

## REVIEW ARTICLE



# Artificial intelligence-based clinical decision support in pediatrics

Sriram Ramgopal<sup>1</sup>✉, L. Nelson Sanchez-Pinto<sup>2,3</sup>, Christopher M. Horvat<sup>4</sup>, Michael S. Carroll<sup>5</sup>, Yuan Luo<sup>3</sup> and Todd A. Florin<sup>1</sup>

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

Machine learning models may be integrated into clinical decision support (CDS) systems to identify children at risk of specific diagnoses or clinical deterioration to provide evidence-based recommendations. This use of artificial intelligence models in clinical decision support (AI-CDS) may have several advantages over traditional “rule-based” CDS models in pediatric care through increased model accuracy, with fewer false alerts and missed patients. AI-CDS tools must be appropriately developed, provide insight into the rationale behind decisions, be seamlessly integrated into care pathways, be intuitive to use, answer clinically relevant questions, respect the content expertise of the healthcare provider, and be scientifically sound. While numerous machine learning models have been reported in pediatric care, their integration into AI-CDS remains incompletely realized to date. Important challenges in the application of AI models in pediatric care include the relatively lower rates of clinically significant outcomes compared to adults, and the lack of sufficiently large datasets available necessary for the development of machine learning models. In this review article, we summarize key concepts related to AI-CDS, its current application to pediatric care, and its potential benefits and risks.

*Pediatric Research* (2023) 93:334–341; <https://doi.org/10.1038/s41390-022-02226-1>

**IMPACT:**

- The performance of clinical decision support may be enhanced by the utilization of machine learning-based algorithms to improve the predictive performance of underlying models.
- Artificial intelligence-based clinical decision support (AI-CDS) uses models that are experientially improved through training and are particularly well suited toward high-dimensional data.
- The application of AI-CDS toward pediatric care remains limited currently but represents an important area of future research.

**INTRODUCTION**

The mass storage of health data continues to transform healthcare delivery. To improve patient care, strategies that allow for the extraction of information and insights from clinical data are paramount. Computerized clinical decision support (CDS) systems provide a way to translate research findings to real-time interventions, promote quality improvement, and decrease variation in care.<sup>1</sup> CDS may be coupled with machine learning (ML) approaches, which use algorithms that leverage statistical methods to learn useful patterns from data. The resulting artificial intelligence (AI)-based CDS (AI-CDS) may allow for improved predictive performance to identify patients with a higher or lower likelihood of developing a disease, suffering a clinical deterioration, or who may benefit from a particular management strategy. AI-CDS represents an exciting opportunity to further improve care delivery. In this narrative review, we summarize key concepts related to AI-CDS: the principles and current state of CDS and AI, the role of AI-CDS in pediatric care, key concepts related to the

development and implementation of AI-CDS, the challenges associated with AI-CDS, and future steps in the field.

**“BIG DATA” IN HEALTHCARE**

While some definitions of “big data” refer to the size of the dataset (e.g., as too large to be stored or analyzed by conventional software solutions), this definition is complicated by the ever-growing capacity of computer systems to work with larger datasets.<sup>2</sup> Another definition by Gartner describes “big data” as “high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making, and process automation”.<sup>3</sup> While “volume” refers to the large size of the dataset, “velocity” refers to the need to be able to rapidly grow. “Variety” refers to the ability to use heterogeneous data, which in a healthcare context may include medical recordkeeping, billing, genomics, social determinants of health, patient-reported outcomes, and genomics. There are increasing efforts to leverage big data for purposes

<sup>1</sup>Division of Emergency Medicine, Ann & Robert H. Lurie Children’s Hospital of Chicago, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>2</sup>Division of Critical Care Medicine, Ann & Robert H. Lurie Children’s Hospital of Chicago, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>3</sup>Department of Preventive Medicine (Health and Biomedical Informatics), Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. <sup>4</sup>Department of Critical Care Medicine, UPMC Children’s Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. <sup>5</sup>Data Analytics and Reporting, Ann & Robert H. Lurie Children’s Hospital of Chicago, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ✉email: [sramgopal@luriechildrens.org](mailto:sramgopal@luriechildrens.org)

Received: 21 March 2022 Revised: 29 June 2022 Accepted: 18 July 2022

Published online: 29 July 2022

beyond recordkeeping to improve the delivery of patient care.<sup>4</sup> AI, defined as a field of study that focuses on how computers learn from data and the development of algorithms that make this learning possible, provides one important way to utilize these technologies to better improve practice.<sup>5</sup>

### CLINICAL DECISION SUPPORT

CDS may be defined as “computer systems designed to impact clinician decision making about individual patients at the point in time that these decisions are made.”<sup>6</sup> CDS generally encompasses three steps: (1) acquiring patient data, (2) summarizing data, and (3) suggesting an appropriate course of action.<sup>7</sup> CDS may include alerts, reminders, order sets, drug-dose calculations, care summary dashboards, and point-of-care information retrieval systems.<sup>8</sup> Within pediatrics, CDS has been used in a variety of applications, including traumatic brain injury (TBI),<sup>9,10</sup> asthma,<sup>11</sup> urinary tract infections (UTIs),<sup>12</sup> screening for developmental disorders,<sup>13</sup> ventilatory support,<sup>14,15</sup> and antibiotic selection.<sup>16</sup>

A systematic review evaluating the impact of CDS tools from 148 randomized controlled trials (primarily from adults) demonstrated that CDS improved outcomes when used for performing preventive services (odds ratio [OR], 1.42; 95% CI 1.27–1.58), ordering clinical studies (OR, 1.72; 95% CI, 1.47–2.00), and prescribing therapies (OR, 1.57; 95% CI 1.35–1.82) compared to care provided without CDS.<sup>8</sup> Few studies, however, have evaluated the balancing measures, including unintended consequences or adverse effects such as false negatives or increased physician workload burden.<sup>17</sup> In pediatrics, traditional rule-based CDS is routinely used to improve patient care but is frequently limited by poor model specificity, often resulting in false positive alerts.<sup>18</sup> These can result in physician dissatisfaction, contribute to burnout, and result in patient harm. One study of drug alerts in the primary care setting demonstrated a decline in the utilization of an alert-based CDS with additional reminders for the same patient.<sup>19</sup> In turn, ignored alerts from a CDS can result in adverse consequences. In one report from a pediatric intensive care unit, clinicians repeatedly overrode alerts in an EHR drug allergy alerting system resulting in the deterioration of a patient over time, emphasizing the potentially harmful effects that can arise from alarm fatigue.<sup>20</sup>

