frontiers | Research Topics

# Artificial intelligence and machine learning in pediatrics

**Edited by**
Kaya Kuru, Noreen Caswell, Gregory R. Hart, Louis Ehwerhemuepha and Jennifer Quon

**Published in**
Frontiers in Artificial Intelligence

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-7763-9
DOI 10.3389/978-2-8325-7763-9

**Generative AI statement**
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Artificial intelligence and machine learning in pediatrics

**Topic editors**

Kaya Kuru — University of Central Lancashire, United Kingdom
Noreen Caswell — University of Central Lancashire, United Kingdom
Gregory R. Hart — Institute for Disease Modeling (IDM), United States
Louis Ehwerhemuepha — Children's Hospital of Orange County, United States
Jennifer Quon — University of Toronto, Canada

**Citation**

Kuru, K., Caswell, N., Hart, G. R., Ehwerhemuepha, L., Quon, J., eds. (2026). *Artificial intelligence and machine learning in pediatrics*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-7763-9

# Table of contents

05 Editorial: Artificial intelligence and machine learning in pediatrics
Kaya Kuru, Noreen Caswell, Gregory R. Hart, Jennifer Quon and Louis Ehwerhemuepha

08 A systematic review of data sources for artificial intelligence applications in pediatric brain tumors in Europe: implications for bias and generalizability
Alberto Eugenio Tozzi, Ileana Croci, Paul Voicu, Francesco Dotta, Giovanna Stefania Colafati, Andrea Carai, Francesco Fabozzi, Giuseppe Lacanna, Roberto Premuselli and Angela Mastronuzzi

16 Identification of platelet-related subtypes and diagnostic markers in pediatric Crohn's disease based on WGCNA and machine learning
Dadong Tang, Yingtao Huang, Yuhui Che, Chengjun Yang, Baoping Pu, Shiru Liu and Hongyan Li

32 AGMA-PESS: a deep learning-based infant pose estimator and sequence selector software for general movement assessment
Ameur Soualmi, Olivier Alata, Christophe Ducottet, Anne Petitjean-Robert, Aurélie Plat, Hugues Patural and Antoine Giraud

41 Improving the quality of life of children and parents with nocturnal enuresis: the role of health education
Jianrong Liao, Lin Zhu, Danfeng Xie, Xiaomeng Wang and Ping Zhou

52 Home spirometry telemonitoring in pediatric patients with asthma: a mixed study
Dingrong Fan, Ao Shen, Qianqian Wang, Lu Peng, Ruixue Xia and Hengyu Zhou

60 Machine learning algorithms to predict feeding practices during diarrheal disease and its determinants among under-five children in East Africa
Tirualem Zeleke Yehuala, Nebebe Demis Baykemagn and Bewuketu Terefe

72 Parents' understanding and attitudes toward the application of AI in pediatric healthcare: a cross-sectional survey study
Yan-Dan Huang, Shu-Long Zeng, Jie Lin, Zhi-Peng Mao, Qi-Liang Zhang, Gao-Zhong Feng and Qiong-Xia Ou

79 Artificial intelligence in ADHD assessment: a comprehensive review of research progress from early screening to precise differential diagnosis
Cuijie Zhao, Yan Xu, Ruixing Li, Huawei Li and Meng Zhang

93 Enhancing bronchopulmonary dysplasia prediction in preterm infants using artificial intelligence and multimodal data integration
Xinkai Zhang, Anping Wang, Rongwei Xu and Dongyun Liu

104 Machine learning model for predicting urinary tract infection risk in febrile children under 3 years of age
Le-zhen Ye, Jian-xin Sun, Jing Chen, Kuan-kuan Cen, Ye Bi and Yun-cong Lu

115 Association of Shanghai air pollution with postoperative infection in adolescent orthopedic patients: a study using a deep learning-based evolutionary model
Yufeng Wang, Yang Yuan, Yang Liu and Haoqi Cai

127 Reliability of ChatGPT answers to common questions on developmental dysplasia of the hip as an information source for parents
Klemens Vertesich, Joachim Ortmayr, Reinhard Windhager and Madeleine Willegger

136 Artificial intelligence and precision medicine: a pilot study predicting optimal ceftaroline dosage for pediatric patients
Maria Frasca, Gianluca Gazzaniga, Agnese Graziosi, Valentina De Nicolo, Costantino De Giacomo, Stefano Martinelli, Michele Senatore, Alessandra Romandini, Chiara Moretti, Giulia Angela Carla Pattarino, Alice Proto, Romano Danesi, Francesco Scaglione, Gianluca Vago, Davide La Torre and Arianna Pani

146 Machine learning strategies for predicting pediatric suicidal behaviors in a Brazilian emergency setting
Isis F. Carvalho, Ana Paula Couto da Silva, Anisio M. Lacerda, Wagner Meira Jr., Danilo Bastos Bispo Ferreira, Lys Malloy Diniz, Alexandre Luiz de Oliveira Serpa, Maria Carolina Lobato Machado, Marco A. Romano-Silva, Debora Marques de Miranda and Gisele Lobo Pappa

frontiers in Artificial Intelligence

TYPE Editorial
PUBLISHED 18 March 2026
DOI 10.3389/frai.2026.1814141

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Thomas Hartung,
Johns Hopkins University, United States

*CORRESPONDENCE
Kaya Kuru
kkuru@uclan.ac.uk

RECEIVED 20 February 2026
REVISED 26 February 2026
ACCEPTED 26 February 2026
PUBLISHED 18 March 2026

CITATION
Kuru K, Caswell N, Hart GR, Quon J and Ehwerhemuepha L (2026) Editorial: Artificial intelligence and machine learning in pediatrics.
*Front. Artif. Intell.* 9:1814141.
doi: 10.3389/frai.2026.1814141

COPYRIGHT
© 2026 Kuru, Caswell, Hart, Quon and Ehwerhemuepha. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Editorial: Artificial intelligence and machine learning in pediatrics

Kaya Kuru[1,2]*, Noreen Caswell[1], Gregory R. Hart[3], Jennifer Quon[4] and Louis Ehwerhemuepha[5]

[1]School of Engineering & Computer Science, University of Central Lancashire, Preston, United Kingdom, [2]Software Engineering Department, Fenerbahce Universitesi, Ataşehir, Türkiye, [3]University of Guam, Mangilao, GU, United States, [4]The Hospital for Sick Children Paediatric Emergency Medicine, Toronto, ON, Canada, [5]Children's Hospital of Orange County, Orange, CA, United States

KEYWORDS
**artificial intelligence, big data analytics, large language models, machine learning, natural language processing, pediatrics**

Editorial on the Research Topic
Artificial intelligence and machine learning in paediatrics

## 1 Introduction

Pediatrics is a branch of medicine that specializes in the medical care of infants, children, and adolescents, who undergo rapid developmental changes and exhibit distinct disease patterns. The development of modern technologies like biosensors (Kuru et al., 2020), wearable sensors (Caswell et al., 2020), medical imaging, well-designed laboratory facilities, and genetics advancements has led to a huge amount of data with multidimensional and complex structures in different formats in pediatrics, which exceeds traditional human analytical capabilities. A massive amount of childcare data is generated every second (Kuru, 2021). These vast volumes of pediatric data indicate the need to develop intelligent technologies to efficiently exploit this wealth of resources. Artificial Intelligence (AI) and Machine Learning (ML) technologies can be used to efficiently process large and complex datasets composed of a diverse set of medical data, such as electronic health records, images, genetic data, and wearable sensor data, to make precise, individualized, and data-driven pediatric care possible.

## 2 Artificial intelligence and machine learning in pediatrics

Recent research shows that AI/ML can discover new approaches to representing multidimensional, nonlinear relationships and patterns in children's data. More specifically, it can help with providing a rather quick and accurate understanding of effective treatment strategies and further improving clinical decision-making. Models based on AI technology show impressive efficacy in early diagnostics of pediatric conditions like congenital heart disease, childhood cancers, and rare genetic disorders. Among the

recent trends in medical imaging (e.g., radiology, pathology, and ophthalmology), deep learning (DL) techniques are used to allow for faster and more accurate identification of anomalies, which often reach or surpass those performed by specialists. However, even the diagnostics have provided a very clear picture of how crucial it is to carry out additional research in employing AI/ML algorithms. There is still a large gap and numerous other obstacles to integrating AI/ML into pediatric health care. The main challenges and guardrails can be summarized as follows

- Age stratification and developmental change (non-stationarity across growth stages).
- Smaller datasets/rarer outcomes and resulting overfitting risk.
- Equity and bias (race/ethnicity, language, disability, access).
- Lack of standardized pediatric data ontologies.
- Consent/privacy complexities for minors and family context.
- Prospective evaluation and deployment (monitoring drift, safety, human factors).

Driven by the exponential growth of complex pediatric data, this Research Topic explores the potential role of AI/ML in pediatrics, particularly concentrating on harnessing big data analytics to enhance children's health outcomes. Finding out how AI/ML may improve the early diagnosis, treatment, and prognosis of pediatric disorders, as well as testing theories about the effectiveness of AI/ML-driven precision medicine in the provision of pediatric care, are among the main objectives. Can AI help develop novel drugs or approaches to cure pediatric diseases? Through asking these questions, the topic attempts to obtain (i) a complete picture of what is possible and (ii) the limitations of AI/ML in pediatric healthcare, ultimately leading the way toward safer, more effective and personalized medical treatment for children.

## 3 Innovative AI studies in pediatric practices

Against the above backdrop, this Research Topic of the journal was convened by both the pediatricians and AI specialists to create a forum which presents conducted scholarly studies focused on recently arising issues in pediatric healthcare. In all, 14 top-notch articles have been included in this Research Topic through rigorous review and selection. Carvalho et al. carry out an investigation on ML techniques regarding their performance in predicting suicide related behaviors among young populations within a lower-middle-income country context- particularly with respect to pediatric psychiatric emergencies in Brazil. The research demonstrates strong potential for identifying suicidal risk in pediatric emergency settings, especially useful in regions with limited specialized psychiatric resources, by integrating clinical data with social determinants. On a separate note, traditional drug dosing approaches are often insufficient to ensure effective therapy, exposing patients to the risk of overdosing, leading to adverse effects with potential therapeutic failure (Anderson and Holford, 2008). To find a practical solution to this common problem, Frasca et al. examine the application of ML to optimize antibiotic dosing in pediatric patients, focusing on predicting the ceftaroline dose. Findings in their research suggest that carefully curated AI/ML models can outperform conventional ceftaroline dosing strategies in pediatric patients. Based on the most recent developments in Large Language models (LLM), Vertesich et al. investigated the quality, accuracy and reliability of inference that can be made from a LLM (ChatGPT) for parental education in a pediatric medical context of developmental dysplasia of the hip (DDH) specifically. The study concludes that LLMs can present high-quality answers to questions about DDH and may serve as a useful supplementary information resource for parents. In another direction, Wang et al. are looking at how much air pollution exposure and surgical site infection (SSIs) are related among adolescents undergoing orthopedic surgery, using Shanghai as a case study. The study demonstrates that DL models can readily identify the dominant predictors, supporting perioperative air-quality interventions. Likewise, genetic disorders are affecting children and their families drastically (Kuru et al., 2014). Zhao et al. present findings about the role and potential of AI, particularly ML/DL, in enhancing the objective measurement of attention deficit hyperactivity disorder (ADHD) treatment outcomes. More specifically, the significant potential of AI in tasks such as early screening and risk prediction, diagnostic assistance and classification, precise differential diagnosis, symptom severity quantification, and the identification of heterogeneous ADHD subtypes was highlighted. Similarly, Tang et al., in their work, describe how to find diagnostic markers and molecular subtypes for pediatric Crohn's disease (PCD) from gene expression data using ML techniques. Around 1.7 billion diarrhea cases in children are reported every year (WHO, 2024). It is the third leading cause of death in children under 5 years old and is responsible for killing around 443,832 children every year (WHO, 2024). Yehuala et al. focus on the factors that determine the child feeding practices during episodes of diarrhea in children in East Africa based on a huge population survey data with the help of ML models. Considering the integration of intelligent devices into child healthcare practices, Fan et al. examine whether home spirometry telemonitoring for children with asthma is possible, trustworthy, and acceptable, and assess real-world issues affecting its clinical integration. Similarly, Soualmi et al. present the development and implementation of an automated instrument (AGMA-PESS) for the purpose of improving the efficiency of General Movement Assessment (GMA) in preterm infants. On the other hand, nocturnal enuresis (NE) is a widespread disease among children, and a proper cure has yet to be found to treat it (Kuru et al., 2024, 2018). Liao et al. discuss how health education improves the quality of life among children receiving treatment for NE and their caregivers using data mining techniques. In the same way, Ye et al. describe an application of ML for urinary tract infection (UTI) risk prediction in febrile children under 3 years of age. In the article of Zhang et al., a review is made concerning the current status and clinical applicability of AI/ML for bronchopulmonary dysplasia (BPD) predictive measures in pre-term infants. Tozzi et al. discuss the position of AI in the contemporary focus on pediatric brain tumors at a European level with respect to data sources and generalisability of AI algorithms. When implementing AI-based

technologies in childcare, it is highly crucial to incorporate the perspectives of patients and their families into all phases (Haan et al., 2019) to make them successful (Caswell et al., 2020), concerning their acceptability. To guide the AI developer in childcare, Huang et al. investigate the perception, knowledge, and attitude of parents regarding AI applications in pediatric practice.

## 4 Conclusion

The continuous evolution of AI is equipping pediatricians with new tools to perform their tasks more efficiently. In this direction, this topic delves into the advancement of AI in pediatric medicine. Based on the findings in this Research Topic, it can be safely concluded that AI/ML applications are increasingly transforming pediatric healthcare by enabling data-driven decision-making, enhancing diagnostic accuracy, personalizing treatments, proactive interventions and streamlining clinical workflows. Further developments in AI will significantly shape the future of pediatrics in the years ahead, paving the way for better healthcare for pediatric patients and their families through informed clinical decision-making. While significant challenges remain, ongoing research, ethical oversight, and interdisciplinary collaboration are essential to realize and exploit the full potential of AI-driven innovations, ultimately improving the health and wellbeing of children worldwide.

While AI is providing invaluable guidance to pediatricians and iincreasingly playing a pivotal role in childcare starting from the prenatal phase, future research should focus on developing explainable and trustworthy AI models, creating large-scale, reliable pediatric-specific datasets and establishing robust validation and regulatory frameworks. Collaborative efforts among clinicians, data scientists, policymakers, and ethicists will be essential to ensure responsible and effective adoption of AI in pediatric healthcare.

## Author contributions

KK: Writing – original draft, Writing – review & editing. NC: Writing – review & editing. GH: Writing – review & editing. JQ: Writing – review & editing. LE: Writing – original draft, Writing – review & editing.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Anderson, B. J., and Holford, N. H. G. (2008). Mechanism-based concepts of size and maturity in pharmacokinetics. *Annu. Rev. Pharmacol. Toxicol.* 48, 303–332. doi: 10.1146/annurev.pharmtox.48.113006.094708

Caswell, N., Kuru, K., Ansell, D., Jones, M. J., Watkinson, B. J., Leather, P., et al. (2020). Patient engagement in medical device design: refining the essential attributes of a wearable, pre-void, ultrasound alarm for nocturnal enuresis. *Pharmaceut. Med.* 34, 39–48. doi: 10.1007/s40290-019-00324-w

Haan, M., Ongena, Y. P., Hommes, S., Kwee, T. C., and Yakar, D. A. (2019). qualitative study to understand patient perspective on the use of artificial intelligence in radiology. *J. Am. Coll. Radiol.* 16, 1416–1419. doi: 10.1016/j.jacr.2018.12.043

Kuru, K. (2021). Management of geo-distributed intelligence: deep insight as a service (DINSaaS) on forged cloud platforms (FCP). *J. Parallel Distrib. Comput.* 149, 103–118. doi: 10.1016/j.jpdc.2020.11.009

Kuru, K., Ansell, D., Hughes, D., Watkinson, B. J., Gaudenzi, F., Jones, M., et al. (2024). Treatment of nocturnal enuresis using miniaturised smart mechatronics with artificial intelligence. *IEEE J. Transl. Eng. Health Med.* 12, 204–214. doi: 10.1109/JTEHM.2023.3336889

Kuru, K., Ansell, D., Jones, M., De Goede, C., and Leather, P. (2018). Feasibility study of intelligent autonomous determination of the bladder voiding need to treat bedwetting using ultrasound and smartphone ML techniques. *Med. Biol. Eng. Comput.* 57, 1079–1097. doi: 10.1007/s11517-018-1942-9

Kuru, K., Ansell, D., Jones, M., Watkinson, B. J., Caswell, N., Leather, P., et al. (2020). Intelligent autonomous treatment of bedwetting using non-invasive wearable advanced mechatronics systems and MEMS sensors. *Med. Biol. Eng. Comput.* 58, 943–965. doi: 10.1007/s11517-019-02091-x

Kuru, K., Niranjan, M., Tunca, Y., Osvank, E., and Azim, T. (2014). Biomedical visual data analysis to build an intelligent diagnostic decision support system in medical genetics. *Artif. Intell. Med.* 62, 105–118. doi: 10.1016/j.artmed.2014.08.003

WHO (2024). *Diarrhoeal Disease*. Geneva: World Health Organization (WHO).

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 27 October 2023
DOI 10.3389/fonc.2023.1285775

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Sabyasachi Dash,
AstraZeneca (United States), United States
Tim Hulsen,
Philips Research (Netherlands), Netherlands

*CORRESPONDENCE
Alberto Eugenio Tozzi
albertoeugenio.tozzi@opbg.net

RECEIVED 30 August 2023
ACCEPTED 16 October 2023
PUBLISHED 27 October 2023

CITATION
Tozzi AE, Croci I, Voicu P, Dotta F, Colafati GS, Carai A, Fabozzi F, Lacanna G, Premuselli R and Mastronuzzi A (2023) A systematic review of data sources for artificial intelligence applications in pediatric brain tumors in Europe: implications for bias and generalizability. *Front. Oncol.* 13:1285775. doi: 10.3389/fonc.2023.1285775

COPYRIGHT
© 2023 Tozzi, Croci, Voicu, Dotta, Colafati, Carai, Fabozzi, Lacanna, Premuselli and Mastronuzzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# A systematic review of data sources for artificial intelligence applications in pediatric brain tumors in Europe: implications for bias and generalizability

Alberto Eugenio Tozzi[1]*, Ileana Croci[1], Paul Voicu[2], Francesco Dotta[3], Giovanna Stefania Colafati[3], Andrea Carai[4], Francesco Fabozzi[5], Giuseppe Lacanna[1], Roberto Premuselli[5] and Angela Mastronuzzi[5]

[1]Predictive and Preventive Medicine Research Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy, [2]Department of Neuroscience and Imaging, "SS Annunziata" Hospital, "G. D'Annunzio" University, Chieti, Italy, [3]Imaging Department, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy, [4]Department of Neurosciences, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy, [5]Department of Hematology/Oncology, Cell and Gene Therapy, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

**Introduction:** Europe works to improve cancer management through the use of artificialintelligence (AI), and there is a need to accelerate the development of AI applications for childhood cancer. However, the current strategies used for algorithm development in childhood cancer may have bias and limited generalizability. This study reviewed existing publications on AI tools for pediatric brain tumors, Europe's most common type of childhood solid tumor, to examine the data sources for developing AI tools.

**Methods:** We performed a bibliometric analysis of the publications on AI tools for pediatric brain tumors, and we examined the type of data used, data sources, and geographic location of cohorts to evaluate the generalizability of the algorithms.

**Results:** We screened 10503 publications, and we selected 45. A total of 34/45 publications developing AI tools focused on glial tumors, while 35/45 used MRI as a source of information to predict the classification and prognosis. The median number of patients for algorithm development was 89 for single-center studies and 120 for multicenter studies. A total of 17/45 publications used pediatric datasets from the UK.

**Discussion:** Since the development of AI tools for pediatric brain tumors is still in its infancy, there is a need to support data exchange and collaboration between centers to increase the number of patients used for algorithm training and improve their generalizability. To this end, there is a need for increased data exchange and collaboration between centers and to explore the applicability of decentralized privacy-preserving technologies consistent with the General Data Protection Regulation (GDPR). This is particularly important in light of using the European Health Data Space and international collaborations.

KEYWORDS

**artificial intelligence, CNS tumors, pediatric oncology, childhood cancer, data sharing**

# 1 Introduction

AI holds the promise of addressing various unmet needs in cancer, although its systematic application is still in progress (1–3). A pivotal requirement for the successful integration of AI into healthcare is that data used to develop algorithms should be representative of the population's diversity, ensuring the avoidance of bias and adverse patient events. Mitigating bias stemming from inadequate representation in the training of healthcare AI tools is a widely acknowledged challenge, demanding targeted strategies for its resolution (4). The cornerstone of these strategies lies in the collection and integration of disparate datasets from various institutions, thereby circumventing the underrepresentation of specific population subsets and the potential ensuing discrimination. Numerous global initiatives have been launched to create openly accessible datasets that encompass a variety of data types essential for training and validating AI systems (5). Concurrently, the feasibility of consolidating datasets from multiple European centers is subject to adhering to GDPR regulations, which could potentially hamper data sharing for scientific purposes (6, 7). Within this context, a notable concern arises: how the current AI tools customized for clinical use in Europe are trained with datasets representing diverse populations.

In 2021, the European Commission introduced the Europe's Beating Cancer Plan, a strategic framework aimed at bridging gaps in cancer prevention, treatment, and care, with a specific emphasis on childhood cancer priorities (8). This plan notably aligns with the European Digital Strategy's endorsement of Artificial Intelligence (9). Building on this perspective, the European Society for Pediatric Oncology envisions harnessing AI technologies to their full potential for the benefit of pediatric cancer patients in Europe (10).

Central nervous system (CNS) tumors constitute nearly 20% of childhood cancers, making them the predominant solid neoplasms in this age group (11). Despite significant progress that has improved the outlook for pediatric CNS tumor patients, challenges persist in fully implementing precision medicine, enabling less invasive interventions, predicting treatment responses, and identifying new therapeutic approaches (10–12).

Childhood CNS tumors significantly differ from their adult counterparts by incidence, histology, molecular biology, treatment strategies, outcomes, and long-term outcomes. Consequently, extrapolating data from adults to children is not appropriate (13). Conversely, the rarity of brain tumors in pediatric patients poses a challenge in amassing a substantial volume of observations to appropriately train algorithms in this disease group.

The WHO stresses the need for AI developers to ensure that AI data is accurate, comprehensive, and representative of diverse age groups, including children (14, 15). Moreover, the proposed EU AI Act requires that "…training, validation and testing data sets shall be relevant, representative, free of errors and complete" (Art. 10) (16). In fact, a substantial portion of existing research on childhood CNS tumors consists of proof-of-concept studies based on limited datasets, which makes them not yet directly applicable into clinical practice (12, 17).

All this considered, we undertook a systematic review of the current literature related to AI tools specifically developed for tackling pediatric brain tumors in the European context. Our objective was to describe the characteristics of the data utilized in their development and the potential associated bias. Our inquiry centered on delineating the sources of data, gauging the scale of the datasets, evaluating their interoperability, ascertaining the presence of external validation, and scrutinizing the geographic representation of the cohorts employed in refining the algorithms. Based on these findings, we discuss potential strategies to accelerate the development and integration of AI tools for pediatric brain tumors into clinical practice in Europe.

# 2 Methods

We performed a systematic review of the current literature by employing a search query based on the terms recommended by the Cochrane collaboration for pediatric tumors (18), including terms specific to pediatric brain cancer. Furthermore, we developed a distinct search query tailored to encompass AI techniques, subsequently merging it with the prior query (Supplemental materials - S1).

Our search spanned the databases of MEDLINE, EMBASE, Web of Science, and Scopus, limiting the inquiry to papers published in English within the period spanning January 2010 to May 31, 2022. The outcomes of this search were imported into the Rayyan software (19) and subjected to a duplicate screening process.

Subsequent to this screening, we manually reviewed the remaining records to ascertain their eligibility against a set of predefined criteria, encompassing: 1) original articles; 2) papers detailing the development of AI tools tailored for pediatric brain tumor diagnosis, prognosis prediction, or therapeutic decision support; 3) at least one author affiliated with a European institution; 4) publications in English. Conversely, we excluded articles: 1) not presenting original data or reviews; 2) published in languages other than English; 3) authored by individuals not affiliated with European institutions; 4) describing studies conducted on animals or simulated environments.

We have included authors from institutions based in the UK, as they continue to be eligible for collaborative projects within established European networks following Brexit.

The selected articles were categorized into two distinct groups: 1) those exclusively comprising observations from patients under 18 years of age for algorithm training; 2) those encompassing both pediatric and adult populations.

Bibliometric details were extracted in a format compatible with bibliometric analytical tools, while three independent reviewers manually extracted specific details from each publication, including: a) the specific brain tumor type under examination; b) the scope of the AI tool; c) the nature of data utilized; d) the data repository employed; e) the count of patients contributing data for algorithm development; e) the use of data standards for interoperability; f) the use of external validation of the algorithms; g) the performance of the AI tool; h) the geographic representation of the cohorts used for developing the algorithm.

The resulting information was subjected to descriptive analysis, with bibliometric data analyzed through the Biblioshiny software (20). Additionally, we evaluated the FAIR Guiding Principles score (21) for each dataset included in the reviewed articles using the SATIFYD online tool (https://satifyd.dans.knaw.nl/). Other statistical analyses were performed utilizing the R software (22).

## 3 Results

Our search strategy yielded a total of 10503 scientific publications. The selection process adhered to PRISMA standards and is visually represented in Figure 1. Upon eliminating duplicates across diverse databases, we identified 7478 unique records. Subsequent review of titles and abstracts narrowed down the selection to 572 records, whose full texts were evaluated. Within this group, we excluded 246 articles solely centered on adult populations, 24 that didn't involve AI methodologies, and 29 that didn't meet the eligibility criteria for various reasons. This left us with 273 remaining studies.

Of these 273 studies, 228 failed to report the age of the patients and were consequently excluded. Ultimately, we selected 45 articles that incorporated pediatric data for the AI algorithm development. Among these, 25 articles exclusively focused on children, while 20 articles used data from both pediatric and adult cohorts (Supplemental materials - S2).

Our review encompassed publications dating back to 2012. Notably, 27 out of the 45 articles (60.0%) included in this study were published from 2020 onward. Additionally, eight studies were published before 2016 and were characterized by relatively modest patient numbers during algorithm training, employing simple AI techniques. The array of AI methods utilized for analysis was diverse, encompassing techniques such as LDA, KNN, Naive Bayes, SVM, Random Forest, and more. Similarly, the choice of software varied, including Python, R, and Mazda Orange, although the preponderance of papers analyzed data within a Python environment. On average, the selected publications garnered 57.8 citations each.

Table 1 provides an overview of the brain tumor types for which AI algorithms were developed, categorized by the included population type. The majority of publications focused their efforts on glial tumors, while other tumor types were addressed less frequently. The category of unclassified tumors encompassed those specifically characterized as brain or posterior fossa tumors, pituitary adenoma, and large B-cell lymphoma. Table 2 shows the scope of the articles included in this review by type of population

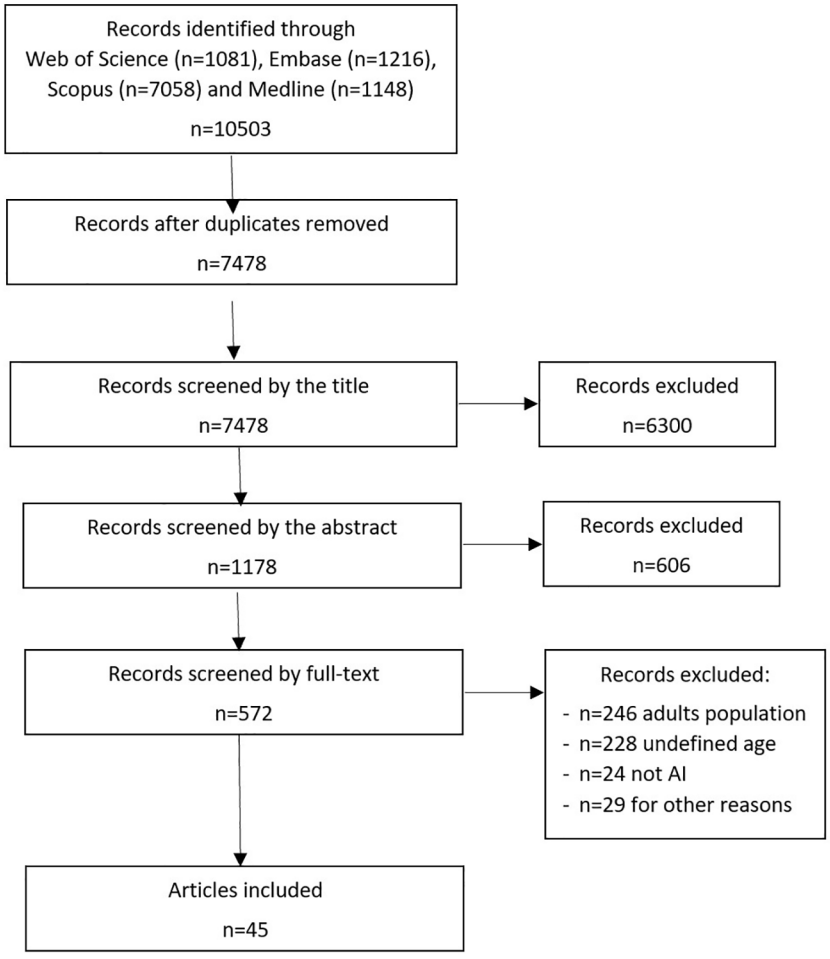


FIGURE 1
Review process of selected publications according to PRISMA.

TABLE 1 Diagnoses included in the selected publications by type of population.

| | Pediatric population only (n=25) | Children and adults (n=20) | Total (n=45) |
|---|---|---|---|
| | n (%) | n (%) | n (%) |
| Glial tumors | 20 (80.0%) | 14 (70.0%) | 34 (75.5%) |
| PNET | 19 (76.0%) | – | 19 (42.2%) |
| Ependymal tumors | 18 (72.0%) | - | 18 (40.0%) |
| Mesenchymal tumors | 1 (4.0%) | 1 (5.0%) | 2 (4.4%) |
| Germ cell tumors | 1 (4.0%) | – | 1 (2.2%) |
| Unclassified | 4 (16.0%) | 6 (30.0%) | 10 (22.2%) |

Multiple diagnoses may be included in the same publications.

included. Notably, the predominant thrust of the selected papers (24 out of 45, accounting for 53.3%) was directed towards algorithmic development for tumor classification. Conversely, fewer papers pertained to prognosis, decision support, or validation tasks.

The majority of publications employed the same dataset for both testing and validating their algorithms, with only 3 out of 45 utilizing external datasets for validation. When reported, the external validation performance of the algorithm was lower than that reported for internal validation.

Table 3 shows the data sources utilized across the reviewed publications. The majority of AI studies examined diagnostic images as primary data sources, specifically employing multiple conventional MRI sequences for classification and prognosis of CNS tumors. However, some studies delved into spectroscopy, a more specialized MRI sequence often utilized at the point of diagnosis (23). Additionally, computer tomography and histopathology images made appearances as data sources. While metabolite profiles, epigenetics, gene expression, and clinical features were used less frequently than images, only 6 out of 45 publications (13.3%) integrated data from multiple sources.

TABLE 2 Scopes of the articles included in the review.

| | Pediatric population only (n=25) | Children and adults (n=20) | Total (n=45) |
|---|---|---|---|
| | n (%) | n (%) | n (%) |
| Classification | 17 (68.0%) | 9 (45.0%) | 26 (57.8%) |
| Diagnosis | 6 (24.0%) | 3 (15.0%) | 9 (20.0%) |
| Prognosis | 3 (12.0%) | 8 (40.0%) | 11 (24.4%) |
| Therapy | 1 (4.0%) | 1 (5.0%) | 2 (4.4%) |

Multiple scopes may be included in the same publications.

Table 4 shows the repositories gathering data from multiple sites used for the development of algorithms and accessible on the web, as reported in the articles included in the review.

Considering data origin, 29 articles (64.4%) drew from single centers, while 16 (35.6%) were the result of multicenter collaborations. Two studies used synthetic data to inform their algorithms' development, and no instances of federated learning were noted. Notably, no selected publication provided insights into the interoperability of their datasets.

As for the FAIR Guiding Principles score, the median value for the datasets used in the articles in this review stood at 23 (IQR: 15-32). This relatively modest score stemmed from the frequent absence of accessible metadata and from the limited accessibility of datasets used for algorithm development.

The number of patients included in the datasets for the development of AI algorithms was quite variable and essentially limited for most data sources. The median observation count used for single-center studies was 89, whereas publications emerging from multicenter collaborations exhibited a median observation count of 120.

Examining geographic scope, 7 out of the 45 reviewed articles tapped into public repositories for data. An additional 14 articles incorporated data from beyond the EU, including countries such as the USA, Canada, China, Iran, and Argentina. In terms of pediatric datasets within the EU, 17 publications (37.8%) harnessed cohorts from the UK, while individual datasets emerged from Belgium, Denmark, Italy, Netherlands, Spain, and Sweden.

Delving into the collaborative landscape, Figure 2 illustrates the interconnected network of institutions represented across the selected publications. Notably, the UK stood out as a frequent contributor, not just in terms of author affiliations, but also due to its extensive collaborations with both European and non-European entities. Moreover, the network reveals consistent partnerships between EU countries and the USA.

# 4 Discussion

Our review unveiled a scarcity of publications regarding AI applied to CNS tumors in children in Europe. Most of the studies emerged from individual centers, indicating a pressing need for improving research in this field. This holds particular significance for the less prevalent tumor types, often overshadowed in the development of AI tools.

Furthermore, our review underscored that several European institutions are engaged in crafting AI tools for identical categories of pediatric brain tumors, with only 16 publications reflecting multicenter data contributions. On the other hand, we did not find any information about data standards for interoperability that can support collaborative research in this field. Notably, collaborative networks, both within Europe and on a global scale, emerged from our analysis, illustrating a fertile ground for reinforcing and expanding such partnerships.

TABLE 3 Data sources used in the selected publications by type of population.

| | Pediatric population only (n=25) | Children and adults (n=20) | Total (n=45) |
|---|---|---|---|
| | n (%) | n (%) | n (%) |
| MRI | 24 (96.0%) | 11 (55.0%) | 35 (77.8%) |
| Histopathology | 1 (4.0%) | 2 (10.0%) | 3 (6.7%) |
| DNA methylation | 1 (4.0%) | 3 (15.0%) | 4 (8.9%) |
| CT | 1 (4.0%) | – | 1 (2.2%) |
| Metabolite profiles | 2 (8.0%) | – | 2 (4.4%) |
| Gene expression | – | 1 (5.0%) | 1 (2.2%) |
| Other | 1 (4.0%) | 2 (10%) | 3 (6.7%) |

Multiple data sources may be included in the same publications.

We also observed that most publications reported AI tools trained with a low number of observations from selected populations. Moreover, few publications reported an external validation of their algorithms.

As in many other medical fields, research on AI on pediatric CNS tumors suffers from fragmented data and studies performed on small sample sizes. Although the performance of the published algorithm in this domain is fair according to reported results for internal validation, the potential for bias is high and their generalizability may be limited.

One of the major obstacles to data-sharing and collaboration among institutions in the development of AI tools that are trained on diverse and representative populations is the concern for privacy. This is particularly relevant in the EU, where the General Data Protection Regulation (GDPR) is in place. Data-sharing agreements require a thorough evaluation of potential risks, which can take a significant amount of time to complete. Additionally, there is still variability in the interpretation of GDPR requirements for data sharing at the country level (24, 25). Identifying the trade-off between privacy preservation and the full development of AI solutions will represent one of the most important topics for discussion at the data governance level.

While compliance with regulations for data sharing is important for scientific research in the field of AI (26), technologies such as federated and swarm learning (27–29) and synthetic data (30) may be prioritized to accelerate the development of AI tools. Notably, none of the publications reviewed in our study used federated or swarm learning and only two studies explored the use of synthetic data.

The datasets examined in the review seldom adhered to all FAIR principles, reinforcing the observation of a limited commitment to reusing and integrating data across different contexts.

Moreover, interoperability and data harmonization (31) are critical for addressing the fragmentation of data. The Observational Health Data Sciences and Informatics (OHDSI) community conducts research to promote the use of standards such as OMOP and standardized vocabularies (32, 33). Another standard that supports interoperability is HL7 FHIR (34). However, our review did not find any information on the interoperability of the datasets used for the development of AI tools.

The effectiveness of artificial intelligence tools is dependent on the quality of data they are trained on. With population selection, misrepresentation, or missing data, generalizability across subgroups is not possible. Our review found publications that used data from both children and adults for the development of algorithms for brain cancer, which may hinder their generalizability. The recent WHO document (13) on a detailed classification of CNS tumors in children provides a valuable tool for distinguishing different types on a molecular basis. AI tools can further improve classification if algorithms are trained with generalizable data.

TABLE 4 Data repositories gathering data from multiple sites used in articles included in this review.

| Name | Type of data | URL | Population | Type of access |
|---|---|---|---|---|
| Children's Cancer and Leukaemia Group - Functional Imaging Database | Functional Imaging of Tumours study with CCLG from a UK network of hospitals. | https://www.cclg.org.uk/ | Children | Restricted |
| eTUMOUR* | MRI, SV 1H MRS data and histology of brain tumors | http://www.etumour.net/ | Children | Restricted |
| Gene Expression Omnibus | Functional genomics data repository | https://www.ncbi.nlm.nih.gov/geo/ | All ages | Public, open |
| The Cancer Genome Atlas | Genomic, epigenomic, transcriptomic, and proteomic data of patients with cancer | https://www.cancer.gov/ccg/research/genome-sequencing/tcga | All ages | Public, open |
| The Cancer Imaging Archive | Imaging, patient outcomes, treatment details, genomics, pathology of patients with cancer | https://www.cancerimagingarchive.net/ | All ages | Public, open |
| Belgian Cancer Registry | Registry of newly diagnosed cases of cancer in Belgium | http://kankerregister.org/Home_en | All ages | Restricted |
| BraTS2020 | MRI images and clinical information from patients with brain tumors | https://www.med.upenn.edu/cbica/brats2020/data.html | All ages | Restricted |

*This repository was developed during a project funded by the European Commission, and it is no longer accessible.

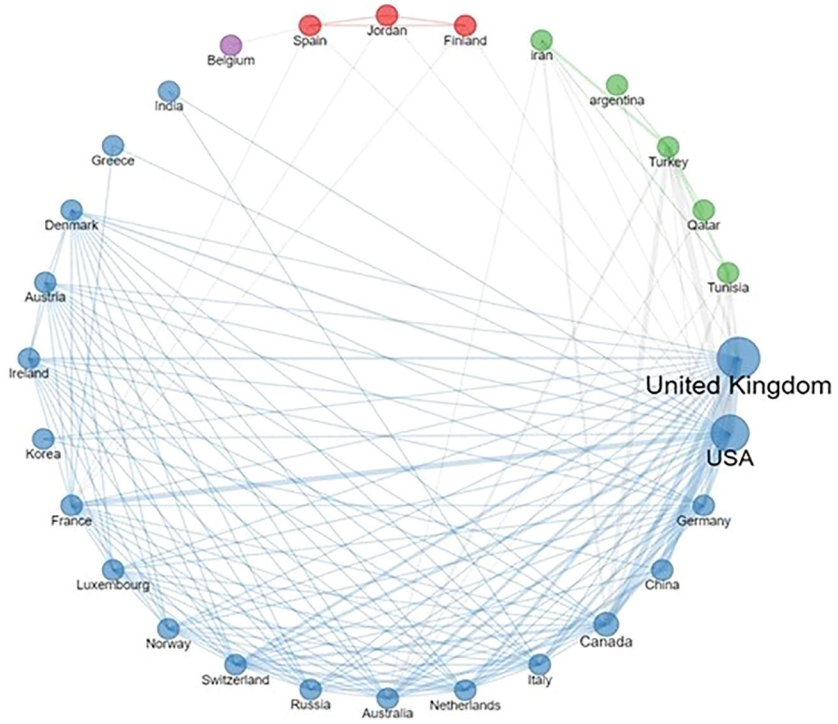


FIGURE 2
Country collaboration network. The thickness of lines reflects the frequency of collaborations.

The majority of AI tools used in the publications reviewed utilized diagnostic images as their data source. Neuroimaging, a well-developed area for AI in oncology, offers detailed analysis of brain microstructures and pathophysiology specific to children (35). However, AI also has the potential for integrating multiple types of data, which could be beneficial for investigating complex patterns and increasing algorithm accuracy. Our review found that only a small number of publications utilized this approach, which is an area that should be given more attention.

The development of AI tools outside medicine typically aggregates massive datasets from multiple sources which correlate with accurate predictions (36). Open repositories for developing AI tools represent an attempt to concentrate large amounts of diagnostic images and overcome the existing barriers to data access. In this review, among the 45 publications included, 7 have used a data repository for the development of their algorithm, of which 3 only were publicly accessible. Several public datasets are available, particularly for neuroimaging, which represents the main area of application of AI in neuro-oncology.

Unfortunately, public datasets for AI purposes are limited in size and do not accurately represent the entire population of children with brain tumors. Additionally, neuroimaging data in these datasets may have inconsistencies due to tumor and scan variations. Furthermore, advances in technology can lead to changes in the quality and features of diagnostic images, known as data drift, which can affect the performance of AI algorithms. This highlights the need for ongoing updates to the datasets used for training AI algorithms (37, 38).

In essence, the success of AI tools depends heavily on the quality and representativeness of the data used for training. Currently, many AI models for pediatric brain tumors are trained on small, single-center datasets. To enhance accuracy and applicability, there is a need for collaboration among research groups working on the same diseases, sharing data to enhance the robustness of AI models. Without this collective effort, AI models developed in isolation may lack the necessary accuracy for clinical decision-making, potentially introducing bias and errors. The absence of such collaboration leaves AI models vulnerable to overfitting, performing well on training data but poorly on external validation from different settings.

Our review has some limitations. Our study aimed to comprehensively review the existing literature on AI and pediatric brain tumor screening by using multiple bibliographic platforms and a structured search strategy with a manual review. However, information about the datasets used for developing AI tools may not be readily available in the literature, particularly if they are part of an industrial process. Therefore, policies that promote wider access and analysis of these datasets should be supported. Additionally, we intentionally focused on publications originating from the EU only. Our aim was to shed light on the challenges faced within the EU in terms of data sharing and AI model development taking into account factors as the General Data Protection Regulation (GDPR) and the issues related to data interoperability of the existing infrastructures. This approach is beneficial in understanding the European context, where there are many scientific collaborations and in light of the future creation of the

European Health Data Space (39). Finally, we focused on data-driven bias only and we did not evaluate potential algorithmic bias in the studies included in this review.

In summary, this review highlights several potential sources of bias in AI tools developed for pediatric CNS tumors that may limit their clinical application. The most important strategy to address this limitation is promoting the use of larger and more diverse datasets through collaboration among different institutions that improve data availability, sharing, standardization, interoperability, continuous updates, and quality. On one hand, it is of paramount importance to promote the adoption of FAIR Principles during the development of AI tools. On the other hand, expanding existing networks that adhere to the same standards could benefit the development of EU-funded research and development projects, and data exchange through the European Health Data Space (39). Collaborative efforts should be supported by data standards such as those of Common Data Model (31, 32). Furthermore, decentralized privacy-preserving technologies such as federated learning and synthetic data may accelerate the development of AI tools based on large populations from different clinical sites while complying with EU privacy regulations (30, 40). Data access remains the biggest challenge in training and developing AI tools, and requires significant research effort and investment to enable the development of AI tools for pediatric brain tumors.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

AT: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. IC: Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. PV: Data curation, Formal Analysis, Writing – review & editing. FD: Data curation, Formal Analysis, Investigation, Writing – review & editing. GC: Data curation, Formal Analysis, Writing – review & editing. AC: Data curation, Supervision, Writing – review & editing. FF: Data curation, Formal Analysis, Investigation, Writing – review & editing. GL: Data curation, Formal Analysis, Writing – review & editing. RP: Supervision, Visualization, Writing – review & editing. AM: Conceptualization, Formal Analysis, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by UNICA4EU, a project funded by the European Union's Call for Pilot Projects and Preparatory Actions (PPPA) under Grant Agreements No 101052609. This work was supported also by the Italian Ministry of Health with Current Research funds.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1285775/full#supplementary-material

## References

1. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. *Front Med* (2020) 7:27. doi: 10.3389/fmed.2020.00027

2. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med* (2022) 28:31–8. doi: 10.1038/s41591-021-01614-0

3. Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell Int* (2021) 21:270. doi: 10.1186/s12935-021-01981-1

4. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA* (2019) 322:2377. doi: 10.1001/jama.2019.18058

5. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. *Cancer Discovery* (2021) 11:900–15. doi: 10.1158/2159-8290.CD-21-0090

6. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)* (2016). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.

7. Hoofnagle CJ, van der Sloot B, Borgesius FZ. The European Union general data protection regulation: what it is and what it means. *Inf Commun Technol Law* (2019) 28:65–98. doi: 10.1080/13600834.2019.1573501

8. *Europe's Beating Cancer Plan. Eur Comm - Eur Comm*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_21_342 (Accessed July 13, 2023).

9. *A European approach to artificial intelligence | Shaping Europe's digital future* (2023). Available at: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence (Accessed July 31, 2023).

10. *SIOP Europe Strategic Plan Update 2021-2026*. Available at: https://siope.eu/media/documents/siop-europes-strategic-plan-update-2021-2026.pdf (Accessed July 24, 2023).

11. Cohen AR. Brain tumors in children. *N Engl J Med* (2022) 386:1922–31. doi: 10.1056/NEJMra2116344

12. Tozzi AE, Fabozzi F, Eckley M, Croci I, Dell'Anna VA, Colantonio E, et al. Gaps and opportunities of artificial intelligence applications for pediatric oncology in european research: A systematic review of reviews and a bibliometric analysis. *Front Oncol* (2022) 12:905770. doi: 10.3389/fonc.2022.905770

13. Horbinski C, Berger T, Packer RJ, Wen PY. Clinical implications of the 2021 edition of the WHO classification of central nervous system tumours. *Nat Rev Neurol* (2022) 18:515–29. doi: 10.1038/s41582-022-00679-w

14. World Health Organization. *Ethics and governance of artificial intelligence for health: WHO guidance* (2021). Geneva: World Health Organization. Available at: https://apps.who.int/iris/handle/10665/341996 (Accessed November 24, 2022).

15. World Health Organization. *Ageism in artificial intelligence for health: WHO policy brief* (2022). Geneva: World Health Organization. Available at: https://apps.who.int/iris/handle/10665/351503 (Accessed November 24, 2022). p.

16. EUROPEAN COMMISSION. *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

17. Huang J, Shlobin NA, Lam SK, DeCuypere M. Artificial intelligence applications in pediatric brain tumor imaging: A systematic review. *World Neurosurg* (2022) 157:99–105. doi: 10.1016/j.wneu.2021.10.068

18. Kremer L, Leclercq E, van Dalen E. *Methods to be used in reviews – CCG standards* . Cochrane Child Cancer Group. Available at: http://childhoodcancer.cochrane.org/sites/childhoodcancer.cochrane.org/files/public/uploads/CCG%20standards_Search%20strategy%20and%20Selection%20of%20studies.pdf (Accessed March 5, 2022).

19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev* (2016) 5:210. doi: 10.1186/s13643-016-0384-4

20. Aria M, Cuccurullo C. bibliometrix : An R-tool for comprehensive science mapping analysis. *J Informetr* (2017) 11:959–75. doi: 10.1016/j.joi.2017.08.007

21. Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* (2016) 3:160018. doi: 10.1038/sdata.2016.18

22. *R: The R Project for Statistical Computing*. Available at: https://www.r-project.org/ (Accessed November 24, 2022).

23. Ross B, Michaelis T. Clinical applications of magnetic resonance spectroscopy. *Magn Reson Q* (1994) 10:191–247.

24. Bentzen HB, Castro R, Fears R, Griffin G, Ter Meulen V, Ursin G. Remove obstacles to sharing health data with researchers outside of the European Union. *Nat Med* (2021) 27:1329–33. doi: 10.1038/s41591-021-01460-0

25. Vlahou A, Hallinan D, Apweiler R, Argiles A, Beige J, Benigni A, et al. Data sharing under the general data protection regulation: time to harmonize law and research ethics? *Hypertens Dallas Tex 1979* (2021) 77:1029–35. doi: 10.1161/HYPERTENSIONAHA.120.16340

26. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med* (2018) 378:981–3. doi: 10.1056/NEJMp1714229

27. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. *J Healthc Inform Res* (2021) 5:1–19. doi: 10.1007/s41666-020-00082-4

28. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med* (2020) 3:119. doi: 10.1038/s41746-020-00323-1

29. Warnat-Herresthal S, Schultze H, Shastry KL, Manamohan S, Mukherjee S, Garg V, et al. Swarm Learning for decentralized and confidential clinical machine learning. *Nature* (2021) 594:265–70. doi: 10.1038/s41586-021-03583-3

30. Rajotte J-F, Bergen R, Buckeridge DL, El Emam K, Ng R, Strome E. Synthetic data as an enabler for machine learning applications in medicine. *iScience* (2022) 25:105331. doi: 10.1016/j.isci.2022.105331

31. Belenkaya R, Gurley MJ, Golozar A, Dymshyts D, Miller RT, Williams AE, et al. Extending the OMOP common data model and standardized vocabularies to support observational cancer research. *JCO Clin Cancer Inform* (2021) 5:12–20. doi: 10.1200/CCI.20.00079

32. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* (2015) 216:574–8. doi: 10.3233/978-1-61499-564-7-574

33. Papez V, Moinat M, Voss EA, Bazakou S, Van Winzum A, Peviani A, et al. Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond. *J Am Med Inform Assoc JAMIA* (2022), 30(1):103–111. doi: 10.1093/jamia/ocac203

34. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. *J BioMed Inform* (2019) 94:103188. doi: 10.1016/j.jbi.2019.103188

35. Nikam RM, Yue X, Kaur G, Kandula V, Khair A, Kecskemethy HH, et al. Advanced neuroimaging approaches to pediatric brain tumors. *Cancers* (2022) 14:3401. doi: 10.3390/cancers14143401

36. Deng J, Dong W, Socher R, Li L-J, Kai Li, Li F-F. (2009). ImageNet: A large-scale hierarchical image database, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL: IEEE. pp. 248–55. doi: 10.1109/CVPR.2009.5206848

37. Quiñonero-Candela J ed. *Dataset shift in machine learning*. Cambridge, Mass: MIT Press (2009). p. 229. p.

38. Chi S, Tian Y, Wang F, Zhou T, Jin S, Li J. A novel lifelong machine learning-based method to eliminate calibration drift in clinical prediction models. *Artif Intell Med* (2022) 125:102256. doi: 10.1016/j.artmed.2022.102256

39. Shabani M. Will the European Health Data Space change data sharing rules? *Science* (2022) 375:1357–9. doi: 10.1126/science.abn4874

40. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell* (2020) 2:305–11. doi: 10.1038/s42256-020-0186-1

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 14 February 2024
DOI 10.3389/fimmu.2024.1323418

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Fei Yuan,
Baylor College of Medicine, United States
Yuan Xiao,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Hongyan Li
lihongyan.de@163.com

RECEIVED 17 October 2023
ACCEPTED 29 January 2024
PUBLISHED 14 February 2024

CITATION
Tang D, Huang Y, Che Y, Yang C, Pu B, Liu S and Li H (2024) Identification of platelet-related subtypes and diagnostic markers in pediatric Crohn's disease based on WGCNA and machine learning.
*Front. Immunol.* 15:1323418.
doi: 10.3389/fimmu.2024.1323418

COPYRIGHT
© 2024 Tang, Huang, Che, Yang, Pu, Liu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Identification of platelet-related subtypes and diagnostic markers in pediatric Crohn's disease based on WGCNA and machine learning

Dadong Tang[1], Yingtao Huang[2], Yuhui Che[1], Chengjun Yang[3], Baoping Pu[1], Shiru Liu[4] and Hongyan Li[4]*

[1]Clinical Medical College, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [2]First Clinical Medical College, Liaoning University of Traditional Chinese Medicine, Shenyang, China, [3]Department of Otorhinolaryngology, Zigong Hospital of Traditional Chinese Medicine, Zigong, China, [4]Anorectal Disease Department, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China

**Background:** The incidence of pediatric Crohn's disease (PCD) is increasing worldwide every year. The challenges in early diagnosis and treatment of PCD persist due to its inherent heterogeneity. This study's objective was to discover novel diagnostic markers and molecular subtypes aimed at enhancing the prognosis for patients suffering from PCD.

**Methods:** Candidate genes were obtained from the GSE117993 dataset and the GSE93624 dataset by weighted gene co-expression network analysis (WGCNA) and differential analysis, followed by intersection with platelet-related genes. Based on this, diagnostic markers were screened by five machine learning algorithms. We constructed predictive models and molecular subtypes based on key markers. The models were evaluated using the GSE101794 dataset as the validation set, combined with receiver operating characteristic curves, decision curve analysis, clinical impact curves, and calibration curves. In addition, we performed pathway enrichment analysis and immune infiltration analysis for different molecular subtypes to assess their differences.

**Results:** Through WGCNA and differential analysis, we successfully identified 44 candidate genes. Following this, employing five machine learning algorithms, we ultimately narrowed it down to five pivotal markers: GNA15, PIK3R3, PLEK, SERPINE1, and STAT1. Using these five key markers as a foundation, we developed a nomogram exhibiting exceptional performance. Furthermore, we distinguished two platelet-related subtypes of PCD through consensus clustering analysis. Subsequent analyses involving pathway enrichment and immune infiltration unveiled notable disparities in gene expression patterns, enrichment pathways, and immune infiltration landscapes between these subtypes.

**Conclusion:** In this study, we have successfully identified five promising diagnostic markers and developed a robust nomogram with high predictive efficacy. Furthermore, the recognition of distinct PCD subtypes enhances our comprehension of potential pathogenic mechanisms and paves the way for future prospects in early diagnosis and personalized treatment.

KEYWORDS

pediatric Crohn's disease, platelet, immune infiltration, machine learning, bioinformatics

# Introduction

Crohn's disease is a chronic inflammatory disorder primarily affecting the gastrointestinal tract, and it is categorized as an inflammatory bowel disease (IBD) along with ulcerative colitis. Crohn's disease can involve any part of the gastrointestinal tract, from the oral cavity to the rectum, but most commonly affects the small intestine and the initial portion of the large intestine (1). In recent years, with the change in people's lifestyles and the influence of environmental factors, the incidence and prevalence of Crohn's disease have gradually increased and shown a significant trend toward younger ages (2). Pediatric Crohn's disease (PCD) patients usually have more extensive involvement, a more severe disease, and more atypical symptoms than adults with Crohn's disease (3). However, PCD is not the early stage of adult Crohn's disease and is quite different from adult IBD etiology and symptoms (4, 5). Presently, the precise etiology of PCD remains partially elusive. Generally, it is attributed to factors such as environmental influences, alterations in intestinal flora, genetic predisposition, anomalous immune responses of the mucosa, and impairment of epithelial barrier function (1, 6).

Given the heterogeneity of PCD patient characteristics, the diagnosis of PCD is particularly difficult in clinical practice and often requires a combination of ultrasound, endoscopy, and pathology to confirm the diagnosis (7). Therefore, it is imperative to find more accurate, convenient, non-invasive, and highly specific diagnostic tools for children with Crohn's disease. In recent years, molecular typing has made great strides in the understanding and treatment of many diseases. Mature molecular typing strategies are able to predict the optimal therapeutic strategy prior to patient treatment, thereby substantially improving patient prognosis (8). In previous reports, the molecular subtyping of diseases such as hepatocellular carcinoma, osteoarthritis, and rheumatoid arthritis, for example, has laid the foundation for their individualized treatment (9–11). Therefore, the development of molecular subtyping of PCD based on its molecular features is crucial for improving the accuracy of clinical treatment decisions and deepening our understanding of PCD.

Platelets are small blood cells derived from bone marrow, and in addition to their well-known role in blood clotting and wound healing, there is growing evidence that they also play a crucial role in autoimmune processes (12, 13). Currently, research on the role of platelets in the pathogenesis of a number of autoimmune diseases, including rheumatoid arthritis and systemic lupus erythematosus, is progressing (14, 15). However, the contribution of platelets to the pathogenesis of PCD, a classic autoinflammatory disease, has not been investigated. Expanding the scope to encompass systems and cell types that have not been extensively studied could furnish a more comprehensive understanding of PCD progression and potentially uncover novel biomarkers.

