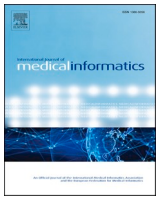
International Journal of Medical Informatics 209 (2026) 106236

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Review article

Synthetic data generation in paediatrics and paediatric nursing: what, how, and why?

Elisabetta Mezzalira [a,*], Maria Paola Boaro [b], Giulia Reggiani [a,b], Riccardo Biondi [c], Gastone Castellani [c], Raffaella Colombatti [a,b]

[a] Department of Women's and Children's Health, University of Padova, Padova, Italy
[b] Pediatric Hematology-Oncology and Bone Marrow Transplant Unit, Department of Woman's and Child's Health, Azienda-Ospedale-Università di Padova, Padova, Italy
[c] IRCCS-Istituto delle Scienze Neurologiche, Bologna, Italy

ARTICLE INFO

Keywords:
Pediatrics
Rare diseases
Paediatric Nursing
Synthetic Data
Synthetic Data Generation
Artificial Intelligence

ABSTRACT

Introduction: This paper explores the potential benefits and limitations of synthetic data (SD) in paediatrics, addressing the challenges of data scarcity and privacy concerns in paediatric research.
Methodology: A narrative literature review was conducted, searching PubMed and Scopus databases for relevant publications up to August 2025. The review focused on studies addressing the use, development, or application of SD in paediatric healthcare settings.
Findings: Synthetic data offers numerous benefits in paediatrics, including enhancing dataset diversity, protecting patient privacy, and enabling AI model development, especially in areas with limited real datasets such as rare diseases. Applications of SD in paediatrics span various fields, including neonatology, oncology, radiology, and neurodevelopmental disorders. However, challenges persist, including potential data bias, ensuring accuracy and quality, privacy concerns, and the lack of standardized guidelines for data generation and validation.
Conclusions and future directions: While SD demonstrates potential in specific paediatric applications, such as improving AI early warning systems and augmenting datasets for rare conditions, its use requires a structured, actionable framework for evaluation. Future efforts should focus through multi-stakeholder engagement, on developing paediatric-specific guidelines, ensuring fair and safe use of SD, and addressing unique aspects of child development in data synthesis.

What was already known on the topic

- Lack of diverse, balanced and large datasets is a challenge for developing AI models in healthcare
- Children and adolescents are underrepresented in research due to ethical, legislative, financial, and relational challenges
- Synthetic data generation can help overcome privacy concerns and data scarcity issues
- What this study adds provides an overview of synthetic data generation methods, applications, and potential benefits in the paediatric context
- Highlights existing examples of using synthetic data in neonatology, medical imaging, nursing and other paediatric areas
- Discusses the importance of developing guidelines, evaluation frameworks, and multi-stakeholder approaches for fair and safe use of synthetic data in paediatrics
- Emphasizes the need to address biases, ensure proper anonymization, and consider the relevant regulations for synthetic data use in paediatrics

1. Introduction

Diverse, balanced and large datasets are necessary to develop and refine best practices in evidence-based medicine involving AI. However, researchers struggle with paucity of annotated medical data in real-world settings. To overcome this challenge, the use of SD is increasing [1]. Their use can be of particular interest in the paediatric population, considering that, as it is highlighted from different authors, children and adolescents (<18) are underrepresented in research, with paediatric studies presenting age-specific challenges that span ethical, legislative, financial, and relational domains [2,3].

2. Methodology

A narrative literature review was conducted to explore the applications, benefits, and limitations of synthetic data (SD) in pediatrics.

* Corresponding author at: Department of Women's and Children's Health, University of Padova, Padova, Italy.
E-mail address: elisabetta.mezzalira@unipd.it (E. Mezzalira).

https://doi.org/10.1016/j.ijmedinf.2025.106236
Received 24 June 2025; Received in revised form 12 December 2025; Accepted 22 December 2025
Available online 27 December 2025

Relevant publications up to August 2025 were identified through searches in PubMed and Scopus databases, using a combination of keywords and MeSH terms related to "synthetic data", "data simulation", "artificial data generation", "healthcare", "medicine", "pediatrics", and "children".

