Pediatrics in Artificial Intelligence Era: A Systematic Review on Challenges, Opportunities, and Explainability

Abstract

Background: The emergence of Artificial Intelligence (AI) tools such as ChatGPT and Bard is disrupting a broad swathe of fields, including medicine. In paediatric medicine, AI is also increasingly being used across multiple subspecialties. However, the practical application of AI still faces a number of key challenges. Consequently, there is a requirement for a concise overview of the roles of AI across the multiple domains of paediatric medicine that the current study seeks to address.

Aim: To systematically assess the challenges, opportunities, and explainability of AI in paediatric medicine.

Methodology: A systematic search was carried out on peer-reviewed databases, PubMed Central, Europe PubMed Central, and grey literature using search terms related to Machine Learning (ML) and AI for the years 2016 to 2022 in the English language. A total of 210 articles were retrieved that were screened with PRISMA for abstract, year, language, context, and proximal relevance to research aims. A thematic analysis was carried out to extract findings from the included studies.

Results: Twenty articles were selected for data abstraction and analysis, with three consistent themes emerging from these articles. In particular, eleven articles address the current state-of-the-art application of AI in diagnosing and predicting health conditions such as behavioural and mental health, cancer, syndromic and metabolic diseases. Five articles highlight the specific challenges of AI deployment in paediatric medicines: data security, handling, authentication, and validation. Four articles set out future opportunities for AI to be adapted: the incorporation of Big Data, cloud computing, precision medicine, and clinical decision support systems. These studies collectively critically evaluate the potential of AI in overcoming current barriers to adoption.

Conclusion: AI is proving disruptive within paediatric medicine and is presently associated with challenges, opportunities, and the need for explainability. AI should be viewed as a tool to enhance and support clinical decision-making rather than a substitute for human judgement and expertise. Future research should consequently focus on obtaining comprehensive data to ensure the generalisability of research findings.

1 Introduction

Science and technology have made significant advancements with the introduction of Artificial Intelligence (AI), and Machine Learning (ML) has been a game-changer. ML has enabled computers to learn without explicit programming by combining computer science and statistics [1]. ML has gained momentum in many fields, including healthcare, thanks to emerging tools like ChatGPT [1], Bard [2] and Glass AI 2.0 [3]. These tools are transforming industries by enabling conversations between humans and machines. ChatGPT, a large language model (LLM), has immense potential to assist in healthcare, including helping patients with mental health issues and aiding healthcare providers in decision-making [2, 3]. Recently, Glass Health introduced Glass AI 2.0, a similar LLM to ChatGPT, but with a clinical knowledge database created and maintained by clinicians to generate differential diagnoses and clinical plan outputs [4].

The integration and scope of such AI tools in healthcare are growing rapidly. Paediatrics is a field with practical challenges like complex comorbidities, increasing emergency admissions, and a lack of access to paediatric care providers, which could hinder the provision of quality and timely care [5]. ML implementations can streamline the paediatric workforce and assist clinical decision-making by enabling physicians to focus on patient-centered care plans by making better use of their clinical knowledge and time [1]. ML techniques can analyse vast datasets and create predictive models that go beyond human cognitive capabilities.

Although technological advancements are expanding the integration and scope of ML in paediatrics, there are challenges associated with the implications of AI, such as unintentional bias from data, like racial segregation and underperforming algorithms, which could jeopardise patient care [6]. To mitigate these issues, it may be better to focus on collective human-support-AI systems instead of complete automation. Additionally, ML could facilitate medical training or provide evidence-based care to patients using AI-based web or mobile applications with the help of human-in-the-loop systems. Therefore, it is crucial to evaluate the explainability of AI models, potential opportunities, and challenges when integrating ML in healthcare, especially for the paediatric population.

[1] https://chat.openai.com/
[2] https://bard.google.com/
[3] https://glass.health/ai

2 Materials and Methods

The field of artificial intelligence (AI) is rapidly growing and evolving, with a variety of buzzwords and terms that can be confusing to navigate. In order to clarify these terms and their relationship to one another, this section provides definitions of key buzzwords related to AI in Table 1. Additionally, a visual representation of the relationship among these buzzwords is presented in Figure 1. By establishing a clear understanding of these terms, the subsequent methodology can be more easily understood and applied.

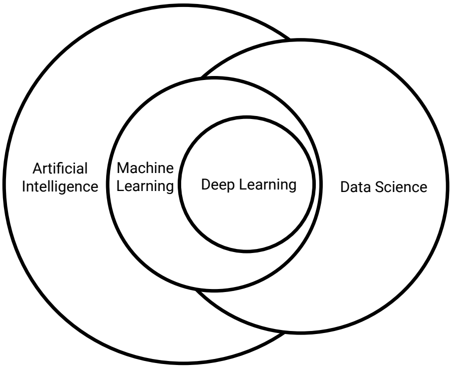

Fig. 1 Relationship between AI, ML, DL and Data science. AI is the creation of computer systems that can perform tasks that typically require human intelligence. Machine learning (ML) is a branch of AI that enables computers to learn from data without being programmed explicitly. Deep learning (DL) is a subset of ML that focuses on neural network algorithms. Lastly, data science deals with everything related to data, including collecting, cleaning, analysing and interpreting it. All of these fields are interconnected and work together to advance our understanding and use of AI.