Although the precise pathogenic mechanisms of PCD remain enigmatic, platelets have garnered considerable attention as potential players in the genesis of autoimmune-related conditions. Our study aims to contribute to the burgeoning body of research by investigating the regulation of platelet-related genes (PRGs) in PCD through an integrated bioinformatics approach. We also endeavor to devise diagnostic models and molecular subtypes grounded in platelet-associated pathways. Our findings hold the promise of illuminating the pathogenesis and prognostic diversity of PCD. Additionally, the identification of fresh biomarkers holds the potential to facilitate early diagnosis, tailored treatment strategies, and risk stratification for PCD in the future. The schematic depiction of our study's methodology is outlined in Figure 1.

# Methods

## Dataset acquisition and processing

In this study, we obtained three datasets from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) to support subsequent analysis: GSE117993 (16), GSE93624 (17), and GSE101794 (18). The GSE117993 dataset comprises transcriptome data from 55 normal individuals and 92 patients with PCD. Similarly, GSE93624 contains transcriptome data from 35 normal individuals and 210 PCD patients. Both GSE117993 and GSE93624 were employed as training sets within this study. Additionally, the GSE101794 dataset includes 304 ileal biopsy samples, with 50 sourced from individuals with normal conditions and 254 obtained from PCD patients. This dataset serves

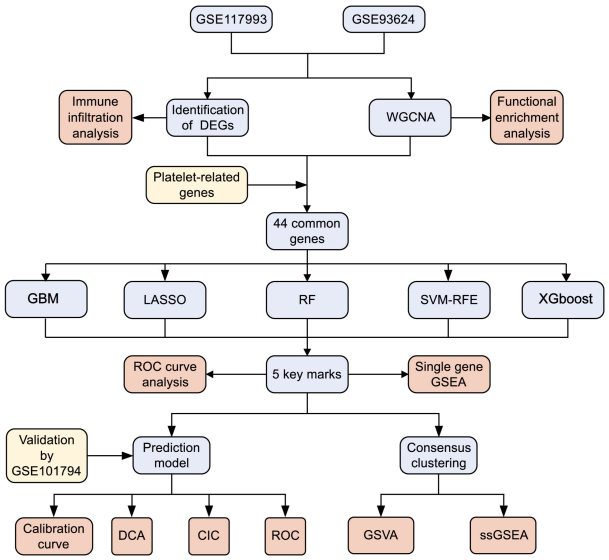


FIGURE 1
Flowchart for research. DEGs, differentially expressed genes; WGCNA,Weighted gene co-expression network analysis; GBM, Gradient boosting machine; LASSO, Least absolute shrinkage and selection operator; RF, random forest; SVM-RFE, support vector machine-recursive feature elimination; XGBoost, Extreme Gradient boosting; ROC, receiver operating characteristic curve; GSEA, gene set enrichment analysis; DCA, decision curve analysis; CIC, clinical impact curves; GSVA, gene set variation analysis.

as the validation set. Detailed specifics regarding these three datasets can be found in Table 1 and Supplementary Table 1. Subsequently, we integrated the two training datasets using the "limma" package in R software (version 4.2.2) (19). Notably, when multiple probes from distinct platforms identified identical genes, we computed the average value to represent the expression level. We then employed the "sva" packages to normalize and correct for batch effects in the training set (20). Additionally, for the acquisition of platelet-related genes (PRGs), we utilized a gene set consisting of 300 PRGs identified in a previous similar study (11).

## Identification of differentially expressed genes

Differential gene analysis was conducted between the samples using the "limma" package on the normalized dataset. We applied significance criteria with adjusted p-values < 0.05 and |log2 fold change (FC)| > 0.3. Subsequently, the differentially expressed genes (DEGs) that met these criteria were visualized using the volcano plot. Additionally, the heat map was generated to display the top 20 up-regulated DEGs and the top 20 down-regulated DEGs.

## Immune infiltration analysis

The single-sample gene enrichment analysis (ssGSEA) algorithm characterizes the state of a cell by assessing the activity levels of specific biological processes and pathway pairs. In this study, we employed the ssGSEA algorithm, relying on the "GSVA" package (21), to assess the relative infiltration abundance and correlation among immune cells in both the PCD and normal group samples. Subsequently, we visualized the results using the "gglot2" package (https://sourceforge.net/projects/ggplot2.mirror/).

## Identification of platelet-related signature genes based on weighted gene co-expression network analysis (WGCNA)

To probe the potential regulatory associations between genes, we used the "WGCNA" package of the R software to construct gene co-expression networks (22). First, in order to ensure the reliability of the constructed network results, the normal-value samples were screened with mean fragments per kilobase million (FPKM) > 0.5 as the filtering criterion. Second, cluster analysis was performed using the "flashClust" toolkit, retaining the samples that best fit in the cluster under a specified threshold. Third, use the "pickSoftThreshold" function to select the soft threshold of the optimal weighting coefficient β value to establish a scale-free network, thereby transforming the similarity matrix into an adjacency matrix. Fourth, convert the adjacency matrix into a topological overlap matrix (TOM) and calculate the corresponding dissimilarity (1-TOM). Fifth, the module clipping height was set to 0.3 and the minimum number of modules was set to 100, and then the dynamic tree cutting method was used to identify modules from the hierarchical clustering tree. In addition, we calculated the module membership (MM) and gene significance (GS) for the modules correlated to the clinical attributes. Finally, the module most closely related to PCD was screened, and the genes in this module were intersected with the DEGs and PRGs for further analysis.

TABLE 1 Characteristics of the microarray datasets for PCD.

| GEO ID | Platform | Year | PCD | Control | Biopsy Tissue Source |
|---|---|---|---|---|---|
| Training set: | | | | | |
| GSE117993 | GPL16791 | 2018 | 106 PCD samples | 55 controls | rectal |
| GSE93624 | GPL11154 | 2017 | 210 PCD samples | 35 controls | ileum |
| Validation set: | | | | | |
| GSE101794 | GPL11154 | 2018 | 254 PCD samples | 50 controls | ileum |

GEO, Gene Expression Omnibus; PCD, PCD.

## Functional enrichment analysis

To investigate the underlying biological mechanisms involved in the development of PCD, we conducted gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses on key module genes, respectively. All analysis were performed by the "clusterProfiler" package of the R software (23), and adjusted q-values < 0.05 were set as significant enrichment thresholds. Subsequently, we obtained candidate diagnostic genes for further analysis by intersecting the key modular genes with DEGs and PRGs.

## Feature genes selection for PCD via machine learning

To identify key markers for PCD, we applied five machine learning algorithms: Support Vector Machine Recursive Feature Elimination (SVM-RFE), Least Absolute Shrinkage and Selection Operator Logistic Regression (LASSO), Gradient Boosting Machine (GBM), Extreme Gradient boosting (XGBoost) and Random Forest. These algorithms were implemented using specific R packages: "e1071" for Support Vector Machine Recursive Feature Elimination (available at https://github.com/johncolby/SVM-RFE), "glmnet" (24) for the LASSO logistic regression, "xgboost" for the XGBoost, (available at https://cran.r-project.org/web/packages/xgboost/) "caret" for the GBM (available at https://cran.r-project.org/web/packages/caret/), and "randomForest" (available at https://cran.r-project.org/web/packages/randomForest/) for Random Forest. It's noteworthy that in this study, both the LASSO regression and SVM-RFE algorithms were evaluated using 10-fold cross-validation to estimate their prediction performance. Additionally, for the Random Forest (RF) algorithm, genes with a relative importance greater than 2.5 were considered key markers. For the "XGBoost" and "GBM" algorithms, we selected the top ten genes in terms of importance for the screening of candidate marker genes. The signature genes were determined by identifying the intersection of genes screened by these five algorithms. To visually represent the relationships among the signature genes, we utilized the "circlize" package to create a heatmap of gene correlations (25). These maps illustrate the interplay between the roles of these genes. Additionally, we generated heat maps based on Spearman correlation analysis to depict the relationships between the signature genes and 23 immune cell types.

## Gene Set Enrichment Analysis (GSEA) for single diagnostic marker

To thoroughly investigate the potential connections between key diagnostic genes and signaling pathways, we conducted a comparative analysis of biosignaling pathways between the disease and control groups. Subsequently, we generated ridge maps utilizing the "clusterProfiler" package to visually represent the top 10 enrichment results for each gene.

## Establishment and validation of the nomogram for PCD

We developed a clinical prediction model for PCD based on these signature genes, utilizing the "gglot2" package. The calibration curve was used to evaluate the nomogram model's predictive accuracy. Then, we performed decision curve analysis and clinical impact curve analysis to estimate the nomogram model's clinical utility. Furthermore, we assessed and validated the predictive performance of each signature gene in both the training and validation sets using receiver operating characteristic (ROC) curves. A larger area under the curve (AUC) in the ROC analysis indicated better predictive performance of the model. Finally, we use the dataset GSE101794 as a validation set to test whether the prediction model is generalizable.

## Identification of platelet-associated molecular subtypes of PCD by consensus cluster analysis

Consensus clustering is an unsupervised algorithm designed to identify and cluster individual samples within a dataset, facilitating the distinction of various subtypes. In this study, we applied consensus clustering to differentiate between distinct subtypes of platelet-related PCD molecules. This differentiation was based on the feature genes identified by five machine learning algorithms, and we utilized the "ConsensusClusterPlus" package for this purpose (26). The optimal number of clusters was determined through the analysis of consensus cumulative distribution function (CDF) plots, principal component analysis (PCA), consensus matrix plots, relative changes in the area under the CDF curve, and tracking plots. To explore variances in platelet-associated pathways across different molecular subtypes, we employed the gene set variation analysis (GSVA) algorithm and visualized the results. Additionally, for the analysis of discrepancies in immune cell infiltration status among subtypes, we conducted the ssGSEA algorithm using R, followed by the Wilcoxon test to assess differences in the immune microenvironment.

# Results

## Data processing and identification of DEGs

The training set consists of GSE117993 and GSE93624 and contains a total of 302 PCD patients and 90 normal patient samples. After the ComBat algorithm to remove the batch effect, all the samples in the training set were centered and well-distributed, indicating the high quality and cross-comparability of this microarray dataset (Figures 2A, B). Subsequent differential analysis eventually screened out 1643 genes, including 887 up-regulated genes and 756 down-regulated genes (Figure 2C and Supplementary Table 1). The heatmap demonstrated the top 20 up-

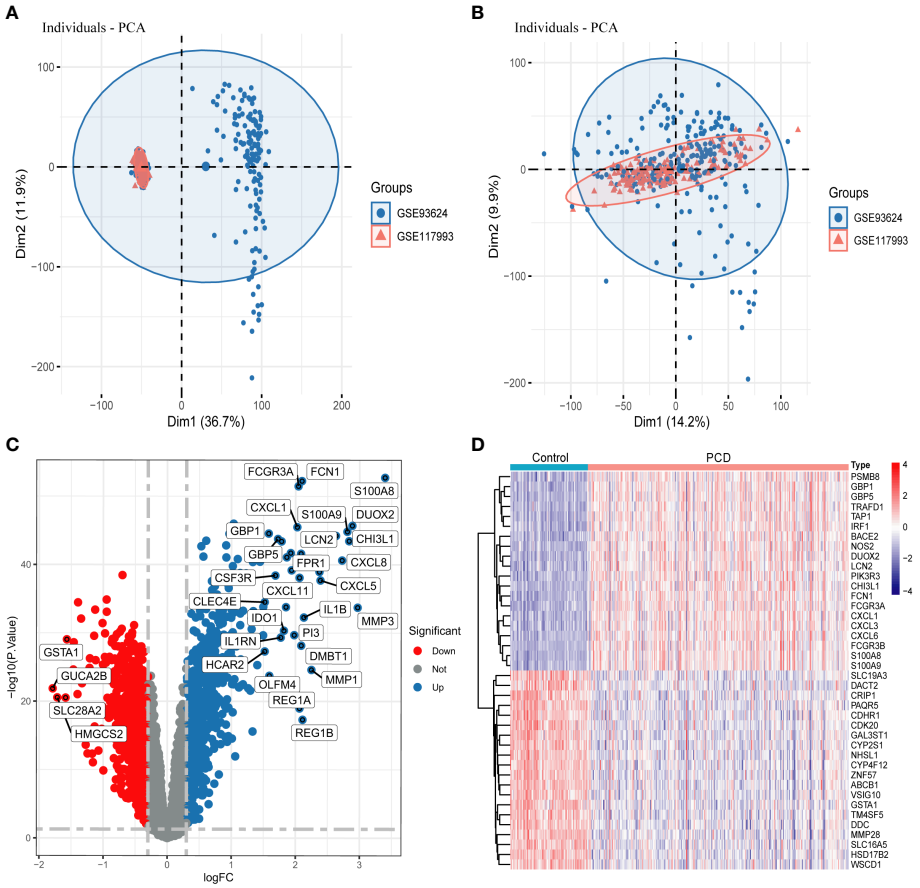


FIGURE 2
Data processing and screening for DEGs. **(A, B)** PCA plots showing the difference in expression profiles of the combined GSE93624 and GSE117993 dataset before and after the de-batch effect. **(C)** Volcano plot of DEGs in the combined dataset. **(D)** Heatmap of DEGs in the combined dataset. DEGs, differentially expressed genes.

and down-regulated DEGs between PCD patients and the control group (Figure 2D).

## Immune infiltration analysis

In this study, we used the ssGSEA algorithm to compare the levels of immune cell infiltration between the PCD patients and the normal group in the combined dataset. Figure 3A depicts the difference in expression of 23 immune cells in PCD samples and normal samples. Box plots were used to show the differences in immune cell infiltration between the PCD and control groups (Figure 3B). The results showed that cells such as activated CD4 T cells, activated B cells, macrophages, and neutrophils were significantly higher in the PCD patient group compared to the control group. In addition, we investigated the relationship between immune cells. As illustrated by the correlation heatmap in Figure 3C, macrophages, immature dendritic cells, mast cells, activated dendritic cells, and type 2 T helper cells exhibited strong correlations with the majority of other immune cell types.

## Identification of platelet-related signature genes based on WGCNA

We used WGCNA to filter the set of genes that are more consistent with PCD and selected the optimal soft threshold based on the results of the "pickSoftThreshold" function of the "WGCNA" package. In addition, in order to make the co-expression network conform to the scale-free principle, β=7 (scale-free $R^2$ = 0.9) was chosen to construct the gene co-expression network in this study (Figure 4A). Finally, we converted the similarity matrix to TOM. Next, based on TOM, we used the previously mentioned criteria to merge the modules that were closer together, and finally obtained a total of 11 modules (Figures 4B, C and Supplementary Table 2). Finally, based on the results of the module-trait association analysis, we selected the turquoise module with the strongest association with PCD (5997 genes in total). In addition, a substantial positive correlation was identified between module membership and gene significance in the turquoise module for PCD samples (r = 0.89, p = 1e-200), as illustrated in Figure 4D.

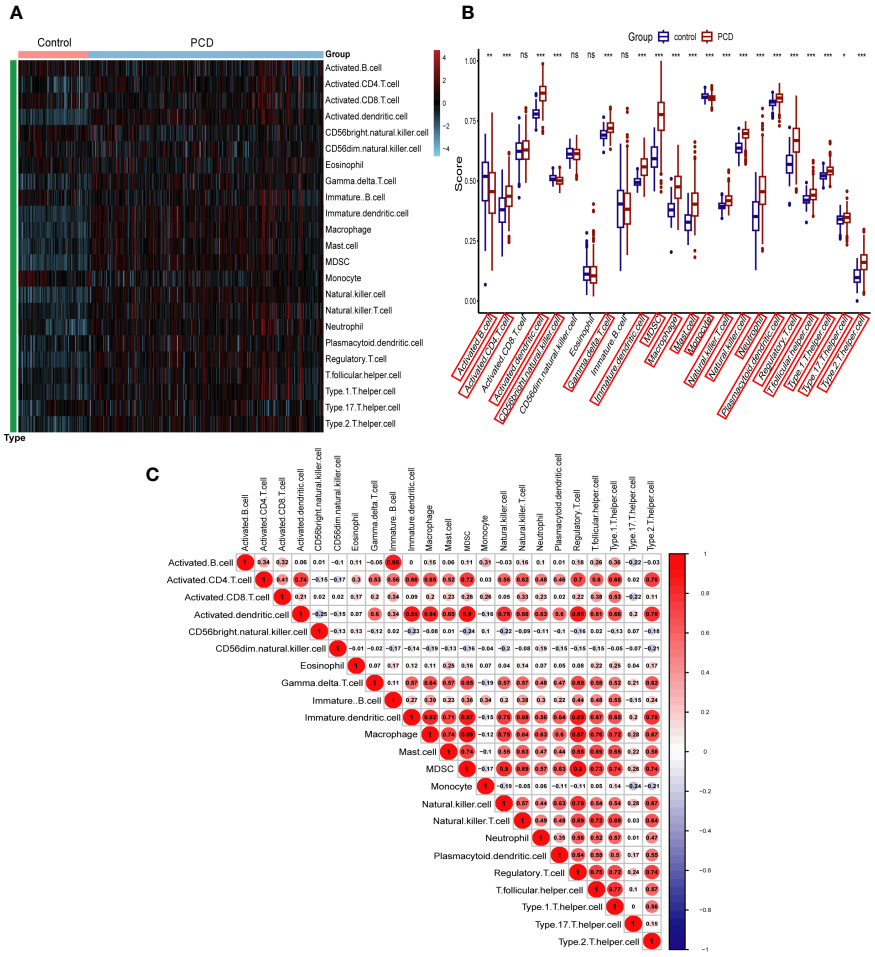


FIGURE 3
Immune cell infiltration analysis in combined datasets. **(A)** Heatmap of all PCD patient and control samples with 23 immune cells. **(B)** Boxplot depicting the level of immune cell infiltration in PCD patients and controls. **(C)** Heatmap depicting the correlations between distinct immune cell compositions. PCD, pediatric Crohn's disease; *p < 0.05, **p < 0.01, ***p < 0.001; ns, no significance.

## Functional enrichment analysis

In order to investigate the biological functions and pathways associated with the turquoise modular gene set, which exhibited the strongest association with PCD as identified through WGCNA, in the context of PCD development, we conducted GO and KEGG analyses on this gene set. This analysis yielded a total of 37 significantly enriched signaling pathways and 1661 biological processes that displayed significant associations (refer to Supplementary Table 3). As depicted in Figure 5A, these genes were predominantly enriched in signaling pathways closely linked to immune and inflammatory responses, including but not limited to the PI3K-Akt, MAPK, TNF, IL-17, and Rap 1 pathways. The GO enrichment analysis indicated that the majority of these genes were primarily involved in cytokine-mediated signaling pathways, positive regulation of cytokine production, responses to molecules of bacterial origin, cell chemotaxis, and reactions to lipopolysaccharides and polysaccharides, among other biological processes, as illustrated in Figure 5B. Furthermore, we employed a Venn diagram, as depicted in Figure 5C, to identify an intersection between the key module gene set and DEGs as well as PRGs. This intersection yielded a set of 44 candidate diagnostic genes for subsequent analysis.

## Screening signature genes by machine learning

Based on the 44 key genes screened in the previous step, we used five machine learning algorithms to further identify potential platelet-related biomarkers for PCD. In the SVM-RFE algorithm, we finally identified 15 featured genes (Figures 6A, B). In the LASSO algorithm, we screened 16 feature genes (Figures 6C, D). In the "XGBoost" and "GBM" algorithms, we filtered out the top 10 feature genes in terms of importance (Figures 6E, F).And in the RF algorithm, we finally screened a total of 16 feature genes with relative importance greater than 2.5 (Figure 6G and Supplementary Figure 1). Subsequently, the Venn diagram showed that there were five identical genes among the five machine learning algorithms: PIK3R3, STAT1, PLEK, GNA15, and SERPINE1 (Figure 6H).

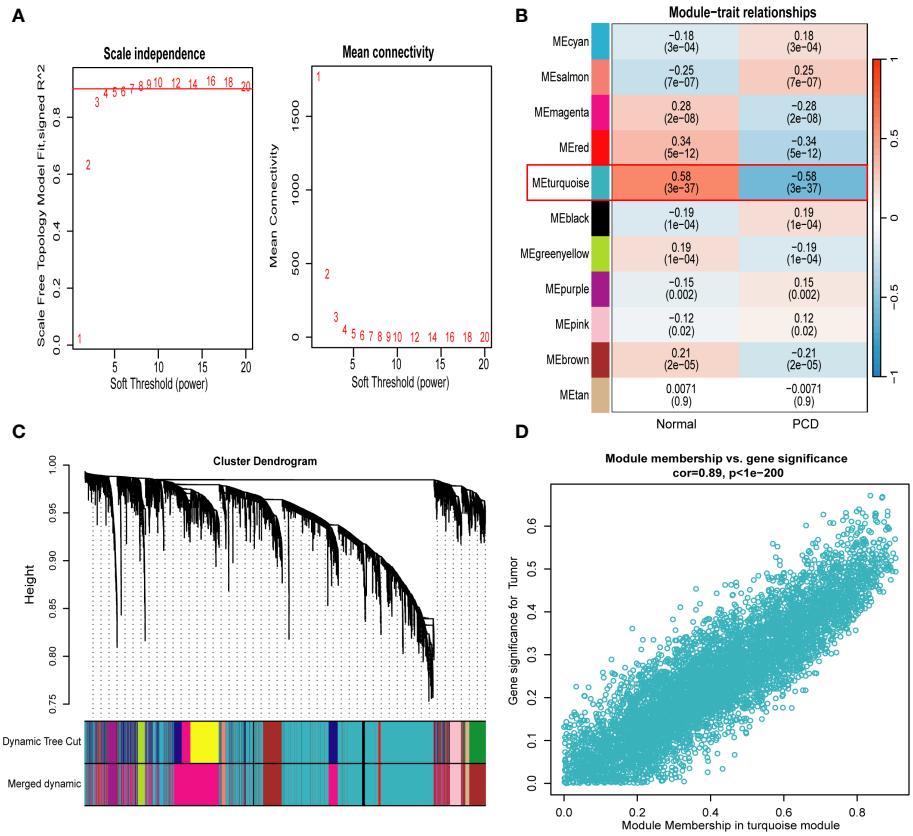


FIGURE 4
Weighted gene co-expression network analysis. **(A)** Identification of soft–threshold powers. $R^2$ = 0.90. **(B)** Relationships between modules and traits in PCD and control. The numbers in each cell represent the correlation coefficient and p-value. **(C)** Cluster tree for all genes in the combined dataset. **(D)** Scatterplot of genes in the turquoise module with the strongest correlation to PCD. PCD, pediatric Crohn's disease.

## The association of key markers with immune cells and GSEA

We conducted Spearman correlation analysis to elucidate the interrelationships among the diagnostic genes and their potential association with immune cell infiltration, thereby enhancing our understanding of the functional role of these pivotal markers in immune infiltration. The correlation analysis revealed a robust positive correlation among all five diagnostic genes (all exceeding 0.4), as illustrated in Figure 7A. Furthermore, the associations between these five diagnostic genes and a majority of the immune cells exhibited a highly consistent pattern. Specifically, all five diagnostic markers displayed varying degrees of negative correlation with activated B cells, CD56 bright natural killer cells, CD56dim natural killer cells, monocytes, eosinophils, T-helper cells type 17, and activated CD8 T cells. Additionally, there existed an overall strong positive correlation between the remaining 16 immune cells and these five diagnostic genes, as depicted in Figure 7B. As demonstrated in Figures 8A–E, GNA15, PIK3R3, PLEK, SERPINE1, and STAT1 exhibited significant enrichment in inflammatory and immune-related pathways, such as cytokine-cytokine receptor interaction, NOD-like receptor signaling pathway, and chemokine signaling pathway.

## Establishment and validation of the nomogram for PCD

Inspired by the results of previous studies, we constructed a clinical risk prediction model for PCD using the five signature genes that were finally screened (Figure 9A). In addition, the ROC-AUCs of the five characterized genes were higher than 0.8, indicating that the model has good discriminative performance (Figure 9B). The calibration curve of the model showed that the nomogram performed close to ideal with excellent predictive consistency (Figure 9C). As shown in Figure 9D, there is a high net clinical benefit for clinicians and patients using this model. The clinical impact curve further confirms the high clinical efficacy of this predictive model (Figure 9E). Finally, we further validated the predictive ability of the model in an external dataset GSE101794. In the validation set, the AUC of the model is 0.839, indicating that the model has some stability and generalizability (Figure 9F).

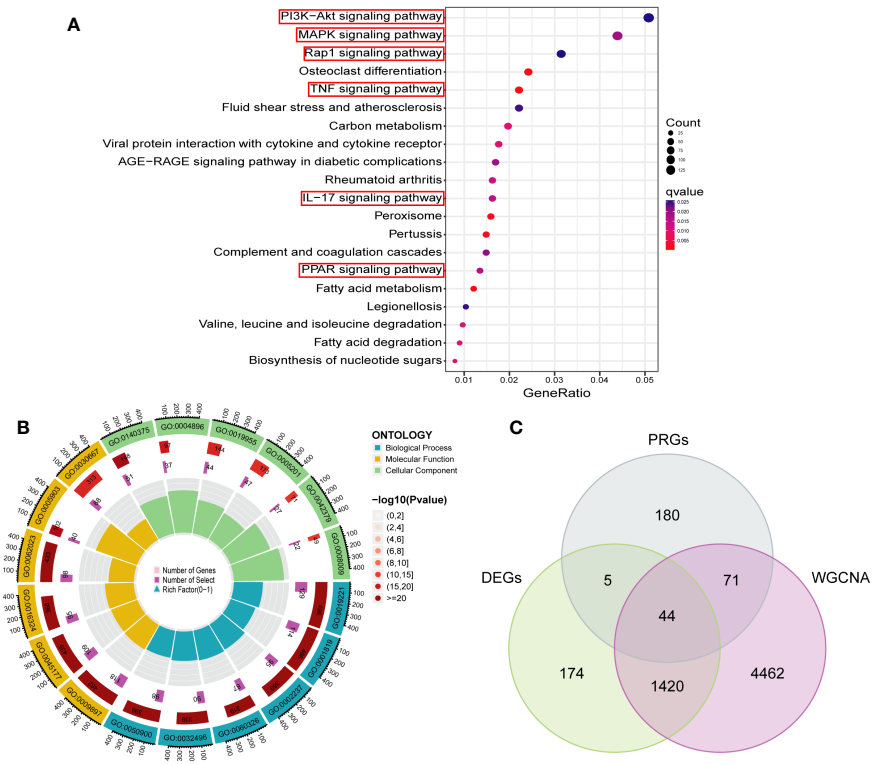


FIGURE 5
Functional enrichment analysis of key module genes and identification of candidate diagnostic genes. **(A)** The bubble plot showing the most enriched KEGG pathways of key module. **(B)** Gene Ontology analysis of key module. **(C)** Candidate diagnostic genes were obtained by overlapping key modules, DEGs and PRGs. DEGs, differentially expressed genes; PRGs, platelet-related genes; KEGG, kyoto encyclopedia of genes and genome; BP, biological process; CC, cellular component; MF, molecular function.

## Identification of platelet-associated molecular subtypes of PCD by consensus cluster analysis

To identify different PCD molecular subtypes, we analyzed the expression profiles of five characteristic genes in all PCD samples in the training set based on a consensus clustering approach. The consensus matrix plot, CDF plot, and the area under the CDF curve change plot finally identified K=2 as the optimal number of subtypes (Figure 10A and Supplementary Figures 2A–C). We named the two platelet molecular subtypes as C1 (n=115) and C2 (n=187) (Supplementary Table 4). Principal component analysis revealed significant heterogeneity in gene expression patterns between the two subtypes (Figure 10B). Subsequently, we also highlighted the expression pattern of PDEGs in the two subtypes with heatmap and boxplot (Figures 10C, D). It was evident that the expression of the five key marks was significantly upregulated in the C2 cluster. In order to delve further into the potential pathogenic mechanisms of each subtype and evaluate disparities in enriched pathways between them, we conducted GSVA for the two subtypes. Our findings revealed notable distinctions. For instance, the C2 cluster exhibited significantly greater enrichment compared to the C1 cluster across the majority of the HALLMARK pathways, encompassing responses to interferon gamma and interferon alpha, xenograft rejection, and inflammatory response. Notably, the fatty acid metabolism pathway displayed heightened enrichment in C1 clusters as opposed to C2 clusters (Figure 11A). Likewise, C2 clusters demonstrated significantly greater enrichment than C1 clusters in pathways involving interleukin 6, 9, 21, 35, and cytokine signaling within the immune system (Figure 11B). Additionally, our analysis based on the KEGG database revealed a significant enrichment in pathways like systemic lupus erythematosus, complement coagulation cascade reaction, and cell adhesion molecules for the C2 cluster (Figure 11C). It is imperative to note that in subsequent immune cell infiltration analysis, we observed a substantially higher abundance of immune infiltration in the C2 cluster across all 23 immune cell types, aligning consistently with the findings from the GSVA (Figures 12A, B).

# Discussion

PCD is a refractory disease that has a major impact on all aspects of a child's life. In terms of nutrition, chronic inflammation and malabsorption can lead to nutritional deficiencies and growth disorders that affect children's development and overall health (27). On a psychological level, living with a chronic disease can take a toll on a child's mental and emotional health, potentially leading to anxiety, depression, and decreased quality of life (28). In terms of school and social activities, frequent symptoms and medical

10.3389/fimmu.2024.1323418

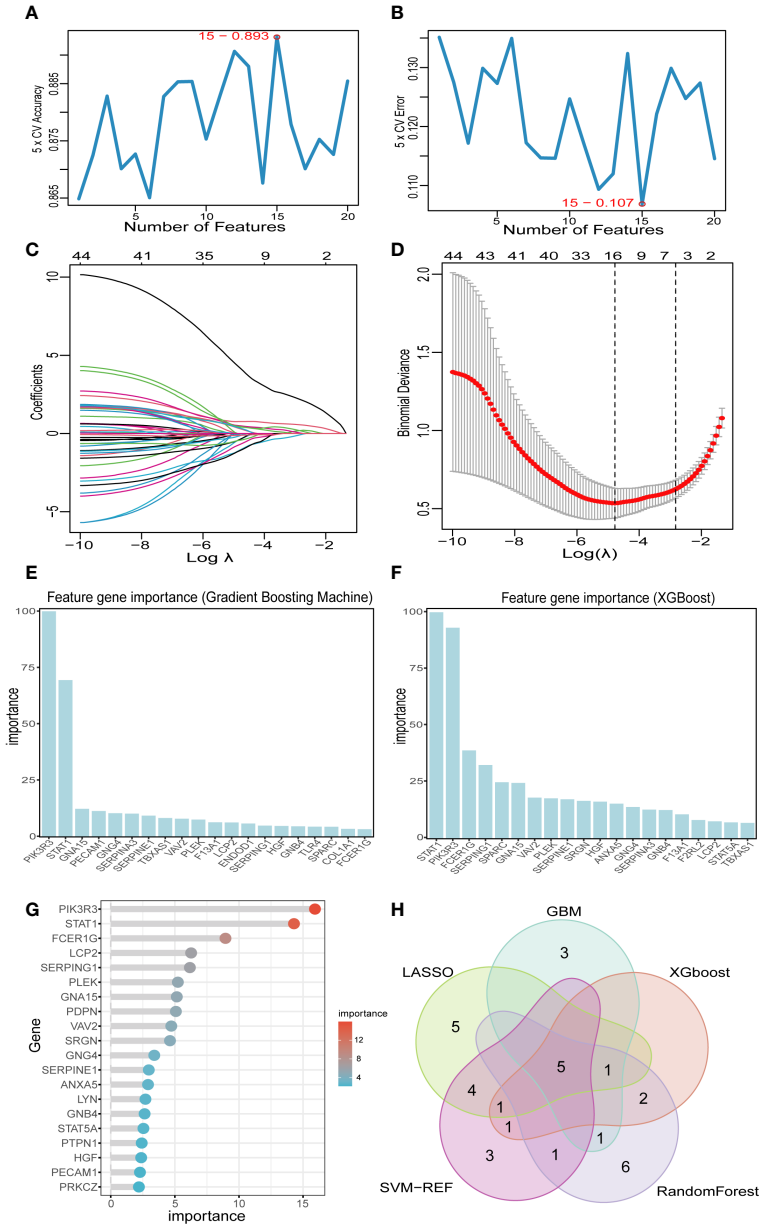


FIGURE 6
Screening key markers by machine learning. **(A, B)** Fifteen feature genes were screened by SVM-RFE algorithm. **(C, D)** Sixteen feature genes were screened by LASSO algorithm. **(E)** The top 20 genes in terms of importance as filtered by the GBM algorithm. **(F)** The top 20 genes in terms of importance as filtered by the XGBoost algorithm. **(G)** Sixteen feature genes were screened by RF algorithm. **(H)** Five key markers were screened by Venn diagram of five algorithms. LASSO, least absolute shrinkage and selection operator; SVM-RFE, support vector machine-recursive feature elimination; GBM, Gradient Boosting Machine; XGBoost, Extreme Gradient boosting; RF, random forest.

appointments can disrupt a child's educational and social life, requiring support from healthcare providers and educators (29). In addition, the vast majority of children with PCD are also at risk for long-term complications, and if not managed properly, PCD can lead to serious complications, including an increased risk of strictures, fistulas, and colorectal cancer (30). This is despite the fact that research into the development of PCD has achieved some milestones in recent years (31). However, issues such as delayed diagnosis, growth and developmental problems, drug safety, psychosocial support, and access to health care remain. Therefore, accurate differentiation of PCD clusters at the molecular level will improve our understanding of PCD heterogeneity and is essential to guiding individualized treatment of PCD. And platelet-related genes have demonstrated excellent diagnostic performance and molecular subtype discrimination in previous studies. However, whether platelets are associated with the development of PCD and the underlying molecular mechanisms of platelets in PCD remain unclear. Therefore, the exploration of new platelet-related diagnostic markers and risk stratification for children with PCD remains an urgent need.

In this study, we identified 887 up-regulated and 756 down-regulated DEGs from 302 PCD cases and 90 normal control

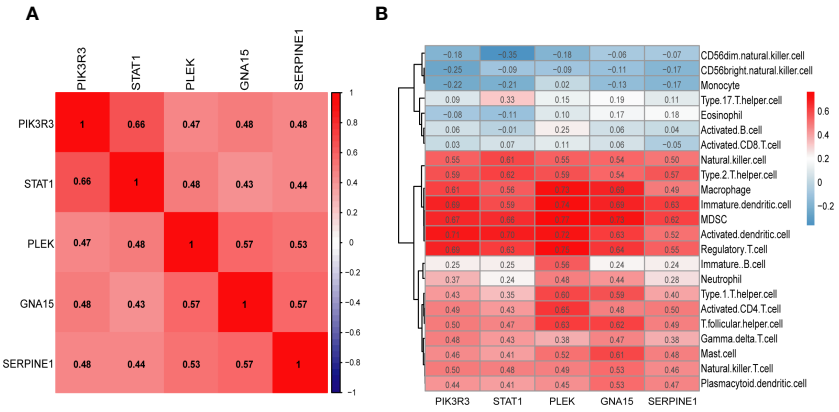


FIGURE 7
**(A)** The heatmap shows the relationship between the key markers. **(B)** Correlation heatmap depicting the relationship between immune cell infiltration and key markers.

samples using data from the GSE93624 and GSE117993 datasets. Employing WGCNA, we isolated 5997 pivotal modular genes closely associated with PCD. Subsequent GO enrichment analysis revealed that these key module genes were primarily enriched in biological processes, including the positive regulation of cytokine production, the negative regulation of immune system processes, and leukocyte migration. Positive regulation of cytokine production means increasing the expression or activity of cytokines, which are important for both normal and abnormal cell-mediated immune responses and intestinal inflammation. Cytokines play a central role in CD as they can either promote or suppress gut inflammation (32, 33). Leukocyte migration, also known as leukocyte transport, plays a pivotal role in the pathogenesis of IBD (34). In IBD, there is a chronic inflammatory cell infiltrate in the intestinal mucosa, and

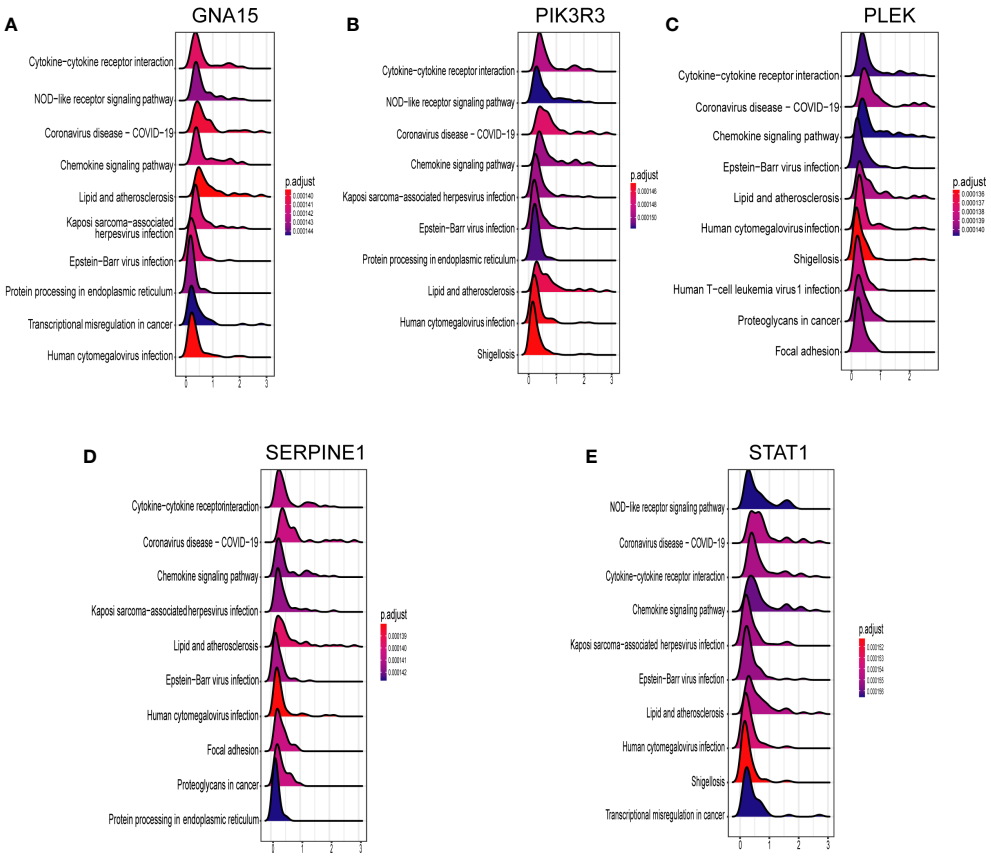


FIGURE 8
GSEA for the single mark. **(A)** GNA15. **(B)** PIK3R3. **(C)** PLEK. **(D)** SERPINNE1. **(E)** STAT1.

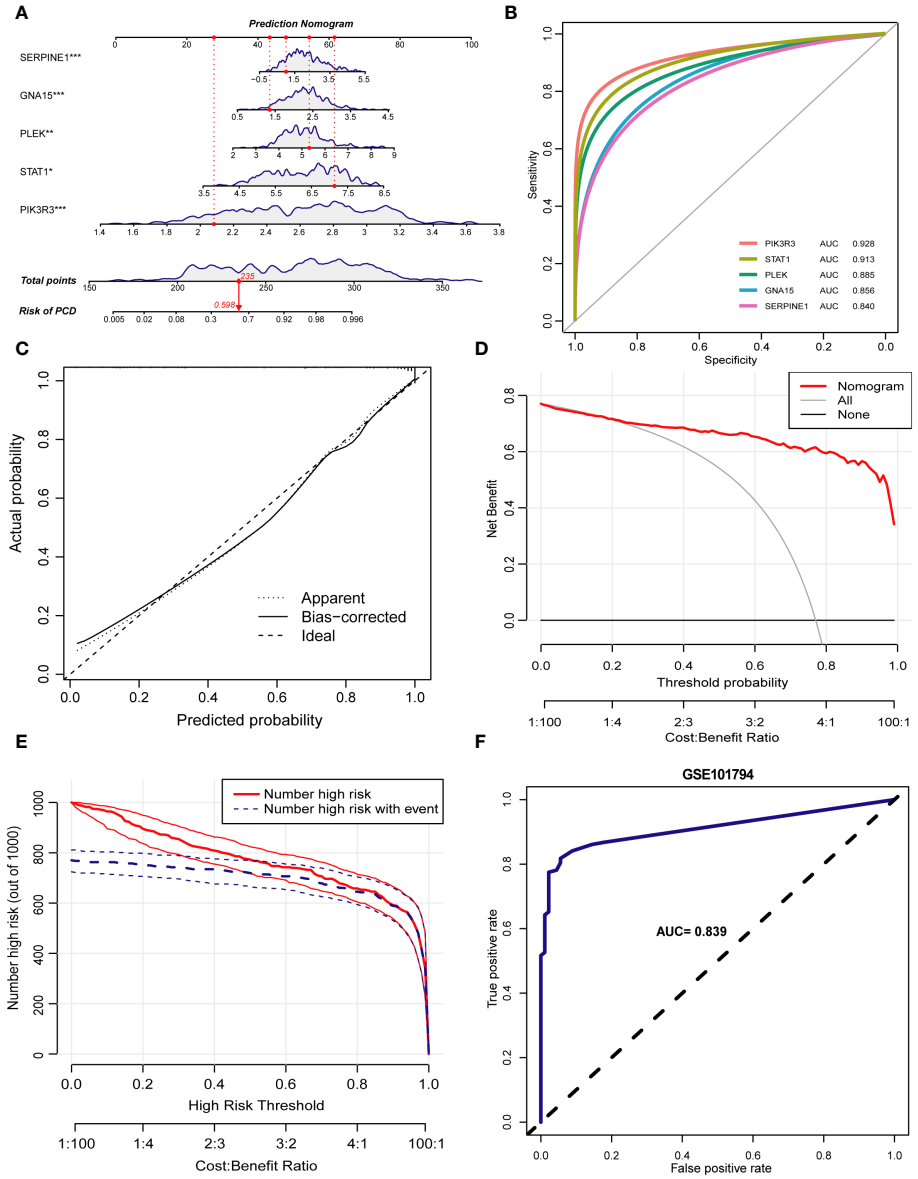


FIGURE 9
Construction of the prediction model. **(A)** Nomogram of PCD patients based on 5 key markers. **(B)** ROC curve of key markers in PCD diagnosis. **(C)** The calibration curve demonstrates nomogram's excellent predictive performance. **(D)** The results of the decision curve analysis showed a good net clinical benefit rate for the model. **(E)** Assessing the clinical impact of the nomogram by clinical impact curves. **(F)** The ROC curve demonstrates the predictive power of the predictive model in the validation set GSE101794. *p < 0.05, **p < 0.01, ***p < 0.001.

immune cells are transported to the gastrointestinal tract through specific mechanisms. Integrins, cell surface receptors, bind cell adhesion molecules, facilitating leukocyte homing and retention. Therapies targeting leukocyte trafficking may play a crucial role in achieving stratified precision medication administration in the care of IBD (35). However, it is important to note that patients may exhibit varying responses to these treatments, underscoring the need for additional mechanistic studies in clinical trials. Furthermore, KEGG enrichment analysis revealed that the development of PCD is primarily associated with signaling pathways such as PI3K-Akt, MAPK, and Rap 1. The PI3K-Akt, MAPK, and Rap 1 pathways are pivotal in various cellular processes, including cell growth, proliferation, differentiation, and survival (36). Aberrant activation of these pathways has been linked to various diseases, including cancer (37). While specific information regarding the relationship between these pathways and IBD is lacking, it is established that inflammation and immune responses play a significant role in IBD. These pathways may potentially influence these processes. For example, the PI3K-Akt pathway is known to regulate immune and inflammatory responses (38), and the MAPK pathway also plays a role in proinflammatory cytokine production (39). As for the Rap 1 pathway, although its role in IBD remains to be fully elucidated, Rap 1 is involved in a variety of cellular processes, including cell adhesion and cell-cell interactions, which could have implications for the pathogenesis of IBD (37). In a subsequent immune cell

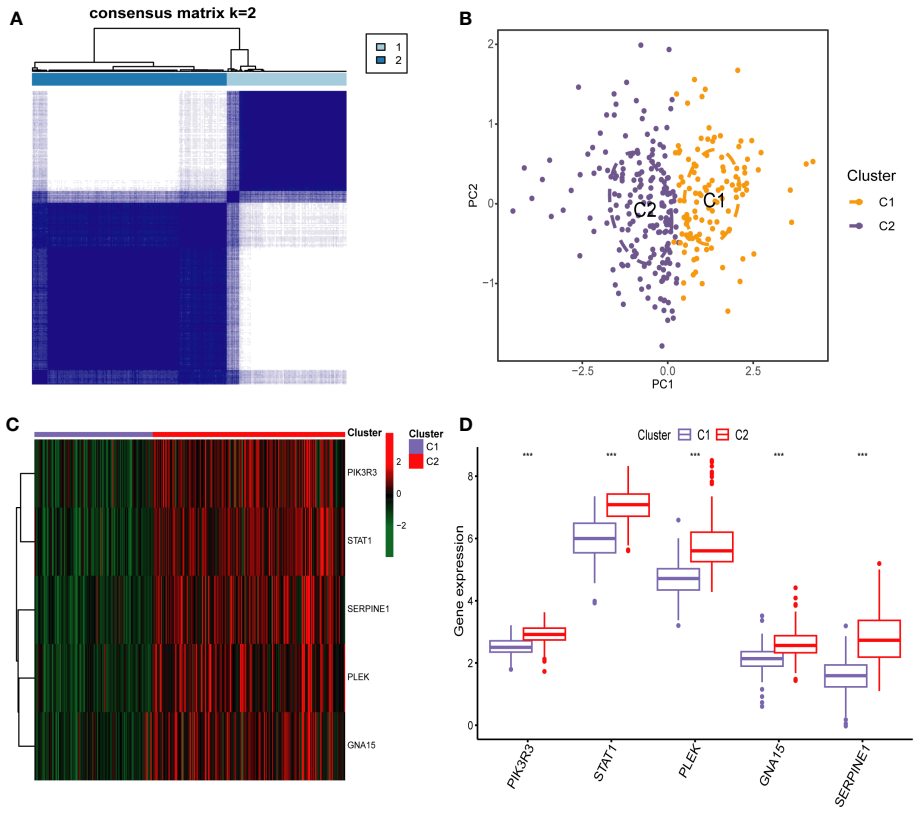


FIGURE 10
Identification of Platelet-Related Molecular Subtypes of PCD by Consensus Cluster anaylsis. **(A)** Consensus matrix heatmap for k = 2. **(B)** Sample distributions of the two subtypes were visualized by principal component analysis. Heatmap **(C)** and boxplot **(D)** demonstrating the expression differences of the five key markers in different subtypes. *** $p < 0.001$.

infiltration analysis, we found that 19 types of immune cells, including activated B cells, activated CD4 T cells, and monocytes, had higher infiltration abundance in PCD patients. Previous studies have found that in patients with Crohn's disease, B cells show signs of chronic stimulation and are localized to granulomatous tissue. These cells also exhibit increased molecular maturation of IgA and IgG1 (40). In addition, there are studies that have identified distinct subpopulations of B cells that are highly infiltrated in the lamina propria of patients with Crohn's disease. These cells express genes related to antigen presentation. This subset of B cells is thought to play a potential promoting role in the pathogenesis of Crohn's disease (41). In previous studies, CD4 T cells were found to play a key role in the pathogenesis of Crohn's disease (42). These cells can differentiate into regulatory and effector T cells, such as Th1, Th2, Th17, follicular helper T cells (Tfh), and regulatory T-cells (Tregs), depending on the cytokine milieu. In particular, CD4+ tissue-resident memory T (T_RM) cells have been found to be expanded in Crohn's disease and are the major source of mucosal tumor necrosis factor α (TNFα), a principal mediator of intestinal injury. There is a unique population of TNFα + IL-17A + CD4 + T_RM cells in Crohn's disease which are largely absent in controls (43). In the context of Crohn's disease, recent developments suggest that there is an impaired monocyte function initiating the disease and an overactivation of monocytes and adaptive immunity maintaining the disease (44). Monocytes and monocyte-derived macrophages have been found to be crucial players in the chronic inflammation seen in Crohn's disease patients (45).

We performed an intersection analysis involving DEGs, key modular genes identified through WGCNA, and PRGs. This resulted in the identification of 44 potential diagnostic markers. Subsequently, we employed five machine learning algorithms to further screen for the top five diagnostic biomarkers associated with PCD: PIK3R3, STAT1, PLEK, GNA15, and SERPINE1, which exhibited the strongest diagnostic capabilities. PIK3R3 serves as the regulatory subunit of phosphatidylinositol kinase, a crucial signaling pathway implicated in cell growth, proliferation, and immune response. This kinase plays a pivotal role in various essential mediators of cellular processes, including inflammation and immune regulation. In a study conducted by Wang et al., they observed that the overexpression of PIK3R3, primarily dependent on SNAI2, triggers a significant intestinal epithelial-mesenchymal transition (EMT). Furthermore, the downregulation of PIK3R3 reversed this process, potentially leading to increased invasion and metastasis of colorectal cancer cells (46). Signal transduction and activator of transcription 1 (STAT1) plays a multifaceted role in the pathogenesis of Crohn's disease, an IBD. Research indicates that STAT1 is involved in the epigenetic regulation of two key genes, Lymphocyte Cytosolic Protein 2 (LCP2) and TNF-α-inducible protein 2 (TNFAIP2), by recruiting EP300, thereby contributing to the development of IBD. Specifically, phosphorylated STAT1

10.3389/fimmu.2024.1323418

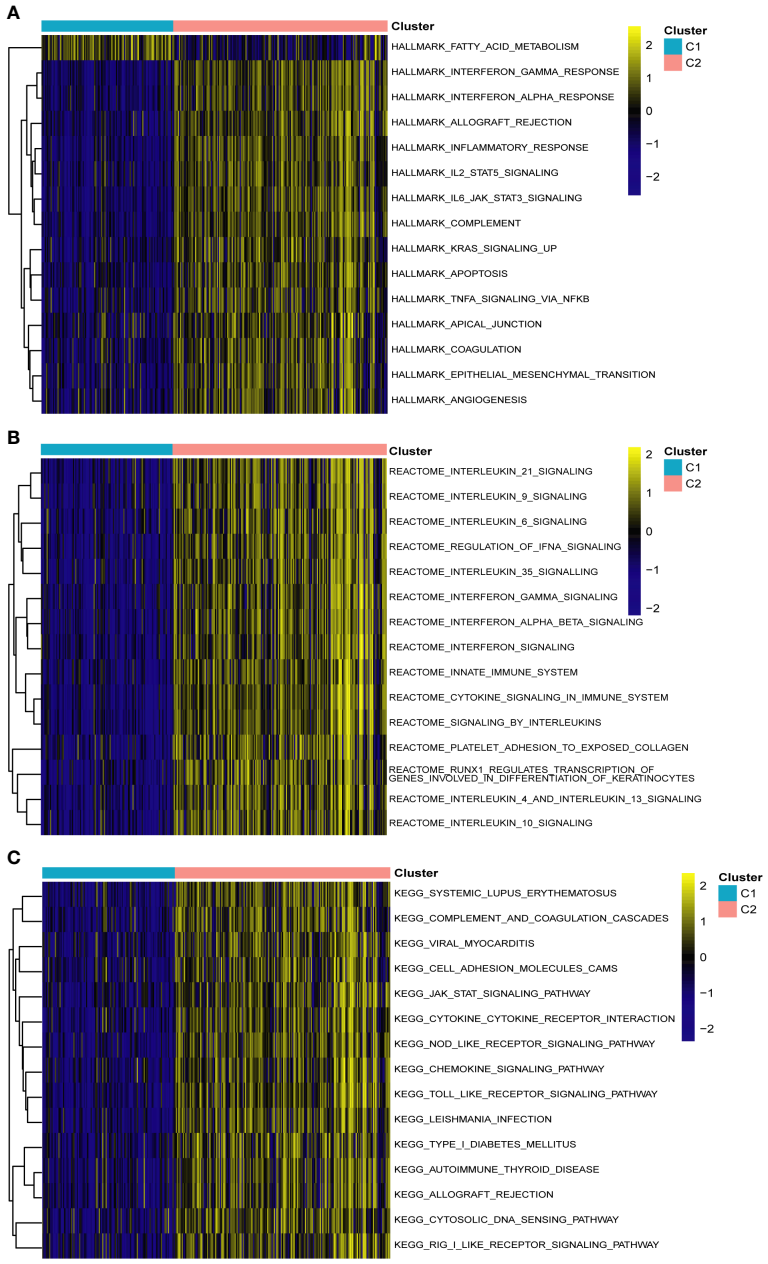


FIGURE 11
Gene set variation analysis between two different subtypes of PCD. **(A)** Enriched pathways based on the hallmark pathway. **(B)** Enriched pathways based on the reactome pathway. **(C)** Enriched pathways based on the KEGG pathway.

(p-STAT1) binds to the enhancer loci of these two genes, where H3K27ac enrichment is notable, leading to subsequent EP300 binding and the regulation of gene expression. This process promotes the expression of TNFAIP2 and LCP2 by augmenting the H3K27ac enrichment on their enhancers, ultimately contributing to the pathogenesis of chronic inflammation (47). Furthermore, a study employing an acute colitis model revealed significant improvements in disease state among STAT1-deficient mice in comparison to wild-type mice. Notably, the induction of highly expressed Ly6c cells in colorectal tissues was notably reduced in STAT1-deficient mice (48). Serpin Peptidase Inhibitor 1 (SERPINE1) has been identified as a potential new disease activity marker in IBD, which includes Crohn's disease. In a study, it was found that the expression of SERPINE1 differed significantly in healthy subjects compared to IBD patients with active disease. After therapy induction, a remarkable decrease was observed in the mucosal SERPINE1 concentration in responders. Moreover, the study found that serum SERPINE1 correlates with disease activity ($p<0,01$, cut-off value: 22 mg/ml, sensitivity = 80%, specificity = 60%, accuracy = 74%), whereas no correlation was observed between the mucosal SerpineE1 concentration and the disease activity ($p > 0.1$, sensitivity = 72%, specificity = 77.8%, accuracy = 73.5%). These findings suggest that SERPINE1 could potentially be used as a marker to monitor disease activity and therapeutic

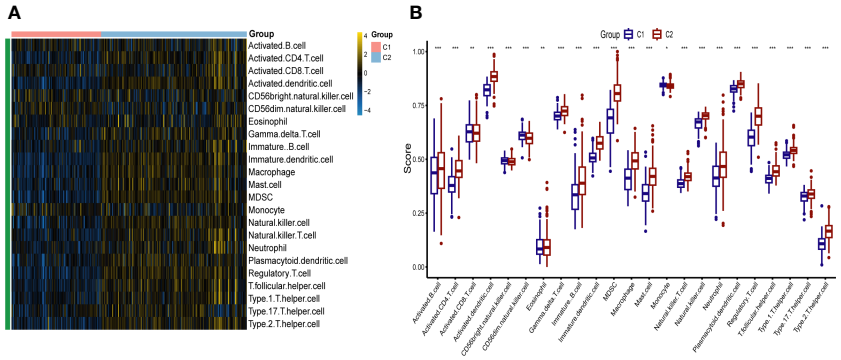


FIGURE 12
The immune landscape of two different platelet subtypes of PCD. Heatmap **(A)** and boxplot **(B)** demonstrating the differences in immune cell infiltration between different subtypes. *p < 0.05, **p < 0.01, ***p < 0.001.

response in IBD, including Crohn's disease (49). PLEK, also known as Pleckstrin, encodes a protein involved in intracellular signaling and the regulation of cell growth, proliferation, and differentiation. It plays a role in diverse cellular processes, including platelet activation and immune response regulation (50). The study by Chen and Medrano et al. reported that aberrant expression of PLEK may be involved in the pathogenesis of IBD by increasing inflammatory factors, which is consistent with our findings (51, 52). GNA15 is a gene that encodes a subunit of a guanine nucleotide-binding protein (G protein). The G proteins, including GNA15, are known to be involved in signal transduction pathways that regulate immune responses. Dysregulation of G protein signaling could potentially contribute to abnormal immune responses and inflammation, which are key features of Crohn's disease (53). However, direct evidence linking GNA 15 to Crohn's disease or IBD is lacking in the current literature.