The eligibility criteria included original studies, reviews, and reports addressing the use, development, or application of SD in pediatric or general healthcare settings, with no restrictions on publication type, study design, or country of origin. Only English-language articles were considered. Titles, abstracts, and full texts were screened to identify studies discussing the generation, validation, or application of SD in pediatric healthcare. Reference lists of included studies were also examined for additional relevant sources. No formal quality assessment or quantitative synthesis was performed. Instead, a narrative synthesis was conducted to provide an overview of the state of knowledge, highlighting the potential benefits, limitations, and future research directions regarding the use of SD in pediatrics, based on the findings from the selected literature.

2.1. Findings from the literature

WHAT: SD have been often defined as "data that has been generated using a purpose-built mathematical model or algorithm, with the aim of solving a (set of) data science task(s)" [4]. In lay terms they could be defined as artificially generated information that mimics real-world data (RWD) statistical properties. Instead of using actual data collected in the real world such as clinical data coming from Electronical Healthcare Records (EHRs), SD are created by computers to look and behave like real data. This can include numbers, text, images, or even video that are produced based on patterns observed in real data but do not contain any actual real-world information. An example of this is constituted by the advent of deep-fake medical videos and images [5]. It is possible to synthetize tabular data, text and media. In terms of general application in practice, they help researchers and developers to test and improve technologies, like AI models, without risking privacy or dealing with the limitations of real-world data. In fact, some estimates are now predicting that within less than 10 years "SD will completely overshadow real data in AI models" [6].

In medicine and healthcare accurate SD can have multiple useful purposes, such as enhance dataset diversity and boost the robustness and adaptability of AI models, as well as augment the potential of a dataset. Moreover, they can be used to protect individuals' privacy and speed up the testing of products [1].

HOW: A typical SD generation pipeline, with its potential application is showed in Fig. 1. This pipeline starts from real data (Fig. 1 A) which should be harmonized; those data are used to train a generative model (Fig. 1 B) to produce a SD. The generation of synthetic data typically involves training a generative model, such as a statistical model, machine learning algorithm, or deep neural network, on a real-world dataset. This generative model learns the underlying statistical patterns and relationships within the training data. Once trained, the model can then generate new synthetic data points that mimic the statistical properties of the original dataset while not containing any actual real-world information, thereby preserving privacy. The synthetic data can be fully artificial, partially synthetic with some real variables retained, or a hybrid combination of real and synthetic data, each approach offering different trade-offs between privacy and analytical utility [7,8]. This dataset could be used alone or in combination with other real data (Fig. 1 C) to train another AI model (Fig. 1 D) to perform a specific task (i.e. classification, regression, etc.).

Rubin introduced fully SD in 1993, which is entirely artificial and offers strong privacy but low analytic value due to data loss [9]. Partially-SD entails the presence of a real dataset, where only sensitive variables are replaced with synthetic versions. This type retains some original data, posing a reidentification risk and excluding the possibility to analyse rare variables or features present in the real dataset [10]. Hybrid synthetic data combines real and SD, balancing privacy with higher utility, but demands more processing time and memory [8]. Hybrid SD generation has now started to be used to tackle the scarcity and fragmentation of data and widen the basis for GDPR-compliant research in situations where the opportunity to collect real data is scarce. An example of their potential lies in rare diseases such as Sickle Cell Disease and many haematological diseases, or in cases with limited amount of data, such as in pre-term babies. The production of SD entails the use of a machine-learning algorithm or a neural network, which analyses a real data set and learns about the statistical relationships within it. It then creates a new data set containing different data points than the original but retaining the same relationships.

WHY: SD generation is expected to have a great impact in the healthcare sector over the next years, because they present potential to overcome many current challenges such as patient privacy, access to databases for research purposes, lack of availability of real data in specific fields (i.e. rare diseases) and ethical challenges. Fragmentation and scarcity of data are common challenges in paediatric care, and they are particularly pronounced in the context of rare diseases and rare complications [11–13]. Several factors contribute to this problem, including the lack of interoperability between hospital and paediatric primary care health records and the ethical complexities of conducting studies in paediatric populations [14,15]. These challenges limit the number of clinical studies and trials involving children [14]. In addition,