Since this study focuses on machine learning in paediatrics; for this purpose, we employ a qualitative approach, expressing research findings and interpretations in terms of non-numeric data. The research thus uses secondary analysis to constitute an evidence-based literature review and associated analysis.

2.1 Search Strategy and Keywords

In this review, the focus is on assessing the challenges and opportunities of machine learning in paediatrics. The search was performed through authentic databases to get the relevant information. The search strategy was based on using a set of keywords and Boolean operators (e.g., using *OR* & *AND*).

Keywords that were used during the search for the desired topic that is challenges and opportunities of machine learning in paediatrics, are: “machine learning AND paediatrics”, “machine learning”, “challenges faced during paediatrics care AND technology”, and “significance of machine learning AND paediatrics care”, “paediatrics AND machine learning history”, “machine learning AND future in paediatrics care”.

Buzzword	Definition
Artificial Intelligence (AI)	This term refers to the creation of intelligent systems that simulate human thinking and behaviour. AI systems can be designed to perform tasks such as speech recognition, decision-making, and problem-solving.
Statistics	Involves making inferences about a population based on a sample of data. It can be used to make predictions and identify patterns within data.
Machine Learning (ML)	A subset of AI, ML involves designing algorithms that can learn from data without being explicitly programmed. It involves finding patterns within data that can be used to make predictions.
Deep Learning (DL)	A subset of ML, DL uses neural network-based methods to generate repeatable predictions by finding patterns within data. It is particularly useful for complex data sets and can be used for tasks such as language, image and speech recognition.
Data Science (DS)	This is the study of data, which involves data preparation, transformation, and analytics. Data scientists use a variety of techniques and tools to make sense of large data sets and to identify patterns that can be used for decision-making.

Table 1 Simplified definitions of the most commonly used AI buzzwords

2.2 Databases, Data Extraction and Selection Criteria

The databases that were used during the searching process include PubMed Central[4] and Europe PubMed Central[5] [7]. PRISMA guidelines for systematic reviews were used [8] for the data extraction process. The selection criterion defining the inclusion of relevant information is peer-reviewed journals from scholarly databases, the English publication language, the timeline of the last 7 years (2016-2022), and the information relevant to the research objectives. The aforementioned criteria are considered while searching for relevant data.

2.3 Data Collection

In this study, for data collection, Preferred Systematic Reviews and Meta-Analysis (PRISMA) guidelines were adopted as shown in Figure 2. The data was collected as per the inclusion and exclusion criteria guidelines in Section 2.2. Initial search results retrieved were N=210 searches (this included scholarly articles and, initially, grey literature). Later, primary screening was performed, and duplicated results were removed, leaving N=86 articles. These were then passed down to secondary screening, where exclusion based on

[4] https://www.ncbi.nlm.nih.gov/pmc/
[5] https://europepmc.org/

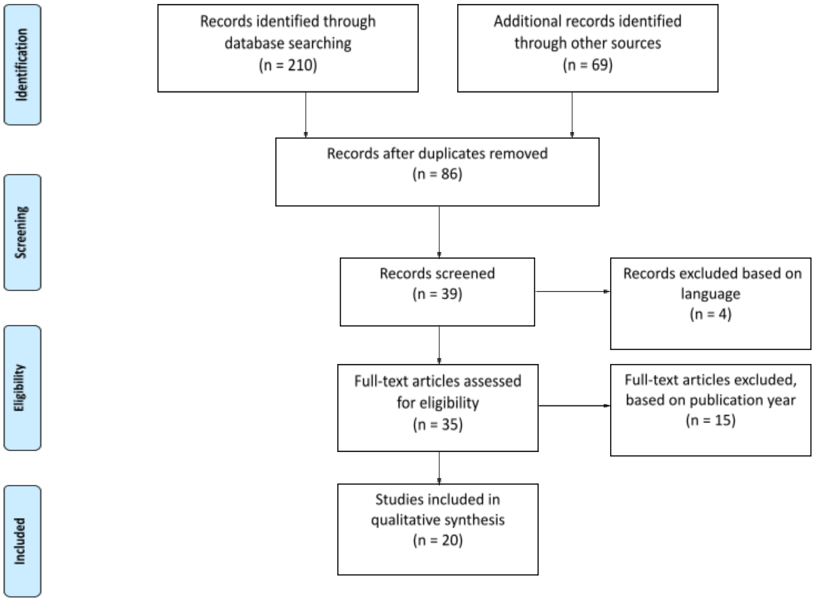

Fig. 2 PRISMA flowchart showing the data extraction and selection criteria.

title/abstracts, language, and timeline was performed, resulting in N=39 articles. Finally, the content analysis was performed for quality check, and N=20 articles were then selected for the review discussion.