### AI AND PREDICTIVE MODELING

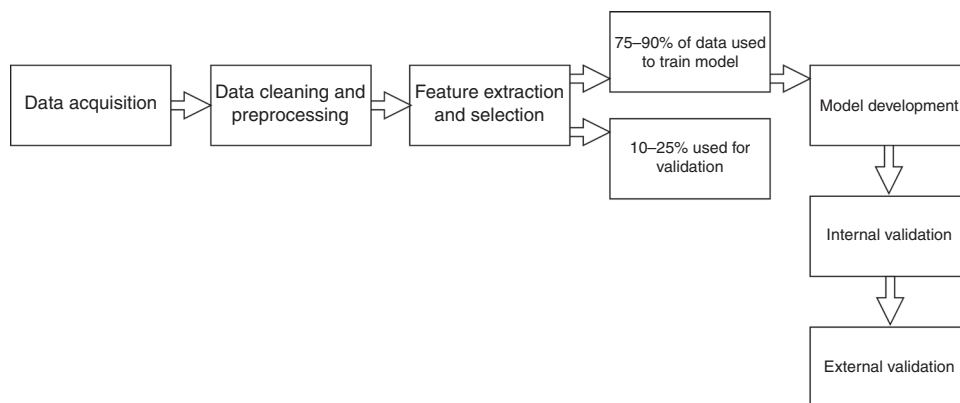
The use of AI may overcome some limitations attributed to traditional CDS. If perceived as ineffective or intrusive, users may ultimately view CDS as a burden toward effective patient care.<sup>21,22</sup>

In addition, CDS may result in inaccurate or poorly individualized recommendations. Prior qualitative work, for example, has cited important concerns with CDS, including inaccuracy of predictions leading to false positive and negative results.<sup>23</sup> AI-CDS represents an important evolution in the development of CDS models.

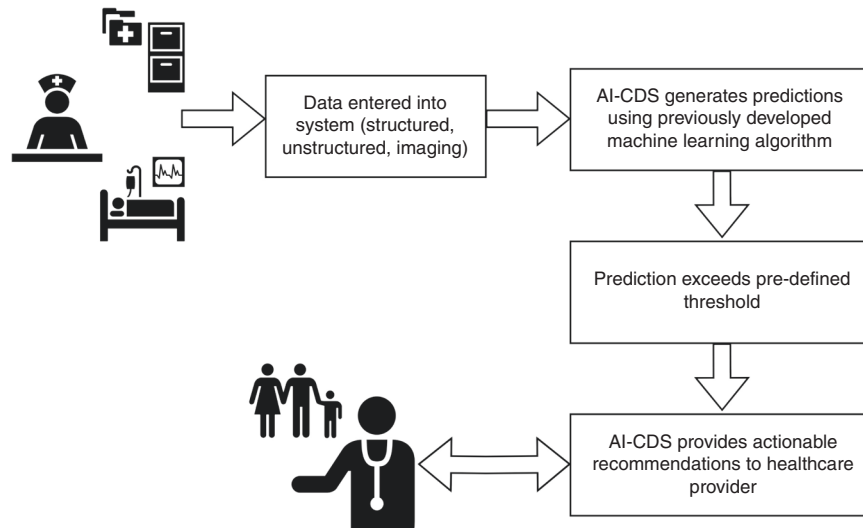
AI-CDS systems are sometimes called “non-knowledge” based AI, as they differ from the “if/then” rules that define rules-based (or “knowledge”) based CDS because the predictions are based on statistical or ML algorithms.<sup>24</sup> ML, considered a domain of AI, is a set of methods for inferring useful relationships in large datasets for which the assumptions and techniques of traditional statistics may be poorly defined. While the validity of traditional statistics comes from attempting to put conservative error bounds on inference, the validity of ML is generally defined with respect to out-of-sample predictive usefulness. ML has origins in statistics but evolved along an independent trajectory as computing power rapidly advanced, fostering the development of computational methods for the analysis of larger and larger datasets. Instead of traditionally used “hard-coded” algorithms or prediction rules, ML algorithms instead experientially improve through training, often entailing dimension reduction steps that particularly suit these methods to high-dimensional data.<sup>25,26</sup> Supervised ML is generally used for predictive modeling (e.g., likelihood of an event happening). In contrast, unsupervised ML is generally used to find natural groupings or clusters in the data (e.g., phenotypes of a disease or syndrome).<sup>5</sup>

In supervised ML, models undergo a period of training in which the ML algorithm is provided with input data in addition to the predefined outcome measure (such as the presence of a disease), to identify relationships between the outcome of interest and the features (or predictors) (Fig. 1).<sup>27</sup> Outcome data may be categorical (called “classification”), such as in-hospital mortality or intensive care unit hospitalization. Alternatively, a “regression” ML task is when a continuous outcome is required (such as hospital length of stay). Hundreds of algorithms for supervised ML have been described. Common methods used in the context of classification include logistic regression, k-nearest neighbors, naïve Bayes, decision trees, random forests, neural networks, and support vector machines. Other algorithms have been described which combine the strengths of base learners. Details of specific ML algorithms are summarized elsewhere.<sup>5</sup>

To date, most predictive models in children have been derived using more classical ML algorithms. For example, models to identify patients at risk of UTIs,<sup>12</sup> pneumonia,<sup>28</sup> bacterial meningitis,<sup>29</sup> clinically important TBI,<sup>9</sup> serious bacterial infections,<sup>30</sup> septic arthritis,<sup>31</sup> or intra-abdominal injury<sup>32</sup> have been derived using readily understandable logistic regression or decision trees. Such approaches can offer familiarity and interpretability to clinicians by



**Fig. 1 Development of a supervised machine learning algorithm.** Datasets frequently require cleaning and/or preprocessing (such as “one-hot” encoding of categorical variables). Following this, initial analyses are performed to identify distinct variables with the strongest association with the study outcome. A portion of the data may be used as a holdout cohort for internal validation. The remainder is used for model training. Following internal validation, the model may be tested in distinct datasets.