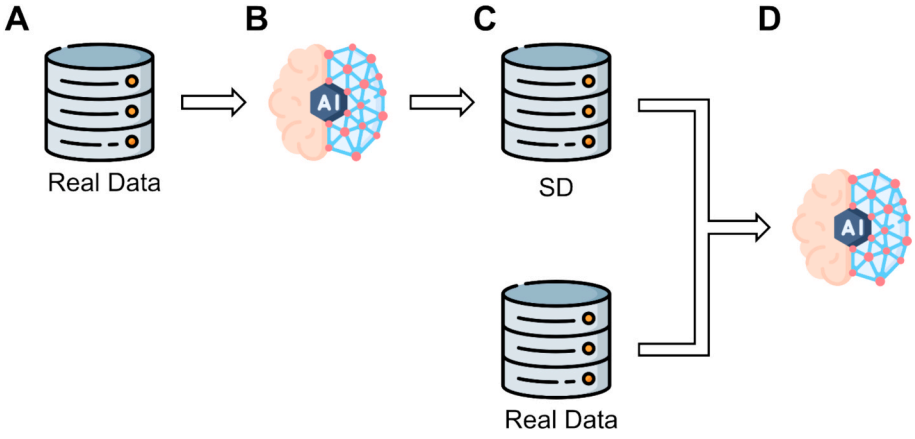


Fig. 1. Schematic of a typical SD generation and application pipeline. Starting from **A)** Real Data, which should be harmonized, it is possible to develop and train a **B)** generative model, which could be a statistical, machine learning or deep learning model. This model is used to gnerate a **C**) Syntetic Dataset (SD) which could be used alone or to integrate real data to train an **D)** AI model, useful in clinical practice.

patients with complex conditions are often dispersed across peripheral hospitals and specialized reference centres, further contributing to data fragmentation [16]. Data concerning minors are also under the legal authority of their parents or guardians, who may choose to opt out of data sharing for various reasons such as concerns about stigma, data security, or the potential burden placed on the child when participating in clinical studies [3,17]. Specialized networks and strategies, such as rare disease data registries and the European Reference Networks, have recently emerged to address these issues [12]. However, the data they collect often remain insufficient to answer key clinical questions, evaluate treatment effectiveness, or support robust predictive modelling [2]. This is why synthetic data represent a crucial area for future development in this field. SD diverse applications in healthcare includes estimating the benefits of screenings, policies, treatments, and interventions [18]. Davis et al. [19] demonstrated its use in policy analysis by creating a synthetic dataset to explore healthcare service effects under demographic aging, simulating scenarios like morbidity and General Practitioner behaviour. Other groups used synthetic data to train Natural Language Processing models for predicting mental health diagnoses from patient discharge reports, while safeguarding sensitive information [20] The existing examples of use of SD in paediatrics cover different fields and show promising results (Fig. 2).

Diagram that defines areas of potential synthetic data integration in pediatric practices In neonatology, Braddon et al. [21], examined whether synthesised data can replicate two prenatal epidemiological associations: between prenatal smoking and lower birthweight, and between prenatal mood disorders and lower birthweight, using data synthesised from de-identified health administrative data collections. Both synthesised datasets performed well in replicating the statistical properties of the original data while addressing privacy issues. Coyner et al.[22] attempted to synthetize highly realistic images to augment the size and diversity of image-based datasets for retinopathy of prematurity (ROP), a potentially blinding disease, applying generative adversarial networks (GANs). This synthetic approach has been proved useful by the authors to increase the availability of retinal fundus image (RFIs) datasets for research while overcoming challenges not only for privacy concerns, but especially in fragile patient populations such as preterm infants. Indeed, Braddon et al., and Coyner et al., highlight the opportunity of using this methodology in rare conditions, that most likely would require for researchers several years to reach meaningful statistical power to run analyses. In another example, Liaqat et al.[23] leveraged both real data and SD to develop an automatic screening tool based on eye gaze data that could identify Autistic Spectrum Disease risk. The AI algorithm reached 67.23% accuracy on the validation dataset. This underlines the opportunity to use synthetic data in different types of vulnerable populations, not only in case of rare diseases but also in neurodevelopmental conditions.