3 Results & Discussions

Three themes have been designed based on the achievable results that are (1) the current state-of-the-art functioning of ML algorithms in paediatric medicine, (2) the challenges of ML algorithm deployment in paediatric medicine, and (3) the future outlook of ML in paediatric medicine. These themes have been sequentially discussed to retain the understanding and integrity of the literary context. The data analysis of the themes is available in the supplementary material 4.

3.1 The Current State-Of-The-Art Functioning of ML Algorithm in Paediatric Medicine

There are many sub-speciality areas in paediatrics, including neonatology, paediatric endocrinology, paediatric emergency, nephrology, neurology, rheumatology, ophthalmology, behavioural medicine, respiratory medicine and many more. ML has integrated into areas where some showed profound outcomes, and others needed improvements. The section will detail the current perspectives on ML application in paediatric medicine.

3.1.1 PTSD Diagnosis

Several studies have investigated the use of deep learning models for neuroimaging to classify PTSD in children who have experienced natural disasters [9, 10]. Ge and colleagues [9] found that property loss and lifestyle deterioration were the most probable variables for predicting PTSD using machine learning algorithms. PTSD is a lasting dysfunctional condition in children, and the ability of ML to perform predictive classification has important implications for early intervention and treatment.

3.1.2 Cancer Diagnosis

ML has also been applied to cancer diagnosis in paediatric care. Fathi *et al.*[11] used the neuro-fuzzy inference system (NFIS) as a DL model for diagnosing paediatric leukemia patients, with the system predicting cases by extracting information from a patient's neutrophil count from blood test records. Neuro-fuzzy has been investigated as a prognostic tool for cancer detection in children, where the prognosis is particularly important, and is therefore highly demanded.

3.1.3 Metabolic Condition Diagnosis

ML has shown promise in detecting genetic metabolic conditions. Zhu *et al.* [12] discussed that even though metabolic data of children are noisy (complex) in a clinical setting, the ML significantly extracts some useful findings from metabolic data to screen PKU in children without a false-positive diagnosis. Moreover, PKU detection in newborns and susceptibility was also reported to be diagnosed with ML models. Rare Disease Auxiliary Diagnosis system (RDAD) is one of the examples of detecting rare phenotype-based metabolic disorders in children [13].

3.1.4 Eye Disorder Diagnosis

ML has also made contributions to ophthalmic medicine, such as predicting myopia development in school-aged children [14, 15]. Lin and colleagues [14] observed the real-time clinical refraction data by applying ML models to predict myopia development, where acceptable prediction among children was found. Another study used a regression model to predict childhood myopia in Chinese children. At a 95% confidence interval (CI), a suitable diagnostic accuracy was reported [15]. In addition, the study also analysed the impact of factors on atropine-treated myopia that control intraocular pressure in children. Different AI models have been used to detect factors that can enhance myopia control by optimising atropine use [16].

3.1.5 Behavioral Disorders

Artificial Intelligence (AI) and Machine Learning (ML) have the potential to revolutionize analysis of behavioural problems in paediatrics, including autism spectrum disorder (ASD), conduct disorder (CD) and attention deficit hyperactive disorder (ADHD) [17, 18]. These conditions lead to the disruptive behaviours of the children. By analysing disruptive behaviours, AI and ML can identify patterns and correlations, which can lead to a more accurate diagnosis of behavioural problems and help in the development of more effective treatment plans [18]. Moreover, AI can also analyse speech patterns and behaviour in children with ASD and identify specific markers that are indicative of ASD. This information can then be used to develop more personalised treatment plans that are tailored to the individual child's needs [19].

3.1.6 Abuse Analysis

Child abuse is an unaddressed public health challenge but causes mental disturbances in children and PTSD [20]. A study proposed the concept that the development of a convolutional neural network (CNN) can facilitate the detection of childhood sexual abuse. ML can perform self-figure drawings to be used as a comparative figure with drawings of non-victims to identify cases of maltreatment in children [21].

Amrit *et al.* [22] applied ML models to cases of abuse in children in the Netherlands using the data extracted from child specialities. They characteristically converted unstructured clinical notes into structured data and used a classification algorithm to characterize the abuse cases. Literature also studied the effects of ML on child abuse detection. ML has progressed in the predictive analysis of childhood abuse cases.

3.1.7 Improving PICU Efficiency

In recent years, the use of AI/ML technologies in paediatric intensive care units (PICU) has greatly advanced patient outcomes for children with severe illnesses. As a result, PICU death rates have significantly decreased, with some studies reporting rates as low as 1 to 2% [23].