In clinical practice, accurate diagnosis of PCD is often accompanied by the performance of invasive procedures, which causes great pain to the affected children. In addition, due to the heterogeneity of PCD, the same therapeutic regimen is not effective for all PCD patients. Therefore, the creation of predictive modeling tools at the genetic level to look for potential molecular subtypes of PCD, while imposing some relatively large burdens in terms of testing costs, nevertheless, the new predictive modeling and molecular typing will not only be effective in reducing the suffering of the children, but also provide for improved diagnostic accuracy, as well as risk stratification and grading of patients with PCD. In this study, we developed a novel predictive model based on five biomarkers for predicting the risk of developing PCD. In subsequent internal and external validations, it demonstrated excellent predictive accuracy and good clinical benefit rates. This suggests that such work will enable accurate assessment and management of patients when encountered in clinical practice. In addition, we identified two platelet molecular subtypes of PCD based on these five markers. In a subsequent pathway enrichment analysis, we found that the marker activity of the C2 cluster was significantly higher than that of the C1 cluster in all pathways except the fatty acid metabolism pathway.In addition, immune infiltration analysis revealed significant differences in the degree of infiltration of both subtypes in 23 immune cells. This provides a new theoretical basis for risk stratification and personalized diagnosis and treatment of PCD. However, this study still has some limitations. First, selection and detection bias could not be completely avoided because the data used in this study were all from the same database. Second, despite the completion of a large number of secondary analysis of previous data, large-scale experiments at the clinical, cellular, and molecular levels, as well as prospective studies, are still needed to further validate our conclusions.

## Conclusions

In summary, this study unveils, for the first time, a potential association between PCD and PRGs through the integration of WGCNA with machine learning algorithms. Furthermore, we identified five novel diagnostic markers: PIK3R3, STAT1, PLEK, GNA15, and SERPINE1. The nomogram constructed using these five markers present the prospect of a non-invasive approach to diagnosing PCD. Additionally, we have, for the first time, classified PCD into two distinct molecular subtypes. This breakthrough may offer fresh perspectives for exploring the heterogeneity of clinical presentations and prognosis of PCD. Moreover, it lays a theoretical foundation for early prevention, risk stratification, and personalized diagnosis and treatment of UC in the future.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/. The accession numbers can be found in the article.

## Author contributions

DT: Formal Analysis, Methodology, Visualization, Writing – original draft. YH: Data curation, Formal analysis, Writing – original draft. YC: Data curation, Methodology, Visualization, Writing – original draft. CY: Data curation, Formal analysis, Visualization, Writing – original draft. BP: Data curation,

Methodology, Validation, Visualization, Writing – original draft. SL: Validation, Writing – review & editing. HL: Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

We acknowledge the contribution of Wei Li in statistics advice and proofreading.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1323418/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Error rate plot for cisualizing Random Forest algorithm results.

SUPPLEMENTARY FIGURE 2
Consensus Cluster anaylsis. **(A)** Consensus CDF delta area curves when k = 2-9. **(B)** Relative alterations in the area under CDF curve. **(C)** Tracking plot showing the sample classification when k = 2–9. CDF, cumulative distribution function.

SUPPLEMENTARY TABLE 1
Differentially expressed genes matrix.

SUPPLEMENTARY TABLE 2
Module cutting results by Weighted gene co-expression network analysis.

SUPPLEMENTARY TABLE 3
Results of Gene Ontology and Kyoto Encyclopedia of Genes and Genome analysis.

SUPPLEMENTARY TABLE 4
Results of consensus cluster anaylsis.

## References

1. Roda G, Chien Ng S, Kotze PG, Argollo M, Panaccione R, Spinelli A, et al. Crohn's disease. *Nat Rev Dis primers* (2020) 6(1):22. doi: 10.1038/s41572-020-0156-2

2. Rogler G, Singh A, Kavanaugh A, Rubin DT. Extraintestinal manifestations of inflammatory bowel disease: current concepts, treatment, and implications for disease management. *Gastroenterology* (2021) 161(4):1118–32. doi: 10.1053/j.gastro.2021.07.042

3. Ricciuto A, Aardoom M, Orlanski-Meyer E, Navon D, Carman N, Aloi M, et al. Predicting outcomes in pediatric Crohn's disease for management optimization: systematic review and consensus statements from the pediatric inflammatory bowel disease-ahead program. *Gastroenterology* (2021) 160(1):403–36.e26. doi: 10.1053/j.gastro.2020.07.065

4. Conrad MA, Rosh JR. Pediatric inflammatory bowel disease. *Pediatr Clinics North America* (2017) 64(3):577–91. doi: 10.1016/j.pcl.2017.01.005

5. Rosen MJ, Dhawan A, Saeed SA. Inflammatory bowel disease in children and adolescents. *JAMA pediatrics* (2015) 169(11):1053–60. doi: 10.1001/jamapediatrics.2015.1982

6. Thia KT, Sandborn WJ, Harmsen WS, Zinsmeister AR, Loftus EV Jr. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. *Gastroenterology* (2010) 139(4):1147–55. doi: 10.1053/j.gastro.2010.06.070

7. Ajbar A, Cross E, Matoi S, Hay CA, Baines LM, Saunders B, et al. Diagnostic delay in pediatric inflammatory bowel disease: A systematic review. *Digestive Dis Sci* (2022) 67(12):5444–54. doi: 10.1007/s10620-022-07452-5

8. Zhang H, Zeng Z, Mukherjee A, Shen B. Molecular diagnosis and classification of inflammatory bowel disease. *Expert Rev Mol diagnostics* (2018) 18(10):867–86. doi: 10.1080/14737159.2018.1516549

9. Li J, Cui Y, Jin X, Ruan H, He D, Che X, et al. Significance of pyroptosis-related gene in the diagnosis and classification of rheumatoid arthritis. *Front endocrinol* (2023) 14:1144250. doi: 10.3389/fendo.2023.1144250

10. Hu X, Ni S, Zhao K, Qian J, Duan Y. Bioinformatics-led discovery of osteoarthritis biomarkers and inflammatory infiltrates. *Front Immunol* (2022) 13:871008. doi: 10.3389/fimmu.2022.871008

11. Li X, Zhao K, Lu Y, Wang J, Yao W. Genetic analysis of platelet-related genes in hepatocellular carcinoma reveals a novel prognostic signature and determines PRKCD as the potential molecular bridge. *Biol procedures online* (2022) 24(1):22. doi: 10.1186/s12575-022-00185-9

12. Machlus KR, Thon JN, Italiano JEJr. Interpreting the developmental dance of the megakaryocyte: a review of the cellular and molecular processes mediating platelet formation. *Br J haematol* (2014) 165(2):227–36. doi: 10.1111/bjh.12758

13. Lesurtel M, Graf R, Aleil B, Walther DJ, Tian Y, Jochum W, et al. Platelet-derived serotonin mediates liver regeneration. *Sci (New York NY)* (2006) 312(5770):104–7. doi: 10.1126/science.1123842

14. Liu Y, Jiang H, Kang T, Shi X, Liu X, Li C, et al. Platelets-related signature based diagnostic model in rheumatoid arthritis using WGCNA and machine learning. *Front Immunol* (2023) 14:1204652. doi: 10.3389/fimmu.2023.1204652

15. El Bannoudi H, Cornwell M, Luttrell-Williams E, Engel A, Rolling C, Barrett TJ, et al. Platelet LGALS3BP as a mediator of myeloid inflammation in systemic lupus erythematosus. *Arthritis Rheumatol (Hoboken NJ)* (2023) 75(5):711–22. doi: 10.1002/art.42382

16. Pelia R, Venkateswaran S, Matthews JD, Haberman Y, Cutler DJ, Hyams JS, et al. Profiling non-coding RNA levels with clinical classifiers in pediatric Crohn's disease. *BMC Med Genomics* (2021) 14(1):194. doi: 10.1186/s12920-021-01041-7

17. Marigorta UM, Denson LA, Hyams JS, Mondal K, Prince J, Walters TD, et al. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet* (2017) 49(10):1517–21. doi: 10.1038/ng.3936

18. Haberman Y, Schirmer M, Dexheimer PJ, Karns R, Braun T, Kim MO, et al. Age-of-diagnosis dependent ileal immune intensification and reduced alpha-defensin in older versus younger pediatric Crohn Disease patients despite already established dysbiosis. *Mucosal Immunol* (2019) 12(2):491–502. doi: 10.1038/s41385-018-0114-4

19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007

20. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinf (Oxford England)* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034

21. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

23. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J Integr Biol* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

24. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software* (2010) 33(1):1–22.

25. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinf (Oxford England)* (2014) 30(19):2811–2. doi: 10.1093/bioinformatics/btu393

26. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinf (Oxford England)* (2010) 26 (12):1572–3. doi: 10.1093/bioinformatics/btq170

27. Balestrieri P, Ribolsi M, Guarino MPL, Emerenziani S, Altomare A, Cicala M. Nutritional aspects in inflammatory bowel diseases. *Nutrients* (2020) 12(2):372.

28. Mackner LM, Crandall WV. Brief report: psychosocial adjustment in adolescents with inflammatory bowel disease. *J Pediatr Psychol* (2005) 31(3):281–5. doi: 10.1093/jpepsy/jsj023

29. Ouahed J, Spencer E, Kotlarz D, Shouval DS, Kowalik M, Peng K, et al. Very early onset inflammatory bowel disease: A clinical approach with a focus on the role of genetics and underlying immune deficiencies. *Inflammatory Bowel Diseases* (2019) 26 (6):820–42. doi: 10.1093/ibd/izz259

30. Annese V, Beaugerie L, Egan L, Biancone L, Bolling C, Brandts C, et al. European evidence-based consensus: inflammatory bowel disease and Malignancies. *J Crohn's colitis* (2015) 9(11):945–65. doi: 10.1093/ecco-jcc/jjv141

31. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nat Rev Immunol* (2008) 8(6):458–66. doi: 10.1038/nri2340

32. Friedrich M, Pohin M, Powrie F. Cytokine networks in the pathophysiology of inflammatory bowel disease. *Immunity* (2019) 50(4):992–1006. doi: 10.1016/j.immuni.2019.03.017

33. Wang L, Hu Y, Song B, Xiong Y, Wang J, Chen D. Targeting JAK/STAT signaling pathways in treatment of inflammatory bowel disease. *Inflammation Res* (2021) 70(7):753–64. doi: 10.1007/s00011-021-01482-x

34. Danese S, Furfaro F, Vetrano S. Targeting S1P in inflammatory bowel disease: new avenues for modulating intestinal leukocyte migration. *J Crohn's colitis* (2018) 12 (suppl_2):S678–s86. doi: 10.1093/ecco-jcc/jjx107

35. Wyatt NJ, Speight RA, Stewart CJ, Kirby JA, Lamb CA. Targeting leukocyte trafficking in inflammatory bowel disease. *BioDrugs* (2021) 35(5):473–503. doi: 10.1007/s40259-021-00496-5

36. Mezynski MJ, Farrelly AM, Cremona M, Carr A, Morgan C, Workman J, et al. Targeting the PI3K and MAPK pathways to improve response to HER2-targeted therapies in HER2-positive gastric cancer. *J Trans Med* (2021) 19(1):184. doi: 10.1186/s12967-021-02842-1

37. Han W, Wu Y, Wang X, Liu L, Ding Y. Dissecting the critical pathway crosstalk mechanisms of thyroid cancer based on drug-target genes and disease genes. *Biologia* (2021) 76(11):3489–99. doi: 10.1007/s11756-021-00869-9

38. Zegeye MM, Lindkvist M, Fälker K, Kumawat AK, Paramel G, Grenegård M, et al. Activation of the JAK/STAT3 and PI3K/AKT pathways are crucial for IL-6 trans-signaling-mediated pro-inflammatory response in human vascular endothelial cells. *Cell communication Signaling CCS* (2018) 16(1):55. doi: 10.1186/s12964-018-0268-4

39. Stefani C, Miricescu D, Stanescu SII, Nica RI, Greabu M, Totan AR, et al. Growth factors, PI3K/AKT/mTOR and MAPK signaling pathways in colorectal cancer pathogenesis: where are we now? *Int J Mol Sci* (2021) 22(19):10260. doi: 10.3390/ijms221910260

40. Timmermans WM, van Laar JA, van der Houwen TB, Kamphuis LS, Bartol SJ, Lam KH, et al. B-cell dysregulation in Crohn's disease is partially restored with infliximab therapy. *PLoS One* (2016) 11(7):e0160103. doi: 10.1371/journal.pone.0160103

41. Shen X, Mo S, Zeng X, Wang Y, Lin L, Weng M, et al. Identification of antigen-presentation related B cells as a key player in Crohn's disease using single-cell dissecting, hdWGCNA, and deep learning. *Clin Exp Med* (2023) 23(8):5255–67. doi: 10.1007/s10238-023-01145-7

42. Schmitt H, Neurath MF, Atreya R. Role of the IL23/IL17 pathway in Crohn's disease. *Front Immunol* (2021) 12:622934. doi: 10.3389/fimmu.2021.622934

43. Bishu S, El Zaatari M, Hayashi A, Hou G, Bowers N, Kinnucan J, et al. CD4+ Tissue-resident memory T cells expand and are a major source of mucosal tumour necrosis factor α in active Crohn's disease. *J Crohn's colitis* (2019) 13(7):905–15. doi: 10.1093/ecco-jcc/jjz010

44. Zhou L, Braat H, Faber KN, Dijkstra G, Peppelenbosch MP. Monocytes and their pathophysiological role in Crohn's disease. *Cell Mol Life Sci* (2009) 66(2):192–202. doi: 10.1007/s00018-008-8308-7

45. Nieto JC, Zamora C, Cantó E, Garcia-Planella E, Gordillo J, Ortiz MA, et al. CSF-1 regulates the function of monocytes in Crohn's disease patients in remission. *Sci Rep* (2017) 7(1):92. doi: 10.1038/s41598-017-00145-4

46. Wang G, Yang X, Li C, Cao X, Luo X, Hu J. PIK3R3 induces epithelial-to-mesenchymal transition and promotes metastasis in colorectal cancer. *Mol Cancer Ther* (2014) 13(7):1837–47. doi: 10.1158/1535-7163.Mct-14-0049

47. Yu YL, Chen M, Zhu H, Zhuo MX, Chen P, Mao YJ, et al. STAT1 epigenetically regulates LCP2 and TNFAIP2 by recruiting EP300 to contribute to the pathogenesis of inflammatory bowel disease. *Clin epigenetics* (2021) 13(1):127. doi: 10.1186/s13148-021-01101-w

48. Kii S, Kitamura H, Hashimoto S, Ikeo K, Ichikawa N, Yoshida T, et al. STAT1-mediated induction of Ly6c-expressing macrophages are involved in the pathogenesis of an acute colitis model. *Inflammation Res* (2022) 71(9):1079–94. doi: 10.1007/s00011-022-01620-z

49. Jójárt B, Molnár T, Szabó V, Varga Á, Resál T, Szántó K, et al. P459 Expression of SerpinE1, a potential new disease activity marker, reflects therapeutic response in Inflammatory Bowel Disease. *J Crohn's Colitis* (2021) 15(Supplement_1):S453–S. doi: 10.1093/ecco-jcc/jjab076.812

50. Wang G, Zhou Q, Xu Y, Zhao B. Emerging roles of pleckstrin-2 beyond cell spreading. *Front Cell Dev Biol* (2021) 9:768238. doi: 10.3389/fcell.2021.768238

51. Medrano LM, Pascual V, Bodas A, López-Palacios N, Salazar I, Espino-Paisán L, et al. Expression patterns common and unique to ulcerative colitis and celiac disease. *Ann Hum Genet* (2019) 83(2):86–94. doi: 10.1111/ahg.12293

52. Chen Y, Li H, Lai L, Feng Q, Shen J. Identification of common differentially expressed genes and potential therapeutic targets in ulcerative colitis and rheumatoid arthritis. *Front Genet* (2020) 11:572194. doi: 10.3389/fgene.2020.572194

53. Jansen P, Müller H, Lodde GC, Zaremba A, Möller I, Sucker A, et al. GNA14, GNA11, and GNAQ mutations are frequent in benign but not Malignant cutaneous vascular tumors. *Front Genet* (2021) 12:663272. doi: 10.3389/fgene.2021.663272

frontiers | Frontiers in Pediatrics

TYPE Technology and Code
PUBLISHED 18 December 2024
DOI 10.3389/fped.2024.1465632

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Yuhanis Yusof,
Universiti Utara Malaysia, Malaysia
Megan K O'Brien,
Shirley Ryan AbilityLab, United States

*CORRESPONDENCE
Antoine Giraud
✉ antoine.giraud@univ-st-etienne.fr

RECEIVED 16 July 2024
ACCEPTED 03 December 2024
PUBLISHED 18 December 2024

CITATION
Soualmi A, Alata O, Ducottet C, Petitjean-Robert A, Plat A, Patural H and Giraud A (2024) AGMA-PESS: a deep learning-based infant pose estimator and sequence selector software for general movement assessment. Front. Pediatr. 12:1465632. doi: 10.3389/fped.2024.1465632

COPYRIGHT
© 2024 Soualmi, Alata, Ducottet, Petitjean-Robert, Plat, Patural and Giraud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# AGMA-PESS: a deep learning-based infant pose estimator and sequence selector software for general movement assessment

Ameur Soualmi[1,2], Olivier Alata[1], Christophe Ducottet[1], Anne Petitjean-Robert[3], Aurélie Plat[4], Hugues Patural[2,4] and Antoine Giraud[2,4]*

[1]Laboratoire Hubert Curien UMR 5516, CNRS, Institut d'Optique Graduate School Université Jean Monnet Saint-Etienne, Saint-Etienne, France, [2]INSERM, U1059 SAINBIOSE, Université Jean Monnet, Saint-Étienne, France, [3]Néonatalogie, Centre Hospitalier Universitaire de Saint-Étienne, Saint-Étienne, France, [4]Service de Réanimation Néonatale, Centre Hospitalier Universitaire de Saint-Étienne, Saint-Étienne, France

The General Movement Assessment (GMA) is a validated evaluation of brain maturation essential to shaping early individual developmental trajectories of preterm infants. To ensure a reliable GMA, preterm infants should be recorded for 30 to 60 min before manually selecting at least three sequences with general movements. This time-consuming task of manually selecting short video sequences from lengthy recordings impedes its implementation within the Neonatal Unit. Moreover, an accurate pose estimation tool for preterm infants is paramount to developing the field of GMA automation. We introduce the AGMA Pose Estimator and Sequence Selector (AGMA-PESS) software, based on the state-of-the-art deep learning infant pose estimation network, to automatically select the video sequences for GMA at preterm and writhing ages and estimate the pose of infants in 2D. Its simplicity and efficiency make AGMA-PESS a valuable tool to promote GMA use within the Neonatal Unit, both for clinical practice and research purposes.

KEYWORDS

**preterm infant pose estimation, automatic sequence selection, general movements, windows software, infant spontaneous movements**

## 1 Introduction

The general movements represent the spontaneous motor activity occurring from nine weeks of gestational age (GA) to the apparition of goal-directed movements around four months of corrected age (1). The General Movement Assessment (GMA) is a validated evaluation based on the qualitative analysis of general movements' complexity, variability, and fluidity (1, 2). The GMA is a functional evaluation of brain maturation that reflects the integrity of extensive cortical-subcortical networks (2). The GMA performed within the Neonatal Unit is essential to shape early individual developmental trajectories (3) to initiate early developmental intervention (4).

To ensure a reliable GMA, preterm infants should be recorded for 30 to 60 min according to Prechtl's method (1) lying in a supine position in the incubator, bed, or on the floor. If the assessment is done without acoustic signals, the assessor must be able to see the infant's face to make sure that rigid movements are not due to crying. The room temperature should be comfortable, and in case of prolonged episodes of

fussing, crying, or distraction, the recording must be stopped. Then, each video must be reviewed to manually select at least three sequences of general movements (1). The manual selection of these sequences is time-consuming and impedes the implementation of the GMA for preterm children within the Neonatal Unit. Over the past decade, multiple studies have focused on automating the assessment of general movements. This task mainly depends on accurately detecting the infant pose through the manually selected video sequences for further analysis. Openpose (5) deep learning model and its retrained version from Chambers et al. (6) were widely used in this context by numerous studies (7–11). However, Openpose was trained on datasets containing only images of adults (5), and the retrained version was generated using 9,039 infant images only, mainly collected from the internet (6), which resulted in poor performance when applied to infants' poses in a clinical environment (12).

Recently, a new architecture was retrained on a dataset of over 88k images of infants in a clinical environment, achieving state-of-the-art results with a PCK@0.2 of 98.30% (12). Based on these results, and to address the raised problems, we developed the AGMA Pose Estimator and Sequence Selector (AGMA-PESS) software, a deep-learning-based software that automatically estimates infants' pose and automatically selects video sequences for the GMA of preterm and term infants.

# 2 Method

AGMA-PESS software was developed using Python 3.7 programming language. The graphical user interface (GUI) was created based on the Pyqt5 library. The source code can be found in this Github repository[1] and the installation .exe file can be downloaded using this link.[2] The AGMA-PESS is compatible with the Windows operating system (Microsoft, WA) with version $\geq$7. It is available in two languages (English and French). The software proposes two main functionalities to the user (see Figure 1A): a pose estimator and a sequence selector.

## 2.1 Pose estimator

Infant pose estimation is the automatic identification of key points on an infant's body, such as joints and limbs, to understand their posture and actions (see Figure 1B). The process typically consists of using deep learning models trained on annotated datasets of infant images or videos. The neural network used in this task is the retrained version of Darkpose32 (12). This network was primarily pretrained on adult images (57 K images and 150 K person instances), then fine-tuned with a dataset containing over 88 k images of 53 preterm infants born before 33 weeks of gestational age (12). The neural network was built as a concatenation of the HRNet network (13) in addition to Darkpose (14), which was used as a plugin that improves the performance of SOTA Human pose estimation models using a new heatmap decoding process. After fine-tuning it, the network achieved a PCK@0.2 of 98.30 on 36,000 infants' test images in the 2D pose estimation task and a mean error of 1.72 cm in the 3D pose infant estimation task. To reconstruct the infant's 3D pose, the network estimates the 2D pose from each view of the stereoscopic images; then, a triangulation is performed to get the real-world 3D coordinates. However, since standard 2D-image acquisition systems are more popular for GMA assessment rather than stereoscopic systems, AGMA-PESS software processes 2D videos only, which does not prevent using it on each view of a stereoscopic setup if needed.

For processing efficiency considerations, AGMA-PESS offers the ability for the user to choose the number of images to process per second instead of all the video images (e.g., 30 FPS). The pose estimator window is a user-friendly interface (see Figure 1B), where one or multiple videos of any duration from the following formats (webm, .mkv, .vob, .avi, .mts, .mov, .wmv, .mp4, .mpg, .3gp, and .flv) can be loaded using only a push-button. The user will be asked to input the pose estimation frame rate as an integer value and to choose whether to save the processed images or not. The image saving format is .jpeg which offers a good balance between file size and quality. Finally, the start button will start the processing, and the user can stop it at any moment by clicking the stop button. At the end of processing, Excel files are generated automatically, where tables of the 17 keypoints coordinates and their respective confidence scores are saved and easily accessible from the "Destination folder" push button.

## 2.2 Sequence selector

The sequence selector automatically selects short video sequences where the infants exhibit the greatest amount of movement. First, the AGMA network (12) is used to estimate the infant pose (17 keypoints) $P = [p_1, p_2, \ldots, p_{17}]$ through all the video ($N$ seconds) at one frame per second rate. Next, the sum of the Euclidean distances between respective keypoints is calculated for every two consecutive poses $P_i$ and $P_{i+1}$, as in the following Equation 1.

$$S_i = \sum_{k=1}^{17} \|P_{i+1,k} - P_{i,k}\| \quad \text{with} \quad i = 1, 2, \ldots, N \tag{1}$$

$S_i$ is a time series that represents the quantity of motion along the video, and the task of finding the starting index $x$ of the most important one-minute sequence of movements from a video can be described mathematically as finding the most

[1]https://github.com/AmeurSoualmi/AGMA-Pose-Estimator-and-Sequence-Selector.

[2]https://drive.google.com/file/d/1y_ReeJUFWt6_V1KUSRPFYD27L-47NB5C/view?usp=sharing.

FIGURE 1
AGMA-PESS user interface. (A) Initialization window. (B) AGMA Pose Estimator window.

important sub-series in $S_i$ (see Figure 2), which can be accomplished by a summing window function $G$ of size 60 s as in Equation 2

$$x = \underset{j\in 0,\ldots,N-60}{\operatorname{argmax}} G(j) = \underset{j}{\operatorname{argmax}} \left( \sum_{i=j}^{j+60} S_i \right) \tag{2}$$

The sequence selector window has more settings than the pose estimation window. The user is asked to input the desired integer number of video sequences and their duration in minutes as a float number. When the start button is pressed, a real-time graph shows the changes in motion quantity as in Equation 1, and a progress bar shows the analysis progression. At the end of processing, the generated video sequences are accessible by clicking the "Destination folder" button.

A video tutorial is available as Supplementary Material to facilitate the use of the software.

# 3 Experiments

In order to validate the effectiveness of the proposed software, comparative experiments with manual sequence selections were conducted on a dataset of videos collected in a clinical environment.

## 3.1 Dataset

The dataset contains six video recordings of one-hour duration collected in the Neonatology department of the Centre Hospitalier Universitaire de Saint-Étienne, France. Six premature infants born before 33 weeks of gestational age (GA) were included with written parental consent. The video recording protocol was controlled as recommended in Prechtl's method of GMA (1). Infants wearing diapers were placed in a radiant heat warmer, lying in a supine position. The room temperature was comfortable, and the recording was stopped in case of prolonged episodes of fussing, crying, or distraction. Videos were recorded for one hour using the left view of ZED2 stereoscopic camera with frames of 1,280 × 720 resolution at 30 FPS. The features of the recorded videos are summarized in Table 1 and some images extracted from the videos are provided in Figure 3.

## 3.2 Manual sequence selection

An expert group composed of $N_e = 4$ experienced General Movements Trust-certified assessors (AS, APR, AP, and AG) reviewed all the video recordings. They were asked to independently select the best five video sequences by manually indicating the starting point of each sequence in seconds. The time each expert took to complete the task for each video was recorded. The results were then compared with AGMA-PESS to check the detection accuracy and validate the method.

## 3.3 Evaluation

In order to evaluate how accurately the AGMA-PESS software selects the $N_s = 5$ GMA one-minute sequences from each of the $N_v = 6$ videos, compared to an expert selector, the maximum dice similarity (MDS) parameter was used. Let $x_{l,i}$ with $l \in \{1, \ldots, N_v\}$ and $i \in \{1, \ldots, N_s\}$ be a one-minute sequence detected by the software with a time interval of $(x_{l,i,\text{start}}, x_{l,i,\text{end}})$.

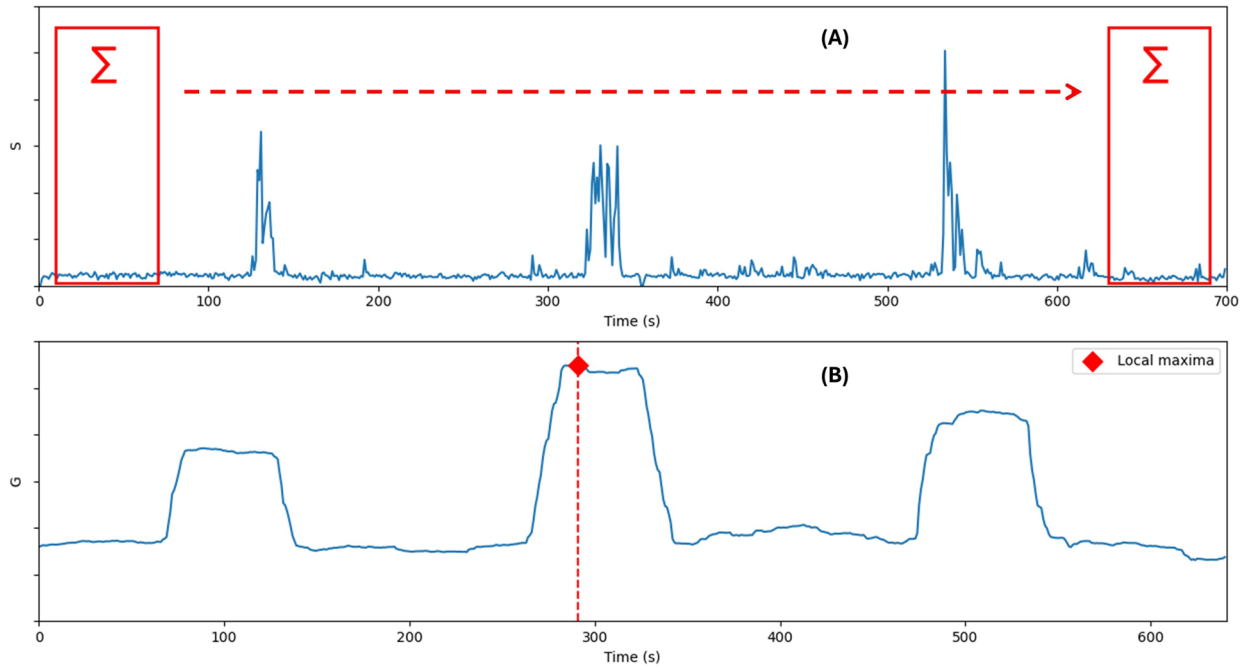


FIGURE 2
(A) The plot of the quantity of motion *S*. (B) The plot of the summing window function *G*.

TABLE 1 Dataset population features.

| Infant | Birth GA (Weeks) | Term at recording (Weeks) | Video duration | Movement types | GMA |
|---|---|---|---|---|---|
| 1 | 31 | 38 | 1 h 07 m 19 s | Writhing | Normal |
| 2 | 26 | 34 | 1 h 00 m 07 s | Preterm | Abnormal |
| 3 | 24 | 34 | 1 h 07 m 39 s | Preterm | Abnormal |
| 4 | 31 | 42 | 1 h 01 m 11 s | Writhing | Normal |
| 5 | 30 | 39 | 1 h 07 m 51 s | Writhing | Normal |
| 6 | 25 | 35 | 1 h 00 m 04 s | Preterm | Normal |

Abbreviations: GA, gestational age; GMA, General Movement Assessment.

FIGURE 3
Dataset population images.

The intersection $I$ of this sequence with another time sequence $y(y_{\text{start}}, y_{\text{end}})$ is defined as in Equation 3.

$$I(x_{l,i}, y) = \max(0,\ \min(x_{l,i,\text{end}}, y_{\text{end}}) - \max(x_{l,i,\text{start}}, y_{\text{start}})) \quad (3)$$

Since the selected intervals are different from one expert to another, for each sequence selected by the software $x_{l,i}$, the corresponding sequence $y_{c,e}$, $c \in \{1, \ldots, N_s\}$, from the five sequences $y_{j,e}$ with $j \in \{1, \ldots, N_s\}$ selected by an expert $e$ with $e \in \{1, \ldots, N_e\}$ is found as in Equation 4:

$$c = \underset{j \in 1,\ldots,N_s}{\operatorname{argmax}} I(x_{l,i}, y_{j,e}) \quad (4)$$

In the case of equality, the minimum index is chosen. Using Equation 3, the intersection of the resulting sequence $y_{c,e}$ with a time interval of $(y_{c,e,\text{start}}, y_{c,e,\text{end}})$ with the sequence $x_{l,i}(x_{l,i,\text{start}}, x_{l,i,\text{end}})$ can be modeled in Equation 5:

$$I(x_{l,i}, y_{c,e}) = \max(0,\ \min(x_{l,i,\text{end}}, y_{c,e,\text{end}}) - \max(x_{l,i,\text{start}}, y_{c,e,\text{start}})) \quad (5)$$

Their dice similarity (DS) is defined as in Equation 6:

$$\text{DS}(x_{l,i}, y_{c,e}) = \frac{2I(x_{l,i}, y_{c,e})}{(x_{l,i,\text{end}} - x_{l,i,\text{start}}) + (y_{c,e,\text{end}} - y_{c,e,\text{start}})} \quad (6)$$

The DS describes the overlap between the sequences selected by the software and each expert. Therefore, checking whether a sequence detected by the software is also selected by one expert at least can be achieved by considering the MDS in Equation 7:

$$\text{MDS}(x_{l,i}) = \max_{e \in \{1,\ldots,N_e\}}(\text{DS}(x_{l,i}, y_{c,e})) \quad (7)$$

The values of MDS vary between 0 and 1. A null value means there is no overlap between the sequence selected by the software and the four sequences selected by the experts. On the other hand, a 1 value means that the selected sequence is also precisely selected at least by one expert. Hence, comparing the MDS to a defined threshold $\tau$ can lead to a binary classification problem, with a software-experts agreement metric defined in Equation 8:

$$\text{Precision}(\tau) = \frac{\sum_{l=1}^{N_v} \sum_{i=1}^{N_s} \delta(\text{MDS}(x_{l,i}) > \tau)}{N_v \times N_s} \quad (8)$$

where $N_v \times N_s$ is the total number of the selected sequences in all the videos.

Based on this metric, an inter-rater precision can also be calculated for each expert compared to the remaining three. This allows to verify how consistently experts make their selections, and the average inter-experts score can be used to test how well the software-experts agreement is positioned compared to the inter-experts agreement.

# 4 Results and discussion

The comparison results between the human experts and the AGMA-PESS automatic selection of five one-minute video sequences from six videos (30 sequences in total) is illustrated in Figure 4. In 28 out of 30 cases, the sequence detected by the AGMA-PESS software intersected with a sequence selected by at least one expert, demonstrating the software's performance relevance. Additionally, the software-expert agreement on selected sequences tended to be higher for videos where the infant was less active, such as in Video 3, which is represented by dispersed peaks in motion quantities (see Figure 4C). Conversely, in videos where the infant was very active, like in video 4, there was more variability in the selections made by the experts, creating a contrast. However, this variability is not a significant issue since all sequences selected by the experts were reliable for GMA, regardless of their location. This justifies the choice of using the maximum dice similarity parameter rather than an average of the four software-expert dice similarities, indicating that a sequence is well detected by the software if at least one expert also selects it.

The sequence selection precision of the AGMA-PESS software is detailed in Figure 5. In the first plot (A), and based on Equation 8, the precision of the AGMA-PESS sequence selector was calculated compared to three experts each time, resulting in four comparable graphs, which shows the relevance of the software performance. In the second plot (B), a high inter-experts agreement can be observed with varying precision for each expert. In the third plot (C), the previous graphs were averaged and compared to inspect how the software-experts agreement was positioned compared to the inter-experts agreement. With a threshold $\tau = 0.5$, which means considering a sequence as detected if it intersects with more than half with another expert-selected sequence, a precision of 86% was achieved, which is very promising since the task of manually selecting the GMA sequences can be very time-consuming (see Table 2). The average time to manually select 5 video sequences was more than 23 min, in addition to the time that will be needed to cut them for later examination. This can be automatically achieved with the AGMA-PESS software with high precision.

Two sequences selected by the AGMA-PESS software had no intersection with the experts' sequences (see Figures 4A,C). One was a valid sequence located at the end of the video. The second corresponded to a scene in which a clinician repositioned the infant, resulting in an important motion quantity. This can be interpreted as a limit of the AGMA-PESS software, which is focused on the infant's movements and does not consider the entire scene. In addition, the software cannot differentiate the crying or fussing scenes since it is only destined for pose estimation.

For pose estimation, AGMA-PESS software achieved an average FPS (Frame per Second) speed of 2.35 (see Table 3) on a machine equipped with a GPU (NVIDIA QUADRO RTX 3000) and a CPU (Intel Core i7-10850H processor). It uses a state-of-the-art retrained network Darkpose32 (12) that has shown very accurate results for 2D pose estimation compared to existing solutions such as Openpose network. This neural network was first used in a stereoscopic framework that allowed 3D pose reconstruction (12). AGMA-PESS software analyses videos from standard 2D cameras,

10.3389/fped.2024.1465632

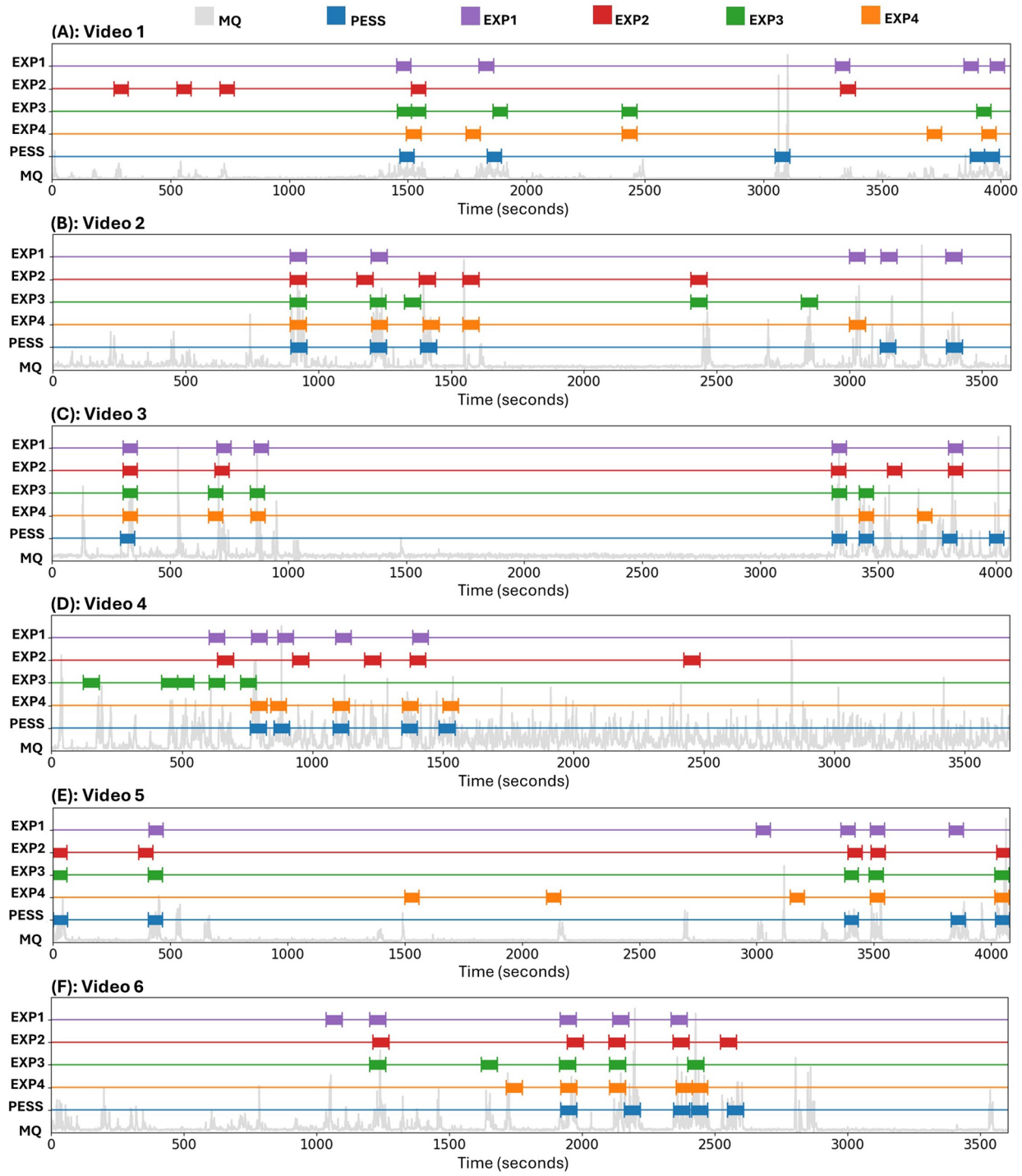


FIGURE 4
Comparison between the AGMA-PESS software and the four experts in the selection of five sequences of one-minute duration from six video recordings. Abbreviations: EXP, expert; MQ, motion quantity; PESS, AGMA-PESS software.

which are more commonly used in clinical practice. However, the AGMA-PESS software can be used on stereoscopic 3D setups by estimating 2D poses on each view, followed by manual triangulation, and selecting sequences using one view.

The AGMA-PESS software is user-friendly, with a video tutorial as Supplementary Material. It was tested with videos of different durations and validated to perform an automated selection of the best sequences of spontaneous motility in order to achieve the GMA of infants hospitalized in the Neonatal Unit at preterm and writhing ages from videos acquired according to Prechtl's method of GMA (1) (infants wearing diapers in supine position). No generalization can be made about the performance of the software outside these conditions. The AGMA-PESS software was not designed to perform an automated selection of

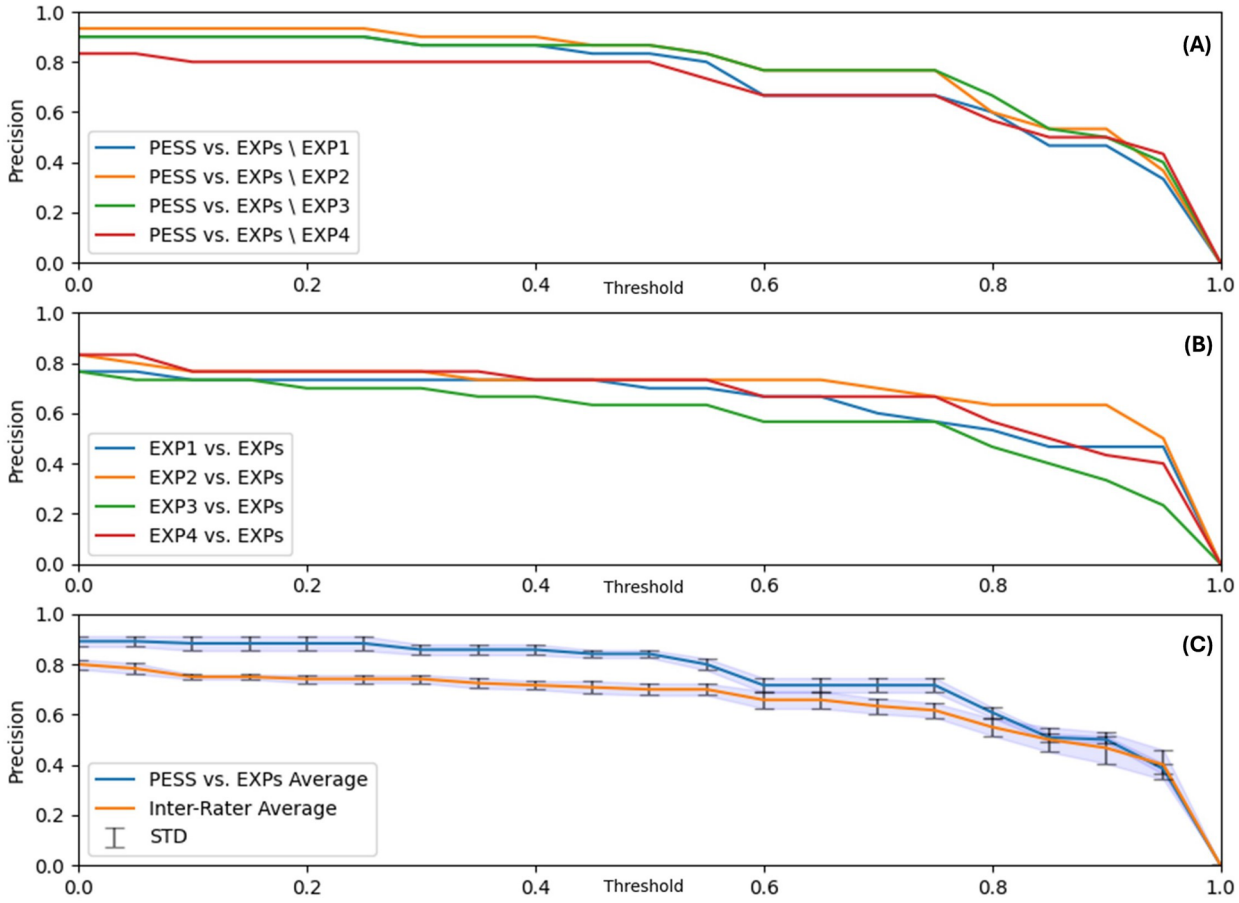


FIGURE 5
Sequence selection precision results. (A) Precision of the AGMA-PESS sequence selector compared to all experts except one each time. (B) Inter-rater agreements. (C) AGMA-PESS sequence selection average precision and average inter-raters agreement. Abbreviations: EXP, expert; STD, standard deviation; PESS, AGMA-PESS software.

TABLE 2 Time spent by each expert to manually perform sequence selection.

| Expert | Video 1 | Video 2 | Video 3 | Video 4 | Video 5 | Video 6 | Total |
|---|---|---|---|---|---|---|---|
| Expert 1 | 25 m 10 s | 20 m 24 s | 18 m 50 s | 32 m 10 s | 17 m 20 s | 30 m 10 s | 2 h 24 m 04 s |
| Expert 2 | 11 m 42 s | 14 m 39 s | 11 m 14 s | 10 m 06 s | 12 m 02 s | 22 m 06 s | 1 h 23 m 37 s |
| Expert 3 | 19 m 14 s | 25 m 51 s | 26 m 46 s | 32 m 53 s | 22 m 08 s | 46 m 30 s | 2 h 53 m 22 s |
| Expert 4 | 30 m | 11 m 49 s | 19 m 42 s | 35 m 53 s | 23 m 14 s | 26 m | 2 h 26 m 38 s |
| Average | 21 m 29 s | 18 m 01 s | 19 m 08 s | 27 m 45 s | 18 m 41 s | 31 m 11 s | 2 h 17 m 24 s |

Abbreviations: h, hours; m, minutes; s, seconds

TABLE 3 AGMA-PESS software processing time to automatically perform sequence selection.

| AGMA-PESS | Video 1 | Video 2 | Video 3 | Video 4 | Video 5 | Video 6 |
|---|---|---|---|---|---|---|
| Processing time | 29 m 31 s | 26 m 23 s | 26 m 52 s | 27 m 01 s | 27 m 05 s | 26 m 45 s |
| FPS | 2.28 | 2.28 | 2.52 | 2.26 | 2.51 | 2.24 |

Abbreviations: FPS, frame per second; h, hours; m, minutes; s, seconds.

the best sequences for GMA at fidgety's age. Nevertheless, fidgety movements should occur frequently at this age (1). The AGMA-PESS software is compatible with the Windows operating system (Microsoft, WA) with version $\geq 7$ and is limited to non-commercial purposes.

## 5 Conclusion

The AGMA-PESS software addresses the time-consuming process of manually selecting video sequences for GMA and offers a simple GUI for estimating an infant's pose with high

accuracy. The experiments proved the software's consistency and precision, comparable to the human level. The software's compatibility with Windows operating systems, user-friendly interface, and availability in multiple languages enhance its accessibility and make it a valuable tool for clinicians and researchers working in neonatal care. It was extensively used in the previous study by Soualmi et al. (15) for selecting 183 one-minute video sequences and estimating the pose of the infants in these videos for automatic classification of infants' movements. While AGMA-PESS excels in a clinical environment, its generalizability to diverse conditions remains a subject for further exploration.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the Comité de Protection des Personnes - Sud-Est II Ethical Committee in February 2021. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

## Author contributions

AS: Conceptualization, Project administration, Resources, Supervision, Validation, Writing – review & editing, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft; OA: Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing; CD: Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – review & editing; AP-R: Investigation, Writing – review & editing; AP: Investigation, Writing – review & editing; HP: Funding acquisition, Resources, Supervision, Validation, Writing – review & editing; AG: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing, Formal Analysis, Investigation, Methodology, Visualization.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by a Contrat Doctoral de l'École Doctorale 488 SIS, and the Agence Nationale de la Recherche (ANR; project ANR-23-SSAI-0005-02). The funding sources were not involved in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; and in the decision to submit the paper for publication.

## Conflict of interest

The authors declare that the research was conducted without any commercial or financial relationships that could potentially create a conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fped.2024.1465632/full#supplementary-material

## References

1. Einspieler C, Cioni HFRPG, Ferrari F, Bos AF. *Prechtls Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants*. London: Springer Science & Business Media B.V. (2008).

2. Hadders-Algra M. Neural substrate and clinical significance of general movements: an update. *Dev Med Child Neurol.* (2018) 60:39–46. doi: 10.1111/dmcn.13540

3. Olsen JE, Cheong JL, Eeles AL, FitzGerald TL, Cameron KL, Albesher RA, et al. Early general movements are associated with developmental outcomes at 4.5–5 years. *Early Hum Dev.* (2020) 148:105115. doi: 10.1016/j.earlhumdev.2020.105115

4. Morgan C, Fetters L, Adde L, Badawi N, Bancale A, Boyd RN, et al. Early intervention for children aged 0 to 2 years with or at high risk of cerebral palsy: international clinical practice guideline based on systematic reviews. *JAMA Pediatr.* (2021) 175:846–58. doi: 10.1001/jamapediatrics.2021.0878

5. Cao Z, Simon T, Wei S-E, Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).

6. Chambers C, Seethapathi N, Saluja R, Loeb H, Pierce S, Bogen D, et al. Computer vision to automatically assess infant neuromotor risk. *IEEE Trans Neural Syst Rehabil Eng.* (2020) 28:2431–42. doi: 10.1109/TNSRE.2020.3029121

7. Doroniewicz I, Ledwoń D, Affanasowicz A, Kieszczyńska K, Latos D, Matyja M, et al. Writhing movement detection in newborns on the second and third day of life

using pose-based feature machine learning classification. *Sensors*. (2020) 20(21):5986. doi: 10.3390/s20215986

8. McCay KD, Hu P, Shum HPH, Woo WL, Marcroft C, Embleton ND, et al. A pose-based feature fusion and classification framework for the early prediction of cerebral palsy in infants. *IEEE Trans Neural Syst Rehabil Eng*. (2022) 30:8–19. doi: 10.1109/TNSRE.2021.3138185

9. Moccia S, Migliorelli L, Pietrini R, Frontoni E. Preterm infants' limb-pose estimation from depth images using convolutional neural networks. In: *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (2019). p. 1–7. doi: 10.1109/CIBCB.2019.8791242

10. Reich S, Zhang D, Kulvicius T, Bölte S, Nielsen-Saines K, Pokorny F, et al. Novel AI driven approach to classify infant motor functions. *Sci Rep*. (2021) 11:9888. doi: 10.1038/s41598-021-89347-5

11. Sakkos D, McCay K, Marcroft C, Embleton N, Chattopadhyay S, Ho E. Identification of abnormal movements in infants: a deep neural network for body part-based prediction of cerebral palsy. *IEEE Access*. (2021) PP:1–. doi: 10.1109/ACCESS.2021.3093469

12. Soualmi A, Ducottet C, Patural H, Giraud A, Alata O. A 3d pose estimation framework for preterm infants hospitalized in the neonatal unit. *Multimed Tools Appl*. (2024) 83:24383–400. doi: 10.1007/s11042-023-16333-6

13. Sun K, Xiao B, Liu D, Wang J. Deep high-resolution representation learning for human pose estimation. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019). p. 5686–96. doi: 10.1109/CVPR.2019.00584

14. Zhang F, Zhu X, Dai H, Ye M, Zhu C. Distribution-aware coordinate representation for human pose estimation. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020). p. 7091–00. doi: 10.1109/CVPR42600.2020.00712

15. Soualmi A, Alata O, Ducottet C, Patural H, Giraud A. Mean 3D dispersion for automatic general movement assessment of preterm infants. In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2023). p. 1–5. doi: 10.1109/EMBC40787.2023.10340961

frontiers | Frontiers in Pediatrics

TYPE Review
PUBLISHED 24 December 2024
DOI 10.3389/fped.2024.1464465

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Gianluca Vergine,
Ospedale Infermi di Rimini, Italy
Maria-Cecilia Lopes,
University of São Paulo, Brazil

*CORRESPONDENCE
Ping Zhou
zhouping3698@aliyun.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 14 July 2024
ACCEPTED 05 December 2024
PUBLISHED 24 December 2024

CITATION
Liao J, Zhu L, Xie D, Wang X and Zhou P (2024) Improving the quality of life of children and parents with nocturnal enuresis: the role of health education.
Front. Pediatr. 12:1464465.
doi: 10.3389/fped.2024.1464465

COPYRIGHT
© 2024 Liao, Zhu, Xie, Wang and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Improving the quality of life of children and parents with nocturnal enuresis: the role of health education

Jianrong Liao[1,2†], Lin Zhu[1,3†], Danfeng Xie[1,3], Xiaomeng Wang[1,3] and Ping Zhou[1,3*]

[1]Department of Pediatric Nephrology and Rheumatology, Sichuan Provincial Maternity and Child Health Care Hospital, Chengdu, Sichuan, China, [2]School of Nursing, Chengdu Medical College, Chengdu, Sichuan, China, [3]Department of Pediatric Surgery and the Pediatric Nephrology Laboratory, Sichuan Clinical Research Center for Pediatric Nephrology, Chengdu, Sichuan, China

The sleep disorder nocturnal enuresis (NE) affects children's health and quality of life, as well as places a heavy burden on their families. Treatment improves the child's quality of life. Unfortunately, some parents do not seek treatment for their children because they are unaware or misinformed about NE. The goal of health education is to enhance or maintain the patient's health status, thereby enhancing the patient's quality of life through positive coping strategies. Educating children and parents about NE is an effective way for nurses to raise their awareness. Furthermore, they can improve the quality of life of children by promoting their active participation in treatment. The purpose of this paper is to review information about NE and explore the role of health education in improving children's and parents' quality of life.

KEYWORDS

nocturnal enuresis, quality of life, health education, children, parents

## 1 Introduction

International Childhood Continence Society (ICCS) defines nocturnal enuresis (NE) as involuntary urination during sleep that lasts at least three months and occurs twice a week (1). NE is a common condition in childhood (2). According to studies, NE prevalence ranges from 4% to 18% (3). It is possible for different countries or regions to have different prevalence rates due to differences in definitions and factors such as geography, environment, health resources, culture, and education. The prevalence of NE in China ranges from 4.07% to 10.03% (4). It is more common among boys before puberty, while it is equally prevalent among girls after puberty (1, 5).

NE is caused by various factors, including genetics, increased nocturnal urine output, reduced bladder capacity or urethral sphincter dysfunction, and sleep-wake disturbances (4, 6). Among them, sleep disorders, especially sleep apnea disorders, interact with NE (7–9). NE increases with the prevalence of OSA, and children with NE are more likely to develop OSA than normal children. And the symptoms of some children with NE may improve after treatment of OSA (9). OSA causes nocturnal hypoxia and carbon dioxide retention, which in turn causes a series of physiological changes, such as sympathetic arousal, elevated blood pressure, and increased cardiac load. These

Abbreviations
NE, nocturnal enuresis; ICCS, International Childhood Continence Society; VR, virtual reality; AR, augmented reality.

physiological changes may affect blood perfusion and urine production in the kidneys, leading to increased nocturnal urine output and thus increasing the risk of NE. In addition, frequent apneas and awakenings can result in sleep fragmentation, making it challenging for children to enter the deep sleep stage of sleep structure disruption, which in turn interferes with the normal function of the bladder, leading to the development of NE (7). Numerous studies have also suggested that excessive psychological pressure and family environment contribute to the development of NE (5). Children who have NE are less likely to develop NE as they get older. Children with NE recover spontaneously in about 10%–15% of cases (10). Chronic bedwetting, however, can negatively affect children's social and psychological well-being (11). Several studies have shown that children with NE have a lower quality of life than those without it (11, 12). In addition, parents of children with NE may experience increased psychological pressure and caregiving burden as a result of frequently changing and washing urine-soaked bedding and clothing (13).

The lack of knowledge and attention about NE, however, leads to some parents not seeking medical consultation or treatment for their children. The survey results show that only a portion of parents, approximately ranging from 2.1% to 55%, will seek treatment for their children (14). Enuresis education, including its causes and effective treatments, has been shown to change parents' views and misconceptions about the condition. This can increase attendance rates at health appointments, and enhance the quality of life for both children and parents (15).

Health education plays a vital role in improving the quality of life for children and parents in NE. Through health education, parents can be informed of NE's negative effects on their children. Treatment is an effective way of dealing with NE. Caldwell et al. Showed that children's quality of life could be boosted by effective treatment (16). At the same time, health education can also enhance children's self-confidence and alleviate anxiety and low self-esteem caused by bedwetting (15). Therefore, providing relevant information on NE is critical for improving children's psychological well-being and quality of life.

Several previous studies have emphasized the importance of correcting families' misconceptions, raising their awareness, and ensuring timely medical attention to improve children's and families' quality of life (4, 14, 17). However, current research focuses on exploring treatment modalities, epidemiology, and pathogenic mechanisms for NE (18). Specific educational methods and strategies are still not fully explored. In this review, we summarize the existing knowledge about NE and to examine the potential benefits of health education for children and families affected by NE. The aim is to provide useful clinical information to children and families affected by NE and to enhance their quality of life.

### 1.1 The impact of NE: on children and families

Numerous studies have found that children and families affected by NE have a significantly lower quality of life than children without the condition (1, 12). Mohamed et al. reported that NE is common and distressing for children and their families (19). Among its symptoms are shame upon awakening, fear of being discovered by friends, and psychological stress caused by parental punishment. In Iscan et al.'s study, it was found that children with enuresis had a poorer quality of life. Among the quality of life domains assessed, self-esteem, emotional well-being, and relationships with family and friends were the most vulnerable (20). In a similar study on the quality of life of children with NE, it was noted that children with NE are prone to low self-esteem, shame and guilt, and fear of being discovered, ridiculed (21). In addition, NE can also lead to sleep disorders, resulting in sleep deprivation at night. Children with NE may experience difficulty concentrating in the classroom, which can result in low grades (22).