**Fig. 2** Functioning of an artificial intelligence clinical decision support (AI-CDS) tool. Electronic health data exists in a variety of formats, including structured (in discrete fields) or unstructured (such as in narrative notes). The machine learning algorithm may then be applied to these test data. When a desired threshold of disease probability is reached, a best practice alert may be provided to the treatment team.

design and have a track record of relatively robust predictive performance in some settings. When models perform similarly, the simpler, more applicable, and transparent model should be considered over the more complicated one.

Within the pediatric acute care setting ML models have been reported for a variety of pediatric indications. Some of these demonstrate superior potential compared to classic ML models. Some examples are provided below:

1. Prediction of children with clinically important TBI: investigators used optimal classification trees to identify children at risk of clinically important TBI and compared it to a frequently utilized model developed through classification and regression trees (CART).<sup>33</sup> The models demonstrated comparable sensitivity to the originally published CART model, but with higher specificity, potentially allowing for a reduction in unnecessary CT scans.
2. Emergency department testing: Singh et al. reported an ML model to predict the need for frequently performed clinical testing (urinary dipstick testing, electrocardiogram, abdominal ultrasonography, testicular ultrasonography, bilirubin level testing, and forearm radiographs) among children presenting to a pediatric emergency department.<sup>34</sup> Using an outcome of testing ordered in triage, the model demonstrated high area under the receiver operating characteristic curve (AUROC) (0.89–0.99, across individual use cases for clinical tests including urinary dipstick testing and electrocardiograms) with a high positive predictive value (0.77–0.94) with results available by a mean time of 165 min. The investigators were further able to characterize model explainability using Shapley Additive Explanation to ensure their clinical relevance.
3. Sepsis prediction: given its complexity with diagnosis, the presence of multiple overlapping diagnoses and the interaction of variables with each other, the early identification of pediatric sepsis may be one that is particularly well suited to an ML application to guide CDS. Several recent studies have demonstrated that ML methods may accurately identify children with sepsis.<sup>35–37</sup> One recent multicenter study that was trained to identify children with septic shock (defined as systolic hypotension and vasoactive use or  $\geq 30$  ml/kg of crystalloid administration within 24 h) using data limited to 2 h of presentation, for example,

demonstrated an AUROC of 0.83–0.85 between the two included validation datasets.<sup>37</sup> This study, which utilized least absolute shrinkage and selection operator with 10-fold cross-validation, demonstrated a specificity of 62% when using a pre-selected sensitivity of 90%. The model was trained on demographic data, chronic conditions, prior visit data, emergency department visit data, clinical (e.g., vital signs), and laboratory data.

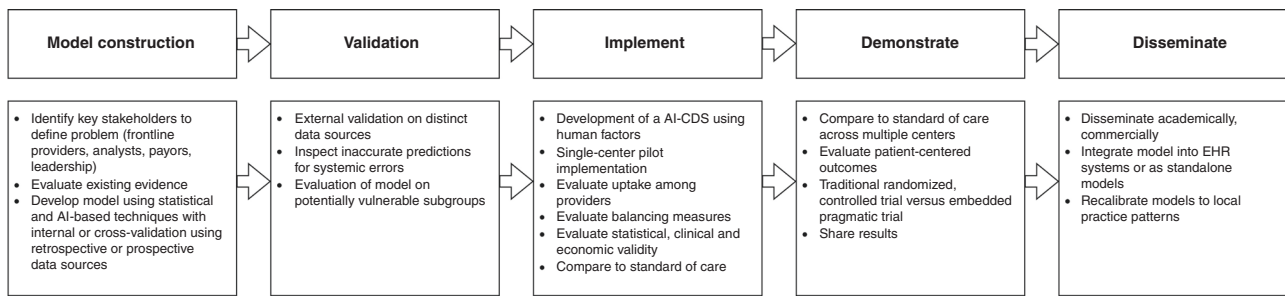
4. Prediction of infants with significant intra-abdominal injury: several investigators have developed ML models to identify children presenting following abdominal trauma who are at risk of intra-abdominal injury requiring intervention, including injuries that resulted in death, therapeutic angiography or laparotomy, blood transfusion, or admission for hospitalization for  $\geq 2$  nights to receive intravenous fluids.<sup>38,39</sup> In one study, the random forest, support vector machine, and generalized linear modeling algorithms demonstrated high predictive performance to identify patients at low risk of this outcome, potentially resulting in a decreased requirement for imaging in this cohort.<sup>39</sup>

Other models have been described to broadly classify pediatric diagnoses,<sup>40</sup> identify young infants with serious bacterial infections,<sup>41</sup> and assist with ventilator support.<sup>42</sup> Nearly all models in pediatric practice have not been externally validated in a peer-reviewed setting.

#### ROLE OF AI-CDS

The integration of computers to assist in medical decision making was first discussed as a possibility in 1959, when the authors postulated that computers may be able to perform complex reasoning tasks, collect and process clinical information, and remind the physician of overlooked diagnoses.<sup>43</sup> As described, several studies constructing or evaluating AI models in children have been reported, suggesting that AI-CDS may have the potential to improve care. Despite this avid interest, less work has been done to externally validate these models, construct CDS tools using these AI models with relevant stakeholder groups, and evaluate their role in implementation.

A template for an AI-CDS model is provided in Fig. 2. Clinical data may be collected from both structured and unstructured data. Structured data exists within predefined fields, such as vital



**Fig. 3 Steps involved in the development of artificial intelligence clinical decision support.** Stakeholders should be recruited early in the process to evaluate existing models, identify key priorities and develop a machine learning model. Models should then externally validated with specific consideration to balancing measures, including false positives and negatives and the performance of the model on minorities and/or socioeconomically disadvantaged subgroups. Models may then be implemented into the electronic health record with subsequent evaluation. Models should be studied and compared to standard of care, and if proven favorable, may then be disseminated.

signs, laboratory results, or diagnosis codes. Unstructured data are less organized and more subject to irregularities, such as information contained within clinician notes and imaging reports. To assist with the interpretation of unstructured data from notes, natural language processing (NLP), a branch of AI concerned with computational interpretation and production of human language, may allow for the translation of free text into results that may be used in risk prediction models.<sup>44</sup> Neural network and deep learning models have also been used to automate the interpretation of unstructured imaging data, the results of which may also feed into a CDS model and have been used for pediatric pneumonia detection from chest radiographs.<sup>45,46</sup>

One recent study demonstrated a role for an NLP-based tool in the interpretation of chest radiographs for pneumonia. Using data from a single center, the authors used NLP-extracted features following a random forest-based ML classifier. The model demonstrated high accuracy (with an AUROC of 0.95). The authors subsequently deployed the model in a CDS tool within the EHR. Chest radiographs with a high probability of pneumonia triggered an interruptive best practice advisory.<sup>47</sup>

Another recent implementation study evaluated the role of AI-CDS in the management of asthma in children in a randomized control study that included 184 participants. For one group, AI-CDS summarized relevant clinical information, provided a prediction of the risk of asthma exacerbations over the following year, and suggested asthma management plans. The AI-CDS was compared to a group managed with usual care. Both groups showed similar declines in asthma exacerbations (12% for the intervention group and 15% for the control group). However, use of the AI-CDS resulted in significant decreases in time spent reviewing medical charts and time to follow-up, and reduced the cost of care compared to the usual care group.<sup>48</sup>

#### DEVELOPMENT AND IMPLEMENTATION OF AI-CDS

The process of AI-CDS development extends from initial model derivation and validation to implementation, study, and dissemination (Fig. 3). Model development is best performed by an interdisciplinary team of stakeholders, including clinicians and other potential end users, data scientists, clinical informaticians, and implementation scientists. For the development and reporting of prediction models, the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) provides a rubric for best practice reporting approaches.<sup>49</sup> In addition, the minimum information about clinical artificial intelligence modeling (MI-CLAIM) criterion provides considerations specific for AI.<sup>50</sup> MI-CLAIM emphasizes the need to perform benchmarking, clearly define input data types, and describe preprocessing data steps (e.g., transformations).

In addition to evaluating a model's overall performance, consideration should be provided to other metrics, including those related to cost, resource utilization, and other balancing measures. Pediatric-specific considerations for balancing measures may include those in the Pediatric Patient Reported Outcomes Measurement Information System rubrics, including those pertaining to physical, mental, and social wellness.<sup>51</sup> Others include those related to painful procedures (i.e., venipuncture), radiation exposure, unnecessary exposure to medications, and harm to the family.

After model development and validation, the next step is for the AI-based algorithm to be implemented in the form of a CDS tool. Models may further be scaled, including through full integration into the EHR using EHR-vendor-specific modules, via application programming interfaces through standards like the Fast Health Interoperability Resources, or as standalone systems.<sup>52</sup> When addressing implementation, interoperability standards within EHRs must be considered.<sup>53</sup> Following local deployment, models require proactive and continued monitoring to evaluate changes in performance over time, retraining the model with additional targeted and augmented data, and identifying barriers and facilitators toward the use of the model by clinicians.<sup>54</sup> Continued vigilance is required after local implementation of a model to evaluate for other balancing effects and a decline in performance over time. CDS tools should have systems in place to report problems and provide feedback.<sup>55,56</sup>

An AI-CDS that performs well locally may indicate the potential for generalizability. Traditional and emerging methods of clinical investigation can be used to evaluate whether the AI-CDS tool results in improved patient-centered outcomes across several centers. Randomized, controlled trials comparing the use of the AI-CDS to existing standards of care are essential to ensure that costs of scaling, implementing, and maintaining a tool more broadly are warranted and to ensure that dissemination does not have unintended effects that might negatively impact patient care. Methods of pragmatic clinical investigation that meld quality improvement techniques such as patient care unit or institution-level process standardization with randomization, such as clustered, stepped-wedge designs, or approaches involving EHR-embedded trial elements, such as randomization at the point-of-care, should be considered in the trial design to aid in the overall efficiency of conducting trials in this space.<sup>57,58</sup> An AI-CDS tool should only be disseminated after it has demonstrated generalizable value through rigorous clinical investigation.

A recent editorial published by Shortliffe on AI-CDS identified important priorities which must be considered during implementation.<sup>59</sup> First, the reasoning behind the AI decision should be transparent in order for the clinician to comprehend the rationale behind the decision. In other words, the "black box" algorithms of AI (in which a decision is made by an algorithm that lacks

interpretability, such as deep learning models) is inadequate for clinical contexts where transparency and clinician trust in the models is paramount. Second, CDS should promote additional efficiency and blend seamlessly with a clinical environment. Third, CDS tools should be intuitively constructed and simple to use so that no major training is required for their use. Fourth, a CDS should reflect an understanding of the pertinent domain and answer clinically relevant questions. Fifth, advice should be offered in a way that recognizes the knowledge of the user to augment—but not replace—decision making. Finally, a CDS tool should be constructed on a rigorous, peer-reviewed scientific basis to establish its reliability and generalizability.

### CHALLENGES OF AI-CDS IN CHILDREN

Though AI-CDS tools may have advantages compared to rule-based CDS in many situations, their use may result in additional challenges. One review of qualitative research on the use of AI in the clinical context that included clinicians, consumers, healthcare executives, and industry professionals, reported positive perceptions about its use, though also identified many reservations including wariness about liability from AI-mediated errors, need for training, reputational harm, worsening communication, privacy concerns, lack of proof regarding efficacy, and a lack of explainability.<sup>60</sup>

Some limitations lie in the development and implementation of AI-CDS carrying greater importance to research in children. These include a need for large datasets, challenges relating to unbalanced data, challenges of generalizability, lack of evidence-based care, maturational variation in children, and ethical issues.

1. Need for large, high-resolution datasets: high-resolution datasets, which contain demographic data combined with detailed clinical data (including relevant data from a patient's past medical history, physical examination, vital signs, monitor data, laboratory results, and outcome data) are ideal for ML applications. From the standpoint of model development and prediction, an important limitation for pediatric research lies in the lack of large datasets. Children account for a smaller proportion of healthcare resources,<sup>61</sup> and datasets in children (either derived from the EHR, prospective, or administrative data) are frequently smaller. For ML to be successful, granular data are required, yet such data are commonly unavailable in administrative datasets (defined as data collected for administrative or billing purposes)<sup>62</sup> and the procurement of more detailed datasets can be expensive. When using smaller datasets, advanced ML methods may offer no advantage or only a slight advantage over classic linear models.<sup>63,64</sup>
2. Imbalanced datasets: pediatric AI research may be challenged by the presence of imbalanced datasets, a term used to describe when one of the categorical outcome variables (such as the lack of a disease state) occurs substantially more commonly than the other (the presence of a disease state). For example, only 0.1% of children presenting to the ED will ultimately be diagnosed with sepsis, making its prediction a challenge.<sup>65</sup> A number of techniques have been described to handle imbalanced datasets. Among these, undersampling (in which the amount of majority class instances is reduced), oversampling (in which the amount of minority class instances is inflated using bootstrapping or a similar approach), or combinations of the two<sup>66</sup> may overcome the limitations of imbalanced datasets and are particularly relevant in ML for pediatric applications. An additional challenge with imbalanced datasets is appropriate measurements of results. Classic performance measures like AUROC may be inappropriate in these cases given that the performance may be overestimated by a model

that tends to over-predict a majority class. Alternative approaches, such as the area under the precision-recall curve (AUPRC) may be more appropriate in this context given that the interpretation of the AUPRC is based on the prevalence of the outcome and represents the relationship between sensitivity and positive predictive value, which are arguably the most important performance measures for rare event prediction.<sup>67</sup>

3. Generalizability: models derived from one hospital system may poorly generalize to another. One recently reported example was the Epic Sepsis Model, a proprietary sepsis prediction model to predict adults with sepsis study that was developed across three health systems. In an external validation of this model in a different hospital system, it demonstrated a decline in performance in a different hospital system, with an AUROC of 0.63 (compared to AUROCs of 0.76–0.83 as originally reported).<sup>68</sup> In addition, challenges exist with interoperability of definitions between EHR systems which need to be reconciled prior to their use.
4. Lack of evidence-based guidelines and variation in care: CDS is most effective when the detection of disease can serve as a prompt to promote evidence-based care. However, a large evidence-base does not currently exist for the optimal management of many common pediatric conditions. As such, the role of a CDS in providing recommendations may be affected by substantial practice pattern variations in care, both between individual providers and across institutions.
5. Developmental variation in children: pediatric AI models must account for the physiologic changes and changes in disease risk that occur through early childhood. Vital signs need to be adjusted for age.<sup>69</sup> ML tools for developmental disorders similarly need to adjust for maturational changes.<sup>70</sup> The automated interpretation of radiographs similarly must account for maturational changes in children, differing appearances of common pathologies, and broader differential diagnoses, which can all create challenges associated with their automated interpretation.<sup>71</sup>
6. Ethical issues: many important ethical challenges have been raised regarding the use of AI-CDS to augment the care of children. Efforts to demonstrate the benefit of AI-CDS among pediatric patients can be expected to encounter a set of challenges already faced by other clinical investigators aiming to conduct trials enrolling pediatric patients. Waived or deferred consent is relatively uncommon in pediatric clinical trials in the United States. This poses a potential challenge during regulatory reviews of the design of embedded, pragmatic studies that seek to evaluate the use of AI-CDS nested within usual care workflows. Other ethical issues surrounding the deployment of AI-CDS relate to the determination of acceptable performance standards for models, questions of whether “black box” tools should be relied on in high-stakes clinical care, the potential infringement of data protection rights of individuals, and whether the use of AI-CDS must be disclosed to patients and their caregivers.<sup>72,73</sup>
7. Disparities in care: some studies have suggested that socio-economic and racial disparities in care may be improved through the standardization in care provided by a CDS system.<sup>74,75</sup> However, AI models have been demonstrated to potentiate systemic racism in healthcare models.<sup>76</sup> Similarly, crisis standards of care during the COVID-19 pandemic have commonly relied on rubrics incorporating regression-based sequential organ failure assessment (SOFA) scores to guide the allocation of scarce critical care resources. Recently highlighted racial disparities in SOFA scores<sup>77–79</sup> have cast substantial doubt on the use of SOFA-based CDS for patients with COVID-19. Similarly, the potential for inequity was noted in a popular web-based system for the identification of young infants at risk of UTI used a dichotomized variable of race,<sup>12</sup> prompting

criticism that such a model may delay the identification of Black children with UTI.<sup>80</sup>

8. Medicolegal issues: the implementation of an AI-guided CDS to partially assist or automate decision making carries medicolegal implications, such as in patients who have a poor outcome due to inaccurate predictions or CDS recommendations. Others have raised concerns about patient privacy or that such a system may be exposed to hacking. In response to stakeholder feedback, the United States Food and Drug Administration proposed a regulatory framework in 2021 to describe a potential approach for the review of AI/ML-driven software modifications, in which AI software would be considered a “medical device”.<sup>81</sup> In it, the FDA described priorities, including a need for “best practice” model development, the importance of a patient-centered approach, consideration of biases, and for continual monitoring of a software product through its post-market life to enable the FDA to have a reasonable assurance of safety and effectiveness as a model continues to iteratively improve over time.
9. Limitations of CDS at large: many limitations associated with CDS system implementation also apply to AI-CDS. One study performed in the primary care setting, for example, noted that providers were likely to use prompts when provided as a means of support and choice, and were resistant when such prompts were perceived as a means of enforcement.<sup>82</sup>

Other issues with AI-CDS that have been proposed include impact on user skill (e.g., human skill deterioration over time), need for continued maintenance, and challenges with data quality.<sup>24</sup> Finally, AI-CDS are expensive to develop and maintain over time.<sup>83</sup>

#### FUTURE STEPS

A well-validated and implemented AI-CDS tool may allow for the improved care of patients at the bedside through the delivery. However, for AI-CDS to benefit children, several critical steps are required. Situations that may benefit from improved predictive performance need to be identified. To construct predictive models, large datasets are needed. These may be potentially generated via the construction of larger, federated datasets, defined as datasets that are mapped from subsidiary datasets that may be interconnected across computer systems.<sup>84</sup> These data may be derived either retrospectively (from hospital medical records) or through prospective trials for use in ML applications. These models need to be published in accordance with reporting guidelines and externally validated in distinct populations and settings. Importantly, more effort is required to ensure generalizability by assessing the performance of models on distinct patient cohorts.

Prior to and in conjunction with data collection and modeling efforts, continued engagement is needed with key stakeholders, including physicians and other healthcare providers, payors, computer scientists, and regulators to identify ways in which AI-CDS models may be most effectively implemented. Critically, patients and their families are a frequently overlooked stakeholder group, though they have an important voice in the conversation about AI-based technologies in healthcare. Two previous studies have suggested that parents may generally be receptive toward the use of AI,<sup>85</sup> though there may be differences in trust by race and age.<sup>86</sup> More data are needed to assess their baseline comfort with the use of AI in the development of healthcare predictions and recommendations: while the public at large appears to have a favorable perspective on the role of AI applications in healthcare,<sup>87–92</sup> important concerns have emerged, such as a loss of humanism in medicine.<sup>93</sup> Other studies have identified racial disparities with respect to the comfort in the role of AI in medicine.<sup>88</sup> Little has been reported on the opinions of caregivers on the role of AI in medicine.

The role of AI in CDS continues to evolve. More work is needed to compare AI-CDS tools to conventional CDS or routine practice, for which little has been reported. There is a growing interest in explainable AI, which is more transparent, allowing the user to evaluate how predictions are constructed.<sup>94</sup> Models that do not require periodic tuning, and that can instead improve over time when exposed to new data, a term called adaptive CDS, also represent an important opportunity for growth.<sup>95</sup>

#### CONCLUSION

Applied in the appropriate clinical context, AI-CDS may provide an opportunity to seamlessly provide evidence-based and individualized care to children. Given the novelty of ML-based models in clinical practice, its current use in any healthcare domain currently remains limited, including in the care of children. AI-CDS may be able to overcome some of the limitations frequently ascribed to CDS, though more research is required. For AI-CDS to positively impact the care of children, a diverse group of stakeholders is required to promote the development of validated tools from high-resolution datasets, with high predictive accuracy, that are transparent and easy to use, and that complement the existing knowledge base of the clinician.

#### REFERENCES

1. Agency for Healthcare Research and Quality. *Clinical Decision Support*. <https://www.ahrq.gov/cpi/about/otherwebsites/clinical-decision-support/index.html>.
2. Manyika, J. et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. (2011).
3. Beyer, M. & Laney, D. The Importance of 'Big Data': A Definition. *Gartner*. 1–9, <https://www.gartner.com/en/documents/2057415> (2012).
4. Bohr, A. & Memarzadeh, K. The rise of artificial intelligence in healthcare applications. *Artif. Intell. Healthc.* 25–60, <https://doi.org/10.1016/B978-0-12-818438-7.00002-2> (2020).
5. Sanchez-Pinto, L. N., Luo, Y. & Churpek, M. M. Big data and data science in critical care. *Chest* **154**, 1239–1248 (2018).
6. Berner, E. & La Lande, T. in *Clinical Decision Support Systems* (ed. Berner, E. S.) 3–22 (Springer, 2007).
7. Goertzel, G. Clinical decision support system. *Ann. N. Y. Acad. Sci.* **161**, 689–693 (1969).
8. Bright, T. J. et al. Effect of clinical decision-support systems: a systematic review. *Ann. Intern. Med.* **157**, 29–43 (2012).
9. Kuppermann, N. et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet* **374**, 1160–1170 (2009).
10. Dayan, P. S. et al. Use of traumatic brain injury prediction rules with clinical decision support. *Pediatrics* **139**, e20162709 (2017).
11. Hoeksema, L. J. et al. Accuracy of a computerized clinical decision-support system for asthma assessment and management. *J. Am. Med. Inform. Assoc.* **18**, 243–250 (2011).
12. Shaikh, N. et al. Development and validation of a calculator for estimating the probability of urinary tract infection in young febrile children. *JAMA Pediatr.* **172**, 550–556 (2018).
13. Carroll, A. E. et al. Use of a computerized decision aid for developmental surveillance and screening: a randomized clinical trial. *JAMA Pediatr.* **168**, 815–821 (2014).
14. Randolph, A. G. et al. Evaluation of compliance with a computerized protocol: weaning from mechanical ventilator support using pressure support. *Comput. Methods Prog. Biomed.* **57**, 201–215 (1998).
15. Hotz, J. C. et al. Real-time effort driven ventilator management: a pilot study. *Pediatr. Crit. Care Med.* **21**, 933–940 (2020).
16. Mullett, C. J., Evans, R. S., Christenson, J. C. & Dean, J. M. Development and impact of a computerized pediatric anti-infective decision support program. *Pediatrics* **108**, e75–e75 (2001).
17. Jankovic, I. & Chen, J. H. Clinical decision support and implications for the clinician burnout crisis. *Yearb. Med. Inform.* **29**, 145–154 (2020).
18. McCoy, A. B. et al. A framework for evaluating the appropriateness of clinical decision support alerts and responses. *J. Am. Med. Inform. Assoc.* **19**, 346–352 (2012).
19. Ancker, J. S. et al. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Med. Inform. Decis. Mak.* **17**, 36 (2017).

20. Carspecken, C. W., Sharek, P. J., Longhurst, C. & Pageler, N. M. A clinical case of electronic health record drug alert fatigue: consequences for patient outcome. *Pediatrics* **131**, e1970–e1973 (2013).
21. Rousseau, N., McColl, E., Newton, J., Grimshaw, J. & Eccles, M. Practice based, longitudinal, qualitative interview study of computerised evidence based guidelines in primary care. *BMJ* **326**, 314 (2003).
22. Zheng, K., Padman, R., Johnson, M. P. & Diamond, H. S. Understanding technology adoption in clinical care: clinician adoption behavior of a point-of-care reminder system. *Int. J. Med. Inform.* **74**, 535–543 (2005).
23. Patterson, E. S. et al. Identifying barriers to the effective use of clinical reminders: bootstrapping multiple methods. *J. Biomed. Inform.* **38**, 189–199 (2005).
24. Sutton, R. T. et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digit. Med.* **3**, 17 (2020).
25. Luo, Y., Wang, F. & Szolovits, P. Tensor factorization toward precision medicine. *Brief. Bioinform.* **18**, 511–514 (2017).
26. Chao, G., Luo, Y. & Ding, W. Recent advances in supervised dimension reduction: a survey. *Mach. Learn. Knowl. Extr.* **1**, 341–358 (2019).
27. Singh, A., Thakur, N. & Sharma, A. A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* 1310–1315 (2016).
28. Ramgopal, S. et al. A prediction model for pediatric radiographic pneumonia. *Pediatrics* **149**, e2021051405 (2022).
29. Nigrovic, L. E. et al. Clinical prediction rule for identifying children with cerebrospinal fluid pleocytosis at very low risk of bacterial meningitis. *JAMA* **297**, 52–60 (2007).
30. Kuppermann, N. et al. A clinical prediction rule for stratifying febrile infants 60 days and younger at risk for serious bacterial infections. *JAMA Pediatr.* **173**, 342–351. <https://doi.org/10.1001/jamapediatrics.2018.5501> (2019).
31. Kocher, M. S., Zurakowski, D. & Kasser, J. R. Differentiating between septic arthritis and transient synovitis of the hip in children: an evidence-based clinical prediction algorithm. *J. Bone Joint Surg. Am.* **81**, 1662–1670 (1999).
32. Holmes, J. F. et al. Identifying children at very low risk of clinically important blunt abdominal injuries. *Ann. Emerg. Med.* **62**, 107–116.e2 (2013).
33. Bertsimas, D., Dunn, J., Steele, D. W., Trikalinos, T. A. & Wang, Y. Comparison of Machine Learning Optimal Classification Trees With the Pediatric Emergency Care Applied Research Network Head Trauma Decision Rules. *JAMA Pediatr.* **173**, 648–656 (2019).
34. Singh, D. et al. Assessment of machine learning–based medical directives to expedite care in pediatric emergency medicine. *JAMA Netw. Open* **5**, e222599–e222599 (2022).
35. Le, S. et al. Pediatric severe sepsis prediction using machine learning. *Front. Pediatr.* **7**, 413 (2019).
36. Lamping, F. et al. Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children—a data-driven approach using machine-learning algorithms. *BMC Pediatr.* **18**, 1–11 (2018).
37. Scott, H. F. et al. Development and validation of a model to predict pediatric septic shock using data known 2h after hospital arrival. *Pediatr. Crit. Care Med.* **22**, 16–26 (2021).
38. Ramgopal, S. & Horvat, C. M. Machine learning approaches for the identification of children at low risk of intra-abdominal injury requiring intervention. *J. Trauma Acute Care Surg.* **90**, e128–e129 (2021).
39. Pennell, C., Polet, C., Arthur, L. G., Grewal, H. & Aronoff, S. Risk assessment for intra-abdominal injury following blunt trauma in children: derivation and validation of a machine learning model. *J. Trauma Acute Care Surg.* **89**, 153–159 (2020).
40. Liang, H. et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat. Med.* **25**, 433–438 (2019).
41. Ramgopal, S., Horvat, C. M., Yanamala, N. & Alpern, E. R. Machine learning to predict serious bacterial infections in young febrile infants. *Pediatrics* **146**, e20194096 (2020).
42. Ghazal, S. et al. Using machine learning models to predict oxygen saturation following ventilator support adjustment in critically ill children: a single center pilot study. *PLoS One* **14**, e0198921 (2019).
43. Ledley, R. S. & Lusted, L. B. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* **130**, 9–21 (1959).
44. Juhn, Y. & Liu, H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *J. Allergy Clin. Immunol.* **145**, 463–469 (2020).
45. Rajaraman, S., Candemir, S., Kim, I., Thoma, G. & Antani, S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Appl. Sci. (Basel, Switzerland)* **8**, 1715 (2018).
46. Chen, Y. et al. Deep learning for classification of pediatric chest radiographs by WHO’s standardized methodology. *PLoS One* **16**, e0253239 (2021).
47. Smith, J. C. et al. Natural language processing and machine learning to enable clinical decision support for treatment of pediatric pneumonia. *AMIA Annu. Symp. Proc.* **2020**, 1130 (2020).
48. Seol, H. Y. et al. Artificial intelligence-assisted clinical decision support for childhood asthma management: a randomized clinical trial. *PLoS One* **16**, e0255261 (2021).
49. Moons, K. G. M. et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann. Intern. Med.* **162**, W1–W73 (2015).
50. Norgeot, B. et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat. Med.* **26**, 1320–1324 (2020).
51. Forrest, C. B. et al. Commentary: The Patient-Reported Outcome Measurement Information System (PROMIS®) for children and youth: application to pediatric psychology. *J. Pediatr. Psychol.* **37**, 614–621 (2012).
52. Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S. & Ramoni, R. B. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J. Am. Med. Inform. Assoc.* **23**, 899–908 (2016).
53. Department of Health and Human Services & National Health Services (England). *UK/US Standards Comparison*. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/465064/Work\\_Stream\\_2.1\\_Final.pdf-AppendixC](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/465064/Work_Stream_2.1_Final.pdf-AppendixC).
54. Luo, Y., Wunderink, R. G. & Lloyd-Jones, D. Proactive vs reactive machine learning in health care: lessons from the COVID-19 pandemic. *JAMA* **327**, 623–624 (2022).
55. Saleem, J. J. et al. Exploring barriers and facilitators to the use of computerized clinical reminders. *J. Am. Med. Inform. Assoc.* **12**, 438–447 (2005).
56. Fung, C. H., Woods, J. N., Asch, S. M., Glassman, P. & Doebbeling, B. N. Variation in implementation and use of computerized clinical reminders in an integrated healthcare system. *Am. J. Manag. Care* **10**, 878–885 (2004).
57. Douthit, B. et al. Designing and building CDS tools for pragmatic trials. *NIH Collaboratory’s Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials* <https://doi.org/10.28929/133> (2017).
58. Douthit, B. et al. Evaluating CDS. *NIH Collaboratory’s Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials* <https://doi.org/10.28929/134> (2017).
59. Shortliffe, E. H. & Sepúlveda, M. J. Clinical decision support in the era of artificial intelligence. *JAMA* **320**, 2199–2200 (2018).
60. Scott, I. A., Carter, S. M. & Coiera, E. Exploring stakeholder attitudes towards AI in clinical practice. *BMJ Heal. Care Informatics* **28**, e100450 (2021).
61. Witt, W. P., Weiss, A. J. & Elixhauser, A. Overview of hospital stays for children in the United States, 2012. In: *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs* [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006 Feb. Statistical Brief #187. (2014).
62. Cadarette, S. M. & Wong, L. An introduction to health care administrative data. *Can. J. Hosp. Pharm.* **68**, 232–237 (2015).
63. Panesar, S. S., D’Souza, R. N., Yeh, F.-C. & Fernandez-Miranda, J. C. Machine learning versus logistic regression methods for 2-year mortality prognostication in a small, heterogeneous glioma database. *World Neurosurg.* **X** 2, 100012 (2019).
64. Lynam, A. L. et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagnostic Progn. Res.* **4**, 6 (2020).
65. Ramgopal, S., Adler, M. D. & Horvat, C. M. Application of the improving pediatric sepsis outcomes definition for pediatric sepsis to nationally representative emergency department data. *Pediatr. Qual. Saf.* **6**, e468–e468 (2021).
66. Yap, B. W. et al. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets BT. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)* (eds. Herawan, T., Deris, M. M. & Abawajy, J.) 13–22 (Springer Singapore, 2014).
67. Dewan, M. & Sanchez-Pinto, L. N. Crystal balls and magic eight balls: the art of developing and implementing automated algorithms in acute care pediatrics. *Pediatr. Crit. Care Med.* **20**, 1197–1199 (2019).
68. Wong, A. et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern. Med.* **181**, 1065–1070 (2021).
69. Lee, B. et al. Development of a machine learning model for predicting pediatric mortality in the early stages of intensive care unit admission. *Sci. Rep.* **11**, 1263 (2021).
70. Wall, D. P., Kosmicki, J., DeLuca, T. F., Harstad, E. & Fusaro, V. A. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl. Psychiatry* **2**, e100 (2012).
71. Padash, S., Mohebbian, M. R., Adams, S. J., Henderson, R. D. E. & Babyn, P. Pediatric chest radiograph interpretation: how far has artificial intelligence come? A systematic literature review. *Pediatr. Radiol.* **52**, 1568–1580. <https://doi.org/10.1007/s00247-022-05368-w> (2022).
72. Rodrigues, R. Legal and human rights issues of AI: gaps, challenges and vulnerabilities. *J. Responsible Technol.* **4**, 100005 (2020).

73. Michelson, K. N., Klugman, C. M., Kho, A. N. & Gerke, S. Ethical considerations related to using machine learning-based prediction of mortality in the pediatric intensive care unit. *J. Pediatr.* S0022–S3476 (2022). <https://pubmed.ncbi.nlm.nih.gov/35038439/>.
74. Mitchell, J. et al. Association between clinical decision support system use and rural quality disparities in the treatment of pneumonia. *J. Rural Heal.* **30**, 186–195 (2014).
75. Lau, B. D. et al. Eliminating healthcare disparities via mandatory clinical decision support: the venous thromboembolism (VTE) example. *Med. Care* **53**, 18 (2015).
76. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
77. Ashana, D. C. et al. Equitably allocating resources during crises: racial differences in mortality prediction models. *Am. J. Respir. Crit. Care Med.* **204**, 178–186 (2021).
78. Tolchin, B. et al. Racial disparities in the SOFA score among patients hospitalized with COVID-19. *PLoS One* **16**, e0257608 (2021).
79. Bhavani, S. V. et al. Simulation of ventilator allocation in critically ill patients with COVID-19. *Am. J. Respir. Crit. Care Med.* **204**, 1224–1227 (2021).
80. Vyas, D. A., Eisenstein, L. G. & Jones, D. S. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N. Engl. J. Med.* **383**, 874–882 (2020).
81. United States Food and Drug Administration. *Software as a Medical Device (SaMD) Action Plan*. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> (2021).
82. McDermott, L. et al. Developing a computer delivered, theory based intervention for guideline implementation in general practice. *BMC Fam. Pract.* **11**, 90 (2010).
83. Kabachinski, J. A look at clinical decision support systems. *Biomed. Instrum. Technol.* **47**, 432 (2013).
84. Heimbigner, D. & Mcleod, D. A federated architecture for information management. *ACM Trans. Inf. Syst.* **3**, 253–278 (1985).
85. Sisk, B. A., Antes, A. L., Burrous, S. & DuBois, J. M. Parental attitudes toward artificial intelligence-driven precision medicine technologies in pediatric healthcare. *Children (Basel, Switz.)* **7**, 145 (2020).
86. Ramgopal, S. et al. Parental perceptions on use of artificial intelligence in pediatric acute care. *Acad. Pediatr.* (2022). <https://pubmed.ncbi.nlm.nih.gov/35577283/>.
87. Miller, S., Gilbert, S., Virani, V. & Wicks, P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. *JMIR Hum. Factors* **7**, e19713 (2020).
88. Yakar, D., Ongena, Y. P., Kwee, T. C. & Haan, M. Do people favor artificial intelligence over physicians? A survey among the general population and their view on artificial intelligence in medicine. *Value Health* **25**, 374–381. <https://doi.org/10.1016/j.jval.2021.09.004> (2022).
89. Spänig, S. et al. The virtual doctor: an interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. *Artif. Intell. Med.* **100**, 101706 (2019).
90. Jutzi, T. B. et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. *Front. Med.* **7**, 233 (2020).
91. Nelson, C. A. et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *JAMA Dermatol.* **156**, 501–512 (2020).
92. Palmisciano, P., Jamjoom, A. A. B., Taylor, D., Stoyanov, D. & Marcus, H. J. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. *World Neurosurg.* **138**, e627–e633 (2020).
93. Tran, V.-T., Riveros, C. & Ravaud, P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. *NPJ Digit. Med.* **2**, 1–8 (2019).
94. Shaban-Nejad, A., Michalowski, M., Brownstein, J. S. & Buckeridge, D. L. Guest editorial explainable AI: towards fairness, accountability, transparency and trust in healthcare. *IEEE J. Biomed. Heal. Inform.* **25**, 2374–2375 (2021).
95. Petersen, C. et al. Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper. *J. Am. Med. Assoc.* **28**, 677–684 (2021).

## AUTHOR CONTRIBUTIONS

S.R. drafted the article. L.N.S.-P., C.M.H., M.S.C., Y.L., and T.A.F. revised the article critically for important intellectual content. All authors provided final approval of the version to be published.

## FUNDING

S.R. is funded by the PEDSnet Scholars Training Program (Department of Pediatrics, Northwestern University Feinberg School of Medicine).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Sriram Ramgopal.

**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.