3.2 Challenges of ML Deployment in Paediatric Medicine

The key challenges to deploying AI in clinical practice are the barriers preventing AI's clinical translation. The safety and timely transformation into practice, accuracy, algorithmic biases, data biases, brittleness, and irregular interpretability are some barriers that affect the AI-based service delivery in clinical settings [24].

3.2.1 Irregularity in Temporality

Clinical patient data are often not recorded at regular intervals causing irregularity. For example, blood pressure samples are only collected as a patient visits his/her clinic, as and when necessary, or else during regular appointments. Irregularities or missing samples are inevitable because a patient can be absent from the appointment, or a clinician may cancel or reschedule the appointment [25]. [26] proposed two important challenges of temporality and irregularity. According to their findings, clinical EHR comprises short- and long-term records for patient health trajectories. However, clinical complexity is present in long-term cases, which may interfere with the interpretation and ML-based structuration of clinical data. [26] showed that the data density of clinical records varies among patients and can produce irregular samples for testing[26].

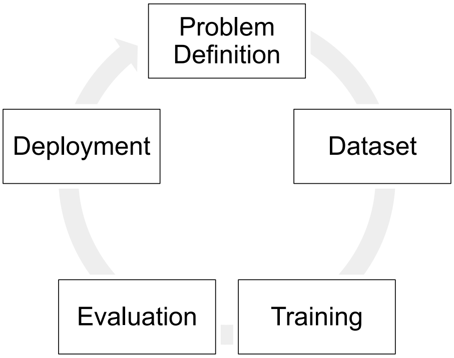

Fig. 3 The ML life cycle is a continuous process that requires ongoing training and adaptation of models to accommodate changes in data and context. This process starts with defining the problem to be solved using ML, followed by gathering and cleaning relevant data. Exploratory analysis provides a basic understanding of the data and helps identify potential biases in the dataset. The next step is to build a model by training and evaluating the dataset, which is a critical phase in the cycle. As new data becomes available, the model needs to be retrained to ensure its accuracy and effectiveness. Additionally, experiential learning is essential in the healthcare industry, where new medicines and treatments are continually being introduced. As such, ML models need to be continuously trained and adapted to incorporate the latest medical knowledge. Finally, the model is deployed for inference/predictions, and its efficiency is continually monitored and improved by adding more data to mitigate bias. In summary, the ML life cycle is an ongoing process that requires continuous training and adaptation to accommodate changes in data and context and to incorporate the latest medical knowledge.

3.2.2 Biases in the Diagnostic Outcomes

The data used to train AI/ML models must be diverse and inclusive of different racial and ethnic groups to avoid biases in the diagnostic outcomes [24]. The limited amount of data available for conditions, such as chromosomal anomalies and genetic malformations, can affect the accuracy of AI/ML models. Approximately 80% of rare diseases (RD) are hereditary in nature, and 75% of them afflict children. Although each case is uncommon, RDs are thought to impact 350 million individuals worldwide cumulatively. Furthermore, the complexity of some syndromic conditions, such as Down Syndrome and Turner Syndrome and the variability in their presentations can make it difficult for AI/ML models to diagnose these conditions accurately. So, the lack of standardisation in the data collected for diagnosis and the use of different diagnostic tools by different healthcare providers can make it challenging for algorithmic models to diagnose syndromic conditions effectively. Moreover, the interpretation and implementation of AI/ML diagnostic results in the clinical setting require careful consideration of ethical and legal considerations, such as patient privacy and informed consent.

3.2.3 Bias and Quality of Data

There is also the risk of bias in AI and ML algorithms. If the algorithms are trained on biased data, they can perpetuate that bias in their predictions and decisions. Especially as PICUs often collect large amounts of patient data, which may be fragmented, inconsistent, or of poor quality, can impact the accuracy of AI and ML algorithms, which can have serious consequences for critically ill children. Integrating AI and ML into PICUs also requires collaboration between healthcare providers, AI and ML experts, and patient representatives to ensure that the technology is used ethically and meets the needs of patients and families [6].

3.2.4 Lack of Variety in Data

Understanding behavioural problems in paediatrics is complex, and AI/machine learning can aid in this process. However, several challenges need to be addressed. Firstly, the data used to train AI models is limited, as it is difficult to collect large amounts of data on children's disruptive behaviour, such as ADHD and CD. This can result in models that are not representative of the population and may lead to incorrect predictions. Moreover, children's behaviour constantly changes, and AI models must be updated frequently to reflect these changes (refer to Figure 3). Likewise, children's behaviour is influenced by many factors, including their environment, genetics, and development stage, making it difficult to predict their behaviour accurately. In addition, there is a risk of AI models reinforcing existing biases, which can lead to unfair treatment of children [18].

