standardised set of keywords and operators, which are listed in the Appendix. No other filtering or restrictions were applied in the search strategy. Additional strategies to identify studies included manual reviews of reference lists from key articles that fulfilled our eligibility criteria and the use of 'related articles' feature in PubMed. The electronic database search was supplemented by searching for grey literature: trial protocols through clinical registers (ISRCTN registry and ClinicalTrials), thesis dissertation (sourced through ND LTD and EthOS), conference proceedings (searched through Web of Science and Embase), and other grey literature databases (OpenGrey and Trip database). Details of the search strategy are presented in the Appendix.

## RESULTS

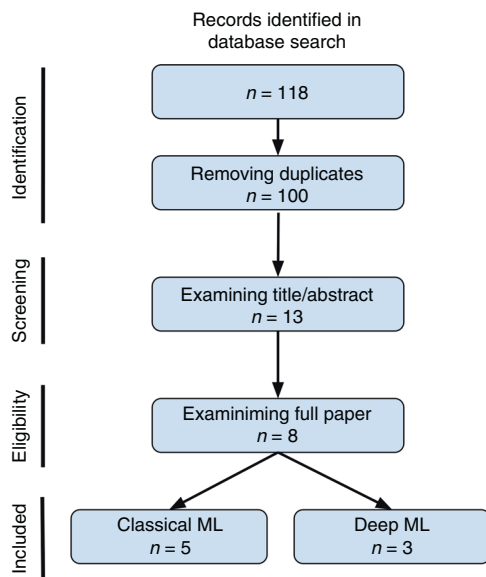
Eight relevant publications were identified. Figure 2 shows a flow diagram of the selection process. We divided the articles into two groups based on the type of ML used.

### Classic ML in decision support in NEC

Classic ML algorithms (such as DTs, SVMs, and LRs) are often utilised for supervised learning using clinical data. Clinical data and other numerical values are often in the form of tabulated values, and classic ML methods are designed for use with such datasets. In a paper published by Mueller et al.,<sup>14</sup> ANN was used to diagnose NEC using a retrospective clinical dataset from a single institution. The authors included a relatively small dataset of 197 premature infants, of whom 67 were diagnosed with NEC. Data were obtained from the Perinatal Information System (PINS) database of the Medical University of South Carolina. Fifty-seven variables from the PINS database were selected as features after performing a literature review and discussion within an expert panel, which features may be relevant to the diagnosis of NEC. The authors investigated the importance of variables that can be used

**Table 1.** Inclusion and exclusion criteria used in literature search.

Included	Excluded
Neonatal population	Population with no neonates
NEC diagnosis	Editorials and expert opinions
Use of AI or ML in NEC diagnosis	
Prospective and retrospective studies	
All languages and dates	

**Fig. 2** Flow diagram of the study selection process. The search identified 118 publications, this number was reduced to only 8 by duplicate removal, screening, and applying eligibility criteria.

to perform accurate decision making in the case of NEC prediction. This study did not include the statistical results for NEC prediction.

Ntonfo et al.<sup>15</sup> presented a different approach for NEC detection. An infrared camera was used for thermal image acquisition instead of collecting clinical data or radiological images. After initial image preprocessing operations, thermal signatures were extracted from newborn abdominal thermal images. The statistical features acquired from the signatures can be fed into a classifier for NEC diagnosis. The features showed different characteristics in children with and without NEC. Unfortunately, this method was only tested in two newborn children and should be further investigated.

Irls et al.<sup>16</sup> developed two estimation models for intestinal perforation (IP) based on a back-propagation ANN: (a) at birth and (b) at birth and during hospitalisation. The study cohort included three groups: (1) control group without NEC ( $N = 27$ ), (2) NEC group ( $N = 23$ ), and (3) IP (Bell's stage IIIB) ( $N = 26$ ). They excluded 15 cases with incomplete clinical information, spontaneous or not associated with NEC IP, as well as digestive tract malformations. For further analysis, they chose 113 variables of maternal and neonatal clinical, feeding, and laboratory parameters from the medical record data of the neonatal intensive care unit (NICU) in a single institution. This study aimed to obtain an ANN-based model to estimate IP associated with NEC diagnosis and investigate key factors for prediction. The regression coefficient between the experimental and predicted data is  $R^2 > 0.97$ . They found that the

male sex was a highly predictive parameter for NEC-associated IP. However, more studies are needed to confirm that the male sex is more likely to progress to IP. These models may allow for quality improvement in medical practice. The main limitation of this study is its single-centre nature and relatively small dataset.