However, to date most efforts in the paediatric field have focused on the generation of synthetic medical images, either Magnetic Resonance Images (MRIs) or CT scans [24–28]. SD images generation has been applied to different specialties. In radiology, it has been used for patients with cystic fibrosis to create synthetic CT scans from Lung MRI with ultrashort echo times (UTEs), in order to enable high-resolution and radiation-free morphologic imaging [25]. Synthetic MRI has also been tested in children with tuberous sclerosis complex, showing that Synthetic MRI enables the detection of cortical tubers and is a developing tool in the quantification of morphometric and tissue alterations in paediatric TSC patients with a rational scanning time [26]. Moreover, in radiotherapy [24,28], it has been researched to enable accurate and improved magnetic resonance imaging (MRI)-based dose calculations. This confirms the potential of SD in the world of medical images, and opportunity to reduce the need of exposing patients to imaging for research purposes.

Among the different healthcare professionals, paediatric nurses could greatly benefit from SD generation. AI early warning systems are starting to reduce adverse events [29]. The system's ability to predict deterioration in a child's health system status hours before visible symptoms is dependent on the magnitude and characteristics of the dataset, and it can lead to more timely interventions. Rare diseases present limitations in terms of data availability, however SD generation can enhance the potential of existing datasets and provide future solutions in terms of AI early warning systems and predicting algorithms. Moreover, performing data augmentation through SD could allow the training of more accurate AI system, improving the automation of routine tasks, enabling healthcare professionals to allocate more time to direct patient care. Notably, AI-based monitoring systems can cut the time spent on administrative tasks, allowing nurses to concentrate more on patient interactions [30,31]. Data Quality Assurance and Bias

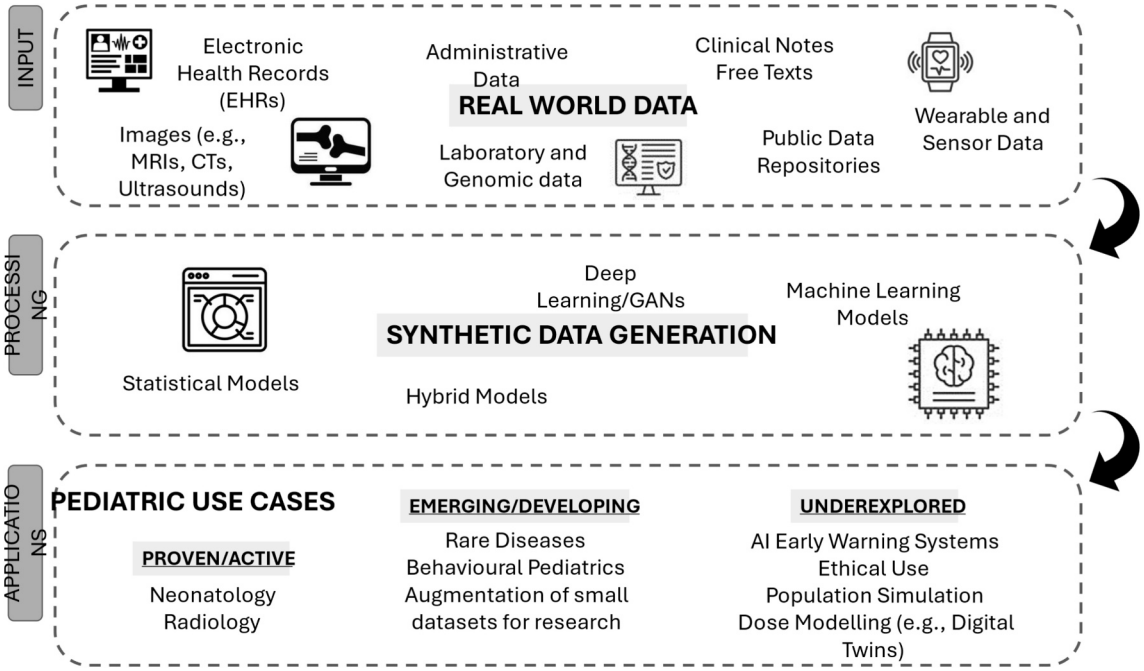


Fig. 2. Existing examples of use of SD in Paediatrics and Paediatric Nursing.

Reduction in Pediatric Synthetic Data.