So, there is a need for interdisciplinary collaboration between AI researchers, paediatricians, and child psychologists to ensure that AI models are developed and used to benefit children. This requires a deep understanding of the complexities of child behaviour and the development of ethical and transparent AI models. Although AI has the potential to play a significant role in understanding behavioural problems in paediatrics, it is important to address the challenges mentioned above. With the right approach, AI can help paediatricians and child psychologists to better understand and treat children with behavioural problems of ADHD, CD and ASD, leading to improved outcomes for children and their families.

3.2.5 Record Duplication

Another significant barrier is record duplication which limits the use of AI in practice. Vogl [27] discussed this barrier and reported that record duplications are prevalent, especially in children's welfare and protection services and prevent processing structured data into AI-based DL methods. AI lacks unique identifiers that may differentiate duplicated records. AI itself merges duplicated data across large datasets where challenges for individual-level detection can be incurred. In addition, where records are inaccurately merged, additional challenges arise as a result. The risk of misattributed information is another risk of AI.

3.2.6 Interpretability of the ML Models

Another challenge is the lack of interpretability of AI and ML algorithms. These algorithms can make predictions based on vast amounts of data, but it can be difficult to understand how they arrived at a particular conclusion, making it difficult for healthcare providers to trust and use the technology effectively [24]. The interpretability and explainability challenge is discussed further in Section 3.4.

A recent study by McCartney *et al.* [28] commented on the practical challenges associated with AI. The example of the Babylon app was an important predictor of the problems. There are still some challenges in the NHS, which first developed the app for the purpose of a symptoms checker in paediatric patients. The problems are mainly associated with the ineffective evaluation of these clinical apps before commercialising. Babylon was not tested for the safest care and treatment. Therefore it was noticeable that technological advancements, specially AI-based models, should be tested for their proper accuracy before introducing them into the clinical practices. Babylon, an AI-driven diagnostic and triage system, claimed 100% accuracy of its app to check symptoms and evaluate results. However, the claim was later proved wrong. Such types of political and legal challenges became more problematic for introducing this AI-based technology.

Additionally, using AI and ML raises ethical concerns about patient privacy and data security. Healthcare providers must ensure that patient data is protected and used only for the purpose of providing care. Furthermore, Davendralingam *et al.* [29] highlighted challenges of data security issues, legal and ethical considerations, and issues with standardising clinical terms. All these challenges must be addressed in the future to improve AI implications in paediatrics.

3.3 Opportunities for ML algorithms in Paediatric Medicine

The previous section discussed the challenges associated with AI in paediatrics. This section will focus on the opportunities presented by AI in paediatric medicine.

Clarke *et al.*[1] found that AI has improved the precision in diagnosis, with ML models able to identify abnormal findings from normal clinical radiographs of paediatric patients[30]. Paediatric ophthalmology could also benefit from

AI in detecting retinopathy of prematurity (ROP) and congenital cataracts, taking into account unique aspects of designing AI applications that differ from adults [31]. Diseases such as asthma require stratification due to heterogeneity in disease severity and response to clinical treatment and trajectories, and ML can improve the classification of such diseases [32]. AI-based algorithms, such as Natural Language Processing (NLP), are also effective in drug development, identifying the most specific druggable targets [33].

One potential application of AI/ML is the prediction of patient deterioration, such as spinal muscular atrophy (SMA), which can help clinicians respond quickly to this potentially life-threatening condition. AI/ML can analyse large amounts of patient data, including vital signs, lab results, and medical history, and use this information to predict which patients are at risk of worsening [34]. Another opportunity for AI and ML in PICUs is in the personalised treatment of patients by analysing patients' data, medical history and current condition to develop individualised treatment plans. For example, AI and ML can be used to determine the most appropriate medications and doses based on a children's age, weight, and medical history.

Computational advancement helps track the progress of children with behavioural problems over time. AI can identify changes in behaviour and determine the effectiveness of different treatments by analysing data from multiple sources, such as medical records, behavioural assessments, and parent reports. This information can then be used to make informed decisions about future treatment plans and to adjust them as needed [35]. The sophisticated deep learning algorithms can analyse data from various sources, including genetic data, family history, and environmental factors; AI can therefore identify children who are at risk of developing behavioural problems of ADHD, CD and ASD. This information can then be used to provide early interventions and support.

AI/ML also has enormous potential in diagnosing syndromic conditions of Turner's syndrome, Down syndrome, Prader-Willi syndrome, and Angelman syndrome in paediatrics. AI processes vast amounts of genetic information about the aforementioned syndromes and identifies patterns that may not be immediately apparent to the human eye [36]. For example, AI can help identify specific patterns in genetic data that may indicate a particular syndrome, such as Down syndrome, allowing for a faster and more accurate diagnosis. This can be particularly useful in diagnosing rare or complex syndromes where traditional methods may be insufficient and give false negative or false positive results. By using AI, paediatricians can make more informed decisions about testing and treatment, leading to improved outcomes for children and their families. Another opportunity is using AI for image analysis, such as in diagnosing cranial facial syndromes. AI algorithms can analyse facial images and identify specific features that are indicative of a particular syndrome, allowing for a more accurate and objective diagnosis. This can be particularly useful in detecting syndromes early and improving treatment outcomes. In addition, AI can help reduce the time and resources required for diagnosis, allowing paediatricians to focus on providing care to children in need. By automating routine tasks, such as data collection and analysis, paediatricians can spend more time with patients and families, improving the overall quality of care.

Filipow [37] reported that ML can potentially diagnose (CRC) chronic respiratory conditions such as chronic obstructive pulmonary disease (COPD) and chronic airway obstruction (CAO). Shu *et al.* [38] discussed several applications of AI. However, it is crucial to improve the predictive architecture of algorithms to improve diagnostic outcomes and precision. In addition, it was well-observed in the literature that the AI model should be introduced in the clinical setting once its characteristic precision, diagnostic efficacy, and reliability are assessed. This would advise the future issues of Babylon apps that can adapt such practices of extensive validation to improve their usage.

3.4 Explainable AI (XAI)

Machine learning (ML) can provide a great benefit to paediatric medicine by assisting in the diagnosis, treatment, and management of diseases. However, it is also essential to explain the decision-making process to both doctors and patients to build trust and provide a better understanding of the reasoning behind a particular diagnosis or treatment recommendation. This need led to the development of Explainable Artificial Intelligence (XAI), which aims to make AI processes more transparent and interpretable. Deep learning algorithms, in particular, can benefit from XAI, as they learn every aspect of the decision-making process on their own, known as "neural weights" [39, 40].

There are two main approaches to XAI [40]: model-based and post-hoc explanation-based. Model-based methods make the ML model directly interpretable to medical practitioners, while post-hoc explanation-based methods translate the model into a more understandable format. The counterfactual explanation is an extension of XAI that can help identify and supplement diagnostic indicators. It allows medical professionals to ask "what if" questions, such as "how would this disease look in an adult?" or "how likely is this diagnosis if the patient were 10 years older?". These methods can be particularly useful in paediatrics, as they can help make connections to general medicine and supplement traditional deep learning methods.

3.5 Limitations

Despite the potential benefits of AI/ML in paediatric medicine and with the recent emergence of AI 2.0 tools such as ChatGPT, Bard, and GLASS A.I 2.0 potentially offering benefits for diagnosing, treating, and managing diseases; deploying these tools in practice has several limitations. One significant limitation is the context-specific nature of machine learning. In paediatric medicine, fewer samples may be available, which can lead to statistical noise in the ML predictions. This can compound other biases related to gender or demographics [41]. For example, when developing an ML model to predict hospital mortality in children admitted to the ICU, it is important to ensure that the model performs well on a validation cohort from a different ICU to demonstrate its ability to *generalise* in different contexts. Similarly, an ML model designed to analyse asthma in a specific population, such as white children, should be able to *generalise* effectively to other diverse populations as well. These examples highlight the importance of testing and validating ML models across different populations and contexts to ensure their reliability and applicability in diverse settings.

Another challenge in using ML in paediatric medicine is the lack of experiential learning with these models. Experiential learning allows clinicians to learn from their experiences and improve their decision-making over time. However, ML models are pre-determined and do not have the ability to learn from new experiences, which can lead to inaccurate predictions or an inability to adapt to changes in the patient's condition over time.

To address these limitations, researchers have developed various strategies, including the use of data augmentation to address category imbalance, interpretation and explanation-based ML methods, and monitoring the deployment context [42]. Additionally, collecting demographic data is crucial for mitigating biases in the data, especially given the potential for compounding biases related to gender and demographics in paediatric medicine [41]. Overall, while ML tools offer tremendous potential for improving paediatric medicine, their deployment comes with significant limitations that need to be addressed. Context-specific testing, validation, and monitoring are crucial to ensure that these tools are effective and accurate in different patient populations. Additionally, researchers must find ways to incorporate experiential learning into these models to ensure that they adapt and improve over time (Figure 3).

4 Conclusion

Many healthcare providers rely on clinical decision support tools in the electronic health record (EHR) for patient safety and improved outcomes. With the advancements in AI and ML, these tools have become even more crucial in the practice of medicine. In particular, AI has been integrated into many areas of paediatrics, such as predicting PTSD among children's survivors, early diagnosis of leukaemia, and detecting PKU cases with high accuracy. However, AI/ML deployment in paediatrics is still facing several challenges, such as data quality issues, unnoticed, hidden clinical variables, complex clinical data, and lack of clinical labels. To ensure the successful integration of AI/ML into paediatric healthcare, specific or generalised improvements in AI design and validation are necessary.

It is important to acknowledge that AI/ML models are context-specific and need testing in diverse populations to ensure their generalisability and identify any biases within the data. While AI tools like ChatGPT, Glass AI 2.0 could provide valuable virtual support to clinicians and parents by answering questions and providing information about symptoms, treatments, and medications, they are limited by the quality of the data they are trained on and may be biased based on the population they are tested on[6]. Additionally, machine learning models lack the ability to adapt to new data or new situations, which limits their potential for experiential learning. It is important to recognise these limitations and to highlight the importance of human judgement and expertise in the use of AI/ML in paediatric healthcare.

In conclusion, it is evident that AI/ML has the potential to significantly enhance patient outcomes in paediatric medicine. Although there are challenges to their deployment, advancements in AI design and validation, and testing in diverse populations can help to overcome these issues. Further research and practical support are recommended to explore areas not yet covered in the current literature. Ultimately, AI should be viewed as a tool to enhance and support clinical decision-making rather than a substitute for human judgement and expertise.

Acknowledgments. We would like to express our sincere gratitude to the anonymous reviewers whose insightful comments and constructive feedback significantly contributed to improving the quality of this paper. Their valuable suggestions and comments have helped us to refine our ideas and strengthen our arguments, and we are truly grateful for their time and effort in reviewing our work.

Competing interests. There is NO Competing Interest.

Supplementary Material. The data analysis is available at: https://github.com/tsantosh7/supplementary_material/blob/main/review_data_analysis.pdf

References

[1] Clarke, S. L., Parmesar, K., Saleem, M. A. & Ramanan, A. V. Future of machine learning in paediatrics. *Archives of Disease in Childhood* **107** (3), 223–228 (2022) .

[2] Biswas, S. S. Role of chat gpt in public health. *Annals of Biomedical Engineering* 1–2 (2023) .

[3] Cascella, M., Montomoli, J., Bellini, V. & Bignami, E. Evaluating the feasibility of chatgpt in healthcare: an analysis of multiple clinical and research scenarios. *Journal of Medical Systems* **47** (1), 1–5 (2023) .

[4] Abdullah, I. S., Loganathan, A. & Lee, R. W. Chatgpt doctors: The medical dream team. *URGENT Matters* (8), 1–5 (2023). URL https://hsrc.himmelfarb.gwu.edu/smhs_URGENT_Matters/8 .

[6] As ChatGPT was trained using data up until September 2021, it may lack the latest medical knowledge, treatments, and advancements that have been made since then.

[5] Jiang, F. *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology* **2** (4) (2017) .

[6] Knake, L. A. Artificial intelligence in pediatrics: the future is now. *Pediatric Research* 1–2 (2022) .

[7] Ferguson, C. *et al.* Europe pmc in 2020. *Nucleic acids research* **49** (D1), D1507–D1514 (2021) .

[8] Selçuk, A. A. A guide for systematic reviews: Prisma. *Turkish archives of otorhinolaryngology* **57** (1), 57 (2019) .

[9] Ge, F., Li, Y., Yuan, M., Zhang, J. & Zhang, W. Identifying predictors of probable posttraumatic stress disorder in children and adolescents with earthquake exposure: a longitudinal study using a machine learning approach. *Journal of affective disorders* **264**, 483–493 (2020) .

[10] Saxe, G. N., Ma, S., Ren, J. & Aliferis, C. Machine learning methods to predict child posttraumatic stress: a proof of concept study. *BMC psychiatry* **17** (1), 1–13 (2017) .

[11] Fathi, E., Rezaee, M. J., Tavakkoli-Moghaddam, R., Alizadeh, A. & Montazer, A. Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* **234** (10), 1051–1069 (2020) .

[12] Zhu, Z. *et al.* Improving the diagnosis of phenylketonuria by using a machine learning–based screening model of neonatal mrm data. *Frontiers in Molecular Biosciences* **7**, 115 (2020) .

[13] Jia, J. *et al.* Rdad: a machine learning system to support phenotype-based rare disease diagnosis. *Frontiers in genetics* **9**, 587 (2018) .

[14] Lin, H. *et al.* Prediction of myopia development among chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study. *PLoS medicine* **15** (11), e1002674 (2018) .

[15] Zhang, M. *et al.* Validating the accuracy of a model to predict the onset of myopia in children. *Investigative ophthalmology & visual science* **52** (8), 5836–5841 (2011) .

[16] Wu, T.-E. *et al.* Evaluating the effect of topical atropine use for myopia control on intraocular pressure by using machine learning. *Journal of Clinical Medicine* **10** (1), 111 (2020) .

[17] Ogundele, M. O. Behavioural and emotional disorders in childhood: A brief overview for paediatricians. *World journal of clinical pediatrics* **7** (1), 9 (2018) .

[18] Aylward, B. S. *et al.* An introduction to artificial intelligence in developmental and behavioral pediatrics. *J. Dev. Behav. Pediatr.* **44** (2), e126–e134 (2023) .

[19] Anagnostopoulou, P. *et al.* Artificial intelligence in autism assessment. *International Journal of Emerging Technologies in Learning (iJET)* **15** (6), 95–107 (2020) .

[20] Elankeerthana, R., Hanshiya, D., Pranavajothi, A., Shanmugapriya, N. & Sujitha, S. *Child abuse mental symptom prediction model using machine learning techniques*, 1459–1465 (IEEE, 2022).

[21] Kissos, L., Goldner, L., Butman, M., Eliyahu, N. & Lev-Wiesel, R. Can artificial intelligence achieve human-level performance? a pilot study of childhood sexual abuse detection in self-figure drawings. *Child Abuse & Neglect* **109**, 104755 (2020) .

[22] Amrit, C., Paauw, T., Aly, R. & Lavric, M. Identifying child abuse through text mining and machine learning. *Expert systems with applications* **88**, 402–418 (2017) .

[23] Gupta, P., Gossett, J. & Rettiganti, M. R. 60: Trends in mortality rates in pediatric intensive care units in the united states from 2004 to 2015. *Critical Care Medicine* **46** (1), 30 (2018) .

[24] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G. & King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine* **17** (1), 1–9 (2019) .

[25] Tirunagari, S., Bull, S. & Poh, N. *Automatic classification of irregularly sampled time series with unequal lengths: A case study on estimated glomerular filtration rate*, 1–6 (IEEE, 2016).

[26] Xiao, C., Choi, E. & Sun, J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association* **25** (10), 1419–1428 (2018) .

[27] M. Vogl, T. *Artificial intelligence and organizational memory in government: the experience of record duplication in the child welfare sector in canada*, 223–231 (2020).

[28] McCartney, M. Margaret McCartney: AI in medicine must be rigorously tested. *BMJ* **361** (2018). https://doi.org/10.1136/bmj.k1752 .

[29] Davendralingam, N., Sebire, N. J., Arthurs, O. J. & Shelmerdine, S. C. Artificial intelligence in paediatric radiology: future opportunities. *The British Journal of Radiology* **94** (1117), 20200975 (2021) .

[30] Chen, K.-C. *et al.* Diagnosis of common pulmonary diseases in children by x-ray images and deep learning. *Scientific reports* **10** (1), 1–9 (2020) .

[31] Reid, J. E. & Eaton, E. Artificial intelligence for pediatric ophthalmology. *Current opinion in ophthalmology* **30** (5), 337–346 (2019) .

[32] Deliu, M. *et al.* Longitudinal trajectories of severe wheeze exacerbations from infancy to school age and their association with early-life risk factors and late asthma outcomes. *Clinical & Experimental Allergy* **50** (3), 315–324 (2020) .

[33] Bravo, À., Piñero, J., Queralt-Rosinach, N., Rautschka, M. & Furlong, L. I. Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. *BMC bioinformatics* **16** (1), 1–17 (2015) .

[34] Ziegler, A. *et al.* Recommendations for gene therapy of spinal muscular atrophy with onasemnogene abeparvovec—avxs-101: Consensus paper of the german representatives of the society for pediatric neurology (gnp) and the german treatment centers with collaboration of the medical scientific advisory board of the german society for muscular diseases (dgm). *Der Nervenarzt* **91**, 518–529 (2020) .

[35] Hunt, X. *et al.* Artificial intelligence, big data, and mhealth: The frontiers of the prevention of violence against children. *Frontiers in artificial intelligence* **3**, 543305 (2020) .

[36] Tornese, G., Pellegrin, M. C., Barbi, E. & Ventura, A. Pediatric endocrinology through syndromes. *European journal of medical genetics* **63** (1), 103614 (2020) .

[37] Filipow, N. *et al.* Implementation of prognostic machine learning algorithms in paediatric chronic respiratory conditions: a scoping review. *BMJ Open Respiratory Research* **9** (1), e001165 (2022) .

[38] Shu, L.-Q., Sun, Y.-K., Tan, L.-H., Shu, Q. & Chang, A. C. Application of artificial intelligence in pediatrics: past, present and future. *World Journal of Pediatrics* **15** (2), 105–108 (2019) .

[39] Adadi, A. & Berrada, M. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access* **6**, 52138–52160 (2018) .

[40] Van Der Velden, B. H., Kuijf, H. J., Gilhuijs, K. G. & Viergever, M. A. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical Image Analysis* 102470 (2022) .

[41] Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366** (6464), 447–453 (2019) .

[42] Vokinger, K. N., Feuerriegel, S. & Kesselheim, A. S. Mitigating bias in machine learning for medicine. *Communications medicine* **1** (1), 25 (2021) .