Lure et al.<sup>17</sup> used random forest and ridge logistic regression to discriminate between NEC and spontaneous IP (SIP). These diseases are difficult to differentiate without bowel visualisation; however, ML algorithms accurately separate NEC and SIP. The risk factors for NEC, including very low gestational age at birth, placental abruption, and asphyxia, were used as explanatory variables (tabulated input values). It has been shown that this method can improve the clinical decision-making process prior to any surgical intervention. The experiments were conducted using a dataset of 40 patients collected from the University of Florida.

Lueschow et al.<sup>18</sup> suggested that no formal comparison between the multiple existing NEC definitions has been performed. They investigated the performance of these definitions and applied ML techniques to test their ability to diagnose NEC. To conduct the experiments, a cohort dataset of >200 patients acquired over 10 years from a single institution was analysed. The features (explanatory variables) selected for the experiments were those required for different NEC diagnosis methods: Bell staging,<sup>5</sup> modified Bell staging,<sup>19</sup> and non-Bell NEC definitions.<sup>20,21</sup> For each NEC definition, six ML classifiers (K-nearest neighbours, simple neural network, NB, random forest, SVM, and DT) were trained on the features required by the definition. NEC diagnosis with ML outperformed traditional criteria in terms of specificity and sensitivity and opened a discussion for further examination and the development of new NEC definitions. Moreover, newer definitions were more accurate than the Bell-based criteria. In addition, feature importance analysis was performed, and the authors suggested that features containing values from the specific range: volume of feeding at NEC onset, and gestational age, can be more informative than simple, binary (yes/no) features.

### Use of DL in decision support

In this section, we describe publications that use DL algorithms for NEC diagnosis. Van Druten et al.<sup>22</sup> proposed a computer-aided diagnosis (CAD) system that consists of an ensemble (outputs of multiple algorithms are combined) of conventional ML and DL algorithms. The authors used AR images of participants with radiological patterns consistent with NEC as well as those without these patterns. Radiologists identified NEC-related patterns on AR images. No information about the study cohort used was noted. The classic ML algorithms include a feature extraction algorithm (e.g., texture-based feature extraction), feature selection, and classification. This study aimed to produce heatmaps for various imaging features to highlight NEC pathology on ARs. In the DL algorithm, ARs were automatically analysed using a deep neural network for their automatic classification. The CAD-based system compares and qualifies the prediction accuracy of conventional ML and DL approaches. As the final output, the algorithm uses a visualisation technique that highlights areas on the AR images with NEC features. The authors did not provide any information on the number of datasets used to develop the methods. However, no quantitative evaluation has been conducted.

Gao et al.<sup>23</sup> proposed a multimodal AI-based system consisting of feature engineering, ML, and DL algorithms. Feature engineering in data science is a process of creating additional features from existing features (e.g., having features  $a$  and  $b$ , we engineer new features  $a/b$ ). A multimodal (as opposed to unimodal) AI model uses different types of data when computing the output (e.g., the probability of NEC). They evaluated the proposed system using ARs and clinical data from a single institution. The study cohort included 2234 infants, including 1201 non-NEC, 622 NEC, and an independent group of 411 NEC patients, including surgical and

medical NEC. Some infants were excluded from the study due to the lack of complete clinical parameters or poor image quality, and 827 infants (342 NEC and 385 non-NEC) were selected for the analysis. They used ARs and clinical data; hence, a multimodal approach. In Gao et al.'s study, the authors identified significant features of ARs and clinical data that were closely related to the AI diagnosis and prediction of the success of surgical intervention for NEC. The authors found that the AI system was capable of predicting which NEC patients would have a higher likelihood of successful surgery. The limitations of this study include the lack of genetic information, microbiome data, and biochemical parameters of the infants. The authors concluded that AI could be used as an auxiliary diagnostic tool to confirm their results in a prospective clinical trial.