It has a serious impact on the social lives of children who suffer from NE. Surveys indicate that bedwetting is the third most disruptive factor in children with NE, behind divorce and parental conflict (1, 23). In a study on children's psychological anxiety and depression, Ylmaz et al. found that bedwetting negatively affects academic performance and interpersonal relationships (24). A study by Eray et al. reported that children with NE tend to feel uneasy and avoid social situations when entering a new environment (25).

As a result of economic development and improved literacy, parents and medical professionals are becoming increasingly concerned about NE (4). Some parents also realize that enuresis is out of control. However, several studies have pointed out that if the child continues to wet the bed with age, parents do not condone it (12, 26). They often blame and penalize the child. Al-Zaben et al. noted that children who were punished for bedwetting had more severe NE than children who were not punished for enuresis, and that children who were punished for frequent and prolonged bodily injury had more severe depressive symptoms and poorer quality of life (27). A survey on NE in Italy found that 54.1% of parents punished their children for bedwetting (28).

NE affects the entire family, not just the child. A long period of changing bed covers and clothes after bedwetting can be exhausting and frustrating for parents. Arguing between parents about who changes the sheets and clothes can be stressful for family members. Tension also adversely affects physical and mental development (3, 17). At the same time, mothers caring for a child with NE may experience a decrease in their quality of life. They may experience feelings of anxiety, stress, and helplessness (26). In addition, families suffering from NE often spend more on laundry, bedding, disposable diapers, and mattress replacements (29).

## 2 Factors affecting the quality of life

### 2.1 Children's factors

Study found that children with NE have a quality of life influenced by gender and age (20, 30). Michel et al. observed that quality of life scores decreased with age, and girls scored lower than boys (31). Kilicoglu and Van et al. came to similar conclusions (32, 33). In

TABLE 1 Advantages and drawbacks of forms of health education and their applicable setting.

| Forms of health education | Advantages | Drawbacks | Applicable setting |
|---|---|---|---|
| Face-to-face education | Receive real-time patient feedback. Personalized answers to patients' questions. It facilitates the building of trust and emotional connection and enhances patient treatment compliance. | High cost of time spent. There are time and space constraints. Information dissemination is limited in speed and scope. | Seniors. Patients with chronic diseases. High-risk groups (heart attack patients, pregnant women, etc.). Children and adolescents. |
| Paper-based brochures | Ease of dissemination and preservation. Reusable. Comprehensive content. | High production costs. Untimely updating of information. Limited audience reach. | Seniors. Patients with chronic diseases. |
| Internet and health education | Wide coverage, no geographical limitations. Convenient and efficient, not limited by time and space. Information is updated in real-time. Variety of formats, including graphic, audio, video, etc. Resource consolidation can be achieved. | The quality of information varies. Privacy and security issues. There is a certain technical threshold. | Young population. Patients with chronic diseases. |
| Virtual reality (VR) and augmented reality (AR) | Immersive experience to deepen understanding and memory. Interactivity. Visual education to visualize some of the scenarios. | High equipment costs. High technical requirements. May bring about adverse reactions such as dizziness and discomfort. | Patients with motor and cognitive disorders requiring rehabilitation. Post-operative recovery training. Patients with chronic disease management, pain management, and psychological governance. Children and adolescents. |

addition, some studies have found older children also report more psychosocial problems, particularly regarding their family and self-concept of appearance (2). There are several possible reasons why this may be the case. First of all, their cognitive level and self-care ability gradually increase with age, which makes them more afraid of being discovered by their friends for bedwetting, and they show lower self-esteem, shame, and exclusion from social activities (34). In addition, it may be related to cultural background. In some countries, especially in Asia, parents tend to pay more attention to boys, believing they will inherit the family's power and duties and have a greater responsibility to support them. Because of this, parents are more willing to spend money, time, and resources on boys than on girls (31).

## 2.2 Family factors

Several factors affect the quality of life of children with NE, including caregiver literacy (28, 35). Research suggests that children from families with lower education levels may be at higher risk of enuresis (36). Furthermore, studies have found that mothers are typically the primary caregivers of children with NE, and their level of education can influence their management style (37). Mothers with higher levels of education are more likely to recognize abnormalities in their children and actively seek medical attention, compared to mothers with lower levels of education. Schlomer et al. found that caregivers with postgraduate education were more likely to seek healthcare for their children (38). The well-being of children with NE is also affected by parents' perceptions and attitudes (2). Less educated parents may have misconceptions about NE, believing that bedwetting is a problem of self-control. Therefore, when bedwetting persists for a long period, they may perceive it as a failure of their education. They may experience anxiety, guilt, and loss of confidence. Finally, they may even blame and scold their children (39).

There is a correlation between economic income and treatment success (24). Families with lower incomes have a higher prevalence of NE, according to studies (3, 28). Children from low-income families with enuresis are at risk of low quality of life (20). Treatment with NE may not be effective in the short term and requires long-term adherence. Thus, the cost of treatment may place more of a burden on less well-off families (17). In addition to treatment costs, managing the child's daily needs will also increase the family's financial burden compared to a normal child.

# 3 Health education approaches

According to the survey, many parents and children are unaware of NE. To ensure that parents and children understand and accept NE treatment, medical professionals must disseminate information about the disease in a health education setting. It will reduce the impact of NE on children and their families, improving their overall quality of life in the long run (16, 24). In the medical system, health education is crucial in improving patients' health literacy, promoting disease recovery, and preventing disease recurrence (40). Traditional health education is delivered through face-to-face teaching, distribution of health education manuals, telephone counseling, or outpatient follow-up visits. As technology advances and medical paradigms change, forms of health education are evolving. For example, virtual reality technology and the internet can now be used together (41, 42).

By integrating modern technologies into health education, we can enhance the effectiveness of health education and thereby improve the satisfaction of our patients (43). Regardless of the form of health education, the most important thing is to convey medical information and knowledge accurately. In order to achieve this goal, nurses must understand the advantages and disadvantages of each type of health education (see Table 1) and how best to communicate disease information to patients and their families.

Face-to-face health education is part of the traditional health education approach (43). Direct face-to-face communication and interaction with the patient are used to convey information about the disease. It is a very popular method of health education. However, in the process of face-to-face health education, some patients and their families are prone to problems such as incomplete information acceptance, insufficient information processing ability, and incorrect information memorization (44). As a result, healthcare professionals may need to inform patients multiple times to achieve an educational effect, resulting in increased time and costs.

A health education manual is a traditional form of health education. The document is comprehensive and can be kept for a long time. Medical terminology, however, may be difficult to understand for some patients and their families (19). In order to assist patients and families in understanding their diseases, healthcare providers can use visual aids, such as pictures.

Globally, 65.6% of the population has access to the Internet (45). As a result, the Internet is a highly beneficial tool for health education. The use of the Internet to facilitate health education helps patients transcend the limitations of time and distance. Online information can be accessed anytime and reviewed multiple times as desired. There is, however, a lack of quality and variability in health information currently available on the internet (46). There may be some patients who have difficulty screening information and are therefore more likely to misjudge false information.

Virtual reality (VR) and augmented reality (AR) are emerging technologies. Technologies like these are used in gaming and entertainment. However, researchers are now also exploring their potential in medical applications, such as critical care medicine, rehabilitation, and pediatrics (41, 47). VR enables users to immerse themselves in a three-dimensional virtual world, while AR allows digital information to be superimposed onto the real world (47). VR and AR offer unique immersion, presence, and interactivity. They can make health education content more interesting and engaging, and help patients and families understand abstract medical knowledge (48–50). It is likely that verbal preaching will not be effective with children due to their limited understanding. VR and AR, however, create a sense of immersion that children can relate to. This makes it easier for them to understand the disease and also reduces the anxiety caused by the disease. VR and AR are effective tools for delivering disease-related information, helping patients understand their disease, and improving their compliance with treatment. There is, however, a limit to the adoption of VR and AR technologies due to their cost, the need for professional equipment, and the lack of technical support.

Researchers in Egypt have observed that implementing a health education learning package can prove to be an effective way for mothers of NE children to improve their knowledge, practices, and attitudes, while at the same time promoting their child's compliance with treatment. However, the health education learning package was based on pre-existing questionnaires designed specifically for the mothers involved. Additionally, the health education learning package employs a variety of health education methods and has been reviewed repeatedly by experts. Despite the high quality of the content, the design and implementation of the health education learning package is time-consuming (16). Video content about enuresis was analyzed on YouTube by Toprak et al. Video quality appears to vary widely based on their findings. In spite of the fact that some content may be helpful to patients and families facing difficulties, the presence of low-quality videos may mislead them (42). Health education for children with NE is currently disseminated through mass media, posters, brochures, maternal and child health centers, and individual counseling. Virtual reality technology has not been studied in children with NE, but it may be explored in the future.

# 4 Implementation strategies for health education

As part of the treatment process for NE, health education plays a complementary but crucial role. To improve compliance and ensure treatment success, children and parents need to have a basic understanding of the principles and methods of treatment before treatment begins, whether medication or behavioral therapy. Meanwhile, healthcare professionals must answer promptly to any problems encountered during treatment. It is well known that health education can be very helpful when it comes to improving treatment and quality of life. Thus, NE treatment methods should be integrated with health education to optimize outcomes and improve compliance.

## 4.1 Behavioral interventions

Behavioral interventions are an initial approach to treating NE. It includes adjusting habitual routines, developing appropriate urination and defecation habits, implementing reward mechanisms, and keeping a urine loss diary (4, 51). These measures can be used alone or in combination with medication or alarms.

One of the etiological factors of NE is a nocturnal reduction in bladder capacity or overactivity of the urethra (1). Therefore, several studies have suggested appropriate solutions to this etiology. Firstly, children drink normally during the day, drinking their daily amount of water. At night they limit their water intake, avoiding all liquids, especially two hours before bedtime. Secondly, empty your bladder before heading to bed. Thirdly, foods or drinks containing theophylline and caffeine can irritate the bladder, so they should be avoided in children, especially at night. Finally, excessive salt intake can lead to increased nocturia, so salt intake should be controlled during dinner (6, 52, 53).

It is possible to treat NE by developing healthy urination and defecation habits (29). It has been found that excess feces in the colon can reduce bladder capacity and that relieving constipation reduces NE (54). To maintain clear stool, children should consume more foods with crude fiber and have regular bowel movements. Children with constipation should also be treated aggressively to avoid interference with NE (55).

Some parents are unaware of enuresis and blame and punish their children for frequent bedwetting. But the truth is that punishment and chastisement to cure bedwetting are counterproductive (6). Therefore, it is very important for healthcare professionals to provide parents with information about the pathogenesis of NE so that they understand that bedwetting is not within the child's control. In addition, studies have shown that encouragement reduces psychological distress and improves treatment adherence (56). Therefore, parents can provide appropriate encouragement to help their children realize that bedwetting is not their fault and reduce psychological pressure. In addition, during the treatment period, parents can reinforce the child's positive behavior and motivation for treatment. This can help the child actively participate in treatment and improve therapy success rates.

It is imperative to keep a urination diary. It is noninvasive and effective, and can objectively assess the child's condition, determine the type of NE, and guide follow-up treatments (57). Additionally, it can help evaluate treatment compliance by the child and family and supervise treatment more effectively. Kyung et al. observed that questionnaires may be subject to memory bias and subjectivity, whereas urinary diary records provide more objective and reliable information (58).

Behavioral interventions play an active role in the treatment of children with enuresis as a complement to first-line treatment. Due to its low cost and lack of risk, behavioral therapy is often preferred by parents who worry about medications' potential negative effects on their children (59).

## 4.2 Alarm therapy

Enuresis alarms are considered one of the first lines of treatment for enuresis (16). It is an electronic device that consists of an alarm and a humidity sensor. There are two main types: body-worn and padded, while the fixed-on-the-body alarm type is the choice for more people (52). The alarm works on the principle that once the humidity sensor attached to the trousers is activated, it will wake up the child via an alarm like a ringing bell (60). This treatment assists children in developing a regular mechanism for urinary arousal during sleep. This is done by enhancing their ability to store urine at night and reducing nocturnal urination. Based on the survey, the success rate of treating enuresis using alarms ranges from 62% to 70% (52). Figure 1 illustrates the alarm therapy treatment process (57).

Even though alarms can effectively treat NE, there are some issues with the treatment process. Alarms can be time-consuming and exhausting for both children and their families (61). According to the ICCS, parents should stay with their child while the alarm sounds, allowing the child to turn off the alarm and go to the bathroom to urinate, thus maximizing the therapeutic effect (60). However, a prolonged alarm can disrupt parents' and children's sleep, resulting in fatigue during the day. Alarm therapy can be interrupted as often as 30% according to a survey (62). Alarm treatment effectiveness can only be determined after at least 6 to 8 weeks of continuous treatment (10). Therefore, if children and parents do not see results in the treatment process, they may lose motivation and discontinue treatment.

Alarms are not suitable for all NE children. Parents and children who are motivated to follow the treatment may be more effective with alarms when bedwetting episodes are regular and urine output exceeds 65% of the expected volume. A better option might be medication for children with NE whose parents don't support them, who are stressed, or who lack motivation (29, 52).

Health education is recommended before alarm therapy to reduce dropout rates and enhance treatment efficiency. For instance, the nurse should illustrate the alarm's operation to ensure that both parents and children comprehend its functionality. Nurses should also advise parents to stay with the child as much as possible during night-time treatments to help the child develop effective treatment habits and achieve therapeutic results. Furthermore, it is essential that parents and children are made aware that the alarm treatment is a process that necessitates time and patience. This is to ensure that they are psychologically prepared for the long-term nature of the treatment, thus preventing them from abandoning it prematurely due to the lack of immediate visible outcomes. Additionally, it is crucial for parents and children to comprehend the efficacy of the treatment approach, which will in turn bolster their confidence in the treatment. During treatment, parents can explain the effects of enuresis and the importance of treatment in a way that the child can understand, so that the child understands why treatment with an alarm is necessary. Parents should encourage their children to play an active role in the treatment process, thereby fostering a sense of responsibility and self-confidence. Concurrently, parents must also be mindful of their children's emotional fluctuations and provide timely encouragement and support when their children encounter difficulties or setbacks.

## 4.3 Medication

### 4.3.1 Desmopress

Desmopressin is another effective option for treating NE. For NE, it is considered one of the first-line drugs (63). Desmopressin is a synthetic analog of an anti-diuretic hormone. It reduces nocturnal urine production by increasing water reabsorption in the renal tubules and regulating renal ion secretion (64). In this way, it is more effective at treating NE caused by antidiuretic hormone deficiency. A complete response is achieved by 20%–30% of patients receiving desmopressin treatment, whereas a partial response is achieved by 20%–40% of patients (65). Figure 2 illustrates how desmopressin therapy works (57).

Desmopressin can be administered in different ways, such as nasal sprays, nasal drops, oral tablet capsules, and frozentablets (66). Hyponatremia and seizures have been reported following the transnasal desmopressin administration. The United States and some European countries have therefore banned its use (66, 67). A study by Keten et al. noted that freeze-dried desmopressin tablets have a higher success rate than other forms of the drug (63).

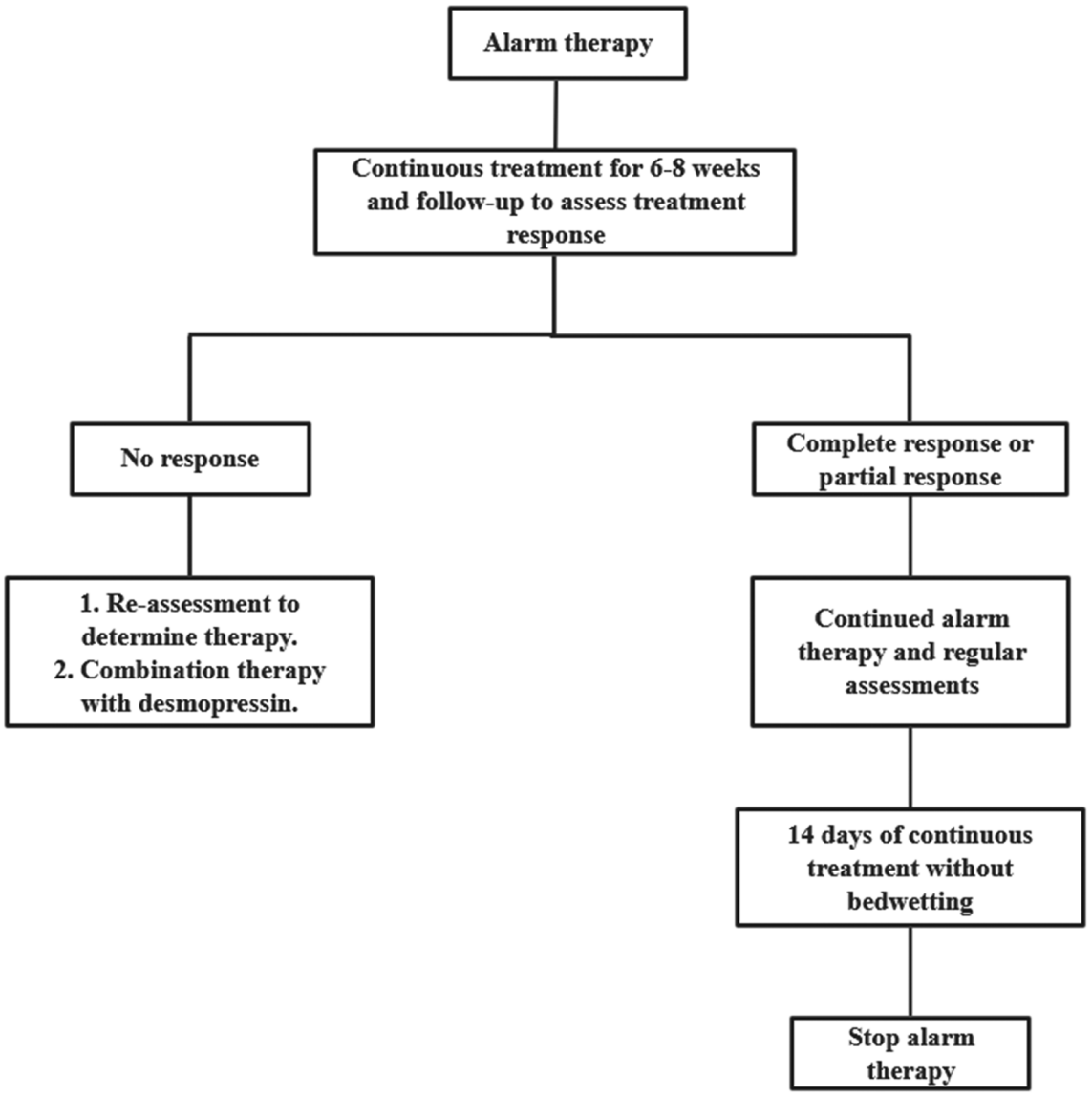


FIGURE 1
Alarm therapy treatment process. By consensus, an expert committee developed a protocol for evaluating and treating pediatric nocturnal enuresis patients in mainland China. No response: 50% reduction in bedwetting frequency; Partial response: 50% to 99% reduction in bedwetting frequency; Complete response: 100% reduction in bedwetting frequency.

Desmopressin should be taken one hour before bedtime or two hours after dinner, while also limiting fluid intake at night to prevent hyponatremia (29, 52, 60). To begin with, Desmopressin should be given at a low dose of 0.1 to 0.2 mg. The dosage could then be adjusted based on the therapeutic effect. Two to four weeks after starting the treatment, it is recommended to evaluate its effectiveness. If the treatment is effective, the dose can be maintained and continued for at least 3 months (52, 66).

Despite being faster-acting than alarm therapy, desmopressin treatment has a significant disadvantage: a high relapse rate. After 12 weeks of desmopressin treatment, the relapse rate was 50% compared to only 12% with alarm therapy (63). A structured discontinuation method, i.e., discontinuing desmopressin in a stepwise manner, has been suggested to reduce the relapse rate (68, 69). However, some studies suggest that there is no difference between the two methods (direct cessation or structured withdrawal) in terms of reducing recurrence rates (69).

The nurse reminds parents to administer medications to their children strictly in accordance with doctors' orders to ensure that children take the correct dose of medication at the correct time. At the same time, supervise children to limit their fluid intake at night and develop good living habits. It is also necessary to teach parents to closely monitor the physical condition of their children during the medication period. If any related complications or abnormalities occur, they should contact health professionals in time. In addition, nurses need to follow up with parents and children on a regular basis to see how the treatment is

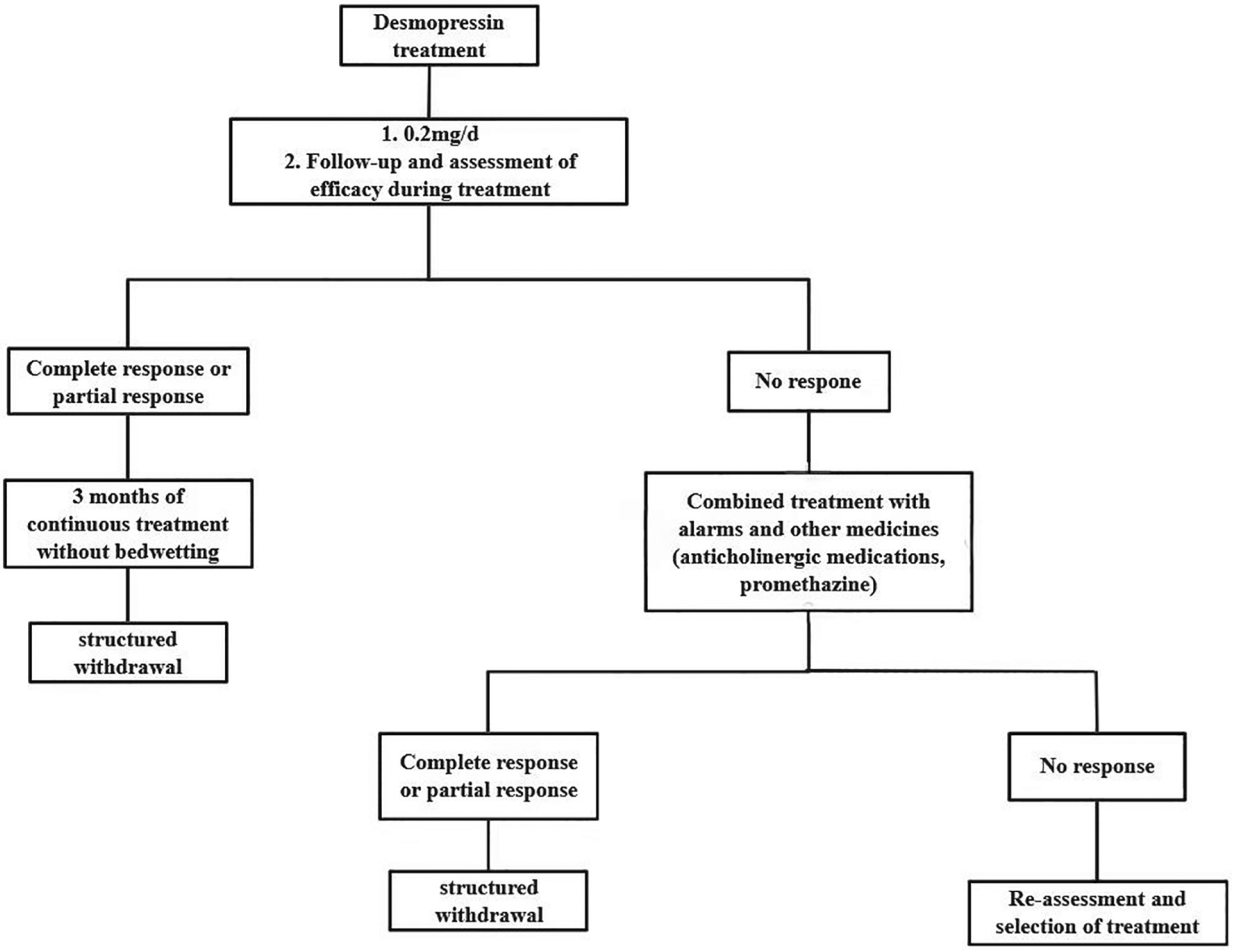


FIGURE 2
Desmopressin therapy treatment process.

progressing and to assess the efficacy of the medication and the child's physical response.

### 4.3.2 Anticholinergic medications

Anticholinergic medications such as tolterodine and oxybutynin can also treat NE, in addition to desmopressin medications. It primarily works by relaxing the bladder muscles, increasing bladder capacity, and improving the overactive or unstable state of the urethral muscles (70). This reduces the frequency of urination and enuresis. Consequently, they are more effective in children with reduced bladder capacity or overactive bladders. When urine loss alarms and desmopressin do not work, anticholinergic drugs are usually used. However, anticholinergic medications are not effective alone, so they are usually combined with desmopressin (52). Some studies have shown that the combination of desmopressin and oxybutynin significantly increases the number of patients with complete and partial responses (71, 72). Additionally, anticholinergic drugs have several potential side effects, including dry mouth, dry eyes, constipation, headaches, and urinary retention (73, 74). It is therefore critical to use them carefully and cautiously.

### 4.3.3 Imipramine

Imipramine is a tricyclic antidepressant that can also treat NE. The mechanism by which it treats NE is unknown. It has been shown that these drugs stimulate adrenergic receptors, modify sleep patterns, and decrease excessive night-time sleep. As a result, patients can wake up once their bladder is full. They also inhibit acetylcholine, which affects bladder contraction and urethral sphincter relaxation. Reducing the frequency of involuntary bladder contractions and increasing urine storage capacity, can help patients improve their nocturnal bedwetting symptoms (56, 75). This treatment benefits 30%–50% of patients with NE (76). Due to its potential cardiotoxicity, imipramine should be used cautiously when treating NE. An overdose can cause cardiac conduction disturbances and myocardial depression (76, 77). In addition, tricyclic antidepressants may be associated with adverse effects such as dry mouth, constipation, blurred vision, tachycardia, weight gain, headaches and dizziness, behavioral changes, irritability, and drowsiness. So for safety reasons, it has been used as a third-line treatment option (77). However, imipramine is relatively inexpensive compared to other treatments, so it is an optional treatment for families with low incomes (52).

### 4.4 Chinese medical treatment

Recent years have seen an increase in the use of traditional Chinese medicine (TCM) for NE clinical treatment (78). In NE treatment, TCM is effective without significant side effects. Treatments include herbal medicine, acupuncture, and tuina (78, 79).

Suoquan Capsule is a herbal medicine commonly used to treat NE. It contains Radix Linderae, Alpinia oxyphylla, and yam, which reduce urethral muscle excitability and concentrate urine. This makes it an effective treatment for NE. In the study by Ma et al, shrinkage capsules along with desmopressin were more effective and had a lower relapse rate than shrinkage capsules alone (79).

Acupuncture stimulates specific points such as Guanyuan, Sanyinjiao, Ashigaru, and Bladder Yu, improving nerve function and bladder control (80). Enuresis can be treated with laser acupuncture, which is less invasive and painful than traditional acupuncture (81). Studies have shown that laser acupuncture is a painless and non-invasive treatment with a low recurrence rate and no side effects. In some cases, it can be considered a viable alternative to NE (82, 83). There is controversy regarding whether laser acupuncture is more effective in treatment than desmopressin. The effectiveness of laser acupuncture and desmopressin for the treatment of enuresis in children was compared by Radmayr et al. (2001). In the study, forty children were randomly divided into two groups, one receiving desmopressin and the other receiving laser acupuncture. Neither modality had a significant difference in treatment success (84). However, in a similar study, Alsharnoubi et al. (2017) found that laser acupuncture was statistically significantly more effective than desmopressin in a similar study (85).

A common treatment for NE in China is pediatric tuina (86). It is similar to the mechanism of action of acupuncture, in specific parts of the body surface, the use of chiropractic, kneading, massage, and other techniques to give a certain amount of stimulation, which increases qi and blood flow and enhances bladder function. The effects of tuina have been shown to affect signaling in various brain reflex zones (86, 87). Treatment for enuresis using tuina can be a non-invasive, painless, and side-effect-free alternative.

TCM treatment has fewer side effects than Western medical therapies, even though there is not enough evidence to conclude that TCM is more effective than Western medicine. Thus, TCM can be considered as an alternative treatment option for enuresis. Before treatment, nurses should explain to parents the specific characteristics, applicable conditions and possible effects of traditional Chinese medicine therapies such as herbal medicine, acupuncture and tuina, and help parents to choose appropriate Chinese therapies or integrative therapies according to their children's specific conditions, such as age, severity of symptoms, physical tolerance, and so on. All therapies require a certain course of treatment to be effective. It is the responsibility of parents to encourage their children to engage actively with the treatment, while simultaneously maintaining a record of their child's progress and any changes in symptomsso that they can provide timely feedback to the doctor.

## 5 Discussions

Behavioral interventions are an initial approach to treating NE, and throughout the course of treatment (51). Some parents may lack an understanding of the necessity of behavioural interventions and are unable to effectively supervise their children in order to implement them, which affects the treatment outcome. Nurses assist parents and children in understanding the principles and importance of behavioural interventions through the provision of health education, thereby increasing the likelihood of parental and child cooperation with changes in lifestyle habits and active participation in treatment.

Although alarm therapy for NE has been demonstrated to be highly effective, it also has a high rate of interruption (52, 62). If the child and parents are not aware of the long-term nature of the treatment and the potential issues that may arise, such as disrupted sleep during treatment and the lack of immediate improvement, they may lose confidence and terminate the treatment prematurely, thereby preventing the treatment from achieving its intended outcomes. Furthermore, a lack of comprehension regarding the correct utilisation of the alarm may result in suboptimal treatment outcomes. Therefore, nurses need to explain to parents and children how the alarm works and how to use it before the treatment is given to ensure that they can use it properly. Parents and children should also be made to understand that alarm treatment is a process that requires time and patience, so that they will not easily give up the treatment because they do not see any results in the short term or encounter difficulties.

It is important to note that pharmacological treatment is not without side effects, and desmopressin medication is also associated with a high relapse rate (63, 73, 74, 77). Consequently, when selecting a medication to treat NE, parents are apprehensive about the potential adverse effects on their children and may be reluctant to cooperate. Through health education, nurses provide parents with comprehensive information about the principles of drug treatment, including the timing of administration, dosage, and potential adverse effects. This enables parents to make informed decisions and enhances their cooperation with the doctor's treatment plan, ultimately improving the adherence and effectiveness of the treatment.

Some parents may have misconceptions about TCM treatments, thinking that their effects are unclear or not believing that they can cure NE. The nurse explains to the parents the specific characteristics, application and possible effects of TCM treatments such as Chinese herbs, acupuncture and tuina (78), helping them to choose the appropriate treatment according to their child's specific condition. At the same time, parents and children should be informed that the side effects of TCM are relatively few, so as to increase their compliance with treatment.

As technology advances, treatments are being updated. From the literature, it has been found that a new product is now being developed for the treatment of NE, a safe, comfortable, non-invasive pre-voidance wearable alarm that measures bladder filling using ultrasonic sensors to trigger an alarm prior to urination (88–91). Traditional alarms can only react to a

bedwetting event after it has occurred and cannot give advance warning (60). The pre-voidance alarm is more proactive and preventive as it can alert the child before he/she is about to urinate, giving him/her enough time to wake up and go to the toilet to urinate. And it can also adjust the threshold and timing of the alarm trigger to suit each child's specific needs through multiple measurements and learning from the bladder (90). In conclusion, by triggering the alert at the pre-urination stage, it helps to develop bladder control, and over time, the child may gradually learn to control his/her urination behaviour, reducing the frequency of nocturnal enuresis and improving the therapeutic outcome. At the same time, warning the child before bedwetting gives him/her the opportunity to control urination on his/her own, avoiding the occurrence of bedwetting incidents and helping to protect the child's self-esteem (91).

This study covers a variety of treatments for enuresis, including behavioural interventions, alarm therapy, multiple medications, and TCM, and elaborates on the importance and specific content of health education in each of these treatments. This provides a more comprehensive guide to clinical practice, enabling healthcare professionals to carry out health education effectively at different stages of treatment according to the conditions of different patients. With the development of technology, this study also explored the potential of new technologies such as VR and AR in health education. Although they are not yet widely used in health education for the treatment of enuresis, it is suggested that these technologies can make the content of health education more interesting and easier for children to understand, providing new directions and ideas for future research and practice.

## 6 Conclusion

Nocturnal enuresis (NE) is one of the most common chronic disorders of childhood, affecting children's health and quality of life, as well as placing a heavy burden on families. Even though some children may recover with age, most may experience long-term effects if they do not receive treatment. Health education plays a vital role in improving the quality of life of NE children and parents. Nurses communicated information about NE to children and parents through health education, which increased their awareness of NE and reduced their anxiety and worry. At the same time, this study summarises the treatment approaches for NE and the role of health education, which can inform healthcare professionals when developing treatment plans and providing patient education. Secondly, this study also highlights the need for further research into more effective health education models and strategies, especially in the context of emerging technologies such as virtual reality and augmented reality. Future research could focus on filling the research gaps identified in this study, such as exploring the long-term effectiveness of different treatments in combination with health education, as well as investigating how to better tailor health education to the individual characteristics of patients and families. In conclusion, this study contributes to a deeper understanding of NE and the importance of its health education. It provides valuable insights for improving the quality of life of children with NE and their parents.

## Author contributions

JL: Writing – original draft, Writing – review & editing. LZ: Writing – original draft, Writing – review & editing. DX: Writing – review & editing. XW: Writing – review & editing. PZ: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Sichuan Provincial Health Commission Popularization Research Project (19PJ256), Key Project of Science and Technology Innovation Fund of Sichuan Provincial Maternally and Child Health Hospital (ZC-202315), DX, 2022 Key R&D Plan of Science and Technology Department of Sichuan Province (2022YFS0149), 2021 Sichuan Medical Research Project of Sichuan Medical Association (S20137), Health Commission of Sichuan Province Technology Projects in 2023, Clinical Research Projects (23LCYJ015), PZ, 2022 Innovation Fund of Sichuan Maternal and Child Health Hospital (CXZD2023-05), XW.

## Acknowledgments

Authors also express gratitude to Ping Zhou for writing assistance.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

# References


1. Huang W, Liang Y, Yang Q, Ma X. Nocturnal enuresis in children: parents' perspectives. *Nurs Open*. (2022) 9(5):2335–41. doi: 10.1002/nop2.1243

2. Tai TT, Tai BT, Chang Y, Huang K. Parents have different perceptions of bed-wetting than children from six to 15 years of age. *Acta Paediatr*. (2015) 104(10): e466–472. doi: 10.1111/apa.13101

3. Alanazi ANH, Alanazi RSM, Alanazi EN, Alanazi RM, Rabbani U. Prevalence of nocturnal enuresis among children and its association with the mental health of mothers in northern Saudi Arabia. *Cureus*. (2022) 14(2):e22232. doi: 10.7759/cureus.22232

4. Hu HJ, Zhang ZW, Liang Y, Luo YY, Dou QF, Song CP, et al. Prevalence, risk factors, and psychological effects of primary nocturnal enuresis in Chinese young adults. *Int Neurourol J*. (2021) 25(1):84–92. doi: 10.5213/inj.2040398.149

5. Sousa E Silva GJ, Sammour SNF, Ferraro AA, Koch VHK. Study of the profile of behavioral problems and quality of life indexes in a pediatric cohort of monosymptomatic enuresis. *J Pediatr (Rio J)*. (2019) 95(2):188–93. doi: 10.1016/j.jped.2017.12.006

6. Walker RA. Nocturnal enuresis. *Prim Care Clin Off Pract*. (2019) 46(2):243–8. doi: 10.1016/j.pop.2019.02.005

7. Pedersen MJ, Rittig S, Jennum PJ, Kamperis K. The role of sleep in the pathophysiology of nocturnal enuresis. *Sleep Med Rev*. (2020) 49:101228. doi: 10.1016/j.smrv.2019.101228

8. Soster LA, Alves R, Fagundes SN, Koch VHK, Bruni O. Sleep disturbances associated with sleep enuresis: a questionnaire study. *Eur J Paediatr Neurol*. (2016) 20(2):282–5. doi: 10.1016/j.ejpn.2015.11.014

9. Soster LA, Fagundes SN, Lebl A, Alves RC, Koch VH, Spruyt K. Beyond bedwetting: how successful treatment is observed in sleep macrostructure. *Sleep Med*. (2024) 124:331–7. doi: 10.1016/j.sleep.2024.09.035

10. Song P, Huang C, Wang Y, Wang Q, Zhu W, Yue Y, et al. Comparison of desmopressin, alarm, desmopressin plus alarm, and desmopressin plus anticholinergic agents in the management of paediatric monosymptomatic nocturnal enuresis: a network meta-analysis. *BJU Int*. (2019) 123(3):388–400. doi: 10.1111/bju.14539

11. Ring IJ, Nevéus T, Markström A, Arnrup K, Bazargani F. Nocturnal enuresis impaired children's quality of life and friendships. *Acta Paediatr*. (2017) 106(5):806–11. doi: 10.1111/apa.13787

12. Rangel RA, Seabra CR, Ferrarez CEPF, Soares JL, Choi M, Cotta RG, et al. Quality of life in enuretic children. *Int Braz J Urol*. (2021) 47(3):535–41. doi: 10.1590/s1677-5538.ibju.2020.0308

13. Quiroz-Guerrero J, Ortega-Pardo A, Maldonado-Valadez RE, De León RG-D, Mercado-Villareal L, Rodea-Montero ER. Maternal anxiety associated with nocturnal childhood enuresis. *Children*. (2022) 9(8):1232. doi: 10.3390/children9081232

14. Tsai H-L, Chang J-W, Chen M-H, Jeng M-J, Yang L-Y, Wu K-G. Associations between psychiatric disorders and enuresis in Taiwanese children: a national population-based study. *Clin Epidemiol*. (2020) 12:163–71. doi: 10.2147/CLEP.S230537

15. Yilmaz EŞ, Büyük ET. Effect of education given to children with enuresis on quality of life. *J Pediatr Urol*. (2021) 17(5):648.e1–e7. doi: 10.1016/j.jpurol.2021.08.001

16. Caldwell PH, Codarini M, Stewart F, Hahn D, Sureshkumar P. Alarm interventions for nocturnal enuresis in children. Cochrane db. *Syst Rev*. (2020) 2021(12):CD002911. doi: 10.1002/14651858.CD002911.pub3

17. Huang H-M, Wei J, Sharma S, Bao Y, Li F, Song J-W, et al. Prevalence and risk factors of nocturnal enuresis among children ages 5–12 years in Xi'an, China: a cross-sectional study. *BMC Pediatr*. (2020) 20(1):305. doi: 10.1186/s12887-020-02202-w

18. Li W, Yang G, Tian W, Li Y, Zhang L, Wang Y, et al. Bibliometric and visual analysis of nocturnal enuresis from 1982 to 2022. *Front Pediatr*. (2022) 10:972751. doi: 10.3389/fped.2022.972751

19. Mohamed RA, Youssef El-Sheikh O, Noaman A. Applying health education learning package for mothers regarding nocturnal enuresis. *Am J Nurs Res*. (2019) 7(4):561–73. doi: 10.12691/ajnr-7-4-18

20. Iscan B, Ozkayın N. Evaluation of health-related quality of life and affecting factors in child with enuresis. *J Pediatr Urol*. (2020) 16(2):195.e1–e7. doi: 10.1016/j.jpurol.2019.12.018

21. Klein MO, Thews A, Schulz-Jürgensen S. Health-related quality of life of children and adolescents with primary nocturnal enuresis who are undergoing therapy. *Urologe*. (2021) 60(9):1175–83. doi: 10.1007/s00120-021-01549-x

22. Shafiek H, Evangelisti M, Abd-elwahab NH, Barreto M, Villa MP, Mahmoud MI. Obstructive sleep apnea in school-aged children presented with nocturnal enuresis. *Lung*. (2020) 198(1):187–94. doi: 10.1007/s00408-019-00304-6

23. Tai TT, Tai BT, Chang Y-J, Huang K-H. The importance of understanding parental perception when treating primary nocturnal enuresis: a topic review and an institutional experience. *Res Rep Urol*. (2021) 13:679–90. doi: 10.2147/RRU.S323926

24. Oztorun ZY, Beser NG, Oztorun K, Arabacı LB. Evaluating the social anxiety depression levels and accompanying psychosocial problems in children diagnosed with enuresis. *Cureus*. (2022) 14(8):e28351. doi: 10.7759/cureus.28351

25. Eray Ş, Tekcan D, Baran Y. More anxious or more shy? Examining the social anxiety levels of adolescents with primary enuresis nocturna: a controlled study. *J Pediatr Urol*. (2019) 15(4):343.e1–e5. doi: 10.1016/j.jpurol.2019.04.002

26. Durmaz O, Kemer S, Mutluer T, Bütün E. Psychiatric dimensions in mothers of children with primary nocturnal enuresis: a controlled study. *J Pediatr Urol*. (2017) 13(1):62.e1–e6. doi: 10.1016/j.jpurol.2016.06.018

27. Al-Zaben FN, Sehlo MG. Punishment for bedwetting is associated with child depression and reduced quality of life. *Child Abuse Neglect*. (2015) 43:22–9. doi: 10.1016/j.chiabu.2014.11.007

28. Ferrara P, Franceschini G, Di Castelbianco FB, Bombace R, Villani A, Corsello G. Epidemiology of enuresis: a large number of children at risk of low regard. *Ital J Pediatr*. (2020) 46(1):128. doi: 10.1186/s13052-020-00896-3

29. Berry AK. Helping children with nocturnal enuresis: the wait-and-see approach may not be in anyone's best interest. *AJN Am J Nurs*. (2006) 106(8):56–63. doi: 10.1097/00000446-200608000-00024

30. Assimamaw NT, Kebede AK, Genetu KB. Effects of sex, toilet training, stress, and caffeine on nocturnal enuresis among school children in gondar town, the metropolitan city of Ethiopia: a community-based study in 2023. *Front Pediatr*. (2024) 12:1366430. doi: 10.3389/fped.2024.1366430

31. The KIDSCREEN group C, Michel G, Bisegger C, Fuhr DC, Abel T. Age and gender differences in health-related quality of life of children and adolescents in Europe: a multilevel analysis. *Qual Life Res*. (2009) 18(9):1147–57. doi: 10.1007/s11136-009-9538-3

32. Kilicoglu AG, Mutlu C, Bahali MK, Adaletli H, Gunes H, Metin Duman H, et al. Impact of enuresis Nocturna on health-related quality of life in children and their mothers. *J Pediatr Urol*. (2014) 10(6):1261–6. doi: 10.1016/j.jpurol.2014.07.005

33. Van Herzeele C, Dhondt K, Roels SP, Raes A, Groen L-A, Hoebeke P, et al. Neuropsychological functioning related to specific characteristics of nocturnal enuresis. *J Pediatr Urol*. (2015) 11(4):208.e1–e6. doi: 10.1016/j.jpurol.2015.04.033

34. Kanaheswari Y, Poulsaeman V, Chandran V. Self-esteem in 6- to 16-year-olds with monosymptomatic nocturnal enuresis. *J Paediatr Child Health*. (2012) 48(10): E178–82. doi: 10.1111/j.1440-1754.2012.02577.x

35. Osman ZHH, Ali SAO, Kamel NMF. Impact of an educational program on mothers' knowledge, attitude and practice regarding their children with nocturnal enuresis. *Int J Adv Res*. (2016) 4(6):771–82. doi: 10.21474/IJAR01/723

36. Elbahnasawy TH, Elnagar AM. Psychological impact of nocturnal enuresis on self-esteem of school children. *Am J Nurs Res*. (2015) 3(1):14–20. doi: 10.12691/ajnr-3-1-4

37. Roccella M, Smirni D, Smirni P, Precenzano F, Operto FF, Lanzara V, et al. Parental stress and parental ratings of behavioral problems of enuretic children. *Front Neurol*. (2019) 10:1054. doi: 10.3389/fneur.2019.01054

38. Schlomer B, Rodriguez E, Weiss D, Copp H. Parental beliefs about nocturnal enuresis causes, treatments, and the need to seek professional medical care. *J Pediatr Urol*. (2013) 9(6):1043–8. doi: 10.1016/j.jpurol.2013.02.013

39. Sá CA, De Souza SAM, Villela MCBVA, Souza VDM, De Souza MHDF, De Figueiredo AA, et al. Psychological intervention with parents improves treatment results and reduces punishment in children with enuresis: a randomized clinical trial. *J Urol*. (2021) 205(2):570–6. doi: 10.1097/JU.0000000000001351

40. Obe RL, Stillman-Lowe C. Health education. *Br Dent J*. (2024) 236(3):181–5. doi: 10.1038/s41415-024-7052-1

41. Urlings J, Sezer S, Ter Laan M, Bartels R, Maal T, Boogaarts J, et al. The role and effectiveness of augmented reality in patient education: a systematic review of the literature. *Patient Educ Couns*. (2022) 105(7):1917–27. doi: 10.1016/j.pec.2022.03.005

42. Wojtara MS. Use of social Media for patient education in dermatology: narrative review. *JMIR Dermatol*. (2023) 6:e42609. doi: 10.2196/42609

43. Martínez-Miranda P, Jiménez-Rejano JJ, Rosales-Tristancho A, Casuso-Holgado MJ. Comparative effect of different patient education modalities on quality of life in breast cancer survivors: a systematic review and network meta-analysis. *Eur J Oncol Nurs*. (2023) 67:102411. doi: 10.1016/j.ejon.2023.102411

44. Herval ÁM, Oliveira DPD, Gomes VE, Vargas AMD. Health education strategies targeting maternal and child health: a scoping review of educational methodologies. *Medicine (Baltimore)*. (2019) 98(26):e16174. doi: 10.1097/MD.0000000000016174

45. Nishizaki N, Hirano D, Oishi K, Shimizu T. YouTube videos in Japanese as a source of information on nocturnal enuresis: a content-quality and reliability analysis. *Pediatr Int*. (2022) 64(1):e15049. doi: 10.1111/ped.15049

46. Toprak T, Tokat E. A quality analysis of nocturnal enuresis videos on YouTube. *J Pediatr Urol*. (2021) 17(4):449.e1–e6. doi: 10.1016/j.jpurol.2021.03.014

47. Bruno RR, Wolff G, Wernly B, Masyuk M, Piayda K, Leaver S, et al. Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective. *Crit Care*. (2022) 26(1):326. doi: 10.1186/s13054-022-04202-x

48. Addab S, Hamdy R, Thorstad K, Le May S, Tsimicalis A. Use of virtual reality in managing paediatric procedural pain and anxiety: an integrative literature review. *J Clin Nurs*. (2022) 31(21–22):3032–59. doi: 10.1111/jocn.16217

49. Chen F-Q, Leng Y-F, Ge J-F, Wang D-W, Li C, Chen B, et al. Effectiveness of virtual reality in nursing education: meta-analysis. *J Med Internet Res*. (2020) 22(9): e18290. doi: 10.2196/18290

50. Hermans ANL, Betz K, Verhaert DVM, Den Uijl DW, Clerx K, Debie L, et al. 360° Virtual reality to improve patient education and reduce anxiety towards atrial fibrillation ablation. *EP Eur*. (2023) 25(3):855–62. doi: 10.1093/europace/euac246

51. Siroosbakht S, Rezakhaniha B. Is renal bladder ultrasound necessary in monosymptomatic primary nocturnal enuresis? *A Case Control Study J Compr Pediatr*. (2018) 9(4):e69006. doi: 10.5812/compreped.69006

52. Netto JMB, Rondon AV, Lima GRMD, Zerati Filho M, Schneider-Monteiro ED, Molina CAF, et al. Brazilian consensus in enuresis-recomendations for clinical practice. *Int Braz J Urol*. (2019) 45(5):889–900. doi: 10.1590/s1677-5538.ibju.2019.0080

53. Rezakhaniha S, Rezakhaniha B, Siroosbakht S. Limited caffeine consumption as first-line treatment in managing primary monosymptomatic enuresis in children: how effective is it? A randomised clinical trial *BMJ Paediatr Open*. (2023) 7(1):e001899. doi: 10.1136/bmjpo-2023-001899

54. Gabr AA, Gad MA, Shalaby A. Quality of life in children with pseudoincontinence after implementing a bowel management program in Egypt. *J Pediatr Surg*. (2020) 55(2):261–4. doi: 10.1016/j.jpedsurg.2019.10.048

55. Palma PL, Marzuillo P, Di Sessa A, Guarino S, Capalbo D, Marrapodi MM, et al. From clinical scenarios to the management of lower urinary tract symptoms in children: a focus for the general pediatrician. *Healthcare*. (2023) 11(9):1285. doi: 10.3390/healthcare11091285

56. Nevéus T, Fonseca E, Franco I, Kawauchi A, Kovacevic L, Nieuwhof-Leppink A, et al. Management and treatment of nocturnal enuresis-an updated standardization document from the international children's continence society. *J Pediatr Urol*. (2020) 16(1):10–9. doi: 10.1016/j.jpurol.2019.12.020

57. Sajith S, Patnaik SK, Kanitkar M. Comparison of a voiding diary with clinical management tool as an outpatient screening tool for childhood functional voiding disorders. *Indian Pediatr*. (2021) 58(12):1147–50. doi: 10.1007/s13312-021-2397-x

58. Kwak KW, Park KH. Clinical inconsistency of lower urinary tract symptoms between questionnaire and bladder diary in children with nocturnal enuresis. *J Urol*. (2008) 180(3):1085–90. doi: 10.1016/j.juro.2008.05.053

59. Jørgensen CS, Kamperis K, Walle JV, Rittig S, Raes A, Dossche L. The efficacy of standard urotherapy in the treatment of nocturnal enuresis in children: a systematic review. *J Pediatr Urol*. (2023) 19(2):163–72. doi: 10.1016/j.jpurol.2022.12.011

60. Chan IH, Wong KK. Common urological problems in children: primary nocturnal enuresis. *Hong Kong Med J*. (2019) 25(4):305–11. doi: 10.12809/hkmj197916

61. Larsson J, Borgström M, Karanikas B, Nevéus T. Can enuresis alarm therapy be managed by the families without the support of a nurse? A prospective study of a real-world sample. *Acta Paediatr*. (2023) 112(3):537–42. doi: 10.1111/apa.16634

62. Önol FF, Guzel R, Tahra A, Kaya C, Boylu U. Comparison of long-term efficacy of desmopressin lyophilisate and enuretic alarm for monosymptomatic enuresis and assessment of predictive factors for success: a randomized prospective trial. *J Urol*. (2015) 193(2):655–61. doi: 10.1016/j.juro.2014.08.088

63. Keten T, Aslan Y, Balci M, Erkan A, Senel C, Oguz U, et al. Comparison of the efficacy of desmopressin fast-melting formulation and enuretic alarm in the treatment of monosymptomatic nocturnal enuresis. *J Pediatr Urol*. (2020) 16(5):645.e1–e7. doi: 10.1016/j.jpurol.2020.07.018

64. Alloussi SH, Mürtz G, Lang C, Madersbacher H, Strugala G, Seibold J, et al. Desmopressin treatment regimens in monosymptomatic and nonmonosymptomatic enuresis: a review from a clinical perspective. *J Pediatr Urol*. (2011) 7(1):10–20. doi: 10.1016/j.jpurol.2010.04.014

65. Radojicic Z, Milivojevic S, Koricanac I, Lazovic JM, Laketic D, Radojicic O, et al. Low compliance contribute to insufficient desmopressin response of primary monosymptomatic nocturnal enuresis and the role of voiding school. *BMC Pediatr*. (2021) 21(1):244. doi: 10.1186/s12887-021-02714-z

66. Gasthuys E, Dossche L, Michelet R, Nørgaard JP, Devreese M, Croubels S, et al. Pediatric pharmacology of desmopressin in children with enuresis: a comprehensive review. *Pediatr Drugs*. (2020) 22(4):369–83. doi: 10.1007/s40272-020-00401-7

67. Robson WLM, Leung AKC, Norgaard JP. The comparative safety of oral versus intranasal desmopressin for the treatment of children with nocturnal enuresis. *J Urol*. (2007) 178(1):24–30. doi: 10.1016/j.juro.2007.03.015

68. Gözüküçük A, Kılıç M, Çakıroğlu B. Desmopressin versus desmopressin +oxybutynin in the treatment of children with nocturnal enuresis. *J Pediatr Urol*. (2021) 17(4):451.e1–e6. doi: 10.1016/j.jpurol.2021.04.001

69. Issi Y, Biçakci U. Does desmopressin withdrawal strategy affect relapse rates in monosymptomatic enuresis treatment? *Eur J Pediatr*. (2021) 180(5):1453–7. doi: 10.1007/s00431-020-03918-8

70. Thiagamoorthy G, Cardozo L, Robinson D. Current and future pharmacotherapy for treating overactive bladder. *Expert Opin Pharmacother*. (2016) 17(10):1317–25. doi: 10.1080/14656566.2016.1186645

71. Ghasemi K, Esteghamati M, Mohammadzadeh M, Zare S. Desmopressin versus oxybutynin for nocturnal enuresis in children in bandar abbas: a randomized clinical trial. *Electron Physician*. (2016) 8(3):2187–93. doi: 10.19082/2187

72. Chua ME, Silangcruz JMA, Chang S-J, Yang SS. Immediate 1-month efficacy of desmopressin and anticholinergic combination therapy versus desmopressin monotherapy in the treatment of pediatric enuresis: a meta-analysis. *J Pediatr Urol*. (2016) 12(3):156.e1–e9. doi: 10.1016/j.jpurol.2015.12.011

73. Meyer SIR, Jørgensen CS, Kamperis K, Andersen RF, Pedersen MJ, Faerch M, et al. Efficacy and safety of multimodal treatment in nocturnal enuresis - A retrospective cohort study. *J Pediatr Urol*. (2021) 17(4):447.e1–e7. doi: 10.1016/j.jpurol.2021.03.005

74. Esteghamati M, Mousavi SE, Zoghi G. Desmopressin plus tolterodine vs desmopressin plus indomethacin for refractory pediatric enuresis: an open-label randomized controlled trial. *Indian Pediatr*. (2023) 60(6):447–51. doi: 10.1007/s13312-023-2906-1

75. Dell'Osso B, Palazzo MC, Oldani L, Altamura AC. The noradrenergic action in antidepressant treatments: pharmacological and clinical aspects. *CNS Neurosci Ther*. (2011) 17(6):723–32. doi: 10.1111/j.1755-5949.2010.00217.x

76. Mello MF, Locali RF, Araujo RM, Reis JN, Saiovici S, Mello LF, et al. A prospective and randomized study comparing the use of alarms, desmopressin and imipramine in the treatment of monosymptomatic nocturnal enuresis. *J Pediatr Urol*. (2023) 19(3):241–6. doi: 10.1016/j.jpurol.2023.01.004

77. Wang T-M, Yang SS-D, Tsai J-D, Yu M-C, Chiou Y-H, Chen K-L, et al. Management of nocturnal enuresis in Taiwan: consensus statements of the Taiwan enuresis expert committee. *J Formos Med Assoc*. (2019) 118(6):965–72. doi: 10.1016/j.jfma.2018.04.014

78. Wenguang W, Huifang Z, Guiwen W, Zhentie Y, Rui H. Effect of traditional Chinese medicine zhiyi decoction and acupuncture on Serum ADH & inflammatory factors in patients of enuresis. *Pak J Pharm Sci*. (2019) 32(1(Special)):465–9.

79. Ma Y, Liu X, Shen Y. Effect of traditional Chinese and western medicine on nocturnal enuresis in children and indicators of treatment success: randomized controlled trial. *Pediatr Int*. (2017) 59(11):1183–8. doi: 10.1111/ped.13417

80. Li J, Hui X, Yao L, Shi A, Yan P, Yao Y, et al. The relationship of publication language, study population, risk of bias, and treatment effects in acupuncture related systematic reviews: a meta-epidemiologic study. *BMC Med Res Methodol*. (2023) 23(1):96. doi: 10.1186/s12874-023-01904-w

81. Kannan P, Bello UM. The efficacy of different forms of acupuncture for the treatment of nocturnal enuresis in children: a systematic review and meta-analysis. *Explore*. (2022) 18(4):488–97. doi: 10.1016/j.explore.2021.11.008

82. Radvanska E, Kamperis K, Kleif A, Kovács L, Rittig S. Effect of Laser acupuncture for monosymptomatic nocturnal enuresis on bladder reservoir function and nocturnal urine output. *J Urol*. (2011) 185(5):1857–62. doi: 10.1016/j.juro.2010.12.068

83. Karaman MI, Koca O, Küçük EV, Öztürk M, Güneş M, Kaya C. Laser acupuncture therapy for primary monosymptomatic nocturnal enuresis. *J Urol*. (2011) 185(5):1852–6. doi: 10.1016/j.juro.2010.12.071

84. Radmayr C, Schlager A, Studen M, Bartsch G. Prospective randomized trial using Laser acupuncture versus desmopressin in the treatment of nocturnal enuresis. *Eur Urol*. (2001) 40(2):201–5. doi: 10.1159/000049773

85. Alsharnoubi J, Sabbour AA, Shoukry AI, Abdelazeem AM. Nocturnal enuresis in children between Laser acupuncture and medical treatment: a comparative study. *Laser Med Sci*. (2017) 32(1):95–9. doi: 10.1007/s10103-016-2090-9

86. Liu M, Li Y, Xian J, Yang W, Gao Q, Yu J. Pediatric tuina (massage) for primary monosymptomatic nocturnal enuresis: a protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. (2020) 99(51):e23738. doi: 10.1097/MD.0000000000023738

87. Van Poecke AJ, Cunliffe C. Chiropractic treatment for primary nocturnal enuresis: a case series of 33 consecutive patients. *J Manipulative Physiol Ther*. (2009) 32(8):675–81. doi: 10.1016/j.jmpt.2009.08.019

88. Kuru K, Ansell D, Jones M, De Goede C, Leather P. Feasibility study of intelligent autonomous determination of the bladder voiding need to treat bedwetting using ultrasound and smartphone ML techniques: intelligent autonomous treatment of bedwetting. *Med Biol Eng Comput*. (2019) 57(5):1079–97. doi: 10.1007/s11517-018-1942-9

89. Kuru K, Ansell D, Jones M, Watkinson BJ, Caswell N, Leather P, et al. Intelligent autonomous treatment of bedwetting using non-invasive wearable advanced mechatronics systems and MEMS sensors: intelligent autonomous bladder monitoring to treat NE. *Med Biol Eng Comput*. (2020) 58(5):943–65. doi: 10.1007/s11517-019-02091-x

90. Caswell N, Kuru K, Ansell D, Jones MJ, Watkinson BJ, Leather P, et al. Patient engagement in medical device design: refining the essential attributes of a wearable, Pre-void, ultrasound alarm for nocturnal enuresis. *Pharmaceut Med*. (2020) 34(1):39–48. doi: 10.1007/s40290-019-00324-w

91. Kuru K, Ansell D, Hughes D, Watkinson BJ, Gaudenzi F, Jones M, et al. Treatment of nocturnal enuresis using miniaturised smart mechatronics with artificial intelligence. *IEEE J Transl Eng Health Med*. (2024) 12:204–14. doi: 10.1109/JTEHM.2023.3336889

frontiers | Frontiers in Pediatrics

TYPE Original Research
PUBLISHED 14 May 2025
DOI 10.3389/fped.2025.1554921

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire, United Kingdom

REVIEWED BY
Arturo Solis-Moya,
Dr. Carlos Sáenz Herrera National Children's Hospital, Costa Rica
Stephanie Bui,
Université de Bordeaux, France

*CORRESPONDENCE
Hengyu Zhou
zhouhengyu@cqmu.edu.cn

RECEIVED 03 January 2025
ACCEPTED 14 April 2025
PUBLISHED 14 May 2025

CITATION
Fan D, Shen A, Wang Q, Peng L, Xia R and Zhou H (2025) Home spirometry telemonitoring in pediatric patients with asthma: a mixed study.
Front. Pediatr. 13:1554921.
doi: 10.3389/fped.2025.1554921

COPYRIGHT
© 2025 Fan, Shen, Wang, Peng, Xia and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Home spirometry telemonitoring in pediatric patients with asthma: a mixed study

Dingrong Fan[1], Ao Shen[1], Qianqian Wang[2], Lu Peng[2], Ruixue Xia[2] and Hengyu Zhou[1]*

[1]School of Nursing, Chongqing Medical University, Chongqing, China, [2]Department of Pediatrics, the Second Affiliated Hospital of Chongqing Medical University, Chongqing, China

**Background:** To evaluate the feasibility and practicality of home spirometry telemonitoring for pediatric patients with asthma, including both motivators and barriers, as well as the requirements for effective implementation.
**Methods:** This single-arm, prospective study involved three phases: outpatient spirometry examination, home spirometry telemonitoring, and semi-structured interviews. A total of 110 children aged 5–12 years, who required spirometry monitoring at the pediatric outpatient clinic of the Second Affiliated Hospital of Chongqing Medical University, were enrolled. The PF286 (referred to as PF286), a home spirometry telemonitoring device was used for this study. Upon enrollment, each child initially underwent spirometry with a clinical-grade spirometer (Jaeger). Subsequently, they were tested using the PF286 under both supervised and unsupervised settings. To assess the consistency between PF286 and the clinical spirometer, we employed Pearson correlation coefficients. Children diagnosed with mild-to-moderate asthma, based on specified tests, participated in a four-week home spirometry telemonitoring program. After the telemonitoring period, semi-structured interviews were conducted with children, their guardians, and healthcare professionals to evaluate their experiences and identify the motivators and barriers in integrating home spirometry telemonitoring into clinical care. Spirometry data were collected using the PF286.
**Results:** The findings suggest that the PF286 is suitable for home spirometry telemonitoring in children, with unsupervised use yielding reliable data for clinical application. Semi-structured interviews with eight groups of children and their guardians, five nurses, and four physicians identified five key themes: "benefits of telemonitoring", "data accuracy and reliability", "barriers", "emotions", and " expectations".
**Conclusion:** The study concludes that home spirometry telemonitoring is feasible and acceptable for pediatric asthma management. However, several challenges, such as cost, insurance coverage, data security, health education, and healthcare workload, need to be addressed prior to its widespread implementation. Future research should focus on leveraging artificial intelligence for early disease detection, treatment guidance, and improving the quality of life for pediatric asthma patients.

KEYWORDS

**home spirometry, telemonitoring, pediatric, asthma, feasibility, practicality**

## 1 Introduction

Asthma is a highly prevalent chronic condition among children with global prevalence estimates projected to exceed 400 million by 2025, and affects approximately 10% of the pediatric population worldwide (1–3). Characterized by recurrent wheezing, airflow limitation, and obstructive ventilatory dysfunction, asthma is linked to significantly

functional impairments, reduced physical activity, school absenteeism, and psychological distress in affected children (4). Current clinical guidelines recommend spirometry for the diagnosis and assessment of asthma control in pediatric patients (5). However, outpatient spirometry faces challenges related to accessibility and continuity of care (6). The COVID-19 pandemic has accelerated the adoption of home-based spirometry telemonitoring, which has shown potential in enhancing self-management strategies for asthma among children (7–11).

The efficacy of home spirometry telemonitoring, despite its widespread adoption, remains a continues issue within the medical community (12). A recent comprehensive review revealed that only three out of ten devices from various manufacturers met the requisite technical standards considered acceptable (13). Consequently, any new portable spirometer must undergo rigorous validation prior to its implemented in clinical practice. Moreover, a consensus on the effectiveness of home spirometry telemonitoring has yet to be reached, with a lack of definitive studies demonstrating equivalence between unsupervised home telemonitoring and supervised clinical spirometry data (14).

Current research on home spirometry telemonitoring for asthmatic school-aged children has predominantly been conducted in Western countries, with a notable absence of studies from Eastern regions (15–17). Cultural differences between Eastern and Western parenting paradigms may significantly impact children's self-management abilities (18). Furthermore, there is a notable lack of research investigating how Chinese families and healthcare providers perceive and utilize this emerging technology in asthma management. This mixed-method study aims to provide new insights into the effects of home-based remote spirometry telemonitoring on pediatric asthma care. To achieve this goal, the concordance between the remote smart spirometer PF286 and clinical assessments was rigorously examined and validated. Additionally, the feasibility and adherence were evaluated to unsupervised home spirometry telemonitoring. We also conducted semi-structured interviews with Chinese families and healthcare professionals to understand their perspectives on this. This study offers compelling evidence that highlights the efficacy of home spirometry telemonitoring in enhancing self-management among Chinese children with asthma.

# 2 Methodology

## 2.1 Experiments

We conducted a single-arm, prospective mixed-method study to evaluate the YouBreathe home spirometry telemonitoring device, PF286, and its associated telemonitoring platform (hereinafter referred to as PF286), both provided by Zhejiang Yiliankang Medical Science and Technology Co., Ltd. We declare no conflicts of interest involving the device manufacturer, members of the ethics committee, guardians of the participants, or in relation to the disclosure of the findings. Ethical approval for this research was granted by Chongqing Medical University (Ethical Approval No. 2023095). Prior to enrollment, informed consent was obtained from all participating children through their guardians. This study was organized into three distinct phases, each aligned with specific research objectives.

## 2.2 Device

PF286 is a handheld home spirometer designed to measure commonly used spirometry parameters, including forced vital capacity (FVC), forced expiratory volume in one second (FEV1), percentage of FEV1/FVC in one second (FEV1/FVC), peak expiratory flow rate (PEF), and forced expiratory flow at 25%–75% of FVC (FEF25-75). Before using the device, the child's information is linked to both the public WeChat terminal and the platform terminal. The device initiates with an automatic zero calibration, and the children are guided to exhale into it following voice prompts. The built-in algorithm conducts intelligent quality control of the results and visual displays. The device uses color coding to indicate the severity of the measurements, with green indicating that no intervention is needed, while yellow or red indicates that the patient's PEF is below 80% of the predicted value, suggesting the need for medical consultation and treatment.

## 2.3 Phase 1: outpatient spirometry tests

### 2.3.1 Population

From September 2023 to August 2024, pediatric patients requiring spirometry assessments were referred to the pediatric outpatient department of the Second Affiliated Hospital of Chongqing Medical University for the diagnostic procedure.

### 2.3.2 Inclusion criteria

Children aged 5–12 who demonstrated the ability to complete acceptable spirometry during routine clinical visits.

### 2.3.3 Exclusion criteria

Patients were excluded from spirometry testing for the following contraindications: (1) Pneumothorax or alveolar abnormalities; (2) Significant cardiac arrhythmias or other medical conditions; (3) Perforated eardrums caused by otitis media; (4) Hemoptysis within the last month; (5) Active tuberculosis or ongoing anti-tuberculosis therapy; (6) Respiratory infections; (7) Recent chest, abdominal, or eye surgeries (within 1–3 months); (8) Seizures requiring pharmacological intervention; (9) Inguinal or umbilical hernias at risk of incarceration; (10) Inability to cooperate with spirometry testing (e.g., cognitive impairments).

### 2.3.4 Data collection

Upon enrollment, children were initially assessed using a large, clinical-grade spirometer (Jaeger) to establish a baseline. Following a mandatory 10-minute rest period, both the children and their guardians received standardized training on the proper operation of the PF286 instrument. Subsequently, under the direct

supervision of a qualified technician, the children completed the PF286 test. After another required 10-minute rest, they were then asked to complete the PF286 test unsupervised—either independently or with their guardians assistance. Throughout these process, measurement techniques and data quality control were rigorously maintained in accordance with current guidelines established by the European Respiratory Society (ERS) and American Thoracic Society (ATS) (19).

## 2.4 Phase 2: home spirometry telemonitoring

### 2.4.1 Participant recruitment

According to the standardized diagnosis and treatment recommendations for pediatric asthma in China and the Childhood Asthma Control Test (C-ACT) (20, 21), families were invited to participate in a 4-week home spirometry program following the diagnosis of mild to moderate asthma in their children after the initial examination.

### 2.4.2 Data collection

During weeks 1–2 after enrollment, children underwent telemonitoring with direct supervision by technicians through Tencent Conference. In weeks 3–4, the children were monitored unsupervised, either with their parents or independently. To ensure compliance, technicians reviewed daily test results at 21:00 and contacted children who missed measurements for three consecutive days to determine the reasons for non- adherence. Throughout the study period, children's asthma control was monitored using the C-ACT questionnaire to prevent adverse events. In the event of illness during the telemonitoring, this would be immediately suspended, and the study team assessed the severity of the adverse event and its potential correlation with remote home pulmonary function telemonitoring. Children who experienced injuries or complications resulting from the treatment would receive free care at the Second Affiliated Hospital of Chongqing Medical University, and any serious adverse events were reported to the Ethics Committee within 48 h.

## 2.5 Phase 3: semi-structured interviews

### 2.5.1 Participant recruitment

After the home spirometry telemonitoring, the child's guardians were contacted to arrange a follow-up appointment at our clinic. Additionally, children, their guardians, and healthcare professionals involved in the consultation were invited to participate in a semi-structured interview gather insights into their experiences and perceptions.

### 2.5.2 Interview framework

The outline for the interviews was initially crafted through a comprehensive literature review and the utilization of the STAR tool from the Behavioral Event Interviewing Method (BEIM) (22). The interview guide was subsequently refined through expert consultation and pilot testing with three families of children with asthma, as well as three healthcare professionals.

For the parent interviews, we focused on key topics, including the ability of home telemonitoring to facilitate earlier detection of changes in their child's lung function, the lifestyle implications of home monitoring compared to traditional hospital visits, and extent to which this approach has empowered them in managing their child's condition. Additional 15 questions were included in interviews.

The healthcare professional interviews focused on issues encountered while using the technology, their views on the effect of remote lung function monitoring on follow-up schedules and check-up frequency for pediatric respiratory patients, its potential to expedite the detection of changes in patients' conditions and assess treatment efficacy, and its potential to reduce the workload of hospitals and physicians. Another 16 questions were included in the interviews.

### 2.5.3 Data collection

By employing a phenomenological approach in the qualitative research, semi-structured in-depth interviews were conducted with family groups engaged in home monitoring and healthcare professionals involved in the consultations. Prior to the interviews, the interviewer introduced themselves and provided a comprehensive explanation of the purpose, methodology, components, and time required for participation. Participants were assured of confidentiality regarding their interview data, and informed consent was obtained prior to conducting face-to-face interviews, which were audio-recorded for accuracy. Each interview, lasting approximately 20–60 min, was conducted in a designated clinic room to ensure privacy and comfort. During the interview, the interviewer attentively listened, observeed, and meticulously recorded the respondent's nonverbal behaviors while probing deeper into specific issues as dictated by the context. After each interview was recorded and transcribed into a written format, the transcript was returned to the respondent for validation, ensuring data accuracy. The process of interviewing continued until data saturation was reached.

## 2.6 Statistical analysis

### 2.6.1 Quantitative analysis

The data analysis was conducted using SPSS 22.0, with descriptive statistics presented as mean and standard deviation ($x \pm s$). To assess changes within the same group over time, paired *t*-tests were employed for pre- and post-comparisons. The ICC consistency along with Bland-Altman plots were used to assess variable relationships. The coefficient of variation was used to evaluate the degree of indicator dispersion. Statistical significance was set at the conventional threshold of $P < 0.05$. Compliance was defined as the number of actual measurements divided by the number of requested measurements, multiplied by 100%.

### 2.6.2 Qualitative analysis

The qualitative analysis of the interview content was analyzed using Colaizzi's 7-step method (22) with Nvivo 11.0 for data

management and support analysis. Two researchers independently conducted the thematic coding. Subsequently, the emergent themes were discussed by the entire research team to achieve consensus.

# 3 Results

## 3.1 Baselinea patient characteristics

A total of 110 children were recruited during the outpatient pulmonary function test phase, 47.27% were female and 52.73% male, the average age was 7.21 ± 2.54 years, with a BMI of 15.56 ± 2.91 kg/m². 52.73% of families had an annual income <100,000 yuan. Parental education levels were predominantly associate degrees (39.09%) and bachelor's degrees (34.55%). 39 children were enrolled in the group of telemonitoring at home and their general information is as follows:the mean age was 7.21 years [standard deviation (SD) ± 1.44], with 25 males (64.1%) and 14 females (35.9%). The mean height was 126.04 cm (SD ± 10.49), and the mean weight was 27.22 kg (SD ± 8.45). Family income distribution showed that 27 families (69.23%) had an annual income of CNY 80,000–150,000, while 12 families (30.77%) earned CNY 150,000–300,000. All participants (39 cases, 100%) were assessed as having "well-controlled" disease status.

## 3.2 Consistency analysis between jaeger, outpatient PF286 supervised and outpatient PF286 unsupervised

The Intraclass Correlation Coefficient (ICC) consistency test was performed to assess the agreement among Jaeger, Outpatient PF286 supervised, and Outpatient PF286 unsupervised. The results demonstrated excellent consistency for FVC% prediction (ICC = 0.912, 95% CI: 0.882–0.936, $P < 0.001$), FEV1% prediction (ICC = 0.946, 95% CI: 0.927–0.961, $P < 0.001$), and PEF% prediction (ICC = 0.974, 95% CI: 0.964–0.981, $P < 0.001$) across all three groups. Detailed results are presented in Table 1.

## 3.3 Consistency analysis between jaeger, PF286 supervised home telemonitoring, PF286 unsupervised home telemonitoring

Consistency analysis of pulmonary function parameters among Jaeger, 14-day home-based PF286 supervised, and 14-day home-based PF286 unsupervised revealed strong agreement across all groups. The intraclass correlation coefficients (ICCs) were 0.947 (95% CI: 0.928–0.962, $P < 0.001$) for FVC%, 0.949 (95% CI: 0.931–0.963, $P < 0.001$) for FEV1%, and 0.967 (95% CI: 0.967–0.983, $P < 0.001$) for PEF%, indicating excellent inter-group consistency. Complete data are shown in Table 2.

## 3.4 Compliance of home spirometry telemonitoring

Participant compliance rates demonstrated a progressive weekly decline: 92.67% in Week 1 (monitoring attempts = 253), decreasing to 87.55% in Week 2 ($n = 239$), 84.25% in Week 3 ($n = 230$), and a significant reduction to 69.23% in Week 4 ($n = 189$). Supervised home monitoring exhibited the highest compliance rate (90.11%, $n = 492$), significantly outperforming unsupervised modalities (76.74%, $n = 419$; $\chi^2 = 34.2$, $P < 0.001$). The overall compliance rate across all monitoring types was 83.42%.

## 3.5 Semi-structured interviews

This qualitative study involved two participant groups: (1) Family Group: 8 families of asthmatic children, including 7 mothers (87.5%) and 1 father (12.5%) as interviewees. The patients were aged 5–10 years (median 6 years), with Childhood Asthma Control Test (C-ACT) scores ranging from 19 to 25 (mean 21.8 ± 2.1), all meeting the asthma control threshold (>19 points). (2) Healthcare Professional Group: 8 respiratory department staff, comprising 5 bachelor's degree holders (62.5%), 2 master's degree holders (25%), and 1 medical doctorate holder (12.5%). Their titles included 3 chief physicians, 1 associate chief

TABLE 1 Consistency analysis between jaeger, outpatient PF286 supervised and outpatient PF286 unsupervised.

| Parameters | Jaeger | Outpatient PF286 supervised | Outpatient PF286 unsupervised | ICC | 95%CI | *P*-value |
|---|---|---|---|---|---|---|
| FVC% | 105.07 ± 11.34 | 104.42 ± 12.24 | 104.01 ± 15 | 0.912 | 0.882–0.936 | <0.001 |
| FEV1% | 106.91 ± 11.56 | 105.85 ± 15.65 | 105.94 ± 15.83 | 0.946 | 0.927–0.961 | <0.001 |
| PEF% | 96.7 ± 15.05 | 95.75 ± 17.84 | 96.14 ± 18.17 | 0.974 | 0.964–0.981 | <0.001 |

TABLE 2 Consistency analysis between jaeger, PF286 supervised home telemonitoring, PF286 unsupervised home telemonitoring.

| Parameters | Jaeger | PF286 supervised home telemonitoring | PF286 unsupervised home telemonitoring | ICC | 95%CI | *P*-value |
|---|---|---|---|---|---|---|
| FVC% | 105.07 ± 11.34 | 103.8 ± 15.45 | 104.11 ± 15.28 | 0.947 | 0.928–0.962 | <0.001 |
| FEV1% | 106.91 ± 11.56 | 105.56 ± 15.8 | 106.11 ± 15.63 | 0.949 | 0.931–0.963 | <0.001 |
| PEF% | 96.70 ± 15.05 | 95.52 ± 17.92 | 95.25 ± 17.46 | 0.976 | 0.967–0.983 | <0.001 |

physician, 1 attending physician, and 4 nursing staff (registered nurses/charge nurses). Professional experience spanned 5–24 years (nursing staff averaged 7.2 years; physician team averaged 17.3 years), with ages ranging from 29 to 52 years, forming a multi-tiered team combining senior and junior members (37.5% ≤35 years old; 37.5% having ≥20 years of experience).These interviews yielded five key themes related to home spirometry telemonitoring were identified: 'Benefits of telemonitoring', 'Data Accuracy and Reliability', 'Barriers', 'Emotional Impact', and 'Future Expectations'.

### 3.5.1 Theme 1: benefits of monitoring

Home spirometry telemonitoring represents a transformative approach in the management of pediatric respiratory diseases, significantly enhancing the efficiency and convenience of healthcare delivery. It fosters closer collaboration and communication between pediatric patients and medical professionals, thereby positively impacting family health management and overall quality of life.

### 3.5.2 Theme 2: data accuracy and reliability

In the home environment, proficiently trained users can obtain test results that closely mirror clinical standards. Although occasional errors may occur due to improper handling, these anomalies do not substantially compromise the overall reliability of the data and can be addressed through standardized usage protocols and comprehensive training. Consequently, these minor discrepancies are unlikely to have a substantial impact on clinical assessments and subsequent treatment decisions.

### 3.5.3 Theme 3: barriers

Non-professionals may encounter challenges in comprehending the specific indicators associated with spirometry monitoring, which necessitates expert guidance and interpretation to assist parents and patients in accurately interpreting the data. Furthermore, continuous public education, enhanced policies and regulations, along with managerial support will be pivotal in advancing the development of remote pulmonary function monitoring technology.

### 3.5.4 Theme 4: emotions

The emotional responses of patients and their parents during the family monitoring program are varied; however, feelings of insecurity and concern persist, necessitating a gradual adaptation process for both parents and medical staff in clinical practice.

| Theme 1 | Sub-theme | Citation explanation |
|---|---|---|
| Benefits of monitoring | Convenience and Flexibility | N6: "This home monitoring method really saves us a ton of time and effort compared to regular hospital check-ups, plus it's super convenient for dealing with minor symptoms. I don't even have to go to the hospital anymore."! |
| | Timeliness and Continuity | N17: "Telemonitoring gives us a great chance to spot any changes in the patient's condition quickly. Parents can keep an eye on things from home, and if anything unusual pops up, the medical team can alert us right away. This way, we can tweak the treatment plan or set up any necessary medical help promptly to prevent the patient's condition from getting worse". |
| | Enhanced Disease Management Autonomy | N8: "After using the tool, I was able to make some independent calls about how my child is doing.". |
| | Avoiding Cross-infection | N5: "A few months back, there was a huge flu outbreak, so we kept an eye on ourselves at home to avoid spreading it and keep our kids safe". |

| Theme 2 | Sub-theme | Citation explanation |
|---|---|---|
| Data accuracy and reliability | Accuracy | N8: "Prior to measuring my child's cough with an instrument, I observed that the recorded value was slightly lower than usual; however, it indicated a green status, leading me to believe it still accurately represented my child's condition". |
| | Reliability | N4: "Before I got started, I was worried about the data quality and thought it wouldn't be fair to compare these numbers with clinical exam results. But looking at the monitoring curve, it turns out that most of the tests actually met the quality control standards". |

| Theme 3 | Sub-theme | Citation explanation |
|---|---|---|
| Barriers | Insufficient patient education | N12: "I'm still having a bit of trouble with these numbers, so I can only tell them apart by color. Red means there's an issue, and green means everything's good". |
| | User acceptance varies | N10: "This tech is still pretty new, and using it at home makes me feel a bit uneasy and unsure". |
| | Policy | N15: "The costs for getting this tool aren't covered by insurance, and we haven't accounted for the extra labor costs that came up while implementing this technology". |
| | Workload | N9: "Each measurement only takes a short amount of time, but it does add an extra layer of planning to family life, and kids don't always adjust to it easily". N6: "I conduct daily assessments to monitor the children's health status, most of this work is completed during breaks and after work, which adds a certain amount of workload to us". |
| | Information Security Management | N5: "I'm a bit worried about data security. Where's the data for kids being uploaded to? Who can access it? And will it be used for anything else?". |

| Theme 4 | Sub-theme | Citation explanation |
|---|---|---|
| Emotions | Positive | N14: "Right now, the relationship between doctors and patients is moving more towards a shared decision-making approach. As healthcare professionals, we need to understand that both the child and their parents are key partners in making decisions, and that family monitoring data should be included in our treatment plans". |
| | Anxiety | N13: "I'm honestly a little concerned that we depend too much on online data interpretation and overlook physical exams and important lab tests. What do we do if there's a disagreement and no clear laws or rules to follow?". |
| | Deficiency in self-assurance. | N12: I monitor my child's progress at home myself, and even though all the numbers look good and the doctor says not to worry, I still can't shake this feeling of insecurity. I'm just not confident in what I know and I'm scared of missing something about my child's condition". |
| | Equanimity | N6: "Well keep an eye on ourselves like the doctor said, and you can't go wrong by following their advice. I'll stick to what the doctor told me and reach out in the group if I run into any issues. No need to stress about it". |

| Theme 5 | Sub-theme | Citation explanation |
|---|---|---|
| Expectation | Optimism | "N14: I think with technology moving forward, it's becoming a trend, and honestly, it's smart to be ready for it ahead of time". |
| | Intelligentization | N7:"Even though I can use colors to get a general idea of a child's health, I'm really looking for more detailed insights. If AI could automatically analyze this data and turn it into easy-to-understand words or specific advice, along with tips on things like weather, diet, and exercise, that would be awesome!". |
| | Forecasting diseases | N16:"The establishment of disease prediction models derived from the analysis of these monitoring data would significantly enhance our physician's ability to make more accurate diagnoses and formulate tailored treatment plans". |
| | Dissemination of scientific knowledge | N8: "I really hope the system can give me personalized educational videos and articles that fit my child's needs. For instance, I'd love to know how to use storage containers or if these medications will impact my child's vaccinations, and so on. There are just so many resources and videos out there, and it takes me forever to find what I'm looking for. If the system could offer tailored recommendations, it would save me a ton of time and effort". |
| | Policy support | N15:"The government and healthcare organizations really need to step up their game by offering more technical support and financial backing, like including this in the health insurance system, setting standard prices, and improving related policies and regulations. That's the only way we can ensure that technology keeps getting better in the future". |

### 3.5.5 Theme 5: expectation

A consensus among the majority of medical professionals and family members of patients express a favorable outlook on remote pulmonary function monitoring, positing that with the ongoing advancements in technology, this approach is poised to become a crucial component in the management of pediatric respiratory diseases. Furthermore, the convergence of artificial intelligence and advanced big data analytics is seen as significant promise for enhancing quality control and analysis within home monitoring of childhood asthma; the integration of large datasets and sophisticated algorithms is expected to catalyze a transformation in disease management paradigms.

## 4 Discussion

The strength of this study lies in its mixed—methods design, which combines quantitative device validation with qualitative interviews. The study not only rigorously validated the consistency of the PF286 device with A clinical—grade spirometer (Jaeger), proving its reliability and effectiveness under both supervised and unsupervised conditions, filling a research gap in this region, it also offers practical suggestions for future promotion. Through semi—structured interviews, it thoroughly explored the experiences, perceptions, facilitators, and barriers of children, guardians, and healthcare professionals regarding remote monitoring technology. Additionally, it examined the acceptance and utilization of emerging technologies by Chinese families and healthcare providers, which is crucial for promoting suitable technologies in developing countries and informing relevant policies.

In this study, the consistency of the home spirometry device PF286 with clinical Jaeger spirometry was validated. There was strong agreement between the PF286 and clinical Jaeger results in children aged ≥5 years with successful spirometry (ICC > 0.9). Similar results were observed during home telemonitoring (ICC > 0.9). In the interviews, both parents and healthcare professionals expressed high acceptance of home spirometry. Parents widely emphasized that home telemonitoring significantly reduces the time costs associated with medical visits. Real-time data transmission facilitated by telemonitorings reshaping intervention models in asthma management, enhancing parental autonomy in managing their children's condition. Additionally, home telemonitoring, as a complementary approach to outpatient treatment, demonstrates unique value in preventing and controlling respiratory infectious diseases. This advantage has drawn significant parental attention post-COVID-19 and has become one of the key drivers for adopting home telemonitoring.so,home spirometry telemonitoring in pediatric patients with asthma is feasible and acceptable,this technology holds potential to improve access to specialized care for asthmatic childrenand alleviate the burden on healthcare systems (23).

Several key obstacles hinder the promotion of home spirometry telemonitoring technologies. These include the absence of a standardized pricing system, exclusion of related equipment from medical insurance, and a lack of legal and regulatory frameworks specifically tailored to telemedicine practices. Currently, there is no domestic or international industry standards for home

spirometry telemonitoring. Participant acceptance varied throughout the monitoring process, with 5 participants dropping out within the first week of home monitoring. The main reasons for dropout included travel, perceived inconvenience, irrelevance of monitoring, and absence of symptoms. In contrast to the independence of Western children, Asian schoolchildren rely heavily on parental supervision to manage their asthma. This suggests that researchers must understand the child's family living conditions when designing optimal monitoring programs. Additionally, researchers should provide consistent health education and scientific outreach to both parents and children, improving parental knowledge of disease management and alleviating concerns about data interpretation. In China, pediatricians face challenges such as shortages and heavy workloads. Therefore, administrators should support job creation, better time allocation, and performance incentives when implementing home spirometry telemonitoring.

It has been reported that anxiety, depression, and helplessness negatively impact parents' ability to effectively manage their child's illness (24). Mandizha J's qualitative survey revealed that participants experienced a wide range of emotions, including worry, uneasiness, indifference, and ambivalence during home spirometry (25). In our interview, uneasiness was the predominant emotion expressed by both the parents and healthcare professionals. This uneasiness among parents predominantly stemmed from an absence of self-assurance in their asthma management knowledge and a concomitant fear that incorrect decisions could exacerbate their child's health condition. Healthcare professionals' anxiety stemmed from uncertainties regarding the clinical application of this emerging technology. This uneasiness may cause them to adopt a conservative approach in promoting and implementing home spirometry telemonitoring. It is recommended that researchers pay close attention to the psychological burden experienced by participants during home spirometry monitoring. Emotional needs should be fully considered, and efforts should be made to help participants build confidence and alleviate uneasiness through regular interviews, psychological counseling, or by incorporating a counselor into the team.

In the current study, we observed a decreasing trend in adherence to home spirometry in an unsupervised environment. Despite implementing system push notifications, WeChat reminders, and telephone inquiries, we were unable to prevent the gradual decline in adherence. Previous studies have also reported that adherence to home spirometry monitoring tends to decrease over time (26, 27). Additionally, factors such as lack of guidance, feedback, a sense of responsibility, and boredom were identified as major reasons for this decline. It is recommended that future research should providing additional remote support or incorporating fun incentives from medical staff when using unsupervised home spirometry. Furthermore, developing passive collection methods for children's spirometry data may be more effective in improving adherence.

Currently, artificial intelligence (AI) shows great potential in interpreting home spirometry data (28, 29). In interviews, both the families of children and healthcare professionals expressed hope that AI could enhance data interpretation accuracy, enable automatic quality control of operations and data, and build predictive models based on big data to actively provide early warnings of the condition. It was also anticipated that AI algorithms could personalize treatment plans and lifestyle guidance by analyzing the child's historical data. Furthermore, healthcare professionals could proactively reach out to parents and children, thereby reducing the caregiving burden on families.

## 5 Conclusion

This study provides the systematic evaluation of home spirometry telemonitoring (using the PF286 device) for pediatric asthma management in China cultural context. By integrating a mixed-methods design that combines rigorous quantitative validation of device consistency with qualitative insights from semi-structured interviews, we demonstrate the feasibility and reliability of unsupervised home spirometry telemonitoring in Chinese children. Furthermore, we identify unique challenges in technology adoption, such as concerns over data security, health education gaps, and policy limitations, which are pivotal for optimizing localized implementation.

## 6 Limitations

This design presents certain limitations. The absence of a conventional monitoring group may hinder precise evaluation of the clinical advantages of home spirometry telemonitoring compared with standard management. Subsequent research will investigate how monitoring data can trigger clinical interventions, thereby validating the practical value of this technology.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Ethics statement

This study has been approved by the Ethics Committee of Chongqing Medical University.

## Author contributions

DF: Conceptualization, Data curation, Investigation, Writing – original draft. AS: Data curation, Writing – original draft. QW: Investigation, Software, Writing – original draft. LP: Investigation, Writing – original draft. RX: Writing – original draft, Formal analysis, Methodology. HZ: Methodology, Conceptualization, Project administration, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Future Medical Youth Innovation Team Support Program Project of Chongqing Medical University, (Grant No. W0008).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. To T, Stanojevic S, Moores G, Gershon AS, Bateman ED, Cruz AA, et al. Global asthma prevalence in adults: findings from the cross-sectional world health survey. *BMC Public Health*. (2012) 12:204. doi: 10.1186/1471-2458-12-204

2. British Thoracic S. *British Guideline on the Management of Asthma*. Scottish intercollegiate guidelines Network (2016). p. 58. Available at: https://www.sign.ac.uk/our-guidelines/british-guideline-on-the-management-of-asthma/ (Accessed July 24, 2019).

3. Global Initiative for A. *Global Strategy for Asthma Management and Prevention*. GINA guidelines (2018). Available at: https://ginasthma.org/2024-report/ (Accessed May 22, 2024).

4. Sangnimitchaikul W, Srisatidnarakul B, Ladores S. Perspectives on managing asthma and facilitators in asthma self-management among Thai school-age children: a qualitative study. *J Health Res*. (2021) 35(3):214–25. doi: 10.1108/JHR-09-2019-0207

5. Rajvanshi N, Kumar P, Goyal JP. Global initiative for asthma guidelines 2024: an update. *Indian Pediatr*. (2024) 61(8):781–86. doi: 10.1007/s13312-024-3260-7

6. Davies B, Kenia P, Nagakumar P, Gupta A. Paediatric and adolescent asthma: a narrative review of telemedicine and emerging technologies for the post-COVID-19 era. *Clin Exp Allergy*. (2021) 51(3):393–401. doi: 10.1111/cea.13836

7. Logie K, Welsh L, Ranganathan SC. Telehealth spirometry for children with cystic fibrosis. *Arch Dis Child*. (2020) 105(12):1203–05. doi: 10.1136/archdischild-2019-317147

8. Moor CC, Mostard RLM, Grutters JC, Bresser P, Aerts JGJV, Chavannes NH, et al. Home monitoring in patients with idiopathic pulmonary fibrosis. A randomized controlled trial. *Am J Respir Crit Care Med*. (2020) 202(3):393–401. doi: 10.1164/rccm.202002-0328OC

9. Kruizinga MD, Essers E, Stuurman FE, Yavuz Y, de Kam ML, Zhuparris A, et al. Clinical validation of digital biomarkers for paediatric patients with asthma and cystic fibrosis: potential for clinical trials and clinical care. *Eur Respir J*. (2022) 59(6):2100208. doi: 10.1183/13993003.00208-2021

10. Nichols A-L, Sonnappa-Naik M, Gardner L, Richardson C, Orr N, Jamalzadeh A, et al. COVID-19 and delivery of difficult asthma services. *Arch Dis Child*. (2022) 107(3):e15–e15. doi: 10.1136/archdischild-2021-322335

11. Beaufils F, Enaud R, Gallode F, Boucher G, Macey J, Berger P, et al. Adherence, reliability, and variability of home spirometry telemonitoring in cystic fibrosis. *Front Pediatr*. (2023) 11:1111088. doi: 10.3389/fped.2023.1111088

12. Richardson CH, Orr NJ, Ollosson SL, Irving SJ, Balfour-Lynn IM, Carr SB. Initiating home spirometry for children during the COVID-19 pandemic–a practical guide. *Paediatr Respir Rev*. (2022) 42:43–8. doi: 10.1016/j.prrv.2021.02.001

13. Wu Z, Huang R, Zhong L, Gao Y, Zheng J. Technical performance analysis of different types of spirometers. *BMC Pulm Med*. (2022) 22:1–7. doi: 10.1186/s12890-021-01752-8

14. Anand R, McLeese R, Busby J, Stewart J, Clarke M, Man WDC, et al. Unsupervised home spirometry versus supervised clinic spirometry for respiratory disease: a systematic methodology review and meta-analysis. *Eur Respir Rev*. (2023) 32(169):220248. doi: 10.1183/16000617.0248-2022

15. Horner SD, Brown A. Evaluating the effect of an asthma self-management intervention for rural families. *J Asthma*. (2014) 51(2):168–77. doi: 10.3109/02770903.2013.855785

16. Friend M, Morrison A. Interventions to improve asthma management of the school-age child. *Clin Pediatr (Phila)*. (2015) 54(6):534–42. doi: 10.1177/0009922814554500

17. Harris K, Kneale D, Lasserson TJ, McDonald VM, Grigg J, Thomas J. School-based self-management interventions for asthma in children and adolescents: a mixed methods systematic review. *Cochrane Database Syst Rev*. (2019) 1:CD011651. doi: 10.1002/14651858.CD011651.pub2

18. Jan R-H, Lee H-TS, Cheng S-C. Parents' views of self-management for children with moderate to severe persistent asthma. *Tzu Chi Med J*. (2014) 26(1):34–9. doi: 10.1016/j.tcmj.2013.09.011

19. Stanojevic S, Kaminsky DA, Miller MR, Thompson B, Aliverti A, Barjaktarevic I, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J*. (2022) 60(1):2101499. doi: 10.1183/13993003.01499-2021

20. Thomas M, Kay S, Pike J, Williams A, Rosenzweig JRC, Hillyer EV, et al. The asthma control TestTM (ACT) as a predictor of GINA guideline-defined asthma control: analysis of a multinational cross-sectional survey. *Prim Care Respir J*. (2009) 18(1):41–9. doi: 10.4104/pcrj.2009.00010

21. Jianguo H. Recommendations for diagnosis and management of bronchial asthma in children(2020). *Chin J Pediatr*. (2020) 9:708–17.

22. McClelland DC. Identifying competencies with behavioral-event interviews. *Psychol Sci*. (1998) 9(5):331–39. doi: 10.1111/1467-9280.00065

23. Wijsenbeek MS, Moor CC, Johannson KA, Jackson PD, Khor YH, Kondoh Y, et al. Home monitoring in interstitial lung diseases. *Lancet Respir Med*. (2023) 11(1):97–110. doi: 10.1016/S2213-2600(22)00228-4

24. Boat TF, Filigno S, Amin RS. Wellness for families of children with chronic health disorders. *JAMA Pediatr*. (2017) 171(9):825–26. doi: 10.1001/jamapediatrics.2017.1682

25. Mandizha J, Lanario JW, Duckworth A, Lines S, Paiva A, Elworthy V, et al. Patient perspectives on home-spirometry in interstitial lung disease: a qualitative co-designed study. *BMJ Open Respir Res*. (2023) 10(1):e001837. doi: 10.1136/bmjresp-2023-001837

26. Moor CC, Wapenaar M, Miedema JR, Geelhoed JJM, Chandoesing PP, Wijsenbeek MS. A home monitoring program including real-time wireless home spirometry in idiopathic pulmonary fibrosis: a pilot study on experiences and barriers. *Respir Res*. (2018) 19:1–5. doi: 10.1186/s12931-018-0810-3

27. Edwards C, Costello E, Cassidy N, Vick B, Russell A-M. Use of the patientMpower app with home-based spirometry to monitor the symptoms and impact of fibrotic lung conditions: longitudinal observational study. *JMIR Mhealth Uhealth*. (2020) 8(11):e16158. doi: 10.2196/16158

28. Sanchez-Morillo D, Fernandez-Granero MA, Leon-Jimenez A. Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: a systematic review. *Chron Respir Dis*. (2016) 13(3):264–83. doi: 10.1177/1479972316642365

29. Topalovic M, Das N, Burgel P-R, Daenen M, Derom E, Haenebalcke C, et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J*. (2019) 53(4):1801660. doi: 10.1183/13993003.01660-2018

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 23 July 2025
DOI 10.3389/fpubh.2025.1513922

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Ameer Ahmad,
Quaid-i-Azam Medical College, Pakistan
Lubaba Shahrin,
International Centre for Diarrhoeal Disease Research (ICDDR), Bangladesh

*CORRESPONDENCE
Tirualem Zeleke Yehuala
✉ sarazeleke3@gmail.com
Bewuketu Terefe
✉ bewuketuterefe@gmail.com

RECEIVED 19 October 2024
ACCEPTED 30 June 2025
PUBLISHED 23 July 2025
CORRECTED 24 July 2025

CITATION
Yehuala TZ, Baykemagn ND and Terefe B (2025) Machine learning algorithms to predict feeding practices during diarrheal disease and its determinants among under-five children in East Africa. *Front. Public Health* 13:1513922. doi: 10.3389/fpubh.2025.1513922

COPYRIGHT
© 2025 Yehuala, Baykemagn and Terefe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Machine learning algorithms to predict feeding practices during diarrheal disease and its determinants among under-five children in East Africa

Tirualem Zeleke Yehuala[1]*, Nebebe Demis Baykemagn[1] and Bewuketu Terefe[2]*

[1]Department Health informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia, [2]Department of Community Health Nursing, School of Nursing, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

**Background:** Diarrhea is the leading cause of childhood malnutrition. Although replacement, continued feeding, and increasing appropriate fluid at home during diarrhea episodes are the cornerstones of treatment packages, food and fluid restrictions are common during diarrheal illnesses in Africa. To fill the methodological and current evidence gaps, this study aimed to build models and predict determinants to increase feeding practices of children in East Africa during diarrheal outbreaks.

**Methods:** We used the most recent demographic and health survey (DHS) statistics from 12 East African nations collected between 2012 and 2023. The analyses included a total weighted sample of 20,059 children aged 5 years. Python software was utilized for data processing and machine learning model building. We employed four ML algorithms, such as Random Forest (RF), Decision Tree (DT), XGB (Extreme Gradient Boosting), and Logistic Regression (LR). In this work, we evaluated the predictive models' performance using performance assessment criteria such as accuracy, precision, recall, and the AUC curve.

**Results:** In this study, 20,059 children aged 5 years were used in the final analysis. Among the proposed machine learning models, random forest performed best overall in the proposed classifier, with an accuracy of 97.86%, precision of 98%, recall of 77%, F-measure of 86%, and AUC curve of 97%. The most significant determinants of increasing feeding practice were richest household, faculty delivery, use of modern contraception method, the number of children 3–5, women's employment status, maternal age is 25–34, having media exposure, and health-seeking decisions made by mothers were associated positively, whereas not using contraception, home delivery, the total number of children is large, and the sex of the household was male, which was associated negatively with feeding practice during diarrhea in East Africa.

**Conclusion:** Machine learning (ML) algorithms have provided valuable insights into the complex factors influencing feeding practices during diarrheal disease in under-five children in East Africa. During diarrhea, only 11 of the 100 children received acceptable child feeding practices. More than one-third of the patients received less than usual or nothing. Reducing diarrhea-related child mortality by improving diarrhea management practices is recommended, particularly focusing on the identified aspects.

KEYWORDS

**feeding practice, diarrhea, determinants, East Africa, machine learning model, prediction**

# Introduction

Diarrhea is defined as three or more loose or watery stools within 24 h or an increase in a person's daily stool fluidity, frequency, or volume compared to what is considered normal (1, 2). Diarrhea, like other infectious illnesses, can be caused by inadequate hygiene, contaminated food, and polluted drinking water passed from person to person (2).

Most recent WHO factsheet (2024) on diarrhea states that diarrhea is the third leading cause of death in children under the age of five (3).

Since 2018, there have been almost 1.7 billion instances of childhood diarrhea per year, and an estimated 5.3 million children under the age of five have died, largely from preventable causes, with nearly half of the deaths occurring within the first month of life (2, 4). Sub-Saharan Africa (SSA) and Southern Asia account for more than 80% of all under-five deaths (5). In Africa, diarrhea is the leading cause of illness among young children, accounting for approximately half of all fatalities in this age group (6).

If not treated properly, childhood diarrhea can result in reduced macronutrient and micronutrient intake, further episodes of infectious disease, malnutrition, impaired physical growth, and impaired cognitive development (7, 8). Optimal feeding practices (both liquid and solid foods) during childhood illnesses are among the most effective global strategies for the integrated management of childhood illnesses, emphasizing the need to increase fluid intake while maintaining feeding during convalescence (9, 10).

The existing literature on rehydration management for Acute Watery Diarrhea (AWD) emphasizes the importance of Oral Rehydration Therapy (ORT) as the primary treatment, particularly in resource-limited settings (11). Oral Rehydration Solution (ORS) is a mixture of electrolytes and sugar that is used to replace lost electrolytes such as potassium, sodium, chloride, and bicarbonate because of diarrhea (12). ORS is usually prescribed for children under five who present with some acute watery diarrhea with no or some dehydration (13). Studies have demonstrated that oral glucose-electrolyte solutions effectively reduce the need for intravenous therapy, especially in cholera patients. However, alternative rehydration methods, such as nasogastric (NG) rehydration, have been found to be equally effective for moderate-to-severe dehydration (14). Intravenous (IV) fluid therapy remains essential for severe dehydration, with the literature distinguishing between repletion and maintenance therapy (the latter aimed at sustaining fluid balance). Additionally, research has explored intraosseous (IO), intraperitoneal (IP), and subcutaneous rehydration, though further evidence is needed to validate their effectiveness (15). A parents and caregivers continue to feed their children through a diarrheal episode, the repercussions of the disease will be less severe and the child's death will be avoided (2, 16, 17).

Children's nutritional condition can occasionally deteriorate during or after illness if they do not receive extra food. Low appetite due to diarrhea in children can contribute to the continuation of the illness and stunting cycle (18). Regardless of the multifaceted benefits of continuous feeding of children with diarrheal disease, some traditional beliefs and practices, such as partial food restrictions, cessation of breast milk, giving foods in a specific composition and amount, and restriction of vegetables and fruits, continue to contribute to inappropriate feeding practices during diarrhea among children (19–21). Furthermore, previous research has found that occupation (17), wealth index (22), maternal age (17, 23), maternal education (22, 24), number of under-five children, and knowledge of child feeding practices (17, 25) are all related to optimal child feeding practices. Evidence suggests that only 35% of under-five children in the region receive sufficient fluid replacement during diarrheal episodes (26). Currently, there is a methodological gap in the previous literatures based on recent data from East African countries (27, 28). Classical statistical methods have limitations in capturing complex and nonlinear relationships between predictors and outcomes. Machine learning algorithms (ML) are typically designed to make accurate predictions by learning from data rather than making prior assumptions. By utilizing ML algorithms, this study addressed uncovered-hidden patterns and interactions within the data, leading to more accurate predictions and a great understanding of the important determinates feeding practice during diarrheal disease. The data-driven insights can inform evidence-based decision-making and resource allocation for treatment related service and policies.

# Methods

## Data sources and sampling procedures

In this study we analyzed the most recent Demographic and Health Survey (DHS) data from 12 East African nations (Burundi, Ethiopia, Comoros, Kenya, Madagascar, Malawi, Mozambique, Rwanda, Tanzania, Uganda, Zimbabwe, and Zambia) collected between 2012 and 2023 to determine the practice of receiving multiple micronutrient powders and the factors that influence it in East Africa. These current datasets were combined to investigate the prevalence and associated determinants of various solid and liquid food feeding practices during diarrhea among children under 5 years of age in Eastern Africa.

The DHS surveys were collected over a 5-year cycle in low- and middle-income countries using structured, pretested, and validated questionnaires. The DHS surveys used the same standard sampling, questionnaires, data collection, and coding procedures, allowing for multi-country analysis. A stratified two-stage cluster sampling was used for this study. At the first stage, a stratified samples of EAs were selected with probability proportional to size in each country of the most recent national census. At the second stage, after a complete household listing was conducted in each of the selected EAs, fixed number of households were selected by equal probability systematic sampling in the selected EAs in each country (29).

## Population, and eligibility criteria

Interviews were conducted with target demographics (women aged 15–49 and males aged 15–64) in selected houses. This study included all under five children who had experienced diarrhea in the 2 weeks preceding the most recent DHS in 12 East African nations. According to the DHS Statistics Guide, missing values and

10.3389/fpubh.2025.1513922

TABLE 1 Countries, sample size, and survey year of Demographic and Health Surveys included in the analysis for 12 East African countries.

| Country | Survey year | Sample size (weighted) | Frequency (weighted) |
|---|---|---|---|
| Burundi | 2016/17 | 2,861 | 14.26 |
| Ethiopia | 2016 | 1,209 | 6.03 |
| Kenya | 2022 | 2,380 | 11.86 |
| Comoros | 2012 | 480 | 2.39 |
| Madagascar | 2021 | 1,048 | 5.22 |
| Malawi | 2016/17 | 3,540 | 17.65 |
| Mozambique | 2023 | 1,126 | 5.62 |
| Rwanda | 2019/20 | 1,129 | 5.63 |
| Tanzania | 2022 | 1,109 | 5.53 |
| Uganda | 2016 | 2,784 | 13.88 |
| Zambia | 2018 | 1,389 | 6.92 |
| Zimbabwe | 2015 | 1,005 | 5.01 |

"don't know" are omitted from the numerator (assumes did not consume or receive). Any missing data for any outcome variable was managed using various missing data management procedures in accordance with the instructions of the DHS statistics guide (30). A total of 20,059 under five children were involved in the study (Table 1).

## Study design and study period

This study adopted a cross-sectional study design for further analysis of DHS, which was conducted from 2012 to 2023. In this study, we design and develop the machine learning model. This involves selecting the appropriate algorithm, pre-processing the data, and engineering features. Finally, we develop a predictive model that predicts the feeding practice and identify it determinants to increase feeding practice during diarrheal disease among under five children.

## Data analysis procedure

Stata version 17 and Python software were used to extract, recode, and analyze the data. Weighting was used throughout the study to ensure representativeness and non-response rates as well as to obtain a suitable statistical estimate (robust standard error) (31). We employed Python software with the most crucial packages are Panda, Scikit Learn, Imblearn, Numpy, and Matplotlib for data preparation, model planning, model construction, model assessment, model deployment, and rule generation in this study. In conclusion, we create a predictive model that predicts the determinants to increase the feeding practice during diarrheal disease among under five children.

## Variables of the study

### The outcome variable

The outcome variable of this study was feeding practice during diarrheal. Then, the outcome variable was categorized as "less than usual" (somewhat less, much less, or none) and "more than usual." The first category less than usual was assigned a value of "0," while the second category (more than usual) was assigned a value of "1." This classification and analysis were conducted according to the guidelines in the DHS Statistics Handbook (30, 32).

### Independent variables

Various maternal and child-related factors were included. This included maternal age, educational status, type of place of residence, marital status, household wealth index, current employment status, mass media exposure, ANC follow-up, place of delivery, PNC check-up, persons usually deciding on mothers' health seeking, number of health visits and total children born, under five children, contraceptive utilizations, age of the child, sex of the child, size at birth, twin status, birth order, sex of the household head, and countries.

## Ethical considerations and data set access

The study was conducted after obtaining a permission letter from www.dhsprogram.com on an online request to access East African DHS data after reviewing the submitted brief descriptions of the survey to the DHS program. The datasets were treated with the utmost confidence. This study was done based on secondary data from East Africa DHS. Issues related to informed consent, confidentiality, anonymity, and privacy of the study participants are already done ethically by the DHS office. We did not manipulate and apply the micro data other than in this study. There was no patient or public involvement in this study, and no ethical clearance was required.

## Data pre- processing

### Data cleaning

We employed data cleaning, which typically includes noise, imbalanced outcome variables, and missing values. Our dataset contains some features that have missing values for data that is categorical data type. We used the mode imputation method to fill in the missing values. In this study among 20,059 records had 114 missing records for critical features, comprising nearly 0.56%. Features like such as place of residence: 0.0061%, marital status: 0.003%, wealth index: 0.001% contraceptive utilizations: 0.02%, age of the child: 0.007, sex of the child: 0.0031, size at birth: 0.053, twin status: 0.046, birth order: 0.3578.

### Data transformation and data integration

In this study, we utilized data transformation, which involves converting the data from one format to another that is suitable for analysis, which involves transforming data types, scaling data,

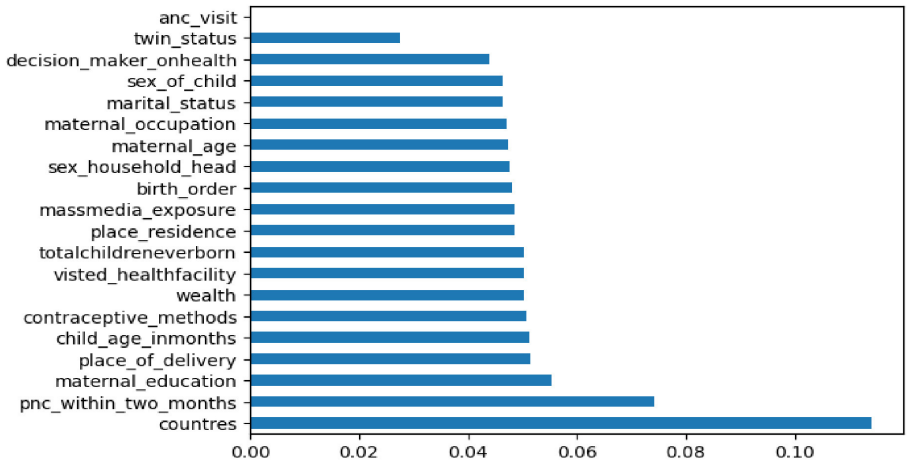


FIGURE 1
Important determinates selected by RFE.

and normalizing data. In this study, the data set has different data formats, like categorical and numeric features. One-hot encoding is a technique employed to convert categorical to numerical data. We integrate the data from 12 East African countries to ensure consistency and standardization; it is suitable for training machine learning models.

## Data discretization

In this study we used data discretization for converting continuous data (numerical values) into discrete categories or intervals (33). In this study, data discretization was used, which limits the impact of outliers and reduces noise by transforming continuous variables into categorical features according to DHS guidelines to make the data easier to understand and analyze.

## Feature selection

In this study we utilized feature selection techniques to removing irrelevant or redundant the number of features during developing a predictive model. The feature selection methods were used in data preprocessing to achieve efficient data reduction. In this study, we employed recursive feature elimination (RFE) to identify the most relevant variables for predicting increased feeding practices during diarrheal disease. Recursive Feature Elimination (RFE) is a feature selection technique that systematically identifies the most relevant predictors for a model by iteratively removing less important features. It begins by training a model on all available features and evaluating their importance using metrics like coefficients or feature importance scores. The least significant features are then eliminated, and the process is repeated until the desired number of predictor's remains. RFE enhances model performance by focusing on essential features, reduces overfitting by excluding irrelevant data, and improves interpretability by simplifying the model. In this study the dataset contains 41 features or variables. To select the most important features, we employed fast recursive feature elimination. This approach offers flexibility in controlling the number of features retained and can effectively handle correlated features. Maternal age, educational status, type of place of residence, marital status, household wealth index, current employment status, mass media exposure, ANC follow-up, place of delivery, PNC check-up, persons usually deciding on mothers' health seeking, number of health visits and total children born, under five children, contraceptive utilizations, age of the child, sex of the child, size at birth, twin status, birth order, sex of the household head, and countries were selected by (fast recursive elimination) RFE, and these determinants were used for model building as shown in Figure 1.

## Class balancing

To develop the model the training data was imbalance. In this study we employed Synthetic Minority Oversampling Technique (SMOTE) this technique produce additional observation. Prior to smote balancing, less than usual and same as usual was 14,282 (89%), while the prevalence more than the usual was 1,765 (11%). We obtained a balanced sample of children who less than the usual and same as usual with counts 14,282 and more than the usual with count 14,282.

## Data splitting

Data splitting is an important step in machine learning to ensure the proper evaluation of a model's performance. The goal is to divide the available dataset into subsets to train and test the model. This helps in assessing how well the model generalizes to unseen data and prevents over fitting to the training set. In this study, we utilized a simple holdout method like 80% (16,047 case) training data for model building and 20% (4,012 case) testing for model performance.

10.3389/fpubh.2025.1513922

TABLE 2 Maternal and child-related socio demographic characteristics of respondent's feeding practice during diarrhea among reproductive-age women in East Africa ($n$ = 20, 059): based on the recent East African countries DHS data.

| Variables | Categories | Less than usual | More than usual | *P*-values |
|---|---|---|---|---|
| Age in years | 15–24 | 6,227 (37.7%) | 793 (36.39%) | $X^2$ = 2.9e+04, *p*-value = 0.000 |
| | 25–34 | 7,933 (45.48%s) | 978 (44.82%) | |
| | 35–49 | 3,284 (28.83%) | 408 (18.73%) | |
| Residence | Urban | 3,857 (22.11%) | 544 (24.97%) | $X^2$ = 8.5e+03, *p*-value = 0.000 |
| | Rural | 13,587 (77.89%) | 1,637 (75.03%) | |
| Mothers' educational status | No education | 4,055 (23.25%) | 403 (18.49%) | $X^2$ = 86.4104, *p*-value = 0.000 |
| | Primary | 9,052 (51.89%) | 1,140 (18.49) | |
| | Secondary and higher | 4,337 (24.86%) | 636 (29.19%) | |
| Mothers employed | No | 6,328 (36.28%) | 740 (36.26%) | $X^2$ = 86.4104, *p*-value = 0.000 |
| | Yes | 11,116 (63.72%) | 1,389 (63.74%) | |
| Wealth index | Poor | 8,561 (49.08%) | 869 (39.88%) | $X^2$ = 2.2e+03, *p*-value = 0.000 |
| | Middle | 3,287 (18.84%) | 435 (19.96%) | |
| | Rich | 5,596 (32.08%) | 875 (40.16%) | |
| Mass media exposure | No | 9,505 (54.48%) | 1,112 (51.03%) | $X^2$ = 638.3930, *p*-value = 0.000 |
| | Yes | 7,941 (45.52%) | 1,067 (48.97%) | |
| ANC follow-up's | No | 585 (37.15%) | 300 (38%) | $X^2$ = 417.876, *p*-value = 0.000 |
| | Yes | 10,161 (62.85%) | 600, 13 (62%) | |
| Place of delivery | Home | 3,465 (19.86%) | 346 (15.88%) | $X^2$ = 1.5e+03, *p*-value = 0.000 |
| | Health facility | 13,979 (80.14%) | 1,833 (84.12%) | |
| Total number of children ever born | 0–2 | 9,101 (52.17%) | 1,184 (54.34%) | $X^2$ = 9.6e+03, *p*-value = 0.000 |
| | 3–6 and above | 8,343 (47.83%) | 995 (45.66%) | |
| Contraceptive method types | No methods | 9,268 (53.13%) | 1,028 (47.18%) | $X^2$ = 1.6e+03, *p*-value = 0.000 |
| | Traditional | 479 (2.75%) | 74 (3.40%) | |
| | Modern | 7,697 (44.12%) | 1,077 (49.43%) | |
| Number of health visits in the past 12 months | Once | 3,918 (21.62%) | 379 (18.22%) | $X^2$ = 3.7e+06, *p*-value = 0.000 |
| | More than one | 12,751 (78.38%) | 379 (18.22%) | |
| Marital status | Never married | 994 (5.70%) | 128 (8.87%) | $X^2$ = 1.1e+045, *p*-value = 0.000 |
| | Married | 11,621 (66.62%) | 1,476 (67.74%) | |
| | Divorced/widowed | 4,829 (27.68%) | 575 (26.39%) | |
| Size of the child | Small | 7,367 (47.22%) | 983 (45.11%) | $X^2$ = 4.3e+03, *p*-value = 0.000 |
| | Average | 6,845 (39.24%) | 816 (37.45%) | |
| | Large | 3,234 (18.54%) | 380 (17.44%) | |
| Sex of the child | Male | 9,101 (52.17%) | 1,184 (54.34%) | $X^2$ = 1.6e+541, *p*-value = 0.000 |
| | Female | 8,343 (47.83%) | 995 (45.66%) | |
| Twin status | Single | 16,978 (97.33%) | 2,118 (97.20%) | $X^2$ = 7.9e+03, *p*-value = 0.000 |
| | Multiple | 466 (2.67%) | 61 (2.80%) | |
| Sex of the household head | Male | 13,196 (76.65%) | 1,637 (75.13%) | $X^2$ = 5.4e+23, *p*-value = 0.000 |
| | Female | 4,248 (24.35%) | 542 (24.87%) | |

### Model selection

The predicted variable in this study was binary prediction, which refers to predicting one of two possible classes (e.g., yes/no, true/false, 1/0). In this study, we utilized four supervised machine learning models, such as random forest (RF), decision tree (DT), logistic regression (LR), and extreme gradient boosting (XGB). The algorithms were chosen in accordance with previous research that used machine-learning methods to classify tasks (34, 35).

### Model interpretation based on SHAP values

A feature's SHAP value indicates the extent to which its value influences the discrepancy between the average model output and the anticipated output of the model (36). In this study we used SHAP values for model interpretation, since features that push the prediction higher compared to the base value have positive SHAP values, while features that push the prediction lower have negative values.

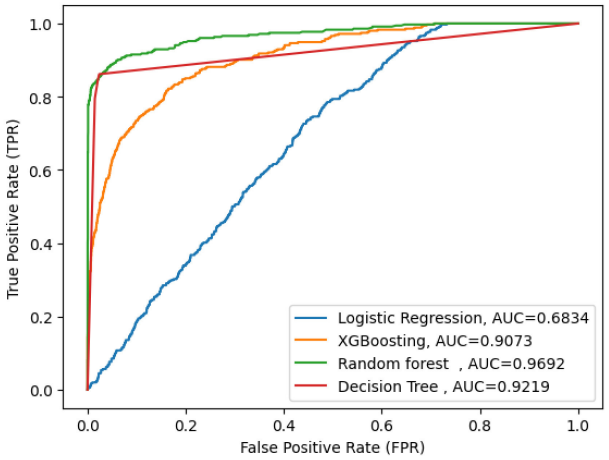


FIGURE 2
AUC curve for selected machine learning model.

## Results

### Socio demographic characteristics of the study participant

In this study, 20,059 children who experienced diarrhea and whose age was from 6 to 59 months old were enrolled in East African countries. Approximately half of them, 9,181 (45.77%), were between 25 and 34 years old within the reproductive age range. Regarding marital status, the majority of the mothers, 13,247 (66.04%), were married. With respect to place of residence types, 15,743 (78.48%), educational status 10,514 (52.42%), wealth index 4,928 (24.57%), place of delivery 16,256 (81.04%), and ANC follow-up 19,174 (95.59%) mothers were from rural, primary educational status, poorest households, institutional delivery, and had at least one ANC follow-up during their pregnancies. Similarly, about 10,655 (53.12%) and 12,948 (64.55%) women had no experience of at least one mass media exposure (listening to radio, watching television, reading magazines or newspapers) and were employed, respectively. About 15,770 (78.62%) women were unable to access healthcare services by themselves when they needed it. Children-related variables, such as weight, age, sex, and birth order, were 8,950 (48.16%), 4,940 (48.59%), 10,617 (52.93%), and 7,393 (36.86%) of children were average weighted, 12–35 months old, male, and 2nd or 3rd birth orders, respectively. Similarly, approximately 16,668 (83.10%) of fewer than three children were found in each household. The majority of 19,516 (97.29%) children were born single, as shown in Table 2.

### Prevalence of feeding practice during diarrhea among under five children in East Africa

In this study, a higher proportion of mothers (53.54) (95% CI, 52.86, 54.24) had provided the foods and liquids just the same as usual. In contrast, about 34.85 (34.19, 35.5) had practiced less than usual while their children had experienced diarrhea.

### Comparing Supervised ML Classifier

The goal of this study was to build a predictive model and identify determinants to increasing feeding practice during diarrhea disease among under five children. We utilized four supervised machine learning algorithms, such as random forest, decision tree (DT), logistic regression (LR), and XGB (Extreme Gradient Boosting). The experiments were done with the same testing parameters. In this study, the predictive performance was evaluated based on the five performance measures (accuracy, precision, F-measure, AUC curve, and recall). After comparing ML models, random forest emerged as the best predictive model, with an accuracy performance of 97.86%, precision of 98%, recall of 77%, F-measure of 86%, and AUC curve of 97%.

In addition, random forests had high specificity (87%) and sensitivity (77%). The true positive rate of the random forest was 96%, the false positive rate was 3.9%, and the AUC curve was high, 97%. Moreover, decision tree (DT) has an accuracy of 96.79%, precision of 83%, recall of 79%, F-measure of 81%, and an AUC curve of 92%. Among the proposed models, logistic regression was the least one with an accuracy of 88%, precision of 58%, F-measure of 19%, recall of 71%, and an AUC score of 68%, since the logistic regression model was highly mis_predicted, as shown in Figure 2 and Table 3.

To evaluate the uncertainty of our findings, we used a 95% confidence interval (CI), which indicates that we have a 95% chance that the true value falls within the computed range. In Table 3 shown for random forest model between 92.5 and 97.2% is the true accuracy with 95% confidence.

Additionally in this study we utilized a calibration plot to evaluate how well the predicted probabilities of a model align with actual outcomes. It's particularly useful for probabilistic classifiers championing random forest models to assess whether the predicted probabilities correspond to the true outcomes as shown in Figure 3.

TABLE 3 Performance comparison of machine learning models using accuracy, precision, recall, and F-measure.

| Machine learning models | Dataset | Accuracy | AUC | Precision | Recall (sensitivity) | F1-score | 95 % confidence interval |
|---|---|---|---|---|---|---|---|
| Decision tree | Balanced (%) | 96.79 | 92 | 83 | 79 | 81 | [0.839, 0.883] |
| Random forest classifier | Balanced (%) | **97.86** | **97** | **98** | **77** | **86** | [0.925, 0.972] |
| Extreme gradient boosting | Balanced (%) | 92.69 | 90.7 | 91 | 20 | 34 | [0.887, 0.925] |
| Logistic regression | Balanced (%) | 88 | 68 | 71 | 15 | 19 | [0.843, 0.895] |

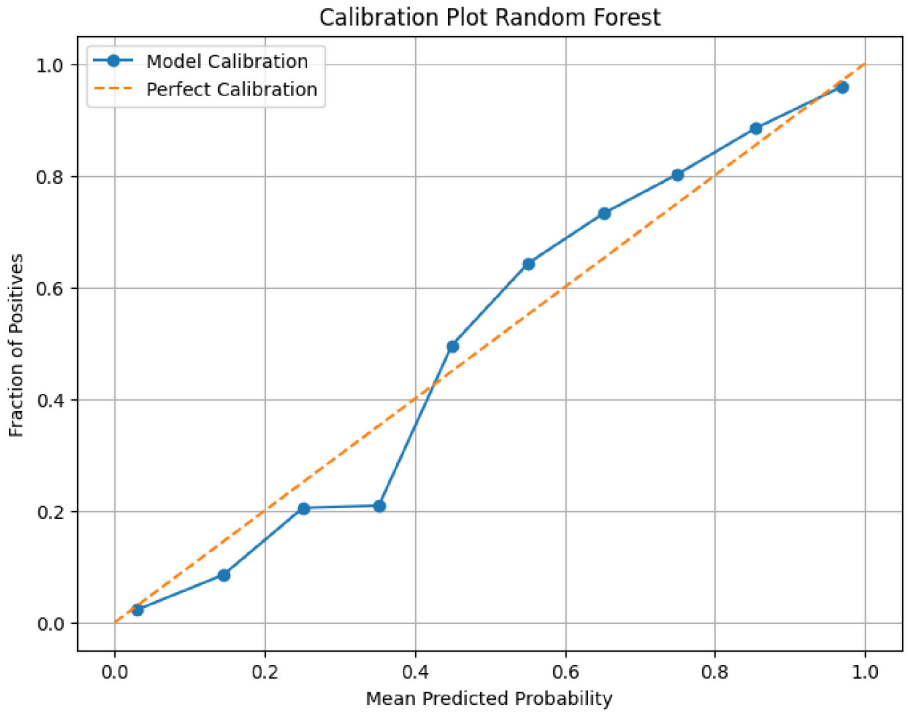


FIGURE 3
Calibration plot for the Random Forest model.

## Model interpretation

In this study we used SHAP values for model interpretation, since features that push the prediction higher compared to the base value have positive SHAP values, while features that push the prediction lower have negative values. Therefore, for the first observation, the combination of the positive contributions (in red) and the negative contributions (in blue) moves the expected value output to the final model output [$f_{(x)} = -2.183$], classified as a negative class less than usual. In this case, features like being 2nd or 3rd birth order, faculty delivery, being female, using modern contraception methods, the number of children 3–5, women's employment status, the wealth status being rich, maternal age being 25–34, having media exposure, and health seeking decisions made by mothers and other features impacting children feeding more than usual during diarrheal disease show in Figure 4.

However, having no PNC checkup within 2 months, delivery at home, a number of birth orders above 5, a mother is not educated, who lives in rural areas, no at least one ANC, and other features has an impact of each predictor (less than usual) on children feeding during diarrheal disease (Figure 5).

## Sensitivity analysis

In this study, we employed sensitivity analysis to understand which features have the greatest influence on the model's predictions and to assess its stability and robustness in the face of changes. Specifically, we utilized global sensitivity analysis as showed Figure 6, which involves evaluating the overall contribution of each feature to the model's predictions and examining how they interact with one another. Features toward the top are the most significant (i.e., the model is most sensitive to them), as Figure 6 illustrates. Red points that are primarily on the right show that the model's prediction is increased by high feature

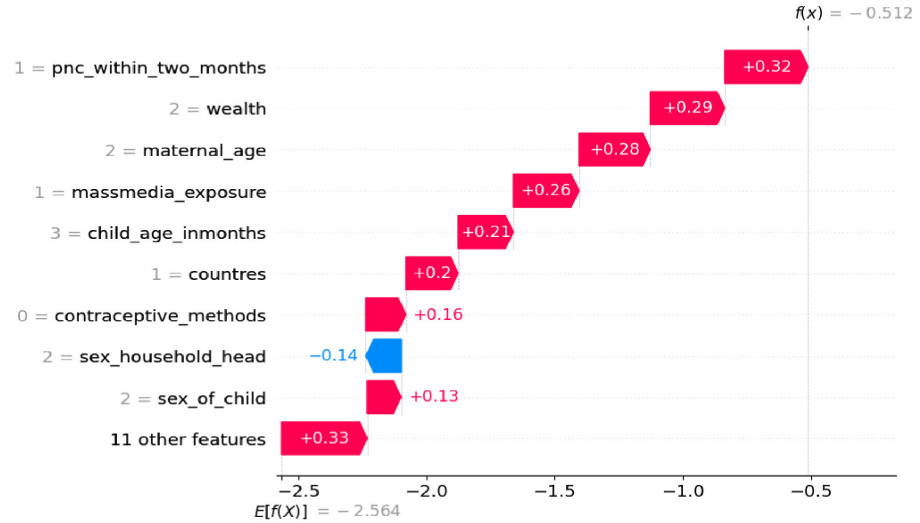


FIGURE 4
Waterfall plot displaying prediction of the zero index observation.

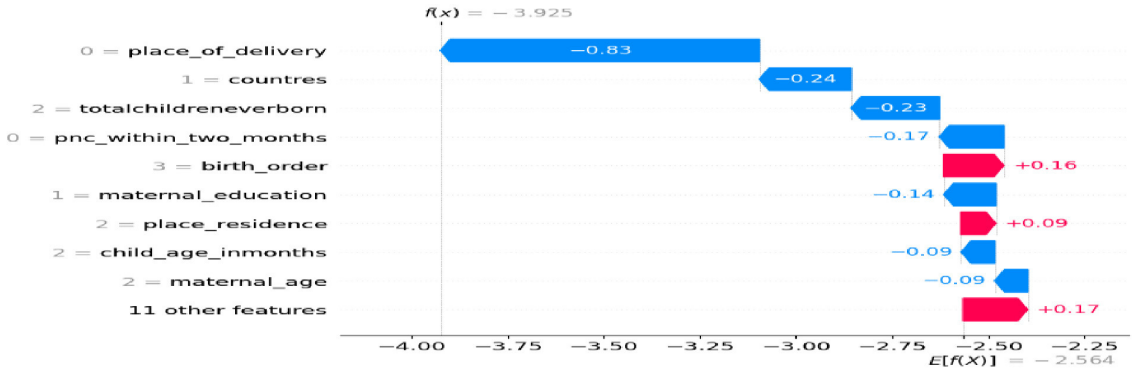


FIGURE 5
Waterfall plot displaying prediction of the 1st observation.

values. We developed the models on the ≤2020 dataset, and then temporally validated them on the ≥2021 dataset. Among the proposed algorithm random forests had high Accuracy with 90.39%, AUC 92%, precision 93.6% and recall 80.13% as shown in Table 4.

## Discussion

According to the World Health Organization's recommendations for continued feeding during diarrhea, as the nutritional complications of diarrhea have become more widely recognized (37–39) more attention has been focused on the appropriate dietary management of this illness (40, 41). To tailor different intervention measures to reduce diarrhea-related mortality among under-five children in East Africa. This study aimed to investigate the current magnitude of appropriate feeding practices and identify the determinants (associated factors) to increase feeding practice among under-five children with diarrheal disease. For this objective, four machine learning models were trained on balanced training data using train-test split. Accuracy, AUC score, precision, recall, and F1 score were used to compare the performance of classifier models. Random forest emerged as the best predictive model, with an accuracy performance of 97.86%, precision 98%, recall 77%, F-measure 86%, and AUC curve 97%. To date, no studies have been conducted utilizing ML algorithms to predict determinants of increasing feeding practice during diarrhea among under-five children. However, the study conducted in Ethiopia, titled Comparative Analysis of Machine Learning Algorithms for Predicting Diarrhea Among Children Under Five in Ethiopia: Evidence from the 2016 EDHS, found that the Random Forest algorithm performed the best, with an accuracy of 93.2%, sensitivity of 98.4%, specificity of 85.5%, and an AUC of 0.916 (42). This study discovered the prevalence of feeding

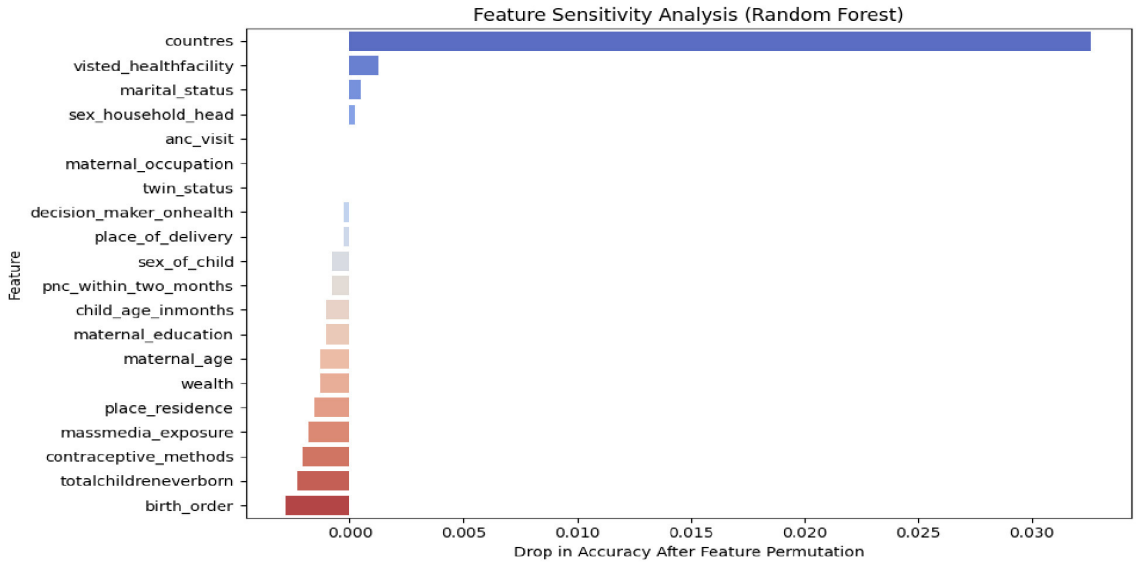


FIGURE 6
Sensitivity analysis with SHAP values.

TABLE 4 Sensitivity analysis of the proposed machine learning algorithms: internal validation on the 2020 dataset and temporal validation on the 2021 dataset.

| Machine learning models | Accuracy (%) | AUC (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| Decision Tree | 89.79 | 90 | 84 | 69 |
| Random forest classifier | **90.39** | **92** | **93.6** | **80.13** |
| Extreme gradient boosting | 90 | 89.9 | 90.6 | 78 |
| Logistic regression | 69.4 | 67 | 59 | 52 |

practice: only 11.60% of children under five with diarrhea received more appropriate food during the diarrheal episode than usual. On the other hand, ~88.4% of them received less than usual or did not take anything during their illness, and the rest are taken as usual. According to the data, a considerable percentage of children in Eastern Africa do not receive appropriate food and fluid during their diarrheal episodes, indicating that this issue is a high priority (23, 43).

In the feature selection, maternal age, children's age, maternal employment, maternal education, visit health facility, place of delivery, media exposure, wealth index, having at least one ANC, mother's health seeking for decision-making on children's health, and child's age were extremely associated with the outcome variable.

Employed women had a 16% of feeding their children more than normal, rather than less than usual, when all other characteristics remained constant. Similarly, when other variables were held constant, mothers from poorer, middle, richer, and richest households had a higher risk of feeding their children more than usual than women from the poorest household wealth index by 24, 45, 67, and 28%, respectively, rather than feeding less than usual. The factor linked with proper child feeding practices in this study was women's employment status, with those with current employment having a higher chance of having adequate child feeding practices than their counterparts. Studies from Nepal and SSA have found similar associations (23, 44).

This could be related to the positive impact of women's employment on household wealth status since working women typically have financial means that allow them to buy appropriate food and feed their children. Their social connections can help improve their knowledge and attitudes regarding eating practices.

Similarly, in earlier studies (22, 23, 45), women with a higher wealth index had a higher likelihood of having a child with acceptable or more than normal feeding practice compared to women with the lowest wealth index. Families with greater financial resources tend to have better access to healthcare, education, and nutritional knowledge, which can lead to appropriate feeding practices during diarrheal episodes. The study conducted sub-Saharan Africa found that children from wealthier households were more likely to receive adequate food and fluids during diarrhea compared to those from lower-income families (46). This is apparent because, in comparison to their counterparts, women with high incomes have no budget limits to purchase various dietary items and fluids for the treatment of diarrhea.

Holding all other variables constant, mothers with 12–35-month-old children and 36–59-month-old children had a higher

relative risk ratio tendency to feed their children more than the usual rather than less than the usual, as compared to 6–11-month-old children, respectively. Similarly, child age was reported as a predictor variable in research conducted in five Asian nations (47), Tanzania (48), and northern Ethiopia. Mothers may be uninformed or they may need to continue breastfeeding their children rather than feeding them more than normal. If these children are breastfed properly, they may have a lower risk of experiencing diarrhea. This may provide opportunities for health program planners to focus on the feeding of younger children. The problem of appropriate feeding of children aged 6–11 months is supported by findings from a previous study conducted in Ethiopia, which revealed a high prevalence of stunting (43%) among children aged 6–11 months, there is a need to place more emphasis on feeding of children aged 6–11 months (49). Additionally studies show that the prevalence of appropriate feeding practices during diarrheal illness among children aged 6–23 months is alarmingly low, particularly in Sub-Saharan Africa. Factors such as maternal education, media exposure, vaccination, and regional differences significantly influence these practices (17).

Holding all other variables constant, having at least one ANC visit during pregnancy and giving birth in a health facility increases the relative chance of feeding more than normal rather than feeding less than usual to 1.49 and 1.19 (increases by 49 and 19%, respectively). Similarly, when all other dependent variables were maintained constant, not using any type of contraception reduced the chance of feeding more than normal rather than feeding less than usual to 0.89 (an 11% reduction). Studies from Ethiopia and SSA have found a similar link between these and other maternal health services (10, 50, 51). Possible arguments include the fact that these services are an excellent point of contact for providing counseling on child feeding practices, in addition to maternity and childcare. Mothers who have ANC and deliver births in a health facility may obtain information, education, and counseling about appropriate infant feeding practices for normal child growth and during an illness, such as diarrhea, from health professionals. Similarly, mothers who use contraception will have a better understanding of feeding practices and fewer children. These interns allowed them to feed their children more than usual.

In this studies emphasize that appropriate practices, such as continuing breastfeeding, having ANC follow up, and offering nutrient-rich foods, but are often underutilized due to cultural beliefs, misinformation, and socioeconomic factors. Harmful practices like unclean food or fluids exacerbate the risk of malnutrition and prolonged illness. Public health strategies, such as educational programs for mothers or caregivers, promoting access to safe water and sanitation, and vaccination have shown promise in addressing these gaps. These strategies can significantly reduce the prevalence of diarrheal diseases and improve child health outcomes.

## Conclusions

In this study, the prevalence of adequate child-feeding practices was found to be quite low in East Africa. In this study, maternal employment status, antenatal care visits, place of delivery, contraceptive use, and child age were all independently related to optimal feeding practices during a diarrheal episode in children under the age of five in East Africa. Thus, intensified intervention programmers and stakeholders working on child health should concentrate on these determinants in order to minimize child death and morbidity and achieve the region's sustainable development goals. Particularly, supporting maternal-related health services and making use of opportunities will have a greater impact on increasing mothers' feeding practices for their infants.

### Strength and limitations

This study aimed to predict feeding practices during diarrheal diseases and identify key predictors. It also utilized the most recent DHS dataset from East African countries, which, due to its large sample size, improves the generalizability of the results. However, the study has certain limitations. Since the data were collected based on mothers' memories, there is a risk of recall bias, and some details may have been overlooked. Additionally, while the study employed mode imputation techniques to handle missing data, these methods have notable limitations, particularly when dealing with high frequencies of missing data. However, the number of missing values in this study was small.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: www.dhsprogram.com.

## Ethics statement

Procedures and questionnaires for standard DHS surveys have been reviewed and approved by the ICF Institutional Review Board (IRB). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because in this study we used secondary data, which were already collected from each East African as a national survey, ethical approval from the IRB is exempted or waived.

## Author contributions

TZY: Conceptualization, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. NB: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. BT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software,

Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare no financial support was received for the research and/or publication of this article.

## Acknowledgments

We would like to acknowledge the DHS program for providing permission for this study following research ethics.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

## Correction note

This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Olopha OO, Egbewale B. Awareness and knowledge of diarrhoeal home management among mothers of under-five in Ibadan, Nigeria. *Univ J Public Health.* (2017) 5:40–5. doi: 10.13189/ujph.2017.050106

2. World Health Organization. *Diarrhoeal Disease: Fact Sheets.* (2017). Available online at: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease#:~:text=Diarrhoeal%20disease%20is%20the%20secondand%20adequate%20sanitation%20and%20hygiene (Accessed October 11, 2024).

3. Rahman AMR, Islam MR, Bashar SJ, Al Fidah MF, Amin R, Rahman MM, et al. Trends in preventive practices against diarrhoeal disease among under-five children: experience from the largest diarrhoeal disease hospital in Bangladesh. *BMJ Paediatrics Open.* (2025) 9:e003259. doi: 10.1136/bmjpo-2024-003259

4. Okoro C, Ikediuwa U, Mgbudem F, Uwabunkonye B, Osondu B. Estimation of levels and trends of under-five mortality in sub-saharan africa: evidence from summary of birth histories of currently married women. *Asian J Prob Stat.* (2020) 7:17–27. doi: 10.9734/ajpas/2020/v7i230176

5. You D, Hug L, Ejdemyr S, Idele P, Hogan D, Mathers C, et al. Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. *Lancet.* (2015) 386:2275–86. doi: 10.1016/S0140-6736(15)00120-8

6. Fischer Walker CL, Perin J, Aryee MJ, Boschi-Pinto C, Black RE. Diarrhea incidence in low-and middle-income countries in 1990 and 2010: a systematic review. *BMC Public Health.* (2012) 12:1–7. doi: 10.1186/1471-2458-12-220

7. Troeger C, Colombara DV, Rao PC, Khalil IA, Brown A, Brewer TG, et al. Global disability-adjusted life-year estimates of long-term health burden and undernutrition attributable to diarrhoeal diseases in children younger than 5 years. *Lancet Global Health.* (2018) 6:e255–69. doi: 10.1016/S2214-109X(18)30045-7

8. Pantenburg B, Ochoa TJ, Ecker L, Ruiz J. Feeding of young children during diarrhea: caregivers' intended practices and perceptions. *Am J Trop Med Hyg.* (2014) 91:555. doi: 10.4269/ajtmh.13-0235

9. Liu L, Hill K, Oza S, Hogan D, Chu Y, Cousens S, et al. Levels and causes of mortality under age five years. *Reprod Matern Newborn Child Health.* (2016) 11:71. doi: 10.1596/978-1-4648-0348-2_ch4

10. Degefa N, Tadesse H, Aga F, Yeheyis T. Sick child feeding practice and associated factors among mothers of children less than 24 months old, in Burayu Town, Ethiopia. *Int J Pediatr.* (2019) 2019:3293516. doi: 10.1155/2019/3293516

11. Abadura SA, Lerebo WT, Kulkarni U, Mekonnen ZA. Individual and community level determinants of childhood full immunization in Ethiopia: a multilevel analysis. *BMC Public Health.* (2015) 15:1–10. doi: 10.1186/s12889-015-2315-z

12. Ofei SY, Fuchs GJ. Principles and practice of oral rehydration. *Curr Gastroenterol Rep.* (2019) 21:1–6. doi: 10.1007/s11894-019-0734-1

13. Alam NH, Ashraf H. Treatment of infectious diarrhea in children. *Pediatr Drugs.* (2003) 5:151–65. doi: 10.2165/00148581-200305030-00002

14. Mahalanabis D, Molla A, Sack DA. *Clinical Management of Cholera.* Cholera: Springer (1992). p. 253–283.

15. Rouhani S, Meloney L, Ahn R, Nelson BD, Burke TF. Alternative rehydration methods: a systematic review and lessons for resource-limited care. *Pediatrics.* (2011) 127:e748–57. doi: 10.1542/peds.2010-0952

16. Gaffey MF, Wazny K, Bassani DG, Bhutta ZA. Dietary management of childhood diarrhea in low-and middle-income countries: a systematic review. *BMC Public Health.* (2013) 13:1–16. doi: 10.1186/1471-2458-13-S3-S17

17. Tsehay CT, Aschalew AY, Dellie E, Gebremedhin T. Feeding practices and associated factors during diarrheal disease among children aged less than five years: evidence from the Ethiopian Demographic and Health Survey 2016. *Pediatr Health Med Ther.* (2021) 12:69–78. doi: 10.2147/PHMT.S289442

18. Paintal K, Aguayo VM. Feeding practices for infants and young children during and after common illness. Evidence from South Asia. *Matern Child Nutr.* (2016) 12:39–71. doi: 10.1111/mcn.12222

19. Jelliffe DB, Jelliffe EFP, World Health Organization. *Dietary Management of Young Children With Acute Diarrhoea: A Manual for Managers of Health Programmes.* Geneva: World Health Organization (1991).

20. Bhutta ZA, Das JK, Walker N, Rizvi A, Campbell H, Rudan I, et al. Interventions to address deaths from childhood pneumonia and diarrhoea equitably: what works and at what cost? *Lancet.* (2013) 381:1417–29. doi: 10.1016/S0140-6736(13)60648-0

21. De Onis M, Dewey KG, Borghi E, Onyango AW, Blössner M, Daelmans B, et al. *The World Health Organization's Global Target for Reducing Childhood Stunting by 2025: Rationale and Proposed Actions.* Hoboken, NJ: Wiley Online Library (2013). p. 6–26.

22. Kassaw MW, Abebe AM, Kassie AM, Abate BB, Tegegne KD, Masresha SA. Evidence from 2016 Ethiopian demographic and health survey data: maternal practice in managing childhood diarrhea at home. *J Pediatr Nurs.* (2020) 55:e250–6. doi: 10.1016/j.pedn.2020.08.017

23. Yeshaw Y, Alem AZ, Ayalew HG, Liyew AM, Tessema ZT, Alamneh TS, et al. Appropriate feeding practice and associated factors among under-five children with diarrheal disease in sub-Saharan Africa: a multi-country analysis. *Trop Med Health.* (2023) 51:13. doi: 10.1186/s41182-023-00503-1

24. Perin J, Carvajal-Velez L, Carter E, Bryce J, Newby H. Fluid curtailment during childhood diarrhea: a countdown analysis. *BMC Public Health.* (2015) 15:1–9. doi: 10.1186/s12889-015-1878-z

25. Kassa S, Egata G, Gobena T, Mekonnen M. Feeding practice during a diarrheal episode and associated factors among mothers/caregivers with 6–59-month-old children visiting health centers in Shanan Dhuggo District, West Hararge Zone, Oromia Region, Ethiopia. *Pediatr Health Med Ther*. (2021) 12:197–204. doi: 10.2147/PHMT.S300035

26. Gupta GR. Tackling pneumonia and diarrhoea: the deadliest diseases for the world's poorest children. *Lancet*. (2012) 379:2123–4. doi: 10.1016/S0140-6736(12)60907-6

27. Aliyu AA, Dahiru T. Factors associated with safe disposal practices of child's faeces in Nigeria: Evidence from 2013 Nigeria Demographic and Health Survey. *Niger Med J*. (2019) 60:198–204. doi: 10.4103/nmj.NMJ_3_19

28. Ogbo FA, Nguyen H, Naz S, Agho KE, Page A. The association between infant and young child feeding practices and diarrhoea in Tanzanian children. *Trop Med Health*. (2018) 46:1–9. doi: 10.1186/s41182-018-0084-y

29. ICF International ( 2012) *Demographic and Health Survey Sampling and Household Listing Manual*. Calverton, MD: MEASURE DHS.

30. Croft T, Marshall AM, Allen CK, Arnold F, Assaf S, Balian, S. *Guide to DHS Statistics: DHS-7 (Version 2)*. Rockville, MD: ICF (2020).

31. Elkasabi M. *Sampling and Weighting with DHS Data*. (2015). Available online at: https://blog.dhsprogram.com/sampling-weighting-at-dhs/ (Accessed October 10, 2024).

32. Croft TN, Marshall AM, Allen CK, Arnold F, Assaf S, Balian, S. *Guide to DHS Statistics*. Rockville: ICF 645 (2018).

33. Liu L, Wong AK, Wang Y. A global optimal algorithm for class-dependent discretization of continuous data. *Intell Data Anal*. (2004) 8:151–70. doi: 10.3233/IDA-2004-8204

34. Yehuala TZ, Derseh NM, Tewelgne MF, Wubante SM. Exploring machine learning algorithms to predict diarrhea disease and identify its determinants among under-five years children in East Africa. *J Epidemiol Global Health*. (2024) 1–11. doi: 10.3389/fped.2024.1388820

35. Deeba F, Patil SR. *Utilization of Machine Learning Algorithms for Prediction of Diseases*. Piscataway, NJ: IEEE (2021). p. 1–7.

36. Mubarak H, Stegen S, Bai F, Abdellatif A, Sanjari MJ. Enhancing interpretability in power management: A time-encoded household energy forecasting using hybrid deep learning model. *Energy Conv Manag*. (2024) 315:118795. doi: 10.1016/j.enconman.2024.118795

37. Martorell R, Habicht J-P, Yarbrough C, Lechtig A, Klein RE, Western KA. Acute morbidity and physical growth in rural Guatemalan children. *Am J Dis Child*. (1975) 129:1296–301. doi: 10.1001/archpedi.1975.02120480022007

38. World Health Organization. *A Manual for the Treatment of Acute Diarrhoea*. Geneva: World Health Organization (1984). p. 25–25.

39. World Health Organization. *WHO Recommendations on the Management of Diarrhoea and Pneumonia in HIV-Infected Infants and Children: Integrated Management of Childhood Illness (IMCI)*. Geneva: World Health Organization (2010).

40. Brown KH, MacLean WC Jr. Nutritional management of acute diarrhea: an appraisal of the alternatives. *Pediatrics*. (1984) 73:119–25.

41. World Health Organization. *The Treatment of Diarrhoea: A Manual for Physicians and Other Senior Health Workers*. World Health Organization (2005).

42. Zemariam AB, Abey W, Kassaw AK, Yimer A. Comparative analysis of machine learning algorithms for predicting diarrhea among under-five children in Ethiopia: evidence from 2016 EDHS. *Health Informatics J*. (2024) 30:14604582241285769. doi: 10.1177/14604582241285769

43. Carvajal-Vélez L, Amouzou A, Perin J, Maïga A, Tarekegn H, Akinyemi A, et al. Diarrhea management in children under five in sub-Saharan Africa: does the source of care matter? A countdown analysis. *BMC Public Health*. (2016) 16:1–14. doi: 10.1186/s12889-016-3475-1

44. Brauner-Otto S, Baird S, Ghimire D. Maternal employment and child health in Nepal: the importance of job type and timing across the child's first five years. *Soc Sci Med*. (2019) 224:94–105. doi: 10.1016/j.socscimed.2019.02.009

45. Misgna HG, Ebessa B, Kassa M. Prevalence of oral rehydration therapy use and associated factors among under-five children with diarrhea in Dangure, Benishangul Gumuz Region, Ethiopia/2018. *BMC Res Notes*. (2019) 12:1–6. doi: 10.1186/s13104-019-4078-6

46. Sood N, Wagner Z. Private sector provision of oral rehydration therapy for child diarrhea in sub-Saharan Africa. *Am J Trop Med Hyg*. (2014) 90:939. doi: 10.4269/ajtmh.13-0279

47. Senarath U, Agho KE, Akram DeS, Godakandage SS, Hazir T, et al. Comparisons of complementary feeding indicators and associated factors in children aged 6–23 months across five South Asian countries. *Matern Child Nutr*. (2012) 8:89–106. doi: 10.1111/j.1740-8709.2011.00370.x

48. Victor R, Baines SK, Agho KE, Dibley MJ. Factors associated with inappropriate complementary feeding practices among children aged 6–23 months in T anzania. *Matern Child Nutr*. (2014) 10:545–61. doi: 10.1111/j.1740-8709.2012.00435.x

49. Mekbib E, Shumey A, Ferede S, Haile F. Magnitude and factors associated with appropriate complementary feeding among mothers having children 6–23 months-of-age in northern Ethiopia; a community-based cross-sectional study. *J Food Nutr Sci*. (2014) 2:36. doi: 10.11648/j.jfns.20140202.13

50. Khan SM, Speizer IS, Singh K, Angeles G, Twum–Danso NA, Barker P. Does postnatal care have a role in improving newborn feeding? A study in 15 sub–Saharan African countries. *J Global Health*. (2017) 7:e020506. doi: 10.7189/jogh.07.020506

51. Fikadu T, Girma S. Feeding practice during diarrheal episode among children aged between 6 to 23 months in Mirab Abaya district, Gamo Gofa Zone, Southern Ethiopia. *Int J Pediatr*. (2018) 2018:2374895. doi: 10.1155/2018/2374895

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 20 August 2025
DOI 10.3389/fpubh.2025.1654482

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Claudio-Alberto Rodríguez-Suárez,
University of Las Palmas de Gran Canaria,
Spain
Hammad Ganatra,
Cleveland Clinic, United States

*CORRESPONDENCE
Qiong-Xia Ou
oqx18006930694@163.com
Gao-Zhong Feng
15060102098@163.com

RECEIVED 27 June 2025
ACCEPTED 08 August 2025
PUBLISHED 20 August 2025

CITATION
Huang Y-D, Zeng S-L, Lin J, Mao Z-P, Zhang Q-L, Feng G-Z and Ou Q-X (2025) Parents' understanding and attitudes toward the application of AI in pediatric healthcare: a cross-sectional survey study.
*Front. Public Health* 13:1654482.
doi: 10.3389/fpubh.2025.1654482

COPYRIGHT
© 2025 Huang, Zeng, Lin, Mao, Zhang, Feng and Ou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Parents' understanding and attitudes toward the application of AI in pediatric healthcare: a cross-sectional survey study

Yan-Dan Huang[1], Shu-Long Zeng[1], Jie Lin[1], Zhi-Peng Mao[1], Qi-Liang Zhang[2], Gao-Zhong Feng[1]* and Qiong-Xia Ou[2]*

[1]Department of Pediatrics, Yongtai County Hospital, Fuzhou, China, [2]Fujian Children's Hospital (Fujian Branch of Shanghai Children's Medical Center), College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, Fuzhou, China

**Objective:** To investigate parents' understanding and attitudes toward the application of Artificial Intelligence (AI) in pediatric healthcare.
**Methods:** An observational, cross-sectional study was conducted using an *ad hoc* questionnaire. Between February and April 2025, 200 family members of children receiving care at our hospital voluntarily participated in the study. Inclusion criteria included being a family member of a child treated at the hospital. Exclusion criteria were an inability to understand the questionnaire and incomplete responses. The AI applications referenced by respondents primarily included large language models such as ChatGPT, DeepSeek, and Kimi, among others. The questionnaire consisted of two sections: demographic information, and attitudes toward the use of AI in pediatric healthcare. Data were analyzed using SPSS (version 25.0). Quantitative variables were expressed as mean and standard deviation, and categorical variables using frequencies and percentages. Group comparisons were performed using chi-square test and t-test ($p$-value $< 0.05$).
**Results:** A total of 185 participants completed the questionnaire. Participants who were unaware of AI applications in pediatric healthcare were more likely to be older, have lower educational levels, and reside in rural areas. The majority of respondents (71.2%) believed that the information provided by AI was partially accurate, while 6.9% considered it partially inaccurate. Regarding perceived benefits, 74% identified convenience as the main advantage of AI in pediatric care, followed by 41.1% who cited high diagnostic efficiency. Key concerns included perceived inaccuracy and the potential for misdiagnosis (52%), as well as uncertainty regarding accountability in the event of an error (44.5%). Most participants (91.1%) believed that AI cannot replace doctors in the future.
**Conclusion:** Although most parents were aware of the use of AI in pediatric healthcare and recognized its convenience and efficiency, they expressed concerns about accuracy, accountability, and data privacy. A notable lack of awareness was observed among older individuals, those with lower levels of education, and residents of rural areas.

KEYWORDS

**AI, children, healthcare, understanding and attitudes, cross-sectional survey**

# Introduction

Artificial Intelligence (AI) is a rapidly emerging computer engineering discipline that possesses human-like intelligence. In today's world, AI has begun to transform multiple industries, including banking, healthcare, and travel, with its influence continually expanding (1). AI has brought unprecedented possibilities to healthcare. It is capable of supporting clinicians in performing a variety of core tasks and facilitating efficient communication (2). Additionally, the application of AI technology helps promote the accessibility of high-quality medical services, reduces the administrative burden on clinicians, optimizes the allocation and utilization of medical resources (3, 4). The research and application of AI are continuously expanding, and its integration with healthcare systems has become an important driving force in the development of medical technology (5). The use of AI based software is becoming increasingly common in all fields. Currently, AI technology has demonstrated superior performance in various areas, including clinical decision-making for pediatric diseases, image diagnosis in pathology, prediction of cardiovascular diseases, prognosis of cancer treatment, prediction of molecular changes in cancer, discovery and development of novel drugs, and control of epidemics (6–12).

When implementing any new technology in healthcare, it is crucial to consider the perspectives of patients (13). Insufficient acceptance of therapeutic measures can undermine patient compliance and worsen outcomes that might otherwise have been successful (14). Therefore, the rational use of AI in healthcare must also ensure its acceptability to patients and their families. However, many studies on the perception of AI in medicine have focused on healthcare professionals, none have specifically focused on parents' understanding and attitudes toward the application of AI in pediatric healthcare. The aim of this study is to investigate parents' understanding and attitudes toward the application of AI in pediatric healthcare and to provide recommendations for the further promotion of AI in this field.

# Methods

## Study design

This was an observational study with a cross-sectional survey. To our knowledge, no study has previously been reported on this topic. Therefore, we conducted this study as an exploratory study. This cross-sectional survey collected data using a *ad hoc* questionnaire. The ad hoc questionnaire was designed by the researchers.

## Setting

In this study, the AI of respondents used were referring primarily to the use of large language models such as ChatGPT, DeepSeek or Kimi.

Abbreviation: AI, artificial intelligence.

## Calculation of study sample size

Based on the results of 78.9% of participants who were aware of the application of AI in pediatric healthcare, and assuming that the power of 0.85, the required number of participants was calculated to be 161. Assuming a 20% missing rate, the total sample size was set as 193.

## Participants

From February to April 2025, a total of 200 family members of children who visited our hospital voluntarily participated in this study. The 200 participants represented 200 unique family units (i.e., caregivers of 200 separate pediatric patients). A total of 185 completed the questionnaire. Inclusion criteria were: (1) family members of children treated in our hospital, (2) agreement to participate in this study. Exclusion criteria were: (1) inability to understand the content of the questionnaire, (2) incomplete questionnaire.

The survey was completed during the children visit to the hospital. The questionnaires were filled out individually under the supervision of the research team. If participants had any confusion regarding the questions, the researchers were available to promptly address their concerns or translate the questions into the local language to provide assistance. After completion of the questionnaires, the data were collected and analyzed separately by other researchers.

## Variables

The *ad hoc* questionnaire collected data in the following two aspects: (1) essential information regarding the demographic status of the participants (5 items) including: age, sex, education level, work and address; (2) attitudes toward the application of AI in pediatric healthcare (8 items).

## Statistical methods

SPSS 25.0 software was used for statistical analysis. The quantitative data were expressed as the mean ± standard deviation, the continuous data were in accordance with normal distribution by normal distribution test, *t*-test was used for statistical analysis. Categorical data were described using frequencies and percentages. Comparisons between groups were performed via the chi-square test. A $p$-value <0.05 was considered statistically significant.

## Ethical considerations

The present study was approved by Yongtai County Hospital, and adhered to the tenets of the Declaration of Helsinki. Prior to participation, the participants were informed of the purpose and content of the study. All participants were provided written informed consent and were informed that their participation was entirely voluntary and they had the right to withdraw for any reason.

10.3389/fpubh.2025.1654482

TABLE 1 Demographic characteristics of the participants.

| Items | *n* = 185 (*n*%) |
|---|---|
| Age (year) | 35.8±9.2 |
| Distribution of age (year) | |
| 20–29 | 33 (17.8%) |
| 30–39 | 111 (60%) |
| 40-49 | 25 (13.5%) |
| ≥50 | 16 (8.7%) |
| Sex | |
| Male | 53 (28.6%) |
| Female | 132 (71.4%) |
| Education level | |
| Junior high school or below | 39 (21.1%) |
| High school or secondary vocational school | 54 (29.2%) |
| Bachelor's degree or college diploma | 87 (47%) |
| Master's degree or above | 5 (2.7%) |
| Work | |
| White-collar workers or civil servants | 15 (8.1%) |
| Professional and technical personnel | 40 (21.6%) |
| Unemployed individuals | 4 (2.1%) |
| Farmers | 28 (15.2%) |
| Other occupations | 98 (53%) |
| Address | |
| Urban areas | 116 (62.7%) |
| Rural areas | 69 (37.3%) |

## Results

A total of 200 family members of children participated in the study, of whom 185 completed the questionnaire. Table 1 showed the demographic characteristics of the participants. There were 53 males and 132 females, with age of 35.8 ± 9.2 years. The age distribution was as follows: 20–29 years (17.8%), 30–39 years (60.0%), 40–49 years (13.5%), and ≥50 years (8.7%). The majority of family members (78.9%) had completed high school or higher education, among whom 49.7% had an Bachelor's degree, college diploma, or higher education level. Those with an educational level of junior high school or below accounted for 21.1%. Among the participants, 8.1% were white-collar workers or civil servants, 21.6% were professional and technical personnel, 2.1% were unemployed, 15.2% were farmers, and 53.0% had other occupations. A total of 62.7% of the participants resided in urban areas, while 37.3% lived in rural areas.

The primary source of information on pediatric healthcare for the participants was the internet, accounting for 48.7% of the participants. AI was the main source of information for 18.4% of participants, while 28.6% participants primarily obtained information from healthcare professionals, and only 4.3% participants relied on books.

A total of 78.9% ($n = 146$) of participants were aware of the application of AI in pediatric healthcare, while the remaining 21.1% ($n = 39$) participants were not aware of the application of AI in pediatric healthcare. The age of participants who were unaware of the application of AI in pediatric healthcare was significantly older than those who were aware of its application ($p < 0.05$). The majority of participants who were unaware of the application of AI in healthcare had an educational level of junior high school or below and had significantly lower educational attainment than those who were aware of AI application ($p = 0.001$). Most participants who were unaware of the application of AI in healthcare resided in rural areas, which was significantly different from those who were aware of AI application ($p = 0.001$). Additionally, most participants who were unaware of the application of AI in healthcare were farmers, which was also significantly different from those who were aware of AI application ($p = 0.001$) (Table 2).

Among the 78.9% ($n = 146$) of participants who were aware of the application of AI in pediatric healthcare, 74% participants had used AI in pediatric healthcare. The majority (50.7%) of participants who had used AI in pediatric healthcare reported occasional use, 19.9% participants reported frequent use, and 3.4% participants reported always using AI. Most participants (71.2%) believed that the information provided by AI was partially accurate, while 6.9% participants thought that the information was partially inaccurate. The main advantages of using AI in pediatric healthcare, as perceived by the majority of participants, were convenience of use (74%) and high diagnostic efficiency (41.1%). The main disadvantages were perceived inaccuracy and the risk of misdiagnosis (52% participants) and unclear responsibility in case of misdiagnosis (44.5% participants). The vast majority participants (91.1%) believed that AI cannot replace doctors in the future (Table 3).

## Discussion

The number of novel applications based on AI is steadily increasing and will be gradually implemented in clinical settings. Therefore, the use of AI-based technologies in healthcare is expected to grow rapidly in the coming years (15). Many AI tools have been widely accepted among healthcare professionals, thereby significantly improving various aspects of patient care and clinical workflow. The implementation of AI in healthcare holds great potential to transform patient outcomes and treatments (16). By leveraging AI-driven predictive analytics, clinical laboratory testing and disease detection can be made more accurate and efficient (17). AI can also assist in population health management by providing real-time relevant information and optimizing medication selection (18). The application of AI in digital health and mental health support has demonstrated its potential to enhance patient prognosis (19).

AI is also increasingly used in pediatrics. Bokov et al. (20) applied machine learning models to adjust the intervention measures for children with high-risk asthma, thereby improving the personalized care outcomes. Mumenin et al. (21) reported that machine learning can be used to predict the seizure status of newborns, thereby promoting early intervention and personalized care. In cardiology, Hunfeld et al. (22) employed machine learning models to identify congenital heart defects and predict outcomes following cardiac events. In gastroenterology, machine learning was used to manage chronic diseases such as Crohn's disease and predict the risk of bacterial gastroenteritis in children (23, 24). Tong et al. (25) applied machine learning models to assess nutritional deficiencies and

TABLE 2 Comparison of the demographic characteristics between participants who were aware of the application of AI in healthcare and those who were not aware.

| Items | Participants were aware of the application of AI in healthcare *n* = 146 | Participants were not aware of the application of AI in healthcare *n* = 39 | *P* |
|---|---|---|---|
| Age (year) | 34.4±7.2 | 38.0±11.5 | 0.009 |
| Distribution of age (year) | | | |
| 20–29 | 33 (17.8%) | 1 (2.5%) | 0.001 |
| 30–39 | 111 (60%) | 9 (23.1%) | |
| 40-49 | 25 (13.5%) | 16 (41.1%) | |
| ≥50 | 16 (8.7%) | 13 (33.3%) | |
| Sex | | | |
| Male | 53 (28.6%) | 7 (17.9%) | 0.170 |
| Female | 132 (71.4%) | 32 (82.1%) | |
| Education level | | | |
| Junior high school or below | 39 (21.1%) | 30(76.9%) | 0.001 |
| High school or secondary vocational school | 54 (29.2%) | 9 (23.1%) | |
| Bachelor's degree or college diploma | 87 (47%) | 0 | |
| Master's degree or above | 5 (2.7%) | 0 | |
| Work | | | |
| White-collar workers or civil servants | 15 (8.1%) | 0 | 0.001 |
| Professional and technical personnel | 40 (21.6%) | 0 | |
| Unemployed individuals | 4 (2.1%) | 2 (5.1%) | |
| Farmers | 28 (15.2%) | 21 (53.8%) | |
| Other occupations | 98 (53%) | 16 (41.1%) | |
| Address | | | |
| Urban areas | 116 (62.7%) | 11 (28.2%) | 0.001 |
| Rural areas | 69 (37.3%) | 28 (71.8%) | |

personalize dietary adjustments for children with chronic gastrointestinal diseases. However, to date, only a limited number of studies have reported professionals' views of AI in healthcare. When implementing any new technology in healthcare, it is crucial to consider the perspectives of patients. There is currently a scarcity of research reporting patients' views on the use of AI in healthcare (26). Moreover, no studies have been reported on parents' understanding and attitudes toward the application of AI in pediatric healthcare. This study investigates and reports on parents' understanding and attitudes toward the application of AI in pediatric healthcare. The AI that the families of the patients have come into contact with mainly refers to the use of large language models such as ChatGPT, DeepSeek or Kimi, etc.

In our study, the majority of participants (78.9%) were aware of the application of AI in healthcare, and a significant proportion (74%) among them had experience using AI in pediatric healthcare. Most individuals in this study believed that application of AI in pediatric healthcare offers the advantages of convenience and high diagnostic efficiency. These findings are similar to those of other studies. Aggarwal et al. (27) investigated the views of 408 participants in London and found that despite patients generally having limited prior knowledge of AI, the majority considered its use to be affirmative, trusted its application, and believed that the benefits outweighed the risks. Fritsch et al. (28) surveyed the attitudes and perceptions of 452 patients toward application of AI in healthcare and found that the majority were in favor of using AI in medicine and healthcare, with 53.18% of respondents rating it as positive or very favorable, and only 4.77% expressing negative or very negative opinions.

The integration of AI in healthcare has the potential to transform patient care, diagnostic procedures, and treatment methods. However, while the majority of people maintain a positive attitude toward the application of AI in healthcare, it still poses some challenges and limitations (29). First and foremost are the issues of information accuracy. The application of AI in medical diagnostics is still in its infancy, and given the complexity of various diseases and their symptoms, establishing accurate diagnostic techniques is highly challenging. In this study, 71.2% of participants believed that the information provided by AI for pediatric healthcare was only partially accurate. Furthermore, 52% of participants identified the main deficiency of AI in pediatric healthcare as insufficient diagnostic accuracy and the risk of misdiagnosis. Moreover, the vast majority of participants (91.1%) believed that even in a more advanced future, AI could never replace physicians. A study by Fritsch SJ et al. found that only a small proportion of patients would agree to be treated directly by AI, while the rest rejected

TABLE 3 Participants' understanding and attitudes toward the application of AI in pediatric healthcare.

| Items | *n* % |
|---|---|
| Q1. Did you aware of the application of AI in healthcare? | |
| Yes | 146 (78.9%) |
| No | 39 (21.1%) |
| Q2. Did you ever used AI in pediatric healthcare? | |
| Yes | 108 (74%) |
| No | 38 (26%) |
| Q3. How often do you use AI software when you encounter pediatric healthcare issues? | |
| Never | 38 (26%) |
| Occasionally | 74 (50.7%) |
| Often | 29 (19.9%) |
| Always | 5 (3.4%) |
| Q4. Do you think the pediatric healthcare information provided by AI is accurate? | |
| Very accurate | 32 (21.9%) |
| Partly accurate | 104 (71.2%) |
| Not accurate | 10 (6.9%) |
| Q5. What do you think are the main advantages of using AI in pediatric healthcare? (Multiple choices allowed) | |
| Convenience of use | 108 (74%) |
| Alleviation of medical resource tension | 29 (19.8%) |
| High diagnostic efficiency | 60 (41.1%) |
| Reduction of human error | 10 (6.8%) |
| Q6. What do you think are the main disadvantages of using AI in pediatric healthcare? (Multiple choices allowed) | |
| Inaccuracy and the potential for misdiagnosis | 76 (52%) |
| Lack of human touch | 36 (24.6%) |
| Data privacy and security concerns | 25 (17.1%) |
| Unclear attribution of responsibility in case of misdiagnosis | 65 (44.5%) |
| Q7. Do you think that AI will be able to replace doctors in the future? | |
| Yes | 13 (8.9%) |
| No | 133 (91.1%) |

AI treatment (28). In their study, all respondents believed that AI should not have too much autonomy and that physicians should closely monitor AI. They hoped that physicians should act as gatekeepers for AI decisions.

Secondly, the application of AI in healthcare also faces challenges related to ethical, privacy, and legal issues (30, 31). A review of AI in pediatric healthcare by Ganatra (32) reports that overcoming obstacles such as data limitations, ethical issues, privacy and model credibility was crucial for its wider application. Our survey results showed that 44.5% participants were concerned about the unclear attribution of responsibility in the event of AI misdiagnosis, and 17.1% participants were worried about data privacy and security issues. Who is responsible for the decisions made through the collaboration between physicians and AI tools? (33). Due to the "black box" nature of most AI algorithms, physicians find it difficult to understand how these algorithms generate their recommendations. Moreover, it remains unclear who should be held accountable and bear legal responsibility if harm is caused due to the use of AI tools by clinicians. Maintaining patient data security and privacy is of utmost importance. However, there is currently no comprehensive privacy protection and regulatory framework for AI in healthcare applications, which is why many participants are concerned about data privacy and security issues.

Our survey results showed that the main characteristics of the 21.1% participants who were unaware of the application of AI in healthcare were: older age, lower educational level, and living in rural areas. Individuals in this group are located in less information-developed regions and have a poorer ability to understand and accept new things. Due to concerns arising from a lack of information, they may be more likely to disagree with the application of AI in pediatric healthcare. Therefore, when promoting the application of AI in pediatric healthcare, special consideration needs to be given to these groups.

There were several limitations of this study. (1) This was a cross-sectional study with small sample size. (2) The data were collected through *ad hoc* questionnaires, and some participants may not have fully understood the questions posed, potentially leading to biases. (3) To our knowledge, there has been no previous evidence on this topic. This study was a exploratory study; therefore, we cannot assess the reliability and validity of the *ad hoc* questionnaire. (4) The survey was conducted in a hospital outpatient setting. The primary caregivers bringing children to the hospital for medical visits were mothers, followed by fathers. As a result, the majority of participants in this study were female (71.4%) and within the age range of 20–40 years. This introduced potential biases related to gender and age.

## Conclusion

This study explored parents' awareness and attitudes toward the application of AI in pediatric healthcare through a questionnaire survey conducted among families of children. The findings indicate that the majority of respondents were aware of the application of AI in healthcare, and most parents recognize its convenience and efficiency. However, significant concerns remained regarding its accuracy, responsibility allocation, and data privacy and security. Additionally, the study revealed that parents from rural areas, with older age, and with lower levels of education had limited awareness of the application of AI in pediatric healthcare. These findings not only revealed the complex attitudes of parents toward the application of AI in pediatric healthcare but also highlighted the weak links in the dissemination of knowledge.

The study filled the gap in research regarding parents' attitudes toward the application of AI in pediatric healthcare and provided important references for practice. When promoting the application of AI technology in pediatric healthcare, it is crucial to pay attention to parents' concerns and needs, especially for those groups with insufficient understanding. Education and publicity should be used to enhance their comprehension and trust in AI technology. It is also necessary to optimize AI technology to improve its accuracy in pediatric healthcare, as well as to establish mechanisms for responsibility allocation and data privacy protection.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the datasets used and analyzed during the current study are available from the corresponding author on reasonable request. Requests to access these datasets should be directed to Gao-Zhong Feng, 15060102098@163.com.

## Ethics statement

The studies involving humans were approved by the Ethics Committee of Yongtai County Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

Y-DH: Methodology, Data curation, Investigation, Formal analysis, Conceptualization, Writing – original draft. S-LZ: Writing – original draft, Formal analysis, Data curation, Investigation, Methodology. JL: Writing – original draft, Formal analysis, Data curation, Investigation. Z-PM: Writing – original draft, Methodology, Formal analysis. Q-LZ: Formal analysis, Writing – review & editing, Methodology, Software. G-ZF: Writing – review & editing, Conceptualization, Methodology, Supervision. Q-XO: Formal analysis, Software, Methodology, Conceptualization, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by Startup Fund for scientific research, Fujian Medical University (grant number: 2022QH1227).

## Acknowledgments

We highly acknowledge all the staff from our unit who participated in this study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Hirani R, Noruzi K, Khuram H, Hussaini AS, Aifuwa EI, Ely KE, et al. Artificial intelligence and healthcare: a journey through history, present innovations, and future possibilities. *Life*. (2024) 14:557. doi: 10.3390/life14050557

2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. (2017) 28:686. doi: 10.1038/nature22985

3. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. (2017) 2:230–43. doi: 10.1136/svn-2017-000101

4. Yu P, Xu H, Hu X, Deng C. Leveraging generative AI and large language models: a comprehensive roadmap for healthcare integration. *Healthcare*. (2023) 11:2776. doi: 10.3390/healthcare11202776

5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7

6. Ramgopal S, Sanchez-Pinto LN, Horvat CM, Carroll MS, Luo Y, Florin TA. Artificial intelligence-based clinical decision support in pediatrics. *Pediatr Res*. (2023) 93:334–41. doi: 10.1038/s41390-022-02226-1

7. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology – new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol*. (2019) 16:703–15. doi: 10.1038/s41571-019-0252-y

8. van Smeden M, Heinze G, Van Calster B, Asselbergs FW, Vardas PE, Bruining N, et al. Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. *Eur Heart J*. (2022) 43:2921–30. doi: 10.1093/eurheartj/ehac238

9. Corti C, Cobanaj M, Marian F, Dee EC, Lloyd MR, Marcu S, et al. Artificial intelligence for prediction of treatment outcomes in breast cancer: systematic review of design, reporting standards, and bias. *Cancer Treat Rev*. (2022) 108:102410. doi: 10.1016/j.ctrv.2022.102410

10. Cifci D, Foersch S, Kather JN. Artificial intelligence to identify genetic alterations in conventional histopathology. *J Pathol*. (2022) 257:430–44. doi: 10.1002/path.5898

11. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. (2019) 18:463–77. doi: 10.1038/s41573-019-0024-5

12. Albalawi U, Mustafa M. Current artificial intelligence (AI) techniques, challenges, and approaches in controlling and fighting COVID-19: a review. *Int J Environ Res Public Health*. (2022) 19:5901. doi: 10.3390/ijerph19105901

13. Haan M, Ongena YP, Hommes S, Kwee TC, Yakar D. A qualitative study to understand patient perspective on the use of artificial intelligence in radiology. *J Am Coll Radiol*. (2019) 16:1416–9. doi: 10.1016/j.jacr.2018.12.043

14. Kleinsinger F. Understanding non-compliant behavior: definitions and causes. *Perm J*. (2003) 7:18–21.

15. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. (2018) 2:719–31. doi: 10.1038/s41551-018-0305-z

16. Chen M, Zhang B, Cai Z, Seery S, Gonzalez MJ, Ali NM, et al. Acceptance of clinical artificial intelligence among physicians and medical students: a systematic review with cross-sectional survey. *Front Med*. (2022) 9:990604. doi: 10.3389/fmed.2022.990604

17. Khan S. Artificial intelligence and machine learning in clinical medicine. *N Engl J Med*. (2023) 388:2398. doi: 10.1056/NEJMc2305287

18. Martin GL, Jouganous J, Savidan R, Bellec A, Goehrs C, Benkebil M, et al. Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. *Drug Saf*. (2022) 45:535–48. doi: 10.1007/s40264-022-01153-8

19. Kim JW, Jones KL, D'Angelo E. How to prepare prospective psychiatrists in the era of artificial intelligence. *Acad Psychiatry*. (2019) 43:337–9. doi: 10.1007/s40596-019-01025-x

20. Bokov P, Mahut B, Flaud P, Delclaux C. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Comput Biol Med*. (2016) 70:40–50. doi: 10.1016/j.compbiomed.2016.01.002

21. Mumenin KM, Biswas P, Khan MA, Alammary AS, Nahid AA. A modified Aquila-based optimized XGBoost framework for detecting probable seizure status in neonates. *Sensors*. (2023) 23:7037. doi: 10.3390/s23167037

22. Hunfeld M, Verboom M, Josemans S, van Ravensberg A, Straver D, Lückerath F, et al. Prediction of survival after pediatric cardiac arrest using quantitative EEG and machine learning techniques. *Neurology*. (2024) 103:e210043. doi: 10.1212/WNL.0000000000210043

23. Jang S, Yu J, Park S, Lim H, Koh H, Park YR. Development of time-aggregated machine learning model for relapse prediction in pediatric Crohn's disease. *Clin Transl Gastroenterol*. (2025) 16:e00794. doi: 10.14309/ctg.0000000000000794

24. Miyagi Y. Identification of pediatric bacterial gastroenteritis from blood counts and interviews based on machine learning. *Cureus*. (2023) 15:e43644. doi: 10.7759/cureus.43644

25. Tong L, Kauer J, Chen X, Chu K, Dou H, Smith ZJ. Screening of nutritional and genetic anemias using elastic light scattering. *Lab Chip*. (2018) 18:3263–71. doi: 10.1039/C8LC00377G

26. Temple S, Rowbottom C, Simpson J. Patient views on the implementation of artificial intelligence in radiotherapy. *Radiography*. (2023) 29:S112–6. doi: 10.1016/j.radi.2023.03.006

27. Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient perceptions on data sharing and applying artificial intelligence to health care data: cross-sectional survey. *J Med Internet Res*. (2021) 23:e26162. doi: 10.2196/26162

28. Fritsch SJ, Blankenheim A, Wahl A, Hetfeld P, Maassen O, Deffge S, et al. Attitudes and perception of artificial intelligence in healthcare: a cross-sectional survey among patients. *Digit Health*. (2022) 8:20552076221116772. doi: 10.1177/20552076221116772

29. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. (2019) 17:195. doi: 10.1186/s12916-019-1426-2

30. Hryciw BN, Fortin Z, Ghossein J, Kyeremanteng K. Doctor-patient interactions in the age of AI: navigating innovation and expertise. *Front Med*. (2023) 10:1241508. doi: 10.3389/fmed.2023.1241508

31. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med*. (2020) 260:113172. doi: 10.1016/j.socscimed.2020.113172

32. Ganatra HA. Machine learning in pediatric healthcare: current trends, challenges, and future directions. *J Clin Med*. (2025) 14:807. doi: 10.3390/jcm 14030807

33. Bærøe K, Gundersen T, Henden E, Rommetveit K. Can medical algorithms be fair? Three ethical quandaries and one dilemma. *BMJ Health Care Inform*. (2022) 29:e100445. doi: 10.1136/bmjhci-2021-100445

frontiers | Frontiers in Artificial Intelligence

TYPE Review
PUBLISHED 02 September 2025
DOI 10.3389/frai.2025.1624485

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Serap Aydin,
Hacettepe University, Türkiye
Xuzhou Li,
Yunnan Normal University, China
Dimitri Marques Abramov,
Oswaldo Cruz Foundation (Fiocruz), Brazil

*CORRESPONDENCE
Ruixing Li
✉ lrxing@hactcm.edu.cn

RECEIVED 07 May 2025
ACCEPTED 14 August 2025
PUBLISHED 02 September 2025

CITATION
Zhao C, Xu Y, Li R, Li H and Zhang M (2025) Artificial intelligence in ADHD assessment: a comprehensive review of research progress from early screening to precise differential diagnosis.
*Front. Artif. Intell.* 8:1624485.
doi: 10.3389/frai.2025.1624485

COPYRIGHT
© 2025 Zhao, Xu, Li, Li and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Artificial intelligence in ADHD assessment: a comprehensive review of research progress from early screening to precise differential diagnosis

Cuijie Zhao[1,2], Yan Xu[1,2], Ruixing Li[1,2]*, Huawei Li[1] and Meng Zhang[1,2]

[1]Pediatrics Hospital, The First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou, Henan, China, [2]School of Pediatrics, Henan University of Chinese Medicine, Zhengzhou, Henan, China

Attention deficit hyperactivity disorder (ADHD) diagnosis traditionally relies on subjective assessments, which lead to challenges like symptom overlap, heterogeneity, and misdiagnosis risk. Artificial intelligence (AI), especially machine learning (ML) and deep learning (DL), offers objective assessment opportunities by processing complex multimodal data (behavioral, neurophysiological, neuroimaging, genetic). This paper reviews AI's current applications in objective ADHD assessment, covering early screening, risk prediction, diagnostic assistance, classification, assistance in precise differential diagnosis, symptom quantification, and heterogeneous subtype identification. While AI models show significant potential in extracting objective biomarkers and improving assessment efficiency, the field faces challenges: insufficient standardized data, limited generalization, interpretability issues, potential biases, and lack of rigorous clinical validation. Future research must establish large-scale, standardized multimodal databases, develop robust, interpretable, and fair AI models, and conduct rigorous clinical translation validation to achieve responsible, precise, objective, and personalized ADHD assessment and management.

KEYWORDS

**attention deficit hyperactivity disorder (ADHD), artificial intelligence (AI), objective assessment, differential diagnosis, biomarkers, neuroimaging, machine learning (ML)**

## 1 Introduction

Attention deficit hyperactivity disorder (ADHD) is characterized by persistent patterns of inattention and/or hyperactivity-impulsivity that are inconsistent with developmental level and significantly interfere with functioning or development (Faraone et al., 2021). It primarily manifests as three presentation types: predominantly inattentive presentation, predominantly hyperactive–impulsive presentation, and combined presentation (Willcutt, 2012). Globally, ADHD affects approximately 5.9% of children and adolescents, with a significant proportion persisting into adulthood, where prevalence rates range from 2.5–2.8% (Faraone et al., 2021; Salari et al., 2023). Gender distribution also shows variation, with ADHD more commonly diagnosed in boys during childhood, though this ratio tends to narrow in adolescence and adulthood, and females often present with predominantly inattentive symptoms (Mowlem et al., 2019; Faraone et al., 2021). While this review encompasses research across all age groups, particular emphasis is placed on pediatric populations (children aged 6–18 years), as this represents the critical period for early identification, intervention, and the establishment of

long-term management strategies. ADHD's clinical presentation, underlying neurobiological mechanisms, and response to interventions can vary significantly across developmental stages, with children often exhibiting more pronounced hyperactivity symptoms compared to adolescents and adults, who may present with more subtle inattentive symptoms and internalizing behaviors (Biederman et al., 2000). This developmental heterogeneity necessitates age-appropriate assessment approaches and biomarker identification strategies, which will be addressed throughout this review.

This disorder significantly impacts academic achievement, peer relationships, and social adaptation during childhood and adolescence. Its symptoms and functional impairments often persist into adulthood, causing profound and lasting adverse effects on educational attainment, career development, interpersonal relationships, daily functioning, and overall mental health (Shaw et al., 2012; Faraone et al., 2021). It may even be associated with increased risk of certain chronic physical diseases in adulthood (Cortese et al., 2016), collectively constituting a substantial personal, familial, and socioeconomic burden (Doshi et al., 2012; Shaw et al., 2012).

Current ADHD diagnosis primarily relies on DSM or ICD diagnostic criteria, which are often applied in conjunction with clinical interviews and standardized behavioral rating scales for comprehensive assessment. This process faces severe challenges due to over-reliance on subjective reports. Scale completion is susceptible to reporter bias, observational environment differences, and varying levels of understanding, with significant discrepancies often existing between different reporters, leading to insufficient objectivity in assessment results (Zysset et al., 2023). More complexly, symptom overlap and differential diagnostic difficulties are particularly prominent. ADHD's core symptoms (especially inattention) significantly overlap with various other mental or developmental disorders, including anxiety, depression, and learning disabilities (Furman, 2005). Notably, some gifted children's "overexcitability traits" (such as high energy, intense curiosity) are behaviorally very similar to ADHD, increasing the risk of misdiagnosis (François-Sévigny et al., 2022). Meanwhile, distinguishing Inattentive presentations from Cognitive Disengagement Syndrome (CDS) neurocognitive characteristics also highlights the ongoing complexity of differential diagnosis (Durak et al., 2025).

Furthermore, ADHD frequently co-occurs with ODD, CD, and anxiety disorders (Kaplan et al., 2001; Lange et al., 2018; Haan et al., 2022), and environmental factors [such as the COVID-19 pandemic (Gosselin et al., 2025)] may also influence disease assessment. Moreover, ADHD itself exhibits high heterogeneity, with patients showing significant individual differences in symptoms, cognitive deficits, and potential neurobiological mechanisms (Nigg et al., 2020), which existing classification criteria cannot fully capture. Finally, limitations in professional resources (such as insufficient experienced clinicians, time-consuming assessments, and uneven resource distribution) also restrict the accessibility of early identification and intervention.

Against this background, artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) technologies, has garnered significant attention due to its immense potential in processing complex, high-dimensional medical data (Handelman et al., 2018; Choi et al., 2020). AI can learn complex nonlinear patterns from multidimensional, multimodal data (such as behavioral observations, cognitive tests, neurophysiological signals EEG/eye movement, neuroimaging MRI, genomics, etc.), potentially discovering more objective and quantifiable disease-related biomarkers (Nayarisseri et al., 2021; Prelaj et al., 2024). Through integrated analysis of these data, AI shows promise in improving the accuracy, efficiency, and consistency of ADHD diagnostic assessment, and may assist in achieving earlier screening, auxiliary in more precise differential diagnosis, and more data-supported personalized assessment methods.

Given AI's potential value and increasing research exploration in ADHD assessment, this paper aims to systematically review and analyze the current applications of AI in objective ADHD assessment. The core objective is to comprehensively demonstrate how AI utilizes multimodal data across various stages of ADHD assessment—from early screening and risk prediction to diagnostic assistance and classification, to assisting precise differential diagnosis, as well as symptom quantification and heterogeneous subtype identification. Compared to existing reviews, the unique contributions of this paper include: (1) explicitly focusing on the paradigm shift from traditional subjective methods to AI-driven "quantitative assessment support"; (2) emphasizing and deeply discussing AI's potential and limitations in "assisting precise differential diagnosis" as a key clinical challenge; (3) conducting systematic, structured analysis and outlook on current research challenges and future directions across multiple dimensions including data, algorithms, validation and translation, ethics and society, while integrating cutting-edge concepts in the field.

In the following sections, this paper will first overview the AI methodology and main data sources used for ADHD assessment, then detail AI applications in specific ADHD assessment scenarios, evaluate the performance, advantages, and limitations of existing research, followed by in-depth discussion of core challenges and future research directions, and finally conclude (Figure 1).

## 2 Data sources and modalities for artificial intelligence analysis

AI-driven objective ADHD assessment relies on diverse data inputs that reflect information from different levels, from macroscopic behavior to microscopic biology. Integrating multimodal data is considered a key strategy for improving assessment accuracy, comprehensiveness, and ultimately achieving more objective and reliable evaluation.

### 2.1 Digital phenotyping data

Digital phenotyping data offers a new, ecological, and continuous approach to ADHD assessment. Digital Phenotyping data utilizes ubiquitous personal computing devices to passively collect users' daily behavioral data, aiming to capture more ecologically valid and objective behavioral patterns. These data include physical activity levels [step count, activity intensity (Park et al., 2023)], sleep patterns and circadian rhythms (Kim et al., 2023), social interaction patterns, screen usage habits, keyboard input patterns, and even voice features. AI can extract digital biomarkers reflecting attention, impulsivity, activity levels, mood fluctuations, and circadian rhythms from these dense, longitudinal, ecologically valid data streams (Lindhiem et al., 2022), providing possibilities for more natural, continuous, and

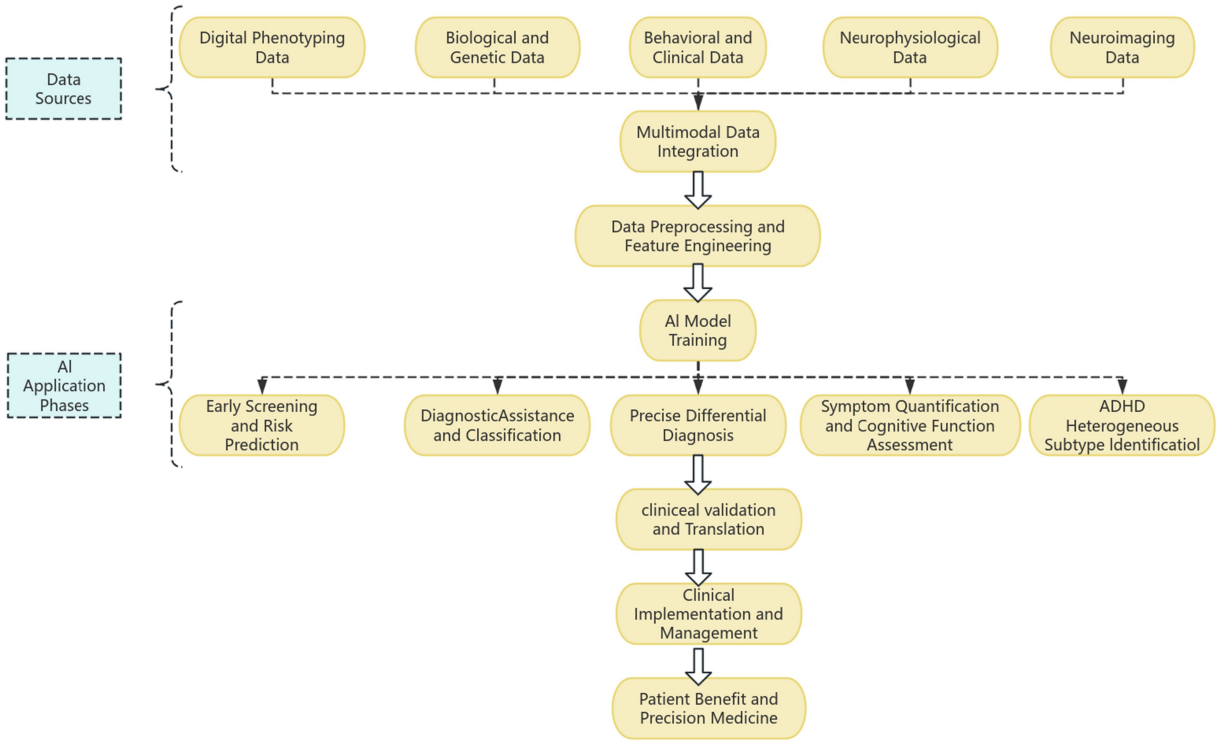


FIGURE 1
Schematic flowchart of AI-driven objective assessment for ADHD.

low-burden ADHD assessment and monitoring (Loh et al., 2022; Loh et al., 2024). Studies often compare digital behavioral patterns between individuals with ADHD and neurotypical controls to identify distinctive digital phenotypes. For instance, deviations in sleep regularity, increased screen time, or characteristic movement patterns can serve as objective indicators. The experimental paradigms typically involve passive data collection through smartphone apps, wearable sensors, or digital platforms, often over extended periods to capture real-world behavior.

However, the application of digital phenotyping data also faces severe challenges, including privacy protection, data noise processing, high variability in individual behavioral patterns, and how to effectively associate digital indicators with clinical symptoms. These data are typically collected passively in individuals' daily lives through "human-machine interfaces" such as smartphone applications and wearable sensors, and then analyzed by AI models, aiming to achieve more natural, continuous, and low-burden ADHD assessment and monitoring.

## 2.2 Biological and genetic data

Biological and genetic data provide molecular-level evidence for ADHD susceptibility and mechanistic research. Deeper objective information comes from biological and genetic data. Genomic data analysis is an important direction, where AI can be used to process high-dimensional genetic data [such as analyzing the association of single nucleotide polymorphisms (SNPs), copy number variations (CNVs), and other genetic variations with ADHD risk (Faraone and Larsson, 2019)] and build Polygenic Risk Score (PRS) models to assess individual genetic susceptibility (Ribasés et al., 2023). Studies in this area often compare genetic profiles between ADHD patients and healthy controls to identify risk variants or polygenic scores associated with the disorder.

Furthermore, some studies have begun to explore using AI to integrate other biomarkers, such as epigenetic data [DNA methylation (Silk et al., 2022)], metabolomic data [metabolite profiles in blood, urine (Predescu et al., 2024)], and even gut microbiome data (Nikolova et al., 2021) to assess ADHD likelihood. These approaches involve comparing biomarker concentrations or patterns between ADHD and control groups, often using advanced analytical techniques to identify subtle differences. Although these areas are still in preliminary exploration stages, they provide potential pathways for understanding ADHD's complex pathophysiological mechanisms and developing new biomarkers.

## 2.3 Behavioral and clinical data

Behavioral and clinical data serve as the most foundational source for ADHD assessment, which AI can enhance for objectivity. The most traditional and fundamental data source is behavioral and clinical data, which includes standardized parent/teacher/self-rating scale scores, structured or semi-structured clinical interview records, computer-based neuropsychological test performance, and direct behavioral observation data obtained through video analysis or motion capture technology. AI models, especially when combined with other modal data, aim to extract more stable and objective

assessment patterns from this information, going beyond the limitations of traditional scoring (Loh et al., 2022; Chiu et al., 2024). In research studies, these data are collected from both ADHD patients and neurotypical controls to identify characteristic behavioral patterns and cognitive deficits. Experimental paradigms often involve administering a battery of neuropsychological tests (e.g., tasks assessing attention, working memory, inhibitory control), structured clinical interviews based on diagnostic criteria (e.g., DSM-5 or ICD-11), and observations of behavior in controlled or naturalistic settings. AI is used to identify subtle deviations and complex interactions within these data, which serve as objective biomarkers distinguishing ADHD from other conditions or identifying specific symptom profiles.

## 2.4 Neurophysiological data

Neurophysiological data provides highly objective biomarkers for ADHD assessment. Neurophysiological data directly reflects brain activity or physiological states and has high inherent objectivity, making it an important source for AI to search for biomarkers. Electroencephalography (EEG) is a typical representative, where different frequency band power and their ratios in resting-state EEG, as well as functional connectivity indicators between brain regions, are widely considered potential biomarkers (Lenartowicz and Loo, 2014; Slater et al., 2022). Studies have explored brain network indices in boys with ADHD-C (Aydin et al., 2022). Additionally, the latency and amplitude of Event-Related Potentials (ERPs) are also extensively studied and used in AI classification models (Kim et al., 2021; Lutz et al., 2021). EEG data plays a crucial role in ADHD diagnosis, classification, and feature extraction, for instance, by utilizing autoencoder feature extraction and ResNet with a double augmented attention mechanism for ADHD classification (Bansal et al., 2025), or through Tunable Q-Factor Wavelet Transform for detection and classification of EEG signals (Joy et al., 2022). The impact of preprocessing and temporal segmentation on classification accuracy has also been highlighted (Garcia-Ponsoda et al., 2024). Recent research has also investigated automated EEG-based characterization of ADHD and CD using explainable deep neural network techniques (Loh et al., 2024), and the effect of methylphenidate treatment on neuro-cortical complexity in children with ADHD through embedding entropy estimations (Cetin et al., 2022).

Eye-Tracking technology indirectly reflects processes such as attentional control, impulse inhibition, and cognitive load by recording eye movement parameters during specific visual tasks (Maron et al., 2021). These parameters include saccadic latency, fixation duration, gaze patterns, and pupil dilation, often measured during sustained attention tasks (e.g., Continuous Performance Test) or visual search paradigms. Using ML to analyze these objective eye movement biomarkers has been successfully applied to assist in ADHD screening and diagnosis, by identifying distinctive patterns in ADHD patients compared to healthy controls (Yoo et al., 2024).

## 2.5 Neuroimaging data

Neuroimaging data offers objective evidence of brain structure and function for ADHD assessment. Neuroimaging data provides macroscopic objective views of brain structure and function. Structural Magnetic Resonance Imaging (sMRI) is used to measure morphological indicators such as volume, cortical thickness, and surface area of different brain regions. AI can be used for automatic segmentation of brain regions and extraction of these features, which have been found to be related to ADHD. Comparative studies between ADHD patients and control groups often reveal significant structural differences in areas like the prefrontal cortex, basal ganglia, and cerebellum, which are then used to build AI classification models (Qureshi et al., 2017; Albajara Sáenz et al., 2019).

Functional magnetic resonance imaging (fMRI), particularly resting-state fMRI (rs-fMRI), is used to study the connectivity patterns of intrinsic functional networks in the brain, such as the default mode network (DMN), salience network (SN), and central executive network (CEN). Task-based fMRI is used to observe brain activation patterns during specific cognitive tasks, like inhibitory control or working memory tasks, often comparing activation differences between ADHD and control groups. AI is commonly used to analyze complex spatiotemporal patterns and functional connectivity matrices extracted from fMRI data to distinguish ADHD patients from controls, identifying neural biomarkers indicative of the disorder (Qureshi et al., 2017; Albajara Sáenz et al., 2019; Pereira-Sanchez and Castellanos, 2021).

Diffusion Tensor Imaging (DTI) is used to assess the microstructural integrity and connectivity of white matter fiber tracts. Studies have shown that ADHD patients have altered white matter microstructure (Aoki et al., 2018; Parlatini et al., 2023), and AI can be used for fiber tract tracking and connectome analysis, utilizing DTI-derived features for classification. Biomedical images are also utilized for analysis in various medical contexts (Khekare et al., 2024).

## 2.6 Experimental paradigms and research designs

Beyond the raw data modalities, the specific experimental paradigms and overall research designs employed significantly influence the type of biomarkers identified and the generalizability of AI models in ADHD assessment. A diverse array of paradigms is utilized to probe different facets of ADHD symptomatology and underlying neurobiology.

For behavioral and neurophysiological data (e.g., EEG, eye-tracking), common paradigms include sustained attention tasks (e.g., Continuous Performance Test, CPT), inhibitory control tasks (e.g., Go/No-Go, Stop-Signal Task), working memory tasks (e.g., N-back), and reward processing tasks (Riccio et al., 2002; Owen et al., 2005; Aron, 2011). These tasks are designed to elicit specific cognitive or motor responses that are often impaired in individuals with ADHD, allowing for the extraction of objective performance metrics (e.g., reaction time variability, error rates) and physiological responses (e.g., ERP components, specific EEG frequency band power changes). The design choice (e.g., block design vs. event-related design in fMRI) influences the temporal resolution and the types of brain activity patterns that can be analyzed (Friston et al., 1998). Resting-state paradigms, where participants are asked to simply relax without engaging in specific tasks, are also widely used, particularly for fMRI and EEG, to assess intrinsic brain network connectivity and spontaneous brain activity (Broyd et al., 2009).

In neuroimaging studies (sMRI, fMRI, DTI), experimental designs typically involve comparing structural volumes, cortical thickness, or functional connectivity patterns between ADHD patients and matched healthy controls (Shaw et al., 2007; Castellanos and Aoki, 2016). Task-based fMRI paradigms in neuroimaging studies are often similar to those used for behavioral and neurophysiological assessments, but focus on mapping brain activation during the performance of these tasks. Longitudinal study designs are crucial for understanding developmental trajectories of ADHD and the long-term effects of interventions, while cross-sectional studies provide snapshots of differences at specific ages (Halperin and Schulz, 2006).

Data diversity and age sensitivity are critical considerations across all paradigms. Research designs must account for the heterogeneity of ADHD presentations across different developmental stages (e.g., childhood, adolescence, adulthood) and varying symptom profiles (e.g., predominantly inattentive, hyperactive–impulsive, combined) (Sonuga-Barke and Fairchild, 2012). Tailoring experimental tasks and data collection protocols to be age-appropriate ensures ecological validity and optimizes the capture of relevant biomarkers. The careful selection and rigorous implementation of experimental paradigms, along with consideration for age-specific manifestations and broader data diversity, are paramount for developing robust and clinically meaningful AI models in ADHD assessment (Figure 2).

# 3 Overview of artificial intelligence methodology for ADHD assessment

In objective ADHD assessment research, various AI are applied to analyze complex data and mine potential patterns. These technologies are primarily categorized into two branches: ML and DL. As model complexity increases, interpretability methods aimed at enhancing model transparency and credibility have gained increasing attention (Khekare et al., 2025).

## 3.1 Machine learning (ML)

Unsupervised learning and feature engineering are key to discovering potential subtypes and improving model interpretability in objective ADHD assessment. Traditional ML methods demonstrate robust performance in processing structured data and physiological or imaging features extracted through feature engineering, laying the foundation for achieving more objective and consistent assessment (Sharma and Chariar, 2024). Supervised Learning is the most commonly used paradigm, aiming to learn mapping relationships from input features to known labels. Common classification algorithms include Support Vector Machine (SVM), often employed for classifying ADHD based on neuroimaging features, and Random Forest, widely utilized for its robustness in handling diverse clinical and behavioral data, alongside XGBoost, and Logistic Regression. These are widely used to build models distinguishing ADHD patients from healthy controls (HC) or differentiating various ADHD subtypes (Jiang et al., 2017). Regression algorithms, such as Support Vector Regression (SVR) or Ridge Regression, can be used to predict continuous variables, for example, predicting specific scores on ADHD rating scales.

Unsupervised Learning aims to discover inherent structures or patterns from unlabeled data. Clustering algorithms (such as K-means, hierarchical clustering) are commonly used to explore potential

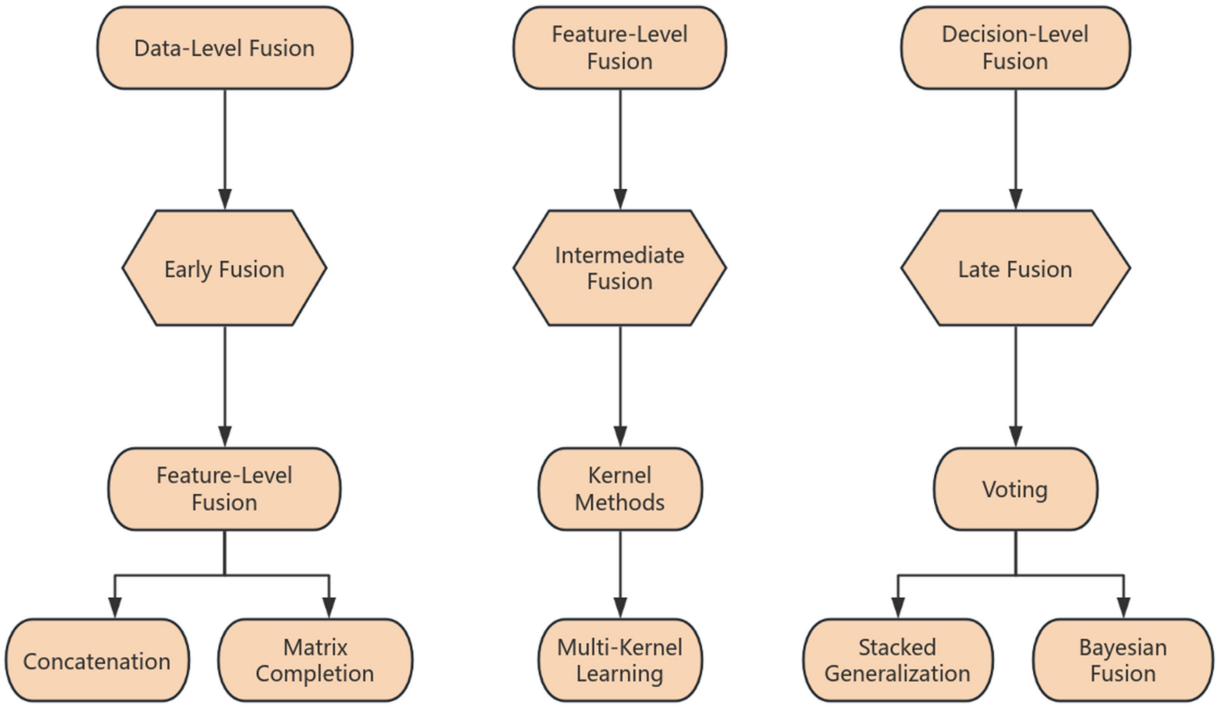


FIGURE 2
Classification framework of multimodal data fusion strategies for AI in ADHD assessment.

subtypes defined by biological or behavioral characteristics within ADHD patient populations, helping to deepen understanding of ADHD's high heterogeneity (Zhang et al., 2022). Additionally, when processing high-dimensional data, it is often necessary to combine Feature Selection (such as LASSO) and Dimensionality Reduction (such as Principal Component Analysis, PCA) techniques to reduce data redundancy, improve model performance, and enhance interpretability.

## 3.2 Deep learning (DL)

DL models have gained prominence for their particular expertise in automatically learning complex, hierarchical representations from raw or low-level feature data, especially suitable for processing high-dimensional, unstructured medical data. Convolutional Neural Networks (CNNs), for instance, are widely applied to analyze neuroimaging data (e.g., fMRI scans for brain activity patterns) and time-frequency representations of EEG signals (e.g., detecting abnormal brain rhythms in ADHD patients) (Quaak et al., 2021), due to their excellent performance in processing data with grid-like topological structures.

For sequence data, Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), can effectively capture temporal dependencies and are thus applied to analyze time series signals from EEG/MEG (e.g., identifying dynamic brain network connectivity associated with ADHD) (Chang et al., 2022; Wang et al., 2022), scan path data from eye tracking recordings, reaction time sequences in continuous cognitive tasks, and even language sequence features extracted from speech or interview texts. Transformer Models, initially successful in Natural Language Processing (NLP), have recently begun to be applied to biomedical sequence data, such as EEG signal analysis (e.g., capturing long-range dependencies in EEG rhythms for ADHD diagnosis) (He et al., 2023), fMRI time series, and clinical text analysis in Electronic Health Records (EHR) (Pillai et al., 2024). Cognitive-based ADHD detection using autoencoder-based hidden Markov models has also shown promising results (Mahesh et al., 2022).

For data representing relationships between entities, such as graph-structured brain functional or structural connectomes, Graph Neural Networks (GNNs) provide a unique analytical perspective, capable of simultaneously learning node features and graph topological structure, helping to understand ADHD-related brain network connectivity abnormalities from a network science perspective (Heck, 2021; Yang et al., 2023).

## 3.3 Machine learning (ML) and deep learning (DL) comparison

ML and DL, while both subsets of AI, possess distinct characteristics that dictate their suitability for various applications in ADHD assessment. DL, a specialized form of ML, utilizes multi-layered neural networks to learn intricate patterns directly from data. This section will compare their respective strengths, weaknesses, and common application scenarios.

Traditional ML models generally perform well with structured or tabular data. Their advantages include requiring less data for training and often offering greater interpretability, which is crucial for clinical trust and validation (Cao et al., 2023; Kerz et al., 2023). They also typically demand less computational power. However, a significant limitation is their reliance on manual feature engineering; for complex, high-dimensional, unstructured data, extensive pre-processing and feature extraction are necessary. In ADHD assessment, ML is widely applied for classification tasks using pre-extracted features from behavioral data, neurophysiological signals, or structural neuroimaging (Sharma and Chariar, 2024). They are also effective for identifying potential subtypes through clustering (Zhang et al., 2022) and for early risk prediction.

DL models excel at automatically extracting hierarchical features from raw, high-dimensional, and unstructured data, such as entire neuroimaging scans, raw EEG time series, or video data (Choi et al., 2020), significantly reducing the need for manual feature engineering. DL models scale exceptionally well with large datasets and have achieved state-of-the-art results in areas like image processing and natural language understanding. Conversely, they typically require substantial data for optimal performance, are computationally intensive, and their "black box" nature can pose challenges for clinical interpretability, though Explainable AI (XAI) techniques are actively addressing this (Amado-Caballero et al., 2023). DL is extensively used for analyzing raw neuroimaging data, processing eye-tracking scan paths, and interpreting clinical texts or speech patterns in ADHD research (Quaak et al., 2021; Chang et al., 2022; Wang et al., 2022; He et al., 2023; Pillai et al., 2024), particularly valuable when large, complex multimodal datasets are available.

Despite their differences, ML and DL are complementary. DL can serve as a powerful feature extractor, with its learned features then utilized by traditional ML models for final classification or regression. This hybrid approach can combine DL's representation learning capabilities with ML's interpretability or efficiency. The optimal choice between ML and DL often depends on the nature and volume of the available data, the specific task, and the importance of model interpretability within the clinical context.

## 3.4 Explainable artificial intelligence (XAI)

With increasing model complexity, especially the "black box" nature of DL models, their clinical application faces challenges in trust and adoption. Consequently, Explainable Artificial Intelligence (XAI) techniques have emerged and become increasingly important in ADHD research. Methods such as SHAP (SHapley Additive exPlanations) (Wicker et al., 2017) or Local Interpretable Model-agnostic Explanations (LIME) aim to reveal the key objective features or data patterns that models rely on for making specific predictions. By enhancing model transparency, XAI helps clinicians understand, validate, and trust model results, which is crucial for promoting reliable AI applications in mental health (Cao et al., 2023; Kerz et al., 2023). XAI not only helps clarify the basis for model decisions but may also assist in discovering new potential biomarkers, thereby deepening understanding of ADHD's objective pathophysiological mechanisms.

# 4 Specific applications of AI in ADHD screening, diagnosis and assessment

AI have been widely applied in multiple aspects of the ADHD assessment process, demonstrating great potential from early warning

to precise classification, with the core goal of improving the objectivity, accuracy, and efficiency of assessment.

## 4.1 AI-driven early screening and risk prediction

In the early stages of the process, AI is used for early screening and risk prediction. Given that early identification of ADHD is crucial for improving patient outcomes, AI models are being explored to integrate known early risk factors [such as premature birth, low birth weight, family genetic risk (Cabana-Domínguez et al., 2023), prenatal exposure factors (Ji et al., 2020)] and behavioral observation indicators during infancy [such as temperament characteristics, motor development milestones, early attention patterns (Gair et al., 2021)]. By building early risk prediction models, AI has the potential to identify high-risk children who need further professional assessment. For example, ML algorithms can analyze data extracted from parent-completed early screening questionnaires or community health records to predict the likelihood of individuals developing ADHD symptoms during school age, thereby facilitating early intervention.

## 4.2 AI-assisted diagnosis and classification

In the diagnostic phase, AI-assisted diagnosis and classification is one of its most widely applied areas. Research aims to build classifiers that can accurately distinguish ADHD patients from healthy controls (HC) based on objective data, typically using single or multimodal data. For instance, ML models like Support Vector Machines (SVMs) and Random Forests have been applied to analyze EEG signal features, achieving high accuracy in distinguishing ADHD from HC (Lenartowicz and Loo, 2014; Tenev et al., 2014; Öztoprak et al., 2017). DL models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) or Transformer Models, are increasingly used for analyzing raw neuroimaging data (fMRI, sMRI) and time-series EEG/fMRI signals, respectively, demonstrating robust performance in identifying ADHD patients and representing a current research hotspot (Riaz et al., 2020; Joy et al., 2022; Zhang-James et al., 2023; Firouzi et al., 2024; Loh et al., 2024; Taspinar and Ozkurt, 2024; Bansal et al., 2025). It is generally believed that models integrating multiple modalities can achieve better classification performance than single modalities and potentially provide more comprehensive and reliable assessment results (Qureshi et al., 2017; Riaz et al., 2018).

## 4.3 AI applications in precise differential diagnosis

AI's application in precise differential diagnosis is particularly prominent, addressing a core challenge in current ADHD clinical assessment. Due to overlapping symptom presentations, accurately distinguishing ADHD from other neurodevelopmental/mental disorders with similar presentations or comorbidities is extremely challenging. AI technology, with its ability to analyze complex patterns, aids in achieving more precise differential diagnosis.

In distinguishing ADHD from gifted children, their behavioral presentations may be very similar, leading to misdiagnosis. AI research is attempting to assist this differentiation process through more refined data analysis. For example, using ML models to analyze parent and teacher behavioral rating scale data to identify subtle but distinctive differences in specific behavioral patterns between the two groups (François-Sévigny et al., 2022). In the future, combining more objective biobehavioral indicators and using AI for comprehensive analysis may provide more objective and reliable evidence.

For distinguishing ADHD from anxiety/depression disorders, inattention is a common symptom, increasing the difficulty of differentiation. AI technology can assist in differentiation by integrating and analyzing multiple data sources. Using AI models to analyze neuroimaging data can help find differential biomarkers (Hettwer et al., 2022). In addition, applying NLP techniques to analyze clinical interview records or social media text can help identify language patterns that reflect different potential emotional states (Thorstad and Wolff, 2019); utilizing AI models to analyze neuroimaging data (such as structural or functional connectivity patterns of the amygdala and prefrontal cortex) can help find differential biomarkers; or analyzing physiological signals such as Heart Rate Variability (HRV) can explore their potential in objective differentiation (Sun et al., 2023).

In distinguishing ADHD from learning disabilities (SLD)/developmental coordination disorder (DCD), careful differentiation is needed due to frequent co-occurrence or similar symptoms. AI models have the potential to train classifiers by integrating multiple aspects of information (such as standardized neuropsychological test results, academic performance records (Czerniak et al., 2023), objective motor coordination assessment based on video analysis (Ouyang et al., 2024), etc.) to identify characteristic presentations under different disorder patterns and improve differentiation accuracy.

Regarding the challenge of distinguishing ADHD from Autism Spectrum Disorder (ASD), although their core symptoms differ, they have high comorbidity rates and behavioral overlap. AI provides multiple approaches, including using ML/DL to analyze neuroimaging data [comparing whole-brain functional connectivity patterns or specific brain region structural differences (Eslami et al., 2020; Song et al., 2020)], or analyzing other behavioral/physiological data (such as quantifying behavioral features in social interaction videos through computer vision, analyzing social scene gaze preferences using eye tracking, etc.), aiming to discover objective and effective biobehavioral markers that distinguish the two (Moreau et al., 2023).

Furthermore, recent research has begun to focus on the challenge of distinguishing ADHD inattentive type (ADHD-RI) from cognitive disengagement syndrome (CDS) (Durak et al., 2025). Using AI to analyze detailed neurocognitive test data or specific objective neurophysiological indicators, may help identify characteristic patterns that distinguish between these two conditions and improve differential diagnostic accuracy.

## 4.4 AI quantification of symptom severity and cognitive function

Beyond classification and differential diagnosis, AI is also applied to quantify symptom severity and cognitive function, which helps more precisely track disease progression and treatment response. For example, by analyzing objective behavioral or physiological data [such as CPT performance metrics, activity levels and sleep patterns collected by wearable devices (Lindhiem et al., 2022; Kim et al., 2023)], AI models have the potential to objectively quantify the severity of

individual inattention, hyperactivity, and impulsivity. This assessment may be more sensitive in capturing subtle changes than traditional subjective rating scales. Some studies also use ML models, based on objective biomarkers or other objective measurement data, to directly predict individual scores on standard ADHD rating scales (such as ADHD-RS) (Cervantes-Henríquez et al., 2022), aiming to provide more objective reference for clinical assessment.

In terms of cognitive function, AI algorithms can integrate and analyze objective performance data from multiple neuropsychological tests (especially those evaluating executive functions) (Duff and Sulla, 2015; Takahashi et al., 2024), potentially going beyond traditional single total score evaluation to more precisely characterize individual specific impairment levels and unique patterns in different cognitive subdomains, providing objective information for understanding individual differences and guiding personalized cognitive interventions.

### 4.5 AI-driven ADHD subtype identification

Finally, to address ADHD's high heterogeneity (Nigg et al., 2020), researchers use unsupervised learning (such as clustering algorithms) and other AI methods to identify ADHD subtypes. By analyzing objective biological or clinical features (such as based on resting-state fMRI functional connectivity patterns (Miranda et al., 2021) or EEG microstate features (Alves et al., 2024)), researchers attempt to identify subgroups within the ADHD patient population that have different brain network abnormalities, cognitive deficit patterns, or treatment responses (Karalunas and Nigg, 2020). These data-driven discovered subtypes aim to go beyond traditional clinical classification and lay the foundation for achieving more precise stratified diagnosis (stratified diagnosis) and personalized interventions.

## 5 Performance evaluation, advantages and current limitations

### 5.1 Overview of existing AI model performance

Based on a comprehensive review of existing literature, AI models demonstrate encouraging performance in specific ADHD assessment tasks. For instance, in binary classification tasks distinguishing ADHD from healthy controls (HC), models utilizing high-quality multimodal neuroimaging data and advanced DL algorithms often achieve accuracy or Area Under the Curve (AUC) values of 0.8 or higher (Quaak et al., 2021; Zhang-James et al., 2023; Sharma and Chariar, 2024; Taspinar and Ozkurt, 2024). However, when facing more challenging differential diagnostic tasks [such as distinguishing ADHD from ASD (Duda et al., 2017)] or ADHD internal subtype identification (Qureshi et al., 2016; Gao et al., 2023), model performance typically decreases, although still showing potential to surpass traditional methods.

### 5.2 Potential advantages of AI in ADHD assessment

AI offers several significant potential advantages in ADHD assessment. Despite this, AI application in ADHD assessment still offers significant potential advantages. First, AI analysis is primarily based on quantifiable objective data, thereby enhancing assessment objectivity, reducing reliance on subjective reports and clinical observations, and potentially improving the stability and reproducibility of assessment results. Second, AI can rapidly process and analyze large amounts of complex data, automating parts of the assessment process (such as signal analysis, image processing), potentially improving assessment efficiency, reducing clinician workload, and shortening assessment cycles. Third, AI's powerful pattern recognition capabilities, particularly DL, help mine potential objective biomarkers or feature combinations from high-dimensional, multimodal data that are difficult to identify using traditional statistical methods. This not only may improve diagnostic accuracy but also helps deepen understanding of ADHD's underlying pathophysiological mechanisms. Additionally, by identifying different biological or clinical subtypes and quantifying individual specific impairment patterns across multiple dimensions, AI provides possibilities for achieving more precise personalized assessment and ultimately stratified intervention strategies (precision medicine). Finally, AI screening or monitoring tools based on mobile computing platforms such as smartphones and wearable devices have advantages of low cost, convenient deployment, and the ability to collect data in natural environments (ecological validity), potentially improving assessment service accessibility, extending to resource-limited areas, or more conveniently integrating into patients' daily lives.

### 5.3 Limitations and challenges in current research

However, the path to applying AI in objective ADHD assessment is not smooth, and it is still in the development stage, facing numerous severe limitations and challenges that hinder the full realization of its potential and transformation into reliable clinical tools. One of the most prominent issues is data limitations and lack of standardization. Most published studies rely on small-scale, single-center datasets, with significant heterogeneity in data collection standards (such as equipment parameters, task paradigms) and preprocessing procedures across different studies, severely limiting model generalization ability and the reproducibility and comparability of research results (Zhang-James et al., 2023; Tian et al., 2024). The lack of large-scale, standardized, publicly available, and representative multicenter multimodal databases is a key bottleneck constraining the development of this field. Notably, most current AI model training datasets primarily focus on diagnosed ADHD patients and typically developing control groups, but lack systematic data collection on healthy variations (neurodiversity) that only exhibit inattentive/hyperactive traits without meeting ADHD diagnostic criteria (i.e., without significant functional impairment). This data deficiency constitutes an important confounding factor, which may lead to AI models struggling to distinguish between normal traits and the disorder, thus posing a risk of pathologizing neurodiversity. Future research urgently needs to construct representative datasets containing such 'intermediate groups' to improve the specificity and differentiation capabilities of AI models.

Closely related to this is the generally insufficient generalization ability and reliability of existing models. Models that perform excellently on specific internal datasets often show significant

performance degradation when applied to new, independent external datasets from different populations, devices, or collection environments. This overfitting problem and domain shift effect severely limit the direct application potential and reliability of models in real-world clinical settings (Dwyer and Koutsouleris, 2022).

Another major obstacle is the poor interpretability of many high-performance AI models (particularly DL models), which are like "black boxes" with opaque internal decision logic (Cao et al., 2023). This not only hinders clinicians' understanding and trust in model output results but also makes it difficult to gain new insights into disease mechanisms. This is one of the main barriers to clinical acceptance and adoption of models, although research on explainable AI (XAI) methods is emerging to address this challenge (Amado-Caballero et al., 2023).

Meanwhile, potential bias risks and fairness issues cannot be ignored. If training data fails to adequately represent the diversity of the target population (such as biases in age, gender, race, socioeconomic status, etc.), models may learn and even amplify these biases, leading to systematic unfairness or inaccuracy in assessment results for certain subgroups (Lee et al., 2021; Elendu et al., 2023), potentially exacerbating health inequalities.

Most critically, there is currently a severe lack of clinical validation. The vast majority of related research remains at the stage of technical exploration, retrospective analysis, or small-scale proof-of-concept, with an extreme lack of well-designed, rigorously executed prospective, multicenter studies with representative samples (such as diagnostic accuracy trials, randomized controlled trials) to evaluate the actual performance, safety, clinical utility, and cost-effectiveness of AI tools in real clinical scenarios (Rashid and Calhoun, 2020; Dwyer and Koutsouleris, 2022). The lack of high-quality clinical evidence is a core obstacle hindering the transformation of AI technology into clinical practice.

Finally, seamless implementation and integration of AI tools into existing clinical workflows also face multiple barriers in terms of technology (such as data interfaces, system compatibility), resources (such as equipment costs, personnel training), and processes (such as assessment time, report interpretation), and some methods relying on expensive equipment or complex processes have limited accessibility and scalability in resource-limited environments, which may lead to application inequality (Rashid and Calhoun, 2020; Dwyer and Koutsouleris, 2022).

A significant challenge lies in how AI evaluates and quantifies the crucial subjective criterion of "significant impairment and suffering" in ADHD diagnosis. Although AI cannot directly understand or feel an individual's internal experiences, it can indirectly infer or quantify the degree of functional impairment by analyzing digital phenotyping data (e.g., daily activity levels, sleep patterns, screen usage habits) (Lindhiem et al., 2022; Kim et al., 2023) and behavioral data (e.g., academic performance records, social interaction frequency, objective indicators of functional impairment in multi-source behavioral reports) (Chiu et al., 2024; Ouyang et al., 2024; Czerniak et al., 2023). However, these data serve only as auxiliary evidence for clinical assessment, and their interpretation must be comprehensively judged by experienced clinicians, integrating non-structured information such as the individual's socio-cultural background, family environment, and narrative reports. AI's value lies in identifying objective patterns that may indicate or accompany significant impairment, rather than replacing the understanding of subjective suffering or the final clinical judgment (Lee et al., 2021; Elendu et al., 2023; Cao et al., 2023).

# 6 Core challenges and future research directions

Future research urgently needs to establish large-scale, standardized multimodal databases through international collaboration, develop robust, interpretable, and fair AI models, and conduct rigorous, transparent clinical translation validation, with the ultimate goal of achieving responsible AI application in precise, objective, and personalized ADHD assessment and management.

## 6.1 Data level: building a solid foundation and promoting secure sharing

In order to facilitate the development of robust AI models, the creation of large-scale, standardized, multimodal databases and ensuring safe, ethical data sharing are necessary measures. First, through international collaboration and multidisciplinary cooperation, we need to jointly build high-quality, large-scale, standardized multimodal databases, which are the cornerstone of the entire field's development. The ideal database should include thousands or even tens of thousands of participants, have longitudinal tracking, multicenter sources, and good population representation, and integrate multidimensional information including behavior, cognition, neurophysiology (EEG, eye movements, etc.), neuroimaging (sMRI, fMRI, DTI, etc.), genetics, environmental exposure, and digital phenotyping. The key is to reach consensus and strictly implement data collection protocols, quality control standards, and annotation specifications to ensure data quality, promote research reproducibility and model comparability, and provide the necessary foundation for developing more powerful and generalizable AI models (Tian et al., 2024). Leveraging emerging technologies such as high-precision motion capture to quantify behavioral phenotypes is also worth exploring (Wu et al., 2025). Meanwhile, under strict compliance with ethical and privacy regulations, it is necessary to actively explore secure and compliant data sharing mechanisms and privacy protection technologies. Privacy-preserving computing technologies such as Federated Learning (Loftus et al., 2022; Brauneck et al., 2023) allow collaborative model training without directly transmitting original sensitive data, potentially breaking down data silos and improving model reliability and generalization ability by utilizing broader and more diverse data. Establishing clear data governance frameworks and access control protocols is also indispensable. Hybrid frameworks integrating blockchain, IoT, and cloud computing also offer secure and scalable solutions for healthcare data management (Kumaran et al., 2024).

## 6.2 Algorithm level: improving performance, reliability, and trustworthiness

At the algorithmic level, the core challenge lies in continuously improving AI model performance, reliability, and trustworthiness through advanced algorithms, multimodal data fusion, and considerations for interpretability and fairness. This means continuously developing AI algorithms that can better handle data heterogeneity, have strong anti-interference capabilities, and perform stably and reliably on unseen data and different populations, to

enhance model reliability and generalization ability. Transfer learning, domain adaptation/generalization, adversarial training, and ML combined with causal inference are important research directions. Meanwhile, Algorithmic Fairness (Lee et al., 2021) must be incorporated as a core consideration in model design, training, and evaluation, adopting bias detection and mitigation techniques to ensure fair performance across different subgroups (classified by gender, race, etc.) and avoid exacerbating health inequalities.

Second, we should deepen multimodal fusion and information integration strategies. Given the complexity of ADHD, single-modal information is limited, requiring the development of more advanced and effective multimodal data fusion strategies to fully utilize the complementarity and synergistic effects of information from different sources, building more comprehensive and precise assessment models.

Third, we need to focus on breakthroughs in interpretability (XAI) and causal inference. Vigorously developing and applying XAI technology is a core link in winning clinical trust, promoting adoption, and potentially discovering new knowledge. Future XAI should provide explanations that conform to clinical logic, are intuitive and easy to understand, and have practical guidance for decision-making. Exploring the combination of ML with causal inference, attempting to move from association to causal exploration, may help deepen understanding of pathological mechanisms.

Finally, to adapt to mobile health applications or resource-limited scenarios, we need to develop lightweight models and explore edge computing applications, reducing model complexity through techniques such as model compression and knowledge distillation, and exploring computation on terminal devices to improve real-time performance, accessibility, and privacy.

### 6.3 Validation and translation level: rigorous evaluation and promoting clinical implementation

For AI tools to be clinically viable, rigorous and transparent clinical validation studies are essential, alongside the establishment of clear regulatory frameworks to ensure their safe and effective translation into practice. This necessitates going beyond simple internal validation and conducting rigorous, transparent, independent clinical validation studies. Large-scale, prospective, multicenter, independent external validation studies and clinical trials following recognized reporting standards [such as TRIPOD (Alewine et al., 2015), STARD-AI (Muehlematter et al., 2023)] should be designed and implemented. Meanwhile, special attention needs to be paid to bridging the gap between laboratory performance and real-world practice performance (performance gap) (Jurva et al., 2025), and conducting calibration evaluation of prediction model probability outputs to ensure their confidence accurately reflects actual risk (Benster et al., 2025).

It is also necessary to develop user-centered clinical decision support systems (CDSS), integrating validated models into user-friendly platforms or EHR systems that conform to clinical workflows, providing clear, interpretable, actionable auxiliary information, and emphasizing human-machine collaboration, with final decision-making authority resting with doctors.

Additionally, establishing clear regulatory science frameworks and approval pathways is crucial. With the development of AI/ML-driven medical devices (including SaMD), regulatory agencies (such as FDA, EMA) are improving corresponding frameworks (Pennello et al., 2021; Muehlematter et al., 2023), clarifying requirements for performance, safety, effectiveness, data quality, algorithm transparency, risk management, and post-market surveillance, providing clear paths for innovative product market access while ensuring patient safety.

### 6.4 Ethical and social level: advocating responsible innovation and ensuring universal accessibility

Addressing ethical and social considerations, such as responsible innovation, universal accessibility, and mitigating potential biases and privacy risks, is paramount for the equitable application of AI in healthcare. The primary principle is to follow strict ethical standards and data governance principles, ensuring full dynamic informed consent throughout the process involving ADHD patient data, respecting privacy rights and data autonomy, and establishing sound data security and governance frameworks to prevent risks.

Proactive identification, assessment, and management of algorithm biases and promotion of fairness (Lee et al., 2021) are necessary. At all stages of the model lifecycle, fairness metrics and bias mitigation techniques must be adopted to ensure AI tools have equal effectiveness and safety across different subgroups, avoiding technology exacerbating inequalities.

Clarification of human-machine collaboration models and responsibility attribution frameworks is needed, clearly defining AI's role as an auxiliary tool, emphasizing that final decisions are made comprehensively by clinicians, and establishing clear accountability mechanisms.

Finally, focus on technology accessibility and bridging the digital divide is necessary. When promoting AI tools, measures should be taken to ensure the universality of technology, avoiding exacerbating health service inequalities due to equipment, technical barriers, or digital skill differences (Elendu et al., 2023).

## 7 Conclusion

AI provides unprecedented opportunities to address the many challenges in current ADHD diagnosis and assessment, particularly in improving assessment objectivity, efficiency, and achieving more precise differential diagnosis and personalized assessment. This paper systematically reviewed the current applications and research progress of AI in objective ADHD assessment, emphasizing how AI leverages multimodal data, including behavioral data, neurophysiological signals, neuroimaging features, genetic information, and emerging digital phenotyping, across various assessment stages. The significant potential of AI in tasks such as early screening and risk prediction, diagnostic assistance and classification, precise differential diagnosis, symptom severity quantification, and the identification of heterogeneous ADHD subtypes was highlighted. By integrating these diverse data sources, AI models offer a pathway to overcome the limitations of traditional subjective assessment methods, paving the way for more objective and consistent diagnostic outcomes.

However, transforming these promising AI research results into reliable clinical practice tools that can truly benefit the vast number of ADHD patients and their families in a safe, effective, and fair manner still faces severe challenges. These challenges span multiple dimensions, including obtaining, standardizing, and securely sharing high-quality data; improving algorithm reliability, generalization ability, and interpretability; conducting rigorous and sufficient clinical validation; developing effective translation strategies; and addressing comprehensive ethical, legal, and social impact considerations. These intertwined complexities necessitate a collaborative effort from the entire research, development, application, and regulatory ecosystem to effectively overcome.

Looking ahead, the future direction for AI in objective ADHD assessment is clear and promising. Substantial and synergistic progress must be made through close interdisciplinary collaboration. This includes continued efforts in building a solid, standardized data foundation, promoting responsible algorithm innovation that prioritizes interpretability and fairness, implementing rigorous and transparent clinical validation studies, and improving sound ethical and governance frameworks. Only through such concerted efforts can AI truly become a powerful, reliable, and trustworthy partner for clinicians in assessing and managing the complexity of ADHD, ultimately promoting a deeper understanding of the disorder and significantly improving the long-term well-being and quality of life for affected individuals and their families.

## Author contributions

CZ: Writing – review & editing, Writing – original draft. YX: Writing – review & editing. RL: Writing – review & editing. HL: Writing – review & editing. MZ: Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China Youth Fund (no. 82205190), the China Postdoctoral Science Foundation General Project (no. 2023M731027), a special grant from the China Postdoctoral Science Foundation (no. 2024T170253), the Henan Province Postdoctoral Project (no. HN2022096), the Henan Provincial Health Commission National Traditional Chinese Medicine Inheritance and Innovation Center Scientific Research Special Project (no. 20237X7X1073), and the Chinese Medicine Scientific Research Special Project of Henan Province (no. 2021JDZX105), the Postdoctoral Research Program in Henan Province (no. 202003096), and the China Postdoctoral Science Foundation (no. 2021M690939).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Albajara Sáenz, A., Villemonteix, T., and Massat, I. (2019). Structural and functional neuroimaging in attention-deficit/hyperactivity disorder. *Dev. Med. Child Neurol.* 61, 399–405. doi: 10.1111/dmcn.14050

Alewine, C., Hassan, R., and Pastan, I. (2015). Advances in anticancer immunotoxin therapy. *Oncologist* 20, 176–185. doi: 10.1634/theoncologist.2014-0358

Alves, L. M., Côco, K. F., De Souza, M. L., and Ciarelli, P. M. (2024). Identifying ADHD and subtypes through microstates analysis and complex networks. *Med. Biol. Eng. Comput.* 62, 687–700. doi: 10.1007/s11517-023-02948-2

Amado-Caballero, P., Casaseca-de-la-Higuera, P., Alberola-López, S., Andrés-de-Llano, J. M., López-Villalobos, J. A., and Alberola-López, C. (2023). Insight into ADHD diagnosis with deep learning on Actimetry: quantitative interpretation of occlusion maps in age and gender subgroups. *Artif. Intell. Med.* 143:102630. doi: 10.1016/j.artmed.2023.102630

Aoki, Y., Cortese, S., and Castellanos, F. X. (2018). Research review: diffusion tensor imaging studies of attention-deficit/hyperactivity disorder: meta-analyses and reflections on head motion. *J. Child Psychol. Psychiatry* 59, 193–202. doi: 10.1111/jcpp.12778

Aron, A. R. (2011). From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. *Biol. Psychiatry* 69, e55–e68. doi: 10.1016/j.biopsych.2010.07.024

Aydin, S., Cetin, F. H., Uytun, M. C., Babadagi, Z., Gueven, A. S., and Isik, Y. (2022). Comparison of domain specific connectivity metrics for estimation brain network indices in boys with ADHD-C. *Biomed. Signal Process. Control.* 76:3626. doi: 10.1016/j.bspc.2022.103626

Bansal, J., Gangwar, G., Aljaidi, M., Alkoradees, A., and Singh, G. (2025). EEG-based ADHD classification using autoencoder feature extraction and res net with double augmented attention mechanism. *Brain Sci.* 15:10095. doi: 10.3390/brainsci15010095

Benster, L. L., Weissman, C. R., Suprani, F., Toney, K., Afshar, H., Stapper, N., et al. (2025). Predictive modeling of response to repetitive transcranial magnetic stimulation in treatment-resistant depression. *Transl. Psychiatry* 15:160. doi: 10.1038/s41398-025-03380-w

Biederman, J., Mick, E., and Faraone, S. V. (2000). Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type. *Am. J. Psychiatry* 157, 816–818. doi: 10.1176/appi.ajp.157.5.816

Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M. M., Bakhtiari, M., Völker, U., Baumbach, J., et al. (2023). Federated machine learning, privacy-enhancing technologies, and data protection Laws in medical research: scoping review. *J. Med. Internet Res.* 25:e41588. doi: 10.2196/41588

Broyd, S. J., Demanuele, C., Debener, S., Helps, S. K., James, C. J., and Sonuga-Barke, E. J. (2009). Default-mode brain dysfunction in mental disorders: a systematic review. *Neurosci. Biobehav. Rev.* 33, 279–296. doi: 10.1016/j.neubiorev.2008.09.002

Cabana-Domínguez, J., Llonga, N., Arribas, L., Alemany, S., Vilar-Ribó, L., Demontis, D., et al. (2023). Transcriptomic risk scores for attention deficit/

hyperactivity disorder. *Mol. Psychiatry* 28, 3493–3502. doi: 10.1038/s41380-023-02200-1

Cao, M., Martin, E., and Li, X. (2023). Machine learning in attention-deficit/hyperactivity disorder: new approaches toward understanding the neural mechanisms. *Transl. Psychiatry* 13:236. doi: 10.1038/s41398-023-02536-w

Castellanos, F. X., and Aoki, Y. (2016). Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: a science in development. *Biol Psychiatry Cogn Neurosci Neuroimaging* 1, 253–261. doi: 10.1016/j.bpsc.2016.03.004

Cervantes-Henríquez, M. L., Acosta-López, J. E., Martinez, A. F., Arcos-Burgos, M., Puentes-Rozo, P. J., and Vélez, J. I. (2022). Machine learning prediction of ADHD severity: association and linkage to ADGRL3, DRD4, and SNAP25. *J. Atten. Disord.* 26, 587–605. doi: 10.1177/10870547211015426

Cetin, F. H., Baris Usta, M., Aydin, S., and Guven, A. S. (2022). A case study on EEG analysis: embedding entropy estimations indicate the decreased neuro-cortical complexity levels mediated by methylphenidate treatment in children with ADHD. *Clin. EEG Neurosci.* 53, 406–417. doi: 10.1177/15500594211064008

Chang, Y., Stevenson, C., Chen, I. C., Lin, D. S., and Ko, L. W. (2022). Neurological state changes indicative of ADHD in children learned via EEG-based LSTM networks. *J. Neural Eng.* 19:016021. doi: 10.1088/1741-2552/ac4f07

Chiu, Y. H., Lee, Y. H., Wang, S. Y., Ouyang, C. S., Wu, R. C., Yang, R. C., et al. (2024). Objective approach to diagnosing attention deficit hyperactivity disorder by using pixel subtraction and machine learning classification of outpatient consultation videos. *J. Neurodev. Disord.* 16:71. doi: 10.1186/s11689-024-09588-z

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., and Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Transl. Vis. Sci. Technol.* 9:14. doi: 10.1167/tvst.9.2.14

Cortese, S., Moreira-Maia, C. R., St Fleur, D., Morcillo-Peñalver, C., Rohde, L. A., and Faraone, S. V. (2016). Association between ADHD and obesity: a systematic review and meta-analysis. *Am. J. Psychiatry* 173, 34–43. doi: 10.1176/appi.ajp.2015.15020266

Czerniak, L. L., Liebel, S. W., Zhou, H., Garcia, G. P., Lavieri, M. S., McCrea, M. A., et al. (2023). Sensitivity and specificity of the ImPACT neurocognitive test in collegiate athletes and US military service academy cadets with ADHD and/or LD: findings from the NCAA-DoD CARE consortium. *Sports Med.* 53, 747–759. doi: 10.1007/s40279-022-01768-y

Doshi, J. A., Hodgkins, P., Kahle, J., Sikirica, V., Cangelosi, M. J., Setyawan, J., et al. (2012). Economic impact of childhood and adult attention-deficit/hyperactivity disorder in the United States. *J. Am. Acad. Child Adolesc. Psychiatry* 51, 990–1002.e2. doi: 10.1016/j.jaac.2012.07.008

Duda, M., Haber, N., Daniels, J., and Wall, D. P. (2017). Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Transl. Psychiatry* 7:e1133. doi: 10.1038/tp.2017.86

Duff, C. T., and Sulla, E. M. (2015). Measuring executive function in the differential diagnosis of attention-deficit/hyperactivity disorder: does it really tell us anything? *Appl. Neuropsychol. Child* 4, 188–196. doi: 10.1080/21622965.2013.848329

Durak, S., Tahıllıoğlu, A., Yazan Songür, Ç., Çoban, M., Varol, B., and Ercan, E. S. (2025). Differentiating pure cognitive disengagement syndrome and attention-deficit/hyperactivity disorder-restrictive inattentive presentation with respect to depressive symptoms, autistic traits, and neurocognitive profiles. *Appl. Neuropsychol. Child* 1-11, 1–11. doi: 10.1080/21622965.2025.2493812

Dwyer, D., and Koutsouleris, N. (2022). Annual research review: translational machine learning for child and adolescent psychiatry. *J. Child Psychol. Psychiatry* 63, 421–443. doi: 10.1111/jcpp.13545

Elendu, C., Amaechi, D. C., Elendu, T. C., Jingwa, K. A., Okoye, O. K., John Okah, M., et al. (2023). Ethical implications of AI and robotics in healthcare: a review. *Medicine (Baltimore)* 102:e36671. doi: 10.1097/md.0000000000036671

Eslami, T., Almuqhim, F., Raiker, J. S., and Saeed, F. (2020). Machine learning methods for diagnosing autism spectrum disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: a survey. *Front. Neuroinform.* 14:575999. doi: 10.3389/fninf.2020.575999

Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., et al. (2021). The world federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. *Neurosci. Biobehav. Rev.* 128, 789–818. doi: 10.1016/j.neubiorev.2021.01.022

Faraone, S. V., and Larsson, H. (2019). Genetics of attention deficit hyperactivity disorder. *Mol. Psychiatry* 24, 562–575. doi: 10.1038/s41380-018-0070-0

Firouzi, M., Kazemi, K., Ahmadi, M., Helfroush, M. S., and Aarabi, A. (2024). Enhanced ADHD classification through deep learning and dynamic resting state fMRI analysis. *Sci. Rep.* 14:24473. doi: 10.1038/s41598-024-74282-y

François-Sévigny, J., Pilon, M., and Gauthier, L. A. (2022). Differences in parents and teachers' perceptions of behavior manifested by gifted children with ADHD compared to gifted children without ADHD and non-gifted children with ADHD using the Conners 3 scale. *Brain Sci.* 12:1571. doi: 10.3390/brainsci12111571

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage* 7, 30–40. doi: 10.1006/nimg.1997.0306

Furman, L. (2005). What is attention-deficit hyperactivity disorder (ADHD)? *J. Child Neurol.* 20, 994–1002. doi: 10.1177/08830738050200121301

Gair, S. L., Brown, H. R., Kang, S., Grabell, A. S., and Harvey, E. A. (2021). Early development of comorbidity between symptoms of ADHD and anxiety. *Res Child Adolesc Psychopathol* 49, 311–323. doi: 10.1007/s10802-020-00724-6

Gao, Y., Ni, H., Chen, Y., Tang, Y., and Liu, X. (2023). Subtype classification of attention deficit hyperactivity disorder with hierarchical binary hypothesis testing framework. *J. Neural Eng.* 20:056015. doi: 10.1088/1741-2552/acf523

Garcia-Ponsoda, S., Mate, A., and Trujillo, J. (2024). Refining ADHD diagnosis with EEG: the impact of preprocessing and temporal segmentation on classification accuracy. *Comput. Biol. Med.* 183:109305. doi: 10.1016/j.compbiomed.2024.109305

Gosselin, M., Dubé, M., Simard, M., Vandal, N., Lunghi, C., and Sirois, C. (2025). Trends in medication use during the COVID-19 pandemic in Quebec, Canada. *Sci. Rep.* 15:14744. doi: 10.1038/s41598-025-98592-x

Haan, E., Westmoreland, K. E., Schellhas, L., Sallis, H. M., Taylor, G., Zuccolo, L., et al. (2022). Prenatal smoking, alcohol and caffeine exposure and offspring externalizing disorders: a systematic review and meta-analysis. *Addiction* 117, 2602–2613. doi: 10.1111/add.15858

Halperin, J. M., and Schulz, K. P. (2006). Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder. *Psychol. Bull.* 132, 560–581. doi: 10.1037/0033-2909.132.4.560

Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Lee, M. J., and Asadi, H. (2018). eDoctor: machine learning and the future of medicine. *J. Intern. Med.* 284, 603–619. doi: 10.1111/joim.12822

He, Y., Wang, X., Yang, Z., Xue, L., Chen, Y., Ji, J., et al. (2023). Classification of attention deficit/hyperactivity disorder based on EEG signals using a EEG-transformer model(∗). *J. Neural Eng.* 20:056013. doi: 10.1088/1741-2552/acf7f5

Heck, T. G. (2021). Cardioprotection during chemotherapy: prospects of antioxidant strategies. *Arq. Bras. Cardiol.* 117, 1159–1160. doi: 10.36660/abc.20210914

Hettwer, M. D., Larivière, S., Park, B. Y., van den Heuvel, O. A., Schmaal, L., Andreassen, O. A., et al. (2022). Coordinated cortical thickness alterations across six neurodevelopmental and psychiatric disorders. *Nat. Commun.* 13:6851. doi: 10.1038/s41467-022-34367-6

Ji, Y., Azuine, R. E., Zhang, Y., Hou, W., Hong, X., Wang, G., et al. (2020). Association of Cord Plasma Biomarkers of in utero acetaminophen exposure with risk of attention-deficit/hyperactivity disorder and autism Spectrum disorder in childhood. *JAMA Psychiatry* 77, 180–189. doi: 10.1001/jamapsychiatry.2019.3259

Jiang, Y., Abiri, R., and Zhao, X. (2017). Tuning up the old brain with new tricks: attention training via Neurofeedback. *Front. Aging Neurosci.* 9:52. doi: 10.3389/fnagi.2017.00052

Joy, R. C., George, S. T., Rajan, A. A., Subathra, M. S. P., Sairamya, N. J., Prasanna, J., et al. (2022). Detection and classification of ADHD from EEG signals using tunable Q-factor wavelet transform. *J Sens* 2022, 1–17. doi: 10.1155/2022/3590973

Jurva, A., Singh, B., Qian, H., Wang, Z., Jacobs, M. L., Dhima, K., et al. (2025). Increased frontoparietal activity related to lower performance in neuropsychological assessment of working memory. *NeuroImage* 313:121240. doi: 10.1016/j.neuroimage.2025.121240

Kaplan, B. J., Dewey, D. M., Crawford, S. G., and Wilson, B. N. (2001). The term comorbidity is of questionable value in reference to developmental disorders: data and theory. *J. Learn. Disabil.* 34, 555–565. doi: 10.1177/002221940103400608

Karalunas, S. L., and Nigg, J. T. (2020). Heterogeneity and subtyping in attention-deficit/hyperactivity disorder-considerations for emerging research using person-centered computational approaches. *Biol. Psychiatry* 88, 103–110. doi: 10.1016/j.biopsych.2019.11.002

Kerz, E., Zanwar, S., Qiao, Y., and Wiechmann, D. (2023). Toward explainable AI (XAI) for mental health detection based on language behavior. *Front. Psych.* 14:1219479. doi: 10.3389/fpsyt.2023.1219479

Khekare, G., Kaliyamoorthi, P., Parikh, S., Rao, K., Kulkarni, S. D., Khekare, U., et al. (2024). Biomedical images utilised to analyse the expression of death receptor ligands. *J. Environ. Prot. Ecol.* 25, 2109–2118.

Khekare, G., Yenurkar, G., Turukmane, A. V., Ameta, G. K., Sharma, P., and Phulre, A. K. (2025). "Artificial intelligence algorithms for better decision-making" in Multi-criteria decision-making and optimum design with machine learning Eds. Van Thanh Tien Nguyen, Nhut T.M. Vo, Van Chinh Truong, Van-Thuc Nguyen (Boca Raton: CRC Press), 252–262.

Kim, W. P., Kim, H. J., Pack, S. P., Lim, J. H., Cho, C. H., and Lee, H. J. (2023). Machine learning-based prediction of attention-deficit/hyperactivity disorder and sleep problems with wearable data in children. *JAMA Netw. Open* 6:e233502. doi: 10.1001/jamanetworkopen.2023.3502

Kim, J. S., Lee, Y. J., and Shim, S. H. (2021). What event-related potential tells us about brain function: child-adolescent psychiatric perspectives. *Soa Chongsonyon Chongsin Uihak* 32, 93–98. doi: 10.5765/jkacap.210012

Kumaran, K. S., Khekare, G., Athitya, M. T., Aakash, A., Pranav, M. A., and Chidambaram, N. H. (2024). "A hybrid framework for secure and scalable healthcare:

integrating Blockchain, IoT, and cloud computing" in Congress on intelligent systems (Springer), 1–13. doi: 10.1007/978-981-96-2694-6_1

Lange, S., Rehm, J., Anagnostou, E., and Popova, S. (2018). Prevalence of externalizing disorders and autism Spectrum disorders among children with fetal alcohol Spectrum disorder: systematic review and meta-analysis. *Biochem. Cell Biol.* 96, 241–251. doi: 10.1139/bcb-2017-0014

Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H. C., et al. (2021). Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging* 6, 856–864. doi: 10.1016/j.bpsc.2021.02.001

Lenartowicz, A., and Loo, S. K. (2014). Use of EEG to diagnose ADHD. *Curr. Psychiatry Rep.* 16:498. doi: 10.1007/s11920-014-0498-0

Lindhiem, O., Goel, M., Shaaban, S., Mak, K. J., Chikersal, P., Feldman, J., et al. (2022). Objective measurement of hyperactivity using Mobile sensing and machine learning: pilot study. *JMIR Form Res* 6:e35803. doi: 10.2196/35803

Loftus, T. J., Ruppert, M. M., Shickel, B., Ozrazgat-Baslanti, T., Balch, J. A., Efron, P. A., et al. (2022). Federated learning for preserving data privacy in collaborative healthcare research. *Digit Health* 8:205520762211344. doi: 10.1177/20552076221134455

Loh, H. W., Ooi, C. P., Barua, P. D., Palmer, E. E., Molinari, F., and Acharya, U. R. (2022). Automated detection of ADHD: current trends and future perspective. *Comput. Biol. Med.* 146:105525. doi: 10.1016/j.compbiomed.2022.105525

Loh, H. W., Ooi, C. P., Oh, S. L., Barua, P. D., Tan, Y. R., Acharya, U. R., et al. (2024). ADHD/CD-NET: automated EEG-based characterization of ADHD and CD using explainable deep neural network technique. *Cogn. Neurodyn.* 18, 1609–1625. doi: 10.1007/s11571-023-10028-2

Lutz, M. C., Kok, R., and Franken, I. H. A. (2021). Event-related potential (ERP) measures of error processing as biomarkers of externalizing disorders: a narrative review. *Int. J. Psychophysiol.* 166, 151–159. doi: 10.1016/j.ijpsycho.2021.06.002

Mahesh, T., Goswami, T., Sriramulu, S., Sharma, N., Kumari, A., and Khekare, G. (2022). Cognitive based attention deficit hyperactivity disorder detection with ability assessment using auto encoder based hidden Markov model. *Int. J. Commun. Networks Inf. Secur.* 14, 53–65. doi: 10.17762/ijcnis.v14i2.5464

Maron, D. N., Bowe, S. J., Spencer-Smith, M., Mellahn, O. J., Perrykkad, K., Bellgrove, M. A., et al. (2021). Oculomotor deficits in attention deficit hyperactivity disorder (ADHD): a systematic review and comprehensive meta-analysis. *Neurosci. Biobehav. Rev.* 131, 1198–1213. doi: 10.1016/j.neubiorev.2021.10.012

Miranda, L., Paul, R., Pütz, B., Koutsouleris, N., and Müller-Myhsok, B. (2021). Systematic review of functional MRI applications for psychiatric disease subtyping. *Front. Psych.* 12:665536. doi: 10.3389/fpsyt.2021.665536

Moreau, C., Deruelle, C., and Auzias, G. (2023). Machine learning for neurodevelopmental disorders. *Mach. Learn. Brain Disord.* 31, 977–1007. doi: 10.1007/978-1-0716-3195-9_31

Mowlem, F. D., Rosenqvist, M. A., Martin, J., Lichtenstein, P., Asherson, P., and Larsson, H. (2019). Sex differences in predicting ADHD clinical diagnosis and pharmacological treatment. *Eur. Child Adolesc. Psychiatry* 28, 481–489. doi: 10.1007/s00787-018-1211-3

Muehlematter, U. J., Bluethgen, C., and Vokinger, K. N. (2023). FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. *Lancet Digit Health* 5, e618–e626. doi: 10.1016/s2589-7500(23)00126-7

Nayarisseri, A., Khandelwal, R., Tanwar, P., Madhavi, M., Sharma, D., Thakur, G., et al. (2021). Artificial intelligence, big data and machine learning approaches in Precision Medicine & Drug Discovery. *Curr. Drug Targets* 22, 631–655. doi: 10.2174/1389450122999210104205732

Nigg, J. T., Karalunas, S. L., Feczko, E., and Fair, D. A. (2020). Toward a revised nosology for attention-deficit/hyperactivity disorder heterogeneity. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5, 726–737. doi: 10.1016/j.bpsc.2020.02.005

Nikolova, V. L., Smith, M. R. B., Hall, L. J., Cleare, A. J., Stone, J. M., and Young, A. H. (2021). Perturbations in gut microbiota composition in psychiatric disorders: a review and Meta-analysis. *JAMA Psychiatry* 78, 1343–1354. doi: 10.1001/jamapsychiatry.2021.2573

Ouyang, C. S., Yang, R. C., Wu, R. C., Chiang, C. T., Chiu, Y. H., and Lin, L. C. (2024). Objective and automatic assessment approach for diagnosing attention-deficit/hyperactivity disorder based on skeleton detection and classification analysis in outpatient videos. *Child Adolesc. Psychiatry Ment. Health* 18:60. doi: 10.1186/s13034-024-00749-5

Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59. doi: 10.1002/hbm.20131

Öztoprak, H., Toycan, M., Alp, Y. K., Arıkan, O., Doğutepe, E., and Karakaş, S. (2017). Machine-based classification of ADHD and nonADHD participants using time/frequency features of event-related neuroelectric activity. *Clin. Neurophysiol.* 128, 2400–2410. doi: 10.1016/j.clinph.2017.09.105

Park, C., Rouzi, M. D., Atique, M. M. U., Finco, M. G., Mishra, R. K., Barba-Villalobos, G., et al. (2023). Machine learning-based aggression detection in children with ADHD using sensor-based physical activity monitoring. *Sensors (Basel)* 23:4949. doi: 10.3390/s23104949

Parlatini, V., Itahashi, T., Lee, Y., Liu, S., Nguyen, T. T., Aoki, Y. Y., et al. (2023). White matter alterations in attention-deficit/hyperactivity disorder (ADHD): a systematic review of 129 diffusion imaging studies with meta-analysis. *Mol. Psychiatry* 28, 4098–4123. doi: 10.1038/s41380-023-02173-1

Pennello, G., Sahiner, B., Gossmann, A., and Petrick, N. (2021). Discussion on "approval policies for modifications to machine learning-based software as a medical device: a study of bio-creep" by Jean Feng, Scott Emerson, and Noah Simon. *Biometrics* 77, 45–48. doi: 10.1111/biom.13381

Pereira-Sanchez, V., and Castellanos, F. X. (2021). Neuroimaging in attention-deficit/hyperactivity disorder. *Curr. Opin. Psychiatry* 34, 105–111. doi: 10.1097/yco.0000000000000669

Pillai, M., Posada, J., Gardner, R. M., Hernandez-Boussard, T., and Bannett, Y. (2024). Measuring quality-of-care in treatment of young children with attention-deficit/hyperactivity disorder using pre-trained language models. *J. Am. Med. Inform. Assoc.* 31, 949–957. doi: 10.1093/jamia/ocae001

Predescu, E., Vaidean, T., Rapciuc, A. M., and Sipos, R. (2024). Metabolomic markers in attention-deficit/hyperactivity disorder (ADHD) among children and adolescents-a systematic review. *Int. J. Mol. Sci.* 25:4385. doi: 10.3390/ijms25084385

Prelaj, A., Miskovic, V., Zanitti, M., Trovo, F., Genova, C., Viscardi, G., et al. (2024). Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review. *Ann. Oncol.* 35, 29–65. doi: 10.1016/j.annonc.2023.10.125

Quaak, M., van de Mortel, L., Thomas, R. M., and van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: systematic review and meta-analysis. *Neuroimage Clin* 30:102584. doi: 10.1016/j.nicl.2021.102584

Qureshi, M. N., Min, B., Jo, H. J., and Lee, B. (2016). Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study. *PLoS One* 11:e0160697. doi: 10.1371/journal.pone.0160697

Qureshi, M. N. I., Oh, J., Min, B., Jo, H. J., and Lee, B. (2017). Multi-modal, multi-measure, and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain MRI. *Front. Hum. Neurosci.* 11:157. doi: 10.3389/fnhum.2017.00157

Rashid, B., and Calhoun, V. (2020). Towards a brain-based predictome of mental illness. *Hum. Brain Mapp.* 41, 3468–3535. doi: 10.1002/hbm.25013

Riaz, A., Asad, M., Alonso, E., and Slabaugh, G. (2018). Fusion of fMRI and non-imaging data for ADHD classification. *Comput. Med. Imaging Graph.* 65, 115–128. doi: 10.1016/j.compmedimag.2017.10.002

Riaz, A., Asad, M., Alonso, E., and Slabaugh, G. (2020). DeepFMRI: end-to-end deep learning for functional connectivity and classification of ADHD using fMRI. *J. Neurosci. Methods* 335:108506. doi: 10.1016/j.jneumeth.2019.108506

Ribasés, M., Mitjans, M., Hartman, C. A., Soler Artigas, M., Demontis, D., Larsson, H., et al. (2023). Genetic architecture of ADHD and overlap with other psychiatric disorders and cognition-related phenotypes. *Neurosci. Biobehav. Rev.* 153:105313. doi: 10.1016/j.neubiorev.2023.105313

Riccio, C. A., Reynolds, C. R., Lowe, P., and Moore, J. J. (2002). The continuous performance test: a window on the neural substrates for attention? *Arch. Clin. Neuropsychol.* 17, 235–272. doi: 10.1093/arclin/17.3.235

Salari, N., Ghasemi, H., Abdoli, N., Rahmani, A., Shiri, M. H., Hashemian, A. H., et al. (2023). The global prevalence of ADHD in children and adolescents: a systematic review and meta-analysis. *Ital. J. Pediatr.* 49:48. doi: 10.1186/s13052-023-01456-1

Sharma, C. M., and Chariar, V. M. (2024). Diagnosis of mental disorders using machine learning: literature review and bibliometric mapping from 2012 to 2023. *Heliyon* 10:e32548. doi: 10.1016/j.heliyon.2024.e32548

Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., et al. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc. Natl. Acad. Sci. USA* 104, 19649–19654. doi: 10.1073/pnas.0707741104

Shaw, M., Hodgkins, P., Caci, H., Young, S., Kahle, J., Woods, A. G., et al. (2012). A systematic review and analysis of long-term outcomes in attention deficit hyperactivity disorder: effects of treatment and non-treatment. *BMC Med.* 10:99. doi: 10.1186/1741-7015-10-99

Silk, T., Dipnall, L., Wong, Y. T., and Craig, J. M. (2022). Epigenetics and ADHD. *Curr. Top. Behav. Neurosci.* 57, 269–289. doi: 10.1007/7854_2022_339

Slater, J., Joober, R., Koborsy, B. L., Mitchell, S., Sahlas, E., and Palmer, C. (2022). Can electroencephalography (EEG) identify ADHD subtypes? A systematic review. *Neurosci. Biobehav. Rev.* 139:104752. doi: 10.1016/j.neubiorev.2022.104752

Song, J. W., Yoon, N. R., Jang, S. M., Lee, G. Y., and Kim, B. N. (2020). Neuroimaging-based deep learning in autism Spectrum disorder and attention-deficit/hyperactivity disorder. *Soa Chongsonyon Chongsin Uihak* 31, 97–104. doi: 10.5765/jkacap.200021

Sonuga-Barke, E. J., and Fairchild, G. (2012). Neuroeconomics of attention-deficit/hyperactivity disorder: differential influences of medial, dorsal, and ventral prefrontal brain networks on suboptimal decision making? *Biol. Psychiatry* 72, 126–133. doi: 10.1016/j.biopsych.2012.04.004

Sun, J., Dong, Q. X., Wang, S. W., Zheng, Y. B., Liu, X. X., Lu, T. S., et al. (2023). Artificial intelligence in psychiatry research, diagnosis, and therapy. *Asian J. Psychiatr.* 87:103705. doi: 10.1016/j.ajp.2023.103705

Takahashi, N., Ono, T., Omori, Y., Iizumi, M., Kato, H., Kasuno, S., et al. (2024). Assessment of executive functions using a 3D-video game in children and adolescents with ADHD. *Front. Psych.* 15:1407703. doi: 10.3389/fpsyt.2024.1407703

Taspinar, G., and Ozkurt, N. (2024). A review of ADHD detection studies with machine learning methods using rsfMRI data. *NMR Biomed.* 37:e5138. doi: 10.1002/nbm.5138

Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Müller, A., and Candrian, G. (2014). Machine learning approach for classification of ADHD adults. *Int. J. Psychophysiol.* 93, 162–166. doi: 10.1016/j.ijpsycho.2013.01.008

Thorstad, R., and Wolff, P. (2019). Predicting future mental illness from social media: a big-data approach. *Behav. Res. Methods* 51, 1586–1600. doi: 10.3758/s13428-019-01235-z

Tian, L., Zheng, H., Zhang, K., Qiu, J., Song, X., Li, S., et al. (2024). Structural or/and functional MRI-based machine learning techniques for attention-deficit/hyperactivity disorder diagnosis: a systematic review and meta-analysis. *J. Affect. Disord.* 355, 459–469. doi: 10.1016/j.jad.2024.03.111

Wang, C., Wang, X., Jing, X., Yokoi, H., Huang, W., Zhu, M., et al. (2022). Towards high-accuracy classifying attention-deficit/hyperactivity disorders using CNN-LSTM model. *J. Neural Eng.* 19:046015. doi: 10.1088/1741-2552/ac7f5d

Wicker, P., Downward, P., and Lopez, F. (2017). Does regional disadvantage affect health-related sport and physical activity level? A multi-level analysis of individual behaviour. *Eur J Sport Sci* 17, 1350–1359. doi: 10.1080/17461391.2017.1376119

Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics* 9, 490–499. doi: 10.1007/s13311-012-0135-8

Wu, J., Guo, W., Luo, K., He, L., and Yang, G. (2025). Research progress in motor assessment of neurodegenerative diseases driven by motion capture data. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* 42, 396–403. doi: 10.7507/1001-5515.202403004

Yang, Y., Ye, C., and Ma, T. (2023). A deep connectome learning network using graph convolution for connectome-disease association study. *Neural Netw.* 164, 91–104. doi: 10.1016/j.neunet.2023.04.025

Yoo, J. H., Kang, C., Lim, J. S., Wang, B., Choi, C. H., Hwang, H., et al. (2024). Development of an innovative approach using portable eye tracking to assist ADHD screening: a machine learning study. *Front. Psych.* 15:1337595. doi: 10.3389/fpsyt.2024.1337595

Zhang, M., Huang, Y., Jiao, J., Yuan, D., Hu, X., Yang, P., et al. (2022). Transdiagnostic symptom subtypes across autism spectrum disorders and attention deficit hyperactivity disorder: validated by measures of neurocognition and structural connectivity. *BMC Psychiatry* 22:102. doi: 10.1186/s12888-022-03734-4

Zhang-James, Y., Razavi, A. S., Hoogman, M., Franke, B., and Faraone, S. V. (2023). Machine learning and MRI-based diagnostic models for ADHD: are we there yet? *J. Atten. Disord.* 27, 335–353. doi: 10.1177/10870547221146256

Zysset, A., Robin, D., Albermann, K., Dratva, J., Hotz, S., Wieber, F., et al. (2023). Diagnosis and management of ADHD: a pediatric perspective on practice and challenges in Switzerland. *BMC Pediatr.* 23:103. doi: 10.1186/s12887-023-03873-x

frontiers | Frontiers in Pediatrics

TYPE Review
PUBLISHED 04 December 2025
DOI 10.3389/fped.2025.1629795

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Kee Hyun Cho,
Kangwon National University,
Republic of Korea
Tng Chang Kwok,
University of Nottingham, United Kingdom

*CORRESPONDENCE
Dongyun Liu
✉ liudongyun006@163.com

†These authors have contributed equally to this work

RECEIVED 16 May 2025
REVISED 07 November 2025
ACCEPTED 18 November 2025
PUBLISHED 04 December 2025

CITATION
Zhang X, Wang A, Xu R and Liu D (2025) Enhancing bronchopulmonary dysplasia prediction in preterm infants using artificial intelligence and multimodal data integration. Front. Pediatr. 13:1629795. doi: 10.3389/fped.2025.1629795

COPYRIGHT
© 2025 Zhang, Wang, Xu and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Enhancing bronchopulmonary dysplasia prediction in preterm infants using artificial intelligence and multimodal data integration

Xinkai Zhang[1,2†], Anping Wang[3†], Rongwei Xu[1,2] and Dongyun Liu[2*]

[1]Qingdao Medical College, Qingdao University, Qingdao, China, [2]Department of Pediatrics Medical Center, The Affiliated Hospital of Qingdao University, Qingdao, China, [3]MS in Computer Information Systems, Boston University Metropolitan College, Boston, MA, United States

Bronchopulmonary dysplasia (BPD) remains a prevalent respiratory condition among preterm infants, with its development influenced by a combination of perinatal and postnatal factors. The development of artificial intelligence (AI) and machine learning (ML) technologies has provided new ideas for building BPD prediction models based on multimodal data (such as clinical information, physiological signals, imaging data, biomarkers, and omics data). This article systematically reviewed the research progress of AI in BPD prediction, analyzed representative models and key tools (such as RTI BPD Outcome Estimator), and assessed their performance and limitations in actual clinical settings. It also sorted out the challenges faced by AI models in clinical translation, including data standardization, model interpretability, system integration capabilities, model update mechanisms, and ethical and legal issues. To address the clinical need of "moving from prediction to intervention", this article discussed the PALM translation framework (Predict–Act–Learn–Monitor) organized around key clinical nodes. In the future, it is necessary to strengthen multi-center data sharing, develop privacy protection technologies such as federated learning, and build a design, validation, integration, regulation, and feedback closed-loop management system to help AI models move from risk prediction to precise intervention, ultimately improving the clinical outcomes of children with BPD.

KEYWORDS

**artificial intelligence, bronchopulmonary dysplasia, machine learning, multimodal data, predictive models**

## 1 Introduction

Bronchopulmonary dysplasia (BPD) represents a significant respiratory condition frequently observed in preterm infants, with its incidence inversely associated with gestational age. Studies have shown that the incidence of BPD in extremely premature infants with a gestational age of 22–28 weeks is as high as 60.7% (1). The introduction and widespread implementation of antenatal corticosteroids (ACS) and pulmonary surfactant (PS) therapies have contributed to improved survival rates among extremely preterm infants—those with birth weights under 1,000 g and gestational ages below 28 weeks. However, this shift has also been accompanied by a rise in cases of "new BPD," a phenotype marked by impaired alveolar development and diminished pulmonary microvascularization, with relatively milder alveolar and airway injury or fibrosis (2).

Infants with BPD frequently exhibit adverse long-term outcomes, including growth and developmental delays, as well as persistent impairments in cardiopulmonary and neurodevelopmental functioning (3, 4). In recent years, artificial intelligence (AI) has emerged as a transformative tool in healthcare, particularly in the analysis and management of complex, high-volume clinical data. This review provides a comprehensive overview of the application of AI and related technologies in the context of BPD prediction and management among preterm infants. Emphasis is placed on improving strategies for early identification and intervention, enhancing clinical decision-making, and supporting favorable long-term outcomes in this vulnerable population. Some previous reviews have addressed the ethical challenges and interpretability issues of AI models in the neonatal field (5, 6), but there is still a lack of a systematic assessment and framework analysis centered on BPD and integrated with specific clinical practice procedures. Therefore, this review, on the basis of summarizing the existing BPD prediction models, focuses on the key obstacles in their clinical translation process. At the same time, this article discussed the PALM translation framework (Predict–Act–Learn–Monitor), which is organized around key clinical nodes, with the aim of providing theoretical basis and practical reference for promoting the clinical implementability of BPD prediction models.

# 2 Overview of the research status of artificial intelligence in BPD prediction

## 2.1 Advantages of integration of multi-source data and AI models

In Neonatal Intensive Care Units (NICUs), preterm infants constitute a population particularly well-suited for AI-based monitoring of BPD. Currently, predictive models frequently incorporate physiological parameters such as heart rate, respiratory rate, oxygen saturation, and ventilator settings (7, 8). While NICU equipment facilitates continuous acquisition of large-scale physiological data, conventional monitoring techniques often lack the capacity to deliver accurate and real-time analyses due to the high variability of these indicators in preterm infants.

AI offers significant advantages in the prediction and management of BPD by enabling the integration and interpretation of complex, multidimensional clinical datasets. For instance, AI models embedded within electronic health record systems have been utilized to identify early risk signals for BPD from vital sign trends, ventilator metrics, and additional physiological variables. These models support clinical decision-making by contributing to early risk stratification, prognostic assessment, and the development of individualized treatment strategies (9, 10). Moreover, AI systems are capable of integrating diverse data sources, including imaging, genomic data, and biochemical profiles, to examine inter-organ interactions, thereby enhancing both the accuracy and comprehensiveness of predictive models (11, 12). The development of clinically applicable BPD risk assessment tools requires minimization of false-positive rates and maximization of clinical interpretability. It remains essential that healthcare professionals synthesize AI-generated outputs with comprehensive clinical evaluations to ensure that decision-making remains clinically grounded and contextually appropriate.

## 2.2 Pathway for AI model development and validation

In the context of BPD prediction, supervised learning is the predominant methodology. That is, based on the data with known labels, the model is trained to achieve the mapping between features and outcomes. In terms of application division of labor: structured clinical variables and laboratory data often adopt machine learning approaches (12–15). Medical imaging and high-dimensional continuous monitoring signals are more suitable for deep learning (13). During the model development phase, hyperparameter optimization is typically used to improve generalization performance. For model performance evaluation, in addition to AUROC (Area Under the Receiver Operating Characteristic Curve), it is recommended to report AUPRC (Area Under the Precision-Recall Curve) simultaneously to better reflect the ability to identify positive cases. Meanwhile, the Brier score is used to assess the consistency between probabilistic predictions and actual outcomes (16). The model validation phase mainly includes: internal validation (e.g., cross-validation, bootstrapping), which is used to evaluate consistency and robustness; external validation (using data from different institutions or time periods), which is used to test transferability across regions and systems (17). At present, most BPD-related AI models are still in the development or internal validation stage, lacking cross-center external validation and real-world data testing, which limits their feasibility for clinical application. Ultimately, evidence still relies on randomized controlled trials (RCTs) to assess their real-world benefits and safety for clinical outcomes, thereby supporting routine application.

## 2.3 Analysis of the progress of AI prediction models, tools and related literature

Currently, RTI BPD Outcome Estimator (https://neonatal.rti.org) is a widely used clinical tool for BPD prediction, featuring clear parameter settings, transparent logic, and a user-friendly interface, making it convenient for bedside rapid application

Abbreviations

BPD, bronchopulmonary dysplasia; AI, artificial intelligence; ML, machine learning; NICHD, National Institute of Child Health and Human Development; ACS, antenatal corticosteroid; NICU, neonatal intensive care unit; DL, deep learning; AUC, area under the curve; RCT, randomized controlled trial; NLP, natural language processing; EHR, electronic health records; VEGF, vascular endothelial growth factor; PLGF, placental growth factor; IL, interleukin; SIRI, systemic inflammatory response index; KL-6, Krebs von den Lungen-6; NT-proBNP, N-terminal pro-B-type natriuretic peptide; MAP, mean airway pressure.

(18). However, its limitation lies in that the model is constructed based on a small number of static variables (such as gestational age, birth weight, and oxygen concentration), and it is difficult to incorporate continuous physiological signals, medical images, and multi-omics data in the NICU. It also lacks a dynamic update mechanism and adaptive capabilities, which limits its adaptability in complex scenarios.

In recent years, the application of AI in BPD research has become increasingly widespread. Lei et al. used a random forest model with 6 variables to predict BPD, achieving an AUC of 0.929 (19). Dai et al. combined clinical and genetic information, and the AUC values for predicting BPD and severe BPD were 0.915 and 0.907, respectively—outperforming the model that used only clinical data (20). A study based on chest x-rays demonstrated that the AI model can accurately predict BPD within 24 h of birth, outperforming expert diagnosis (21); another transcriptomics study also showed that a small number of genes can effectively predict BPD, expanding the data dimension (22). However, a high AUC alone is insufficient to support clinical translation. First, in the context of class imbalance, AUC tends to be overestimated. It is recommended to report AUPRC simultaneously to reflect the value of identifying minority classes (patients with diseases). Second, features must be accessible at the time of prediction to avoid data/label leakage. Recent reviews and empirical studies have shown that leakage significantly exaggerates effect estimates, especially in small sample sizes (23); third, external validation must be completed across three dimensions—geography, time, and equipment—and the sample size, distribution of gestational age/body weight, definition of BPD, and its sensitivity analysis should be transparently disclosed; fourth, in terms of evaluation dimensions, the focus should expand from discrimination to calibration and clinical net benefit: beyond AUC/AUPRC, regularly report the Brier score, expected calibration error (ECE), and calibration slope/intercept, use Decision Curve Analysis (DCA) to quantify net benefits at preset adoption thresholds (24), and simultaneously provide the PPV (Positive Predictive Value), NPV (Negative Predictive Value), and alert rate at these thresholds to align with the action thresholds of NICUs; the aforementioned practices collectively enhance the model's robustness and adoptability, serving as a prerequisite for its entry into real-world evaluation and clinical trials.

## 3 Multimodal data sources supporting AI modeling: current status and integration potential

The integration of diverse data modalities is anticipated to significantly enhance both the understanding and predictive accuracy of AI models for BPD. Figure 1 illustrates the connections between various data sources and BPD: clinical informatics (gestational age, birth weight, duration of oxygen therapy/ventilation, and structured EHR variables) provides patients' baseline information and disease progression context; continuous physiological monitoring (HR, RR, $SpO_2$, and blood pressure) reflects dynamic changes in respiratory/circulatory status; imaging (CXR, LUS, and CT/MRI) offers objective signs of lung aeration and interstitial changes; laboratory biomarkers (inflammatory factors, KL-6, NT-proBNP, VEGF/PLGF, etc.) indicate inflammation, epithelial/interstitial damage, and pulmonary vascular load; and multi-omics (metabolomics, microbiomics, genomics/transcriptomics) reveal pathways and molecular-level mechanisms. On this basis, cross-modal synergy

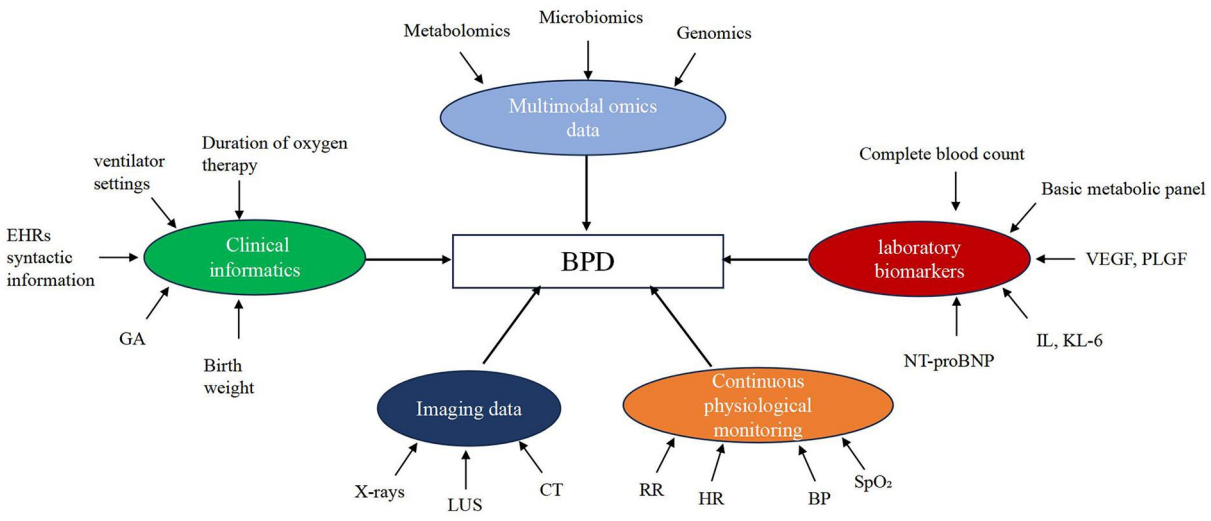


FIGURE 1
Multiple omics data sources used for modeling in BPD prediction by AI. BPD, bronchopulmonary dysplasia; EHRs, electronic health records; GA, gestational age; CT, computed tomography; RR, respiratory rate; HR, heart rate; BP, blood pressure; $SpO_2$, peripheral capillary oxygen saturation; VEGF, vascular endothelial growth factor; PLGF, placental growth factor; IL, interleukin; KL-6, Krebs von den Lungen-6; NT-proBNP, N-terminal pro-B-type natriuretic peptide; LUS, lung ultrasound.

has clear practical value: the combination of imaging and physiological monitoring can be used for early warning and ventilation strategy optimization; the integration of physiological monitoring and laboratory biomarkers helps identify "re-elevation of inflammation-pulmonary vascular load," guiding fluid management or pharmacological interventions (25); existing studies have shown that models integrating clinical information with multi-omics data yield better performance than single-modal models, suggesting that this direction has feasibility and scalable potential (20).

## 3.1 Clinical informatics data

BPD is recognized as a multifactorial disorder, and reliance on individual predictors is insufficient for accurate risk estimation, multivariate integration offers greater advantages (26). Observable clinical information corresponding to alveolar simplification is mainly reflected in the distribution and fluctuation of $FiO_2$ and $SpO_2$, the frequency and duration of intermittent hypoxemia (IH), and the oxygen saturation index (OSI). Phenotypes associated with increased pulmonary vascular load often manifest as persistently high oxygen demand under similar ventilation settings and difficulty in weaning from ventilation. In NICUs, high-resolution data (e.g., ventilator parameters, continuous $SpO_2$, blood gas results, OSI) and low-resolution data (e.g., demographics, medical history, medication records) together form longitudinal evidence, enabling early identification of ventilation-oxygenation mismatch and trends in respiratory work (27). Based on this, it is recommended to extract robust temporal features (median, variability, trend slope, etc.) from EHRs across three clinical time windows: ≤24 h after birth, 3–7 days after birth, and approximately 28 days after birth, with the features limited to those accessible at the time of prediction. The model output should provide calibrated individualized risk and report clinical net benefit under preset thresholds, facilitating direct alignment with ventilator parameter fine-tuning, oxygen therapy target setting, and weaning assessment.

It is important to emphasize that despite the richness of available clinical data, multiple challenges persist, including the labor-intensive nature of manual data entry, missing values, changes in disease definitions over time, difficulties in capturing complete longitudinal records, potential biases in data collection, and concerns related to data governance and privacy. These limitations underscore the necessity of advancing data acquisition and storage methodologies to enhance the robustness of BPD-focused research and model development.

## 3.2 The role of imaging and imaging AI in BPD

It complements the sequential variables such as EHR and enables repeatable quantification. In the NICU, chest X-ray (CXR) and bedside lung ultrasound (LUS) are the most accessible: CXR is convenient for longitudinal follow-up of lung gas volume and interstitial changes, while LUS can be frequently performed bedside and sensitively reflect lung gas volume and edema status (28). CT and MRI are more suitable for evaluating the pulmonary microstructure and ventilation/perfusion conditions in critical decision-making or research scenarios, but they are often limited by factors such as radiation exposure, time requirements, and the need for sedation (29).

The imaging AI technology is mainly divided into radiomics based on manual features and end-to-end deep learning. Usually, it needs to be combined with lung field segmentation, quality control and clinical variables for multimodal fusion to enhance the robustness and interpretability of the model (30). Research shows that the deep learning model based on neonatal CXR can output risk signals comparable to or even better than those of expert interpretations at an extremely early stage of birth. However, cross-center external validation is still insufficient, and it is urgently necessary to evaluate its generalizability under a unified annotation and standardized collection framework (21). Based on this, this article further proposes: The imaging AI evidence should be integrated into specific clinical time windows and intervention nodes, rather than just being a comparison of methodological performance. Thus, the experimental "imaging-AI" is transformed into a clinical pathway of "imaging-AI-guided intervention pathway".

## 3.3 Structured biomarker data and the prospects of AI integration

Blood biomarkers are easy to standardize and structurally input, making them important potential input features for AI models. Existing evidence shows the following: Angiogenic factors (VEGF, PLGF) are crucial for pulmonary vascular and lung development (31). Serum VEGF/PLGF is expected to be used for early risk identification and intervention stratification (32). Inflammatory factors (IL-4/6/8, TNF-α, IL-33) reflect the persistence of inflammation and are associated with a high risk of BPD (33, 34). Systemic immune-inflammatory indices (e.g., NLR, SIRI) are significantly elevated at birth and at 36 weeks of gestational age in infants with moderate-to-severe BPD, indicating an increased systemic inflammatory load (35, 36). Among lung injury biomarkers, KL-6 indicates alveolar epithelial and pulmonary interstitial damage (37) and its level at 14 days after birth shows good discriminative power for BPD (AUC = 0.88) (38). NT-proBNP reflects volume load/pulmonary vascular stress (39); in very low birth weight infants, NT-proBNP > 2.264 ng/L at 14 days after birth can early indicate the risk of BPD (sensitivity: 100%, specificity: 86%, AUC = 0.93) (40).

For clinical application, biomarker modeling should align with disease course time windows and pathological pathways: Within ≤24 h after birth, prioritize indicators reflecting alveolar/interstitial damage and early inflammation (e.g., KL-6, IL-6/IL-8) for extremely early risk stratification; from 3 to 7 days after birth, combine NT-proBNP with inflammatory profiles to identify the possible "re-elevation of inflammation-pulmonary vascular load," so as to indicate the timing of respiratory

support and fluid/pharmacological intervention; at approximately 28 days after birth, integrate persistent oxygen dependence, nutritional/growth indicators, and VEGF/PLGF to assess phenotypes of persistent damage and microvascular development disorders and plan follow-up. Integration with other modalities should follow the principle of "starting with the easy and moving to the difficult": When sample size is limited or modalities are frequently missing, adopt late fusion (train each modality independently, then perform confidence weighting or stacking integration after outputting calibrated probabilities)—this facilitates external validation and allows the use of Decision Curve Analysis (DCA) to demonstrate clinical adoptability; after multi-center data and annotations are improved, transition to mid-level fusion (jointly model "pathway-level" biomarkers with imaging and temporal physiological data within the same time window) to obtain cross-modal information gain and mechanistic interpretability (41).

Although the above studies mostly employed traditional statistical models, their data structure is highly suitable for being incorporated into AI systems for multi-feature modeling. Although the blood biomarkers have shown certain potential initially, due to the small sample size of the research and the high clinical heterogeneity, they are currently still in the research stage and have not yet entered the routine clinical application process. Larger-scale multi-center studies are needed to verify their efficacy (36).

### 3.4 A large number of biological information data and future AI modeling directions

Multi-omics provides a key supplement for revealing the molecular and pathway-level mechanisms of BPD. At the metabolomics level: Premature infants with BPD show elevated levels of alanine, serine, taurine, and citrulline in bronchoalveolar lavage fluid on the first day after birth (42). Meanwhile, studies have shown that the levels of hydroxy fatty acids and oxidized fatty acids in the amniotic fluid of premature infants with BPD are higher than those in the non-BPD group (43, 44). At the microbiomics level: Research indicates that compared with the non-BPD group, the pulmonary microbial community of the BPD group shows a trend of increasing proportion of Firmicutes and decreasing proportion of Proteobacteria over time (45). Additionally, gut microbiota may affect BPD development through modulation of systemic immunity and inflammatory responses.

At the genomics/transcriptomics level: Variations in angiogenesis pathways (e.g., VEGF) and abnormalities in airway exosomal miRNAs (such as miR-876–3p, miR-20a-5p) are associated with BPD susceptibility (43, 46).

To advance omics evidence from "description" to "computational evidence usable for prediction and stratification," two types of alignment need to be achieved. First is methodological alignment: perform batch effect correction and scale unification for multi-center/multi-platform data, and reduce high-dimensional omics features to pathway/module-level representations (e.g., aggregating genes/metabolites into activity scores for pathways such as "inflammation, oxidative stress, and angiogenesis") to lower dimensionality and facilitate mechanism-based interpretation. Second is temporal alignment: align omics data with concurrent imaging data and continuous physiological signals according to clinical prediction windows (0–24 h, 3–7 d, ≈28 d) for joint modeling, ensuring features are accessible at the time of prediction and avoiding data/label leