



# From Clinic to Computer and Back Again: Practical Considerations When Designing and Implementing Machine Learning Solutions for Pediatrics

Sujay Nagaraj, BScH<sup>1,2,\*</sup> 

Vinyas Harish, BCompH<sup>1,3</sup>

Liam G. McCoy<sup>1,3</sup>

Felipe Morgado, MSc<sup>1,4</sup>

Ian Stedman, BA(Hons), MA, LLB, LLM, PhD<sup>5</sup>

Stephen Lu<sup>6</sup>

Erik Drysdale, BA, MA, MSc<sup>6</sup>

Michael Brudno, BA, MS, PhD<sup>2,6,7,8</sup>

Devin Singh, BMSch, MBBS<sup>2,6</sup>

## Address

<sup>1</sup>Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

<sup>2</sup>Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

Email: s.nagaraj@mail.utoronto.ca

<sup>3</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

<sup>4</sup>Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

<sup>5</sup>School of Public Policy and Administration, York University, Toronto, Ontario, Canada

<sup>6</sup>Paediatric Emergency Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

<sup>7</sup>University Health Network, Toronto, Ontario, Canada

<sup>8</sup>Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada

Published online: 15 September 2020

© Springer Nature Switzerland AG 2020

Vinyas Harish, Liam G. McCoy and Felipe Morgado were listed as co-second authors, as their contributions to the work was equivalent. This article is part of the Topical Collection on *Patient Safety*

Published online: 15 September 2020

**Keywords** Machine learning · Artificial intelligence · Pediatric medicine · Translation · Design · Ethics

## Abstract

*Purpose of review* Machine learning (ML), a branch of artificial intelligence, is influencing all fields in medicine, with an abundance of work describing its application to adult practice. ML in pediatrics is distinctly unique with clinical, technical, and ethical nuances limiting the direct translation of ML tools developed for adults to pediatric populations. To our knowledge, no work has yet focused on outlining the unique considerations that need to be taken into account when designing and implementing ML in pediatrics.

*Recent findings* The nature of varying developmental stages and the prominence of family-centered care lead to vastly different data-generating processes in pediatrics. Data heterogeneity and a lack of high-quality pediatric databases further complicate ML research. In order to address some of these nuances, we provide a common pipeline for clinicians and computer scientists to use as a foundation for structuring ML projects, and a framework for the translation of a developed model into clinical practice in pediatrics. Throughout these pathways, we also highlight ethical and legal considerations that must be taken into account when working with pediatric populations and data.

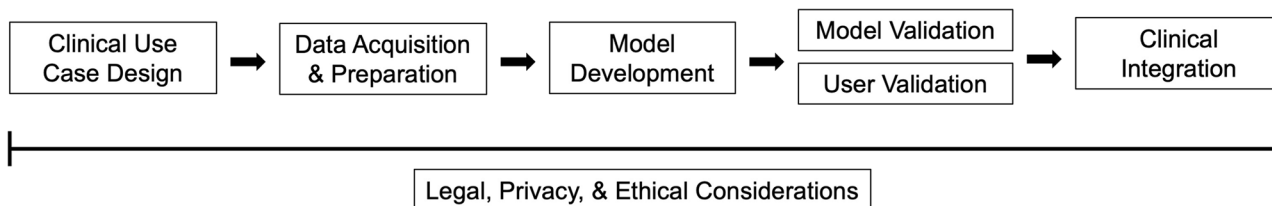
*Summary* Here, we describe a comprehensive outline of special considerations required of ML in pediatrics from project ideation to implementation. We hope this review can serve as a high-level guideline for ML scientists and clinicians alike to identify applications in the pediatric setting, generate effective ML solutions, and subsequently deliver them to patients, families, and providers.

## Introduction

Artificial intelligence (AI) has the potential to drastically reshape medicine. The uncertainty associated with how this will unfold contributes to mixed reactions of enthusiasm and concern. Despite this, most healthcare providers agree that there exists an increasing need for improved efficiency, enhanced patient safety, and equitable access to care that is free from geographic, financial, and racial barriers. When developed to integrate with existing clinical workflows and with sound ethical principles in mind, AI has the potential to address each of these concerns while adding value to healthcare systems at scale [1•]. Traditional human workflows do not generally scale seamlessly in response to spikes in patient volume and demand. The strain that healthcare systems are facing globally in response to the COVID-19 pandemic, for example, is evidence of the fragility inherent to human-based workflows and highlights the need for innovation. AI can also help reduce human practice variation,

which is known to be associated with patient harm, and aid in democratizing medicine for improved equity in care delivery [2], while simultaneously reducing healthcare costs [3]. Given this potential, a surge in machine learning for healthcare (ML4H) applications can be seen in both academic and private sectors [4].