Lin et al.<sup>24</sup> proposed a novel, interpretable neural network-based architecture solution for independent microbiota DNA sequences. Multicentre datasets came from two centres and contained 3595 stool samples from 261 at-risk infants, 75 of whom developed NEC. The data were collected noninvasively. In addition, 10 clinical metadata features collected and reported in both of the two historical studies, including maternal information, such as age, parity, and biometric data, and details about the birth, were used. In this study, the authors used DL neural networks to estimate NEC risk. They introduced a novel ML method called the 'growing bag' analysis, which models time evolution. They found that NEC predictions can be made on an average of 8 days before disease onset. The system described in this study generates a longitudinal NEC risk score from a limited set of bacterial taxa and basic clinical metadata. It allows early and accurate NEC prediction, with a mean sensitivity and specificity of 86% and 90%, respectively.

## DISCUSSION

The number of publications on the use of ML in NEC diagnosis is small, as we discovered in this study. We found only eight publications using the inclusion and exclusion criteria. Some of those publications were conference papers, which typically underwent a less strict peer-review process, with missing key information, such as the number of patients. One commonality of all the papers that we found is the limitations, sometimes severe, in the size of the data that was used. ML models, particularly DL models, require high-quality large datasets for model training and testing. Obtaining such datasets for NEC is difficult owing to the rarity of the disease that, in our opinion, will hinder the development of such models in the future compared with other AI applications for more common diseases. Medical institutions differ in the ways NEC patients are diagnosed and cared for, which makes the process of developing robust and generalisable AI models even more difficult. This heterogeneity of data from different sources puts even more demand on the amount needed to create an algorithm that can be successfully used across different institutions.

Notwithstanding the difficulties associated with the lack of data, we also found results that encourage the pursuit of AI as a decision support tool in NEC diagnosis. The best example is the recent work of Lin et al.<sup>24</sup> The authors predicted the future occurrence of NEC in patients with impressive sensitivity and specificity of 86% and 90%, respectively. These results are encouraging and demonstrate the potential of AI in improving the care of patients with NEC.

In the reviewed work, AI was used to generate inferences from data acquired at the neonatal stage to diagnose NEC. AI can be used before conception and during the perinatal and neonatal stages. ML algorithms can help predict the success rate of in vitro fertilisation treatment and outcomes,<sup>25,26</sup> as well as live births from embryo data.<sup>27</sup> Using images of embryos, AI can increase the chances of successful implantation and development into

pregnancy<sup>28</sup> during the selection process. In the perinatal period, AI solutions help improve maternal and foetal care, thereby increasing the chances of successful delivery. The most common solutions are preterm birth risk assessment,<sup>29–31</sup> foetal biometry measurement,<sup>32–34</sup> foetal heart disease detection,<sup>35–37</sup> and computer-assisted fetoscopic surgical treatment.<sup>38,39</sup> In NICUs, AI is useful for monitoring patient vital signs and has been shown to predict life-threatening situations, including birth asphyxia, seizures, respiratory distress symptoms, or sepsis.<sup>40,41</sup>

NEC has an unknown aetiology, and factors present in the preconception, perinatal, and neonatal stages leading to the development of NEC in newborns may contribute to the disease. As such, an AI model capable of accurately diagnosing NEC may need to consider a whole range of factors and various types of retrospective data from all stages. As described in the previous paragraph, there are efforts to construct AI models applied to different tasks, and the idea of combining all those models and using the data on the entire spectrum leading to NEC may be a future solution to create a robust AI system for NEC diagnosis. However, we note that there is even more demand for data to create an AI model.

Recommendations from decision support systems are largely opaque, which means that recommendations are provided; however, they do not explain why. This is a serious limitation, especially when used in high-stake decision making, such as in medicine. Considering that there is an AI tool diagnosing NEC in newborn children and evaluating whether a specific child with severe symptoms does not have to undergo the surgery, how can a physician reconcile such a recommendation if he/she is of a different opinion? This is especially difficult if a recommendation comes without an explanation.<sup>42</sup>

Explaining the complex computations behind AI recommendations is a subject of worldwide research efforts and will be important for successful AI solutions used for NEC diagnosis. A new research area devoted to reliable AI systems was created and called explainable artificial intelligence (XAI).<sup>43</sup> XAI focuses on developing new algorithms or DL architectures that would help understand the decisions made by AI systems or generate a proper explanation. For instance, gradient-weighted class activation mapping<sup>44</sup> is a method for generating explanations (heatmaps) that can be superimposed over images analysed by AI. If AI provides an assessment of AR for patterns of PI or others consistent with NEC, it highlights an image region where these structures are present. Other algorithms have also been designed to explain these decisions. These include model-agnostic explanations,<sup>45</sup> which use linear relationships to simplify more complex models, or Shapley additive explanations, which generate explanations based on the game theory.<sup>46</sup> Unfortunately, as of 2022, all XAI methods have significant drawbacks and cannot be sufficiently generalised to make them more widely used, particularly in medicine.

There are opportunities to utilise AI decision support systems for the diagnosis and treatment of NEC. Improvements in the interpretation of the available data during the diagnostic process are a natural avenue for using AI, which is also reflected in the literature. NEC diagnosis is time and resource consuming and is often performed with AR imaging that uses ionising radiation. Non-ionising imaging using US is an optional imaging technique for NEC with no harmful side effects. The value of US for NEC diagnosis has been widely recognised in recent years.<sup>47</sup> However, it is difficult to perform and interpret, and AI may help physicians in performing high-quality examinations as well as in interpreting US images.

AI has the potential to be invaluable, especially for clinical practice with little experience in NEC diagnosis. The papers reviewed here demonstrate the potential of this approach. However, the road to clinical implementation is unclear, and no studies have documented successful AI implementation in a

clinical setting. We found only eight relevant papers on the topic of this review, which is an important limitation of this study. Drawing strong conclusions about trends in the literature based on such a small sample size is impossible.

## CONCLUSIONS

In this narrative review, we present the currently available literature on the use of AI and ML to diagnose NEC in newborns. Only a small number of publications relevant to this topic were found. We recognise that there is a substantial need for further research to fill this gap. AI, especially DL, has the potential to improve NEC diagnosis and provide predictions of treatment outcomes, as shown by reviewed work; however, no literature exists showing its clinical impact. We emphasise that the opaque predictions of DL models (black-box predictions) and the lack of large multi-institutional datasets evident from the review will hinder and slow down the development and implementation of clinical AI systems for NEC diagnosis in the near future.

## REFERENCES

- Kuzma-O'Reilly, B. et al. Evaluation, development, and implementation of potentially better practices in neonatal intensive care nutrition. *Pediatrics* **111**, e461–e470 (2003).
- Malin, S. W., Bhutani, V. K., Ritchie, W. W., Hall, M. L. & Paul, D. Echogenic intravascular and hepatic microbubbles associated with necrotizing enterocolitis. *J. Pediatr.* **103**, 637–640 (1983).
- Neu, J. & Walker, W. A. Necrotizing enterocolitis. *N. Engl. J. Med.* **364**, 255–264 (2011).
- Epelman, M. et al. Necrotizing enterocolitis: review of state-of-the-art imaging findings with pathologic correlation. *RadioGraphics* **27**, 285–305 (2007).
- Walsh, M. C. & Kliegman, R. M. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr. Clin. North Am.* **33**, 179–201 (1986).
- Merritt, C., Goldsmith, J. & Sharp, M. Sonographic detection of portal venous gas in infants with necrotizing enterocolitis. *Am. J. Roentgenol.* **143**, 1059–1062 (1984).
- Leonidas, J. C., Hall, R. T. & Amoury, R. A. Critical evaluation of the roentgen signs of neonatal necrotizing enterocolitis. *Ann. Radio. (Paris)* **19**, 123–132 (1976).
- Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med.* **28**, 31–38 (2022).
- Sitek, A. et al. Artificial intelligence in PET. *PET Clin.* **16**, 483–492 (2021).
- Soofi, A. A. & Awan, A. Classification techniques in machine learning: applications and issues. *J. Basic Appl. Sci.* **13**, 459–465 (2017).
- Wiering, M. & van Otterlo, M. (eds) *Reinforcement Learning* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, Accessed 28 March 2022). <http://link.springer.com/10.1007/978-3-642-27645-3>.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems* (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) (Curran Associates, Inc., 2012). <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.
- Vaswani, A. et al. Attention is all you need. In: *Advances in Neural Information Processing Systems* (eds Guyon, I. et al.) (Curran Associates, Inc., 2017). <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fdb053c1c4a845aa-Paper.pdf>.
- Mueller, M., Taylor, S. N., Wagner, C. L. & Almeida, J. S. Using an artificial neural network to predict necrotizing enterocolitis in premature infants. In: *2009 International Joint Conference on Neural Networks* 2172–2175 (IEEE, Atlanta, GA, USA, 2009, Accessed 28 March 2022). <http://ieeexplore.ieee.org/document/5178635/>.
- Ntonfo, G. M. K., Frize, M. & Barciak, E. Detection of necrotizing enterocolitis in newborns using abdominal thermal signature analysis. In: *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings* 36–39 (IEEE, Torino, Italy, 2015, Accessed 28 March 2022). <http://ieeexplore.ieee.org/document/7145168/>.
- Irls, C. et al. Estimation of neonatal intestinal perforation associated with necrotizing enterocolitis by machine learning reveals new key factors. *Int. J. Environ. Res. Public Health* **15**, 2509 (2018).
- Lure, A. C. 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.* **56**, 1703–1710 (2021).
- Lueschow, S. R., Boly, T. J., Jasper, E., Patel, R. M. & McElroy, S. J. A critical evaluation of current definitions of necrotizing enterocolitis. *Pediatr. Res.* **91**, 590–597 (2022).
- Bell, M. J. et al. Neonatal necrotizing enterocolitis: therapeutic decisions based upon clinical staging. *Ann. Surg.* **187**, 1–7 (1978).
- Battersby, C., Longford, N., Costeloe, K. & Modi, N., for the UK Neonatal Collaborative Necrotising Enterocolitis Study Group. Development of a gestational age-specific case definition for neonatal necrotizing enterocolitis. *JAMA Pediatr.* **171**, 256 (2017).
- Gephart, S. M. et al. Changing the paradigm of defining, detecting, and diagnosing NEC: perspectives on Bell's stages and biomarkers for NEC. *Semin. Pediatr. Surg.* **27**, 3–10 (2018).
- van Druten, J., Sharif, M. S., Chan, S. S., Chong, C. & Abdalla, H. A deep learning based suggested model to detect necrotizing enterocolitis in abdominal radiography images. In: *2019 International Conference on Computing, Electronics & Communications Engineering (ICCECE)* 118–123 (IEEE, London, United Kingdom, 2019, Accessed 28 March 2022). <https://ieeexplore.ieee.org/document/8941615/>.
- Gao, W. et al. Multimodal AI system for the rapid diagnosis and surgical prediction of necrotizing enterocolitis. *IEEE Access* **9**, 51050–51064 (2021).
- Lin, Y. C., Salleb-Aouissi, A. & Hooven, T. A. Interpretable prediction of necrotizing enterocolitis from machine learning analysis of premature infant stool microbiota. *BMC Bioinforma.* **23**, 104 (2022).
- Endo, S. et al. Association of maternal factors with perinatal complications in pregnancies complicated with diabetes: a single-center retrospective analysis. *J. Clin. Med.* **7**, 5 (2018).
- Liu, B. et al. Predicting pregnancy using large-scale data from a women's health tracking mobile application. In: *The World Wide Web Conference on – WWW '19* 2999–3005 (ACM Press, San Francisco, CA, USA, 2019, Accessed 28 March 2022). <http://dl.acm.org/citation.cfm?doid=3308558.3313512>.
- Goyal, A., Kuchana, M. & Ayyagari, K. P. R. Machine learning predicts live-birth occurrence before in-vitro fertilization treatment. *Sci. Rep.* **10**, 20925 (2020).
- Khosravi, P. et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *Npj Digit. Med.* **2**, 21 (2019).
- Gao, C. et al. Deep learning predicts extreme preterm birth from electronic health records. *J. Biomed. Inf.* **100**, 103334 (2019).
- Włodarczyk, T. et al. Spontaneous preterm birth prediction using convolutional neural networks. In: *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis* (eds Hu, Y. et al.) 274–283 (Springer International Publishing, Cham, 2020, Accessed 28 March 2022). [https://link.springer.com/10.1007/978-3-030-60334-2\\_27](https://link.springer.com/10.1007/978-3-030-60334-2_27).
- Włodarczyk, T. et al. Machine learning methods for preterm birth prediction: a review. *Electronics* **10**, 586 (2021).
- Plotka, S. et al. FetalNet: multi-task deep learning framework for fetal ultrasound biometric measurements. In: *Neural Information Processing* (eds Mantoro, T., Lee, M., Ayu, M. A., Wong, K. W. & Hidayanto, A. N.) 257–265 (Springer International Publishing, Cham, 2021, Accessed 28 March 2022). [https://link.springer.com/10.1007/978-3-030-92310-5\\_30](https://link.springer.com/10.1007/978-3-030-92310-5_30).
- Plotka, S. et al. Deep learning fetal ultrasound video model match human observers in biometric measurements. *Phys. Med. Biol.* **67**, 045013 (2022).
- Bano, S. et al. AutoFB: automating fetal biometry estimation from standard ultrasound planes. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (eds de Bruijne, M. et al.) 228–238 (Springer International Publishing, Cham, 2021, Accessed 28 March 2022). [https://link.springer.com/10.1007/978-3-030-87234-2\\_22](https://link.springer.com/10.1007/978-3-030-87234-2_22).
- Budd, S. et al. Detecting hypo-plastic left heart syndrome in fetal ultrasound via disease-specific Atlas maps. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (eds de Bruijne, M. et al.) 207–217 (Springer International Publishing, Cham, 2021, Accessed 28 March 2022). [https://link.springer.com/10.1007/978-3-030-87234-2\\_20](https://link.springer.com/10.1007/978-3-030-87234-2_20).
- Zhao, Z. et al. DeepFHR: intelligent prediction of fetal Acidemia using fetal heart rate signals based on convolutional neural network. *BMC Med. Inf. Decis. Mak.* **19**, 286 (2019).
- Chen, J. et al. EllipseNet: anchor-free ellipse detection for automatic cardiac biometrics in fetal echocardiography. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (eds de Bruijne, M. et al.) 218–227 (Springer International Publishing, Cham, 2021, Accessed 28 March 2022). [https://link.springer.com/10.1007/978-3-030-87234-2\\_21](https://link.springer.com/10.1007/978-3-030-87234-2_21).
- Casella, A. et al. A shape-constraint adversarial framework with instance-normalized spatio-temporal features for inter-fetal membrane segmentation. *Med. Image Anal.* **70**, 102008 (2021).
- Bano, S. et al. Deep placental vessel segmentation for fetoscopic mosaicking. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (eds Martel, A. L. et al.) 763–773 (Springer International Publishing, Cham, 2020, Accessed 28 March 2022). [https://link.springer.com/10.1007/978-3-030-59716-0\\_73](https://link.springer.com/10.1007/978-3-030-59716-0_73).

40. Adegboro, C. O., Choudhury, A., Asan, O. & Kelly, M. M. Artificial intelligence to improve health outcomes in the NICU and PICU: a systematic review. *Hosp. Pediatr.* **12**, 93–110 (2022).
41. Moccia, S., Migliorelli, L., Carnielli, V. & Frontoni, E. Preterm infants' pose estimation with spatio-temporal features. *IEEE Trans. Biomed. Eng.* **67**, 2370–2380 (2020).
42. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
43. Gunning, D. et al. XAI—explainable artificial intelligence. *Sci. Robot* **4**, eaay7120 (2019).
44. Selvaraju, R. R. et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *2017 IEEE International Conference on Computer Vision (ICCV)* 618–626 (IEEE, Venice, 2017, Accessed 5 April 2022). <http://ieeexplore.ieee.org/document/8237336/>.
45. Ribeiro, M. T., Singh, S. & Guestrin, C. "Why Should I Trust You?": explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135–1144 (ACM, San Francisco, California, USA, 2016, Accessed 5 April 2022). <https://dl.acm.org/doi/10.1145/2939672.2939778>.
46. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems* (eds Guyon, I. et al.) (Curran Associates, Inc., 2017). <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>.
47. van Druuten, J., Khashu, M., Chan, S. S., Sharif, S. & Abdalla, H. Abdominal ultrasound should become part of standard care for early diagnosis and management of necrotizing enterocolitis: a narrative review. *Arch. Dis. Child Fetal Neonatal Ed.* **104**, F551–F559 (2019).

## ACKNOWLEDGEMENTS

We would like to thank Editage ([www.editage.com](http://www.editage.com)) for English language editing.

## AUTHOR CONTRIBUTIONS

A.S. and J.S.-S. conceptualised and designed the review. K.W. performed literature search. All authors wrote and contributed to all sections. A.S. produced the final version and all authors approved the final version of the manuscript.

## FUNDING

This publication is partly supported by the European Union's Horizon 2020 research and innovation programme under grant agreement Sano No. 857533 and the International Research Agendas programme of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Arkadiusz Sitek or Joanna Seliga-Siwecka.

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

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## APPENDIX

Search strategy for narrative review.

### PUBMED

("artificial intelligence"[tiab] OR "artificial intelligences"[tiab] OR AI[tiab] OR "machine learning"[tiab] OR "machines learning"[tiab] OR "deep learning"[tiab] OR "decision tree"[tiab] OR "decision trees"[tiab] OR "neural network"[tiab] OR "neural networks"[tiab] OR "neural net"[tiab] OR "neural nets"[tiab] OR "medical image processing"[tiab] OR "Artificial Intelligence"[Mesh]) AND ("necrotizing enterocolitis"[tiab] OR NEC[tiab] OR "Enterocolitis, Necrotizing"[Mesh]).

### EMBASE

('artificial intelligence':ti,ab,kw OR 'artificial intelligences':ti,ab,kw OR ai:ti,ab,kw OR 'machine learning':ti,ab,kw OR 'machines learning':ti,ab,kw OR 'deep learning':ti,ab,kw OR 'decision tree':ti,ab,kw OR 'decision trees':ti,ab,kw OR 'neural network':ti,ab,kw OR 'neural networks':ti,ab,kw OR 'neural net':ti,ab,kw OR 'neural nets':ti,ab,kw OR 'medical image processing':ti,ab,kw OR 'artificial intelligence'/exp) AND ('necrotizing enterocolitis':ti,ab,kw OR nec:ti,ab,kw OR 'necrotizing enterocolitis'/exp).

### ARXIV

"necrotizing enterocolitis".

### IEEE Xplore

("artificial intelligence" OR "artificial intelligences" OR AI OR "machine learning" OR "machines learning" OR "deep learning" OR "decision tree" OR "decision trees" OR "neural network" OR "neural networks" OR "neural net" OR "neural nets" OR "medical image processing") AND "necrotizing enterocolitis".