While synthetic data generation holds promise in healthcare, it is important to highlight that SD generation presents several risks and limitations [18].Table 1 presents an outline of the challenges, their impact for clinical practice and possible mitigation strategies. One major concern is the potential for data bias, as SD may inadvertently replicate biases present in the original datasets, leading to skewed outcomes and impacting the fairness of AI models. Additionally, there are challenges in ensuring the quality and accuracy of SD, which may not fully capture the complexity and variability of real-world medical scenarios. Privacy concerns also arise, as SD might still allow for re-identification of individuals if not properly anonymized. Furthermore, the lack of standardized guidelines for generating and validating SD can result in inconsistencies across different applications, limiting its reliability and widespread adoption in clinical settings [1,22]. In the application of SD, ensuring high data quality and minimizing bias are crucial considerations. Generation models will perpetuate or exacerbate existing biases in original datasets, including those related to gender, ethnicity, socioeconomic factors, and disease distribution [32]. To address these challenges, transparency throughout the data synthesis process is essential, including detailed documentation of the generation methods, algorithms, and source datasets [33]. Effective strategies for enhancing data quality and fairness include utilizing diverse and representative real-world data sources, implementing rigorous validation processes, and engaging multidisciplinary expert panels for review [33]. Validation should encompass statistical measures such as distribution similarity, correlation maintenance, and predictive consistency between the synthetic and original datasets [33]. In the context of pediatric health data, special consideration must be given to the unique aspects of child development. This includes appropriate age stratification, accounting for various developmental stages, and capturing the wide variability in health conditions specific to children [34].

Table 1
Key Challenges and Mitigation Strategies for Synthetic Data in Pediatrics.

CHALLENGE	IMPACT	MITIGATION
Data Bias and Representativeness	Potential for biased AI models that perpetuate health disparities	Employ diverse data sources and implement fairness-aware algorithms (32)
Limited Data Availability	Insufficient training data, especially for rare conditions	Utilize data augmentation techniques and synthetic data generation (32)
Privacy Concerns	Risk of re-identification and breach of patient confidentiality	Apply differential privacy methods and conduct thorough privacy assessments (33)
Lack of Standardization	Inconsistent data quality and difficulty in comparing studies	Develop and adhere to standardized evaluation metrics for synthetic data (33)
Model Explainability	Reduced trust and difficulty in clinical adoption	Focus on developing explainable models and transparent reporting of methodologies (33)
Regulatory Uncertainty	Barriers to implementation in clinical settings	Engage with regulatory bodies to establish clear guidelines for synthetic data use (33,34)

3. Conclusions and future directions

SD generation has the potential to be an important resource in the field of paediatrics research, due to the potential of effectively addressing privacy concerns, ethical issues and paucity of real-world-data existing datasets as well as clinical trial recruitment. However, despite its potential, SD carries risks and limitations, including the possibility of replicating biases from original datasets, which could skew AI model outcomes and reduce fairness and potential improper anonymization, which might still lead to the re-identification of individuals. For this reason, it will be important to develop and use SD in the context of standardized guidelines specific for the paediatric context, ensuring their fair and safe use in paediatrics. Future efforts in synthetic data evaluation should focus on developing a structured, actionable framework that defines both the content and mechanisms of SDG. This framework should include clear criteria for assessing data fidelity, representativeness, and bias mitigation specific to pediatric populations, recognizing developmental, and physiological differences across age groups. Additionally, a multi-stakeholder roadmap, involving clinicians, data scientists, and ethicists should guide the design, validation, and continuous monitoring of synthetic datasets. Concrete mechanisms such as age-stratified benchmarking metrics, transparency protocols for model provenance, and bias detection pipelines ensure that in the future SD will preserve privacy, accuracy and equity across diverse pediatric subgroups, taking into account also regulatory frameworks such as the AI act [26], which came into force in June 2025.

CRediT authorship contribution statement

Elisabetta Mezzalira: Writing – review & editing, Writing – original draft, Methodology. **Maria Paola Boaro:** Writing – review & editing, Conceptualization. **Giulia Reggiani:** Writing – review & editing. **Riccardo Biondi:** Writing – review & editing, Visualization, Methodology. **Gastone Castellani:** Writing – review & editing, Supervision. **Raffaella Colombatti:** Supervision, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: R. C. received grants from Vertex and Agios (to institution); participated on advisory boards/consultancy for Vertex, Pfizer, Novo Nordisk, Forma Therapeutics, Global Blood Therapeutics, AddMedica, and Agios..

The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This narrative review has been conducted as an ancillary activity of the project SYNTHEMA, an initiative funded by the European Union's Horizon Europe Research and Innovation programme under grant agreement N° 101095530.

References

[1] R.J. Chen, M.Y. Lu, T.Y. Chen, D.F.K. Williamson, F. Mahmood, Synthetic data in machine learning for medicine and healthcare, Nat. Biomed. Eng. 5 (6) (2021) 493–497.
[2] E.M. Speer, L.K. Lee, F.T. Bourgeois, D. Gitterman, W.W. Hay, J.M. Davis, et al., The state and future of pediatric research—an introductory overview: the state and future of pediatric research series, Jan 24 [cited 2025 Dec 11]; Available from: Pediatr. Res. (2023) https://www.nature.com/articles/s41390-022-02439-4.
[3] V. Muralidharan, A. Burgart, R. Daneshjou, S. Rose, Recommendations for the use of pediatric data in artificial intelligence and machine learning ACCEPT-AI, Npj Digit. Med. 6 (1) (2023) 166.
[4] Q. Liu, M. Khalil, R. Shakya, Jovanovic J. Scaling While Privacy Preserving: A Comprehensive Synthetic Tabular Data Generation and Evaluation in Learning Analytics [Internet]. arXiv; 2024 [cited 2025 Dec 11]. Available from: https://arxiv.org/abs/2401.06883.
[5] F. Prezja, J. Paloneva, I. Pölönen, E. Niinimäki, S. Äyrämö, DeepFake knee osteoarthritis X-rays from generative adversarial neural networks deceive medical experts and offer augmentation potential to automatic classification, Sci. Rep. 12 (1) (2022) 18573.
[6] D. Shanley, J. Hogenboom, F. Lysen, L. Wee, A. Lobo Gomes, A. Dekker, et al., Getting real about synthetic data ethics: are AI ethics principles a good starting point for synthetic data ethics? EMBO Rep. 25 (5) (2024) 2152–2155.
[7] A. Gonzales, G. Guruswamy, S.R. Smith, Synthetic data in health care: a narrative review. Johnson a, editor, PLOS Digit. Health 2 (1) (2023) e0000082.
[8] H. Surendra, H. Mohan, A review of synthetic data generation methods for privacy preserving data publishing, Int. J. Sci. Technol. Res. 6 (3) (2017) 95–101.

[9] T. Raghunathan, J. Reiter, Rubin D. Multiple imputation for statistical disclosure limitation. J. Off. Statistics. 1(1):19.
[10] J. Reiter, Inference for partially synthetic, public use microdata sets, Surv. Methodol. 29 (2) (2003) 181–188.
[11] A. Collado, M.P. Boaro, S. Van Der Veen, A. Idrizovic, B.J. Biemond, D. Beneitez Pastor, et al., Challenges and opportunities of precision medicine in sickle cell disease: novel european approach by genoMed4All Consortium and ERN-EuroBloodNet, HemaSphere 7 (3) (2023) e844.
[12] M.D.M. Mañú Pereira, R. Colombatti, F. Alvarez, P. Bartolucci, C. Bento, A. L. Brunetta, et al., Sickle cell disease landscape and challenges in the EU: the ERN-EuroBloodNet perspective, Lancet Haematol. 10 (8) (2023) e687–e694.
[13] S. Chatzimatthaiou, F. Bonifazi, A.C. Gimbert, R. Colombatti, F. Cremonesi, A. Glenthøj, et al., HELIOS Action: advancing research, education, and equity in hemoglobinopathies across Europe and beyond, HemaSphere 9 (12) (2025) e70258.
[14] S.L. Furth, Trials and Tribulations — The Challenges of Clinical Trials in Children. NEJM Evidence [Internet]. 2023 Nov 28 [cited 2025 Dec 11];2(12). Available from: https://evidence.nejm.org/doi/10.1056/EVIDe2300280.
[15] F. Mollerus, C. Lynch, H. Bruining, Data interoperability for a systems approach to developmental conditions, Neurosci. Biobehav. Rev. 176 (2025) 106245.
[16] E.A. Flasch, Health equity and children with medical complexity/children and youth with special health care needs: a scoping review, J. Pediatr. Health Care 38 (2) (2024) 210–218.
[17] M.J.S. Beauvais, B.M. Knoppers, Coming out to play: privacy, data protection, children's health, and COVID-19 research, Front. Genet. 12 (2021) 659027.
[18] M. Giuffrè, D.L. Shung, Harnessing the power of synthetic data in healthcare: innovation, application, and privacy, Npj Digit. Med. 6 (1) (2023), 186.
[19] P. Davis, R. Lay-Yee, J. Pearson, Using micro-simulation to create a synthesised data set and test policy options: the case of health service effects under demographic ageing, Health Policy 97 (2–3) (2010) 267–274.
[20] J. Ive, N. Viani, J. Kam, L. Yin, S. Verma, S. Puntis, et al., Generation and evaluation of artificial mental health records for Natural Language Processing, Npj Digit Med. 3 (1) (2020), 69.
[21] A.E. Braddon, S. Robinson, R. Alati, K.S. Betts, Exploring the utility of synthetic data to extract more value from sensitive health data assets: a focused example in perinatal epidemiology, Paediatric Perinatal Epid. 37 (4) (2023) 292–300.
[22] A.S. Coyner, J.S. Chen, K. Chang, P. Singh, S. Ostmo, R.V.P. Chan, et al., Synthetic medical images for robust, privacy-preserving training of artificial intelligence, Ophthalmol. Sci. 2 (2) (2022) 100126.
[23] S. Liaqat, C. Wu, P.R. Duggirala, S. Cheung, S. Ching, C.N. Chuah, S. Ozonoff, et al., Predicting ASD diagnosis in children with synthetic and image-based eye gaze data, Signal Process. Image Commun. 94 (2021 May) 116198.
[24] M. Maspero, L.G. Bentvelzen, M.H.F. Savenije, F. Guerreiro, E. Seravalli, G. O. Janssens, et al., Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy, Radiother. Oncol. 153 (2020 Dec) 197–204.
[25] A. Longuefosse, Raoult J, Benlala I, Denis De Senneville B, Benkert T, Macey J, et al. Generating High-Resolution Synthetic CT from Lung MRI with Ultrashort Echo Times: Initial Evaluation in Cystic Fibrosis. Radiology. 2023 July 1;308(1): e230052.
[26] G. Coban, E. Gumeler, S. Parlak, B. Konuskan, J. Karakaya, D. Yalnizoglu, et al., Synthetic MRI in children with tuberous sclerosis complex, Insights Imaging. 13 (1) (2022 Dec) 115.
[27] C. Andica, A. Hagiwara, M. Hori, K. Kamagata, S. Koshino, T. Maekawa, et al., Review of synthetic MRI in pediatric brains: basic principle of MR quantification, its features, clinical applications, and limitations, J. Neuroradiol. 46 (4) (2019 July) 268–275.
[28] A. Szmul, S. Taylor, P. Lim, J. Cantwell, I. Moreira, Y. Zhang, et al., Deep learning based synthetic CT from cone beam CT generation for abdominal paediatric radiotherapy, Phys. Med. Biol. 68 (10) (2023 May 21) 105006.
[29] L.I. Veldhuis, N.J.C. Woittiez, P.W.B. Nanayakkara, J. Ludikhuize, Artificial intelligence for the prediction of in-hospital clinical deterioration: a systematic review, Crit. Care Explor. 4 (9) (2022 Aug 26) e0744.
[30] N. Bienefeld, E. Keller, G. Grote, AI interventions to alleviate healthcare shortages and enhance work conditions in critical care: qualitative analysis, J. Med. Internet Res. 13 (27) (2025 Jan) e50852.
[31] W. Glover, Z. Li, D. Pachamanova, The AI-Enhanced Future of Health Care Administrative Task Management. Catalyst non-issue content. NEJM Catalyst. 2022.
[32] C. Xiao, E. Choi, J. Sun, Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review, J. Am. Med. Inform. Assoc. 25 (10) (2018 Oct 1) 1419–1428.
[33] S.M. Bellovin, P.K. Dutta, N. Reitinger, Privacy and Synthetic Datasets [Internet]. Law Archive; 2018 [cited 2025 Dec 11]. Available from: https://osf.io/bfqh3_v1.
[34] The EU Artificial Intelligence Act , European Commission. https://artificialintelligenceact.eu/.