Despite the promise, medical specialties have yet to realize the true potential of AI as evidenced by very little integration into clinical practice. This is especially apparent in the field of pediatrics. Reasons for this are multifactorial and include the challenges associated with bridging the gap between pediatric medicine and computer science. In this review, we present a framework for both computer scientists and pediatric specialists to outline key considerations and nuances encountered when conceptualizing, building, and integrating machine learning (ML) models into pediatric workflows.



**Fig. 1.** A high-level pipeline that can be used to structure ML4H projects in pediatrics. From start to finish: clinical use case design, data acquisition and preparation, model development, model and user validation, ending with clinical integration. We place special emphasis on legal, privacy, and ethical considerations throughout the entirety of the pipeline.

## Pediatric ML4H pipeline

At a high-level, the overall pipeline for developing ML4H tools is common across most fields of medicine and can be seen in Fig. 1. In order to help focus projects on maximizing clinical utility, a patient-centered clinical use case should anchor ML4H initiatives. The scope of clinical use cases can involve any aspect of care as long as the necessary data is available to fuel the development of ML models. Of note, “big data” is not always required to build sophisticated ML models, and the absence of large datasets should not be the sole reason to not search for ML-based solutions. The amount of data required for building a successful ML model is influenced by the complexity of the problem and the clarity of the signal within a dataset, which is often not known until an initial proof-of-concept trial is completed [5]. After model development, statistical validation is required, and the method and degree of rigor of this evaluation should be influenced by the intended clinical use case and implementation [6]. For example, a tool optimizing physician scheduling will require a different method of validation than an AI system developed to automate treatment decisions for children.

This common pipeline is valuable for clinicians and computer scientists to use as a foundation for structuring ML4H projects. However, a wide array of unique considerations also arises when we are specifically considering the pediatric context. Each stage of the development pipeline from use case design to implementation contains many clinical, technical, and ethical nuances limiting the direct translation of ML4H applications developed for adults to pediatric populations. Understanding these differences when compared with adult medical specialties is essential for the successful development and implementation of AI in pediatric medicine.

## Pediatric clinical use case design

Asking the right questions is critical to the success of any ML4H project, and identifying these questions is no trivial task. This is complicated further in pediatrics by the nature of varying developmental stages and the prominence of family-centered care [7]. Different patients may be involved in vastly different data-generating processes and have different abilities to interact with

technology based on their developmental age. For example, while it may be possible to have a mental health assessment tool for use by adolescents that is patient-facing, an equivalent tool for younger children may have to primarily target caregivers or parents—a difference which subsequently has a substantial influence on the data gathering, machine learning, and user experience design processes.

The task of clinical use case design must not be rushed, as careful consideration of these factors at the beginning of the project will inform the specifics of every subsequent aspect of the ML pipeline. Design thinking methodology is one process that provides an excellent framework to approaching clinical use case design as it is particularly well-suited to the context of pediatrics. Success with this framework has been seen across an array of use cases, from adolescents with cancer reporting on their pain to developing collaborative decision-making around pediatric asthma care [8]. Design thinking focuses on patient needs and prioritizes engagement with diverse stakeholders (families, children, clinical providers, administrators, etc.) in order to come to an understanding of the root causes of a problem, including social, political, economic, and organizational factors [9]. These factors, along with an awareness of current clinical workflows and how an AI solution will integrate, are essential to mapping a strategy for final implementation [10].

## Data acquisition and preparation

Data is the essential lifeline required for developing and maintaining all ML4H systems. The field of pediatrics in general suffers from a lack of pediatric-specific data due to various practical and ethical challenges when gathering data in children [11] [12]. ML4H has largely been pushed forward through the common use of large centralized databases, upon which numerous algorithms are developed and validated. In the adult critical care world, one of the largest such databases is MIMIC-III, which has been cited by more than 1300 projects [13] (although most of the MIMIC-III research is focused on adults, it should be noted that some data from neonates are present). No clear equivalent exists in the context of pediatric data science, although the recently released PIC database [14] contains physiological signal data from a large cohort of Chinese pediatric intensive care units and the ACS' pediatric surgical outcomes database contains more than 600,000 operations [15]. Many of the other databases that exist consist largely of unstructured electronic health record (EHR) data, such as PEDSnet [16] and EHR4CR [17] and may only be useful for projects of a specific nature or may otherwise be limited owing to lack of structure. If we are to address this research gap that currently exists in ML4H, high-quality pediatric databases are required.

A particular challenge when working with pediatric data is that children have unique physiologic features compared with adults. This has a direct and meaningful impact on the data collected and the associated data pre-processing steps required prior to training an ML model. For example, the median normal heart rate in children ranges from 140 beats per minute (bpm) for neonates to 70–80 bpm for adolescents [18]. A similar trend can be seen for respiratory rates in children and many other physiological parameters and lab measurements. Again, this large continuum of normal values based on age does not occur in

adults. In order to apply meaningful clinical context, unique pre-processing steps may be required to bin pediatric patients into the relevant categories to assist algorithms in understanding what is normal for a given age. In addition, the difference in the probability of a diagnosis between a 1-year-old and a 10-year-old child is generally going to be far greater than the difference between a 50- and 60-year-old adult. This variation imposed by age and the increased number of subgroups within pediatrics can sometimes require significantly more data to sufficiently power models when compared with adult ML4H projects.

We must also pay close attention to some of the unique challenges that arise when the patient (i.e., the child) is not the only person providing information about their symptoms and overall health. Pediatric patients typically co-report health outcomes alongside their caregivers. It is generally acknowledged however that proxies can sometimes be poor at reporting health-related information [19]. Specifically, data can be missing, wrong, or incomplete. The degree of involvement of a caregiver in the presentation of symptoms and other health data also differs with age, family circumstances, and cultural context, creating a spectrum of variability that can be difficult to control while also adding bias and noise to datasets.

## Model development

The development of an ML model involves passing pre-processed data (cleaned and prepared) into an algorithm that learns a task by undergoing a variety of mathematical optimization techniques that differ based on the type of ML model being utilized. Table 1 highlights some common ML models with their associated clinical utility.

Novel ML techniques are being developed to synthesize or otherwise supplement our current data. Although there should be an emphasis on building and growing high-quality pediatric datasets, certain techniques, such as transfer learning, can enable ML models to perform well in data-constrained environments [28]. This involves training an algorithm in one domain and exploiting commonalities between the data in the training domain (e.g., adult chest X-ray images) and the target domain (e.g., pediatric chest X-ray images) to build a model that generalizes between them. As a demonstration of the usefulness of this approach, transfer learning was leveraged by Liang et al. to improve pediatric pneumonia classification [29]. This field of research is growing, but it is limited in that it requires enough underlying similarities between the training and target domains.

## Model validation

In order to implement a model at the bedside after model development, a series of prospective trials are required to assess and validate model performance across multiple domains. We propose the following framework illustrated in Fig. 2 to serve as a high-level approach for the translation of a developed model into clinical practice in pediatrics.

**Table 1. Descriptions of ML tasks and examples of associated clinical use cases**

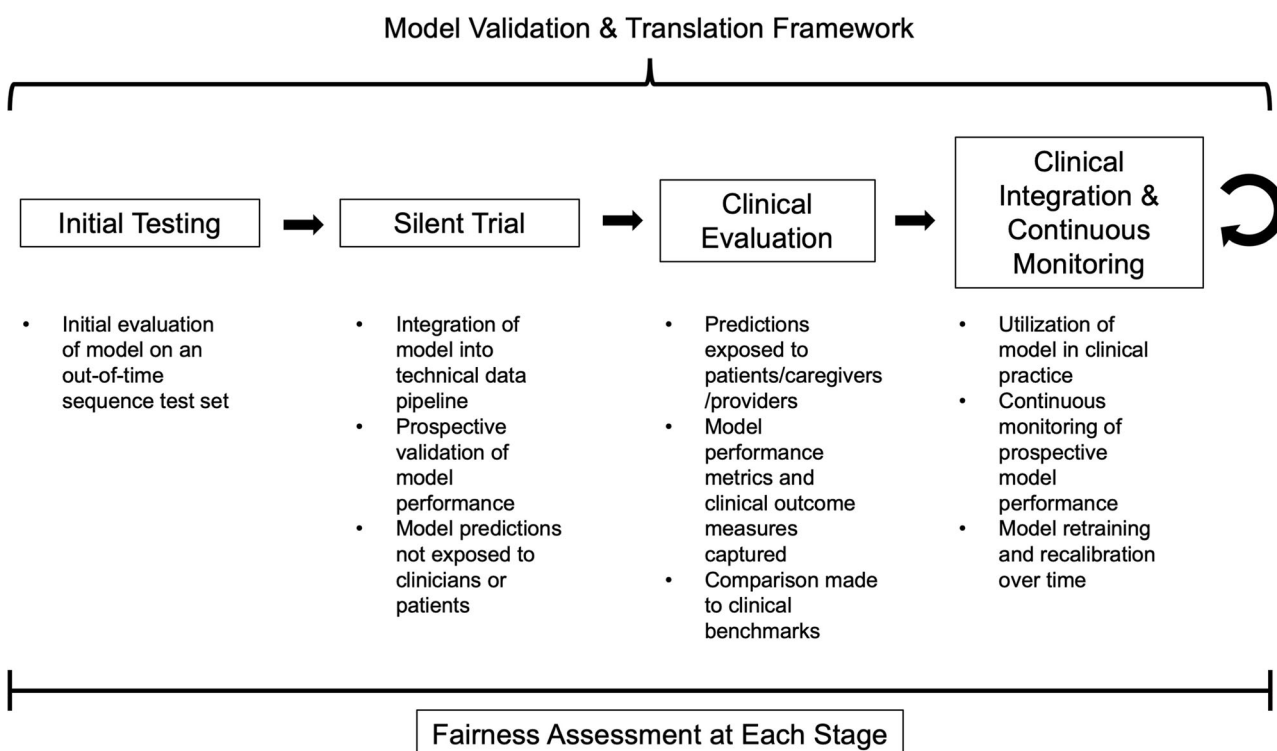
Machine learning tasks	Description	Common types of models*	Clinical example
Classification	Classify data into pre-specified groups	Logistic regression, support vector machines, random forests, neural networks	Predict critical care usage and hospitalization in children presenting to emergency departments [20]
Regression	Predict a continuous variable outcome	Linear regression, random forests, neural networks	Predicting fracture healing time in children [21]
Clustering	Completed without labeled data in order to categorize patients into either known groupings or novel subgroups	K-means clustering, Gaussian mixture models	Discovering phenotypes of pediatric asthma on the basis of asthma control [22]
Outlier detection	Unsupervised approaches to detecting deviations from distributions in data	One-class support vector machines, Bayesian networks, hidden Markov models, isolation forests	Detecting copy number variations in patients with Tetralogy of Fallot [23]
Time-series models	Modeling longitudinal data sampled over multiple time points	Recurrent neural networks, long short-term memory networks	Prediction of cardiac arrest from continuous physiological signals over time in pediatric intensive care unit patients [24]
Reinforcement learning	Modeling the optimal action to take in order to maximize a reward in response to environmental changes	Markov decision processes, Q-learning	Determining optimal and individualized treatment suggestions for septic patients in intensive care units [25]
Image-based models	Models involving medical image data (i.e., X-rays, MRIs, and CT)	Convolutional neural networks, generative adversarial networks	Diagnosing and providing treatment suggestions for congenital cataracts [26]
Natural language processing	Processing and analyzing language-based data (i.e., electronic medical records)	Latent Dirichlet allocation, long short-term memory networks, transformer models	Extracting text from an EMR to predict a patient's diagnosis, assisting in triage or aiding in complex cases [27]

\*Does not include all types of models for each category as new modeling techniques are rapidly evolving

## Initial testing

The statistical outcome metrics used throughout these stages will vary widely depending on the clinical prediction task being evaluated. These metrics include, but are not limited to, the following: precision, recall, area under the receiver operator curve, sensitivity, specificity, and accuracy [30]. Each has their own unique advantages and disadvantages that should be viewed holistically when evaluating ML models as opposed to giving sole value to any one “universal” metric [31] [32].

Initial statistical outcome metrics used to evaluate an ML model should ideally be generated on a non-random, out-of-time sequenced set of data. If using a dataset that contains a single year's worth of EHR patient data, a model might be trained using data from patients who presented between January and



**Fig. 2.** A framework of prospective studies to consider when focusing on model validation in pediatrics. From start to finish: initial testing, silent trial, and clinical evaluation, ending with clinical integration and continuous monitoring. We suggest fairness assessments occur throughout each stage of the pipeline.

October, then tested on subsequent patients who presented in November and December. This approach would allow for an assessment of how the model might behave when making predictions with respect to future patients. Most importantly, out-of-time validation allows for an initial estimation of model performance when factoring in seasonality effects and environmental shifts in the distribution of patient data.

### Silent trial

Conducting a silent trial enables further prospective validation and is the next safe step forward toward translation. A silent trial involves integrating a developed ML model into a data pipeline (e.g., EMRs) in real-time such that data can be passed into the model and predictions can be made at a frequency that directly represents how the model will be utilized in clinical practice. These predictions are made in the background (i.e., silently) without being disclosed to patients and/or their providers and does not influence current patient care. Statistical outcome metrics of the model's ongoing prospective performance should be captured and repeatedly evaluated to ensure that it maintains performance over time. During this phase, the integrity of data streams, network speeds, computational capacity, and model latency can also be evaluated. These technical considerations are important as they directly impact the evaluation and usability of ML tools in practice.



## Clinical evaluation

After a silent trial demonstrates that a model behaves well prospectively, clinical evaluation can be undertaken as needed to determine the impact the model will have on patient and provider outcomes. The majority of ML4H research consists of proof-of-concept models or systems built on retrospective cohorts [33] that are beneficial for rapid prototyping and development. However, retrospective analysis does not offer researchers the same insights as well-designed prospective studies based on local cohorts. Prospective studies are vital for ensuring that retrospective validity translates into real clinical impact. The structure of the clinical evaluation needed will depend on the clinical use case in question and may involve both traditional research (e.g., prospective cohort study, randomized control trial, RCT) and quality improvement (QI) methodologies. The procedures and study designs appropriate in fulfilling this task will vary based on the complexity of the task and the level of risk associated with the model's implementation [34].

Prior to conducting a clinical evaluation, patient risk should be reassessed based on the outcome metrics obtained during the silent trial—with these results directly informing approval from research ethics boards. Issues to consider with clinical trials in ML4H largely mirror traditional trials: studies must be sufficiently powered for clinical endpoints, comparisons must be made to best available practice (e.g., current standard of care), and the objective of the trial (e.g., demonstrating superiority, non-inferiority, or equivalence) should align with the design and analytic methods used. Finally, researchers have pointed out that randomization, done to balance known and unknown confounders between treatment groups, can be difficult to implement with ML applications that change clinical workflows [35•]. Such challenges can be ameliorated with pragmatic, stepped-wedge cluster designs that allow for an increase in the number of clusters that are exposed to an intervention over time [35•], [36].

Prospective cohort studies of ML4H applications in pediatrics are more common than RCTs, the “gold-standard” in clinical medicine, but are still rare compared with retrospective studies. Examples of these studies include predicting disease trajectory in children with juvenile idiopathic arthritis, identifying neuroanatomical vulnerability in youth at high risk for psychosis, and detecting autism from home videos [37], [38], [39]. Only one RCT of an ML4H application had occurred in pediatrics by early 2020—a Chinese study on a previously published system for diagnosing and providing treatment recommendations for cataracts [40].

The lack of clinical trials in ML4H may be partly explained by the different publishing strategies that apply in computer science as compared with those in medicine. Computer science places greater emphasis on publishing in conference proceedings than do most other academic disciplines [41]. Whereas prospective studies and RCTs can take several years to design, recruit for, and publish in peer-reviewed journals, conference publication cycles take place every few months. As outlined throughout this piece, the development of ML4H systems is a collaborative effort, and discussions need to be had between stakeholders across disciplines on the advantages of publishing in different venues depending on the stage of the project.



Once clinical trials for ML4H tools are established and begin to move into later phases requiring more participants, a unique consideration in pediatrics is the concentration of patients in highly specialized, often urban, tertiary care centers [42]. To bolster recruitment, prospective studies may need to become multicenter, which has important implications due to dataset shift. Special attention must be paid to the training data used, and the generalizability of the model when an ML4H application becomes used across different centers, as changes in underlying statistical distributions can substantially decrease model performance [43]. External model validation has been commonly completed with various clinical risk scores created using traditional epidemiological methods, and is becoming more commonplace in ML4H [44]. This external model validation is especially important because a major critique of ML methods is that there is a risk of “overfitting” or memorizing training data such that the accuracy of a model may not be sustained across sites containing variations in patient distributions.

### Continuous monitoring

In order to ensure that model performance is maintained, statistical outcome metrics should continue to be assessed even after an ML tool is implemented [45]. The interval of how frequently this is done should reflect the potential risk associated with implementation of the model. Ideally, a software is developed to continuously monitor relevant outcome metrics, and alarms/flags are raised when performance declines. Failure to undertake ongoing assessment could lead to a decline in model accuracy because patient features, including corresponding distributions of data and trends in children, may change over time. Continuous auditing of the model’s performance is a proactive way to address this concern while simultaneously gathering information about how frequently model retraining and recalibration should be completed [46].

### Fairness assessment

An assessment of outcome metrics across gender, age, ethnicity, socioeconomic status, and geography is strongly advocated for at each stage of model validation to ensure equitable model performance across all subgroups. Failure of a model to perform on a select subgroup may reflect an underlying deficiency or bias within the dataset. Implementing a model without accounting for these performance inequities may unintentionally contribute to socioeconomic disparities in pediatric healthcare rather than improve upon them [47].

## User validation

It is essential for user validation testing to be incorporated into the pipeline when building an ML model for clinical integration, in order to ensure that associated clinician and patient user experiences are positively impacted [48] [49]. From a human computer interaction perspective, the needs of the end-user should be heavily factored into the evaluation of the clinical utility of an AI tool. Many medical innovations fail to adequately consider these needs and cannot be effectively integrated into clinical practice as a result [50]. Similar to the engagement in *Pediatric Clinical Use Case Design*, circling back to design thinking methodology at this time provides an excellent framework for re-engagement of

all stakeholders. This will help to ensure the solutions developed are usable and yield both quantitative and qualitative improvements in patient care. Machine learning scientists, interprofessional clinical staff, children, and families all have a role to play in effectively designing and implementing useful AI tools in a pediatric setting [35•].

## Clinical integration

Integrating a successfully validated ML model into clinical practice represents the final hurdle to overcome prior to attaining meaningful clinical impact from an ML4H project [45•]. Augmenting clinical workflows such that patients, their families, and clinicians each obtain value from the new process is key to user uptake and satisfaction [51•]. Ignoring this concern at the use case design stage and again at time of clinical integration is known to contribute to the failure of technology innovation in healthcare [52] [53]. Features associated with successful integration of new technology include the following: automation of use, providing customizable and specific recommendations rather than just alerts, and providing information at the time and location of decision-making [54].

The implementation of effective change management strategies also contributes to success by proactively addressing issues associated with provider resistance [55]. Indifference of healthcare professionals and lack of motivation is known to contribute to poor organizational adoption of new technologies [56]. This often stems from lack of confidence in the tool's performance and workflow disruption. Anticipating these challenges and addressing them head-on can improve ease of integration.

Successful clinical integration is also associated with a hospital's ability to effectively execute QI initiatives [57]. Having a QI focus at this stage enables ML4H project teams to iterate through *plan, do, study, act* (PDSA) cycles, in order to measure the impact of clinical integration on both primary outcomes and counterbalancing measures [58]. PDSA cycles also allow for review and re-adjustment of integration approaches as needed until target levels of engagement and success are achieved.

## Legal, privacy, and ethical considerations

As the technical science continues to advance, researchers are also working to identify and address the ethical and legal challenges that arise when using AI in various healthcare settings. Since much of the work in ML4H is taking place within adult healthcare settings, so too has the bulk of the related social scientific work. Among the ethical and legal issues being explored are:

- concerns about how data is collected
- whether that data contains biases
- fairness and equity regarding who will benefit
- how to adequately and ethically test and regulate ML tools
- where liability should lie for harm that results from reliance on ML

- whether and when healthcare institutions might have a moral or legal duty to inform patients, staff, and/or hospital users about monitoring, data collection, and the use of predictive analytics to inform administrative and/or clinical decision-making

In the pediatric context, concerns about privacy and consent in particular are more nuanced and take on greater significance. This includes complex issues around surrogate decision-making. Given the data-intensive nature of modern medicine, how we collect pediatric data and obtain consent for its secondary use is very important. This is particularly true if we wish to work toward building larger local or site-specific pediatric datasets. Obtaining blanket authorization for secondary use of data from a surrogate, although legally acceptable, is qualitatively different than actually obtaining a patient's informed consent. It is for this reason that the ethical and legal norms that govern research generally maintain that consent is an ongoing process [59]. Furthermore, our normative and legal frameworks work from the premise that data should only be shared with a deep "respect for the context in which it was collected" (e.g., to help advance research into a particular disease). Machine learning challenges this premise because it looks for things we cannot see or predict. If we want informed consent to remain meaningful in a world of big data, we must find ways to explain what analytics are expected or likely to do [60].

Until it is feasible to provide a specific and meaningful explanation to patients and their proxies about what we expect from data analytics, re-contacting children (e.g., once they reach legal adulthood or otherwise gain the requisite capacity) for ongoing permission to use of their data shows respect for that child's autonomy and their evolving maturity [61]. Researchers should ideally address the topic of re-contact when children are first enrolled or provide consent for their data to be used in research [59]. That being said, there remains some debate in the literature about whether re-contact is always appropriate given logistical challenges, the scope of parental authority, and what the actual justification for the re-contact is [61]. Regardless of how one chooses to tackle the challenge of re-contact, ensuring that children retain the right to withdraw consent for the use of their data is an ethically meaningful practice that should be undertaken whenever possible [61].

Some creative solutions to this challenge of re-consent in pediatric data sharing have also been proposed. One possible approach could be to move away from using the language of property law when we talk about EHR data and to re-think whose data it is that we are referring to. We might re-imagine EHR data as being *about* patients instead of *belonging to* patients and consider this data to be co-constructed "through a collaborative process involving the patient and the clinician, with support from other professionals within the health system" [62]. Under such a re-imagining, an alternative approach to consent might involve exploring different models of collective data governance that include patients, families, healthcare professionals, and stakeholders from different relevant communities. That governance community could make collective decisions about how individual data sets can be used. Patients and/or their proxies could be told about this data governance model at the time consent is sought for the collection and use of their data, and this infrastructure

could help allay concerns about the need to re-contact and re-consent individuals as they gain capacity.

## Conclusion

The application of AI and ML in pediatric medicine presents a range of unique considerations, from project ideation to implementation. In this paper, we highlight the different stages of effectively building and implementing ML models in pediatrics. Having a robust understanding of how ML is different in pediatrics will allow for the effective design of solutions by clinicians and data scientists in collaboration with patients, families, and caregivers.

## Availability of Data and Material

Not applicable.

## Code Availability

Not applicable.

## Compliance with Ethical Standards

### Conflict of Interest

Sujay Nagaraj declares that he has no conflict of interest.  
 Vinyas Harish declares that he has no conflict of interest.  
 Liam G. McCoy declares that he has no conflict of interest.  
 Felipe Morgado declares that he has no conflict of interest.  
 Ian Stedman declares that he has no conflict of interest.  
 Stephen Lu declares that he has no conflict of interest.  
 Erik Drysdale declares that he has no conflict of interest.  
 Michael Brudno declares that he has no conflict of interest.  
 Devin Singh declares that he has no conflict of interest.

## References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. • Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;
2. Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. *Nat. Med.* 2020.
3. M. H, I. B, C. H, Q. N, G.F. C, S. V. Outlier-based detection of unusual patient-management actions: an ICU study. *J. Biomed. Inform.* 2016.
4. Wachter RM, Cassel CK. Sharing health care data with digital giants: overcoming obstacles and reaping benefits while protecting patients. *JAMA - J Am Med Assoc.* 2020.
5. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-learning-based, computer-

- aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol*. 2019;180:373–81.
6. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
  7. Eichner JM, Johnson BH, Betts JM, Chitkara MB, Jewell JA, Lye PS, et al. Patient- and family-centered care and the pediatrician's role. *Pediatrics*. 2012.
  8. Jibb LA, Cafazzo JA, Nathan PC, Seto E, Stevens BJ, Nguyen C, et al. Development of a mHealth real-time pain self-management app for adolescents with cancer: an iterative usability testing study. *J Pediatr Oncol Nurs*. 2017;34:283–94.
  9. McLaughlin JE, Wolcott MD, Hubbard D, Umstead K, Rider TR. A qualitative review of the design thinking framework in health professions education. *BMC Med Educ*. 2019;
  10. Altman M, Huang TTK, Breland JY. Design thinking in health care. *Prev Chronic Dis*. 2018;
  11. Goulooze SC, Zwep LB, Vogt JE, Krekels EHJ, Hankemeier T, van den Anker JN, et al. Beyond the randomized clinical trial: innovative data science to close the pediatric evidence gap. *Clin Pharmacol Ther*. 2019.
  12. Joseph PD, Craig JC, Caldwell PHY. Clinical trials in children. *Br J Clin Pharmacol*. 2015;79:357–69.
  13. Johnson AEW, Ghassemi MM, Nemati S, Niehaus KE, Clifton D, Clifford GD. Machine learning and decision support in critical care. *Proc IEEE*. 2016;
  14. Zeng X, Yu G, Lu Y, Tan L, Wu X, Shi S, et al. PIC, a paediatric-specific intensive care database. *Sci Data*. 2020;7:14.
  15. Raval MV, Dillon PW, Bruny JL, Ko CY, Hall BL, Moss RL, et al. Pediatric American College of Surgeons National Surgical Quality Improvement Program: feasibility of a novel, prospective assessment of surgical outcomes. *J Pediatr Surg*. 2011;46:115–21.
  16. Forrest CB, Margolis PA, Charles Bailey L, Marsolo K, Del Beccaro MA, Finkelstein JA, et al. PEDSnet: A national pediatric learning health system. *J Am Med Informatics Assoc*. 2014;
  17. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform*. 2015.
  18. Fleming S, Thompson M, Stevens R, Heneghan C, Plüddemann A, MacOnochie I, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *Lancet*. 2011;377:1011–8.
  19. Lipstein EA, Brinkman WB, Fiks AG, Hendrix KS, Kryworuchko J, Miller VA, et al. An emerging field of research: challenges in pediatric decision making. *Med Decis Mak*. 2015;35:403–8.
  20. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw open*. 2019;
  21. Malek S, Gunalan R, Kedija SY, Lau CF, Mosleh MAA, Milow P, et al. Random forest and Self Organizing Maps application for analysis of pediatric fracture healing time of the lower limb. *Neurocomputing*. 2018;
  22. Ross MK, Yoon J, Van Der Schaar A, Van Der Schaar M. Discovering pediatric asthma phenotypes on the basis of response to controller medication using machine learning. *Ann Am Thorac Soc*. 2018;
  23. Bansal V, Dom C, Grunert M, Klaassen S, Hetzer R, Berger F, et al. Outlier-based identification of copy number variations using targeted resequencing in a small cohort of patients with tetralogy of fallot. *PLoS One*. 2014;
  24. Tonekaboni S, Mazwi M, Laussen P, Eytan D, Greer R, Goodfellow SD, et al. prediction of cardiac arrest from physiological signals in the pediatric ICU. *Mlhc-2018*. 2018;
  25. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*. 2018;
  26. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nat Biomed Eng*. 2017;
  27. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med*. 2019.
  28. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22:1345–59.
  29. Liang G, Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput Methods Programs Biomed*. 2020;
  30. Weng W-H. *Machine Learning for Clinical Predictive Analytics*.
  31. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019.
  32. Thomas R, Uminsky D. The problem with metrics is a fundamental problem for AI. 2020 [cited 2020 Mar 27]; Available from: <http://arxiv.org/abs/2002.08512>
  33. Beaulieu-Jones B, Finlayson SG, Chivers C, Chen I, McDermott M, Kandola J, et al. Trends and focus of machine learning applications for health research. *JAMA Netw Open*. 2019;2:e1914051.
  34. Maddox TM, Rumsfeld JS, Payne PRO. Questions for artificial intelligence in health care. *JAMA - J. Am. Med. Assoc*. 2019.
  35. • Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;
- A roadmap for machine learning in healthcare, with a focus on responsible development at all stages of the pipeline. Important considerations for translating machine learning to clinical practice.



36. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350.
37. Eng SWM, Aeschlimann FA, van Veenendaal M, Berard RA, Rosenberg AM, Morris Q, et al. Patterns of joint involvement in juvenile idiopathic arthritis and prediction of disease course: a prospective study with multilayer non-negative matrix factorization. *PLoS Med*. 2019;16:e1002750.
38. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, et al. Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. *JAMA Psychiatry*. 2018;
39. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: a development and prospective validation study. *PLoS Med*. 2018;15.
40. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine*. 2019;9:52–9.
41. Vrettas G, Sanderson M. Conferences versus journals in computer science. *J Assoc Inf Sci Technol*. 2015;66:2674–84.
42. Casimir G. Why children's hospitals are unique and so essential. *Front. Pediatr*. 2019
43. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics*. 2019.
44. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafiyan H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;
45. • Sendak M, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, et al. A path for translation of machine learning products into healthcare delivery. *EMJ Innov*. 2020; **Review highlighting examples of machine learning products that have been translated in healthcare. Specifically, models that input data from EHRs applied to decision support tasks. Focusing on design, development, validation, and scaling. Paper seeks to unify the translational path to inform future translation efforts by identifying common challenges and suggestions for improvement.**
46. Lee J, Kramer BM. Analysis of machine degradation using a neural network based pattern discrimination model. *J Manuf Syst*. 1993;12:379–87.
47. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern. Med*. 2018, Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data.
48. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016.
49. Dekker FW, Ramspek CL, Van Diepen M. Con: Most clinical risk scores are useless. *Nephrol. Dial. Transplant*. 2017.
50. Poncette AS, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, et al. Clinical requirements of future patient monitoring in the intensive care unit: qualitative study. *J Med Internet Res*. 2019.
51. • Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes. *ACM*; 2019 [cited 2020 Mar 27];11. Available from: <https://doi.org/10.1145/3290605.3300468>. **This paper provides insights into how to effectively design AI-based clinical decision tools for the end-user. They provide lessons on prototyping AI systems as a situated experience.**
52. Devaraj S, Sharma SK, Fausto DJ, Viernes S, Kharrazi H. Barriers and facilitators to clinical decision support systems adoption: a systematic review. *J Bus Adm Res*. 2014.
53. Gravel K, Légaré F, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: a systematic review of health professionals' perceptions. *Implement. Sci*. 2006.
54. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Br Med J*. 2005;330:765.
55. Ranjan S, Jha VK, Pal P. Literature review on ERP implementation challenges. *Int. J. Bus. Inf. Syst*. 2016
56. Elwyn G, Scholl I, Tietbohl C, Mann M, Edwards AG, Clay C, et al. "Many miles to go.": A systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC Med Inform Decis Mak*. 2013.
57. Tyagi RK, Cook L, Olson J, Belohlav J. Healthcare technologies, quality improvement programs and hospital organizational culture in Canadian hospitals. *BMC Health Serv Res*. 2013;13.
58. Jones B, Vaux E, Olsson-Brown A. How to get started in quality improvement. *BMJ*. 2019;
59. Knoppers BM, Sénécal K, Boisjou J, Borry P, Cornel MC, Fernandez C V., et al. Recontacting pediatric research participants for consent when they reach the age of majority. *IRB Ethics Hum Res*. 2016;
60. Froomkin AM. Big data: destroyer of informed consent. 2019.
61. Rahimzadeh V, Schickhardt C, Knoppers BM, Sénécal K, Vears DF, Fernandez C V., et al. Key implications of data sharing in pediatric genomics. *JAMA Pediatr*. 2018.
62. Ballantyne A. How should we think about clinical data ownership? *J Med Ethics*. 2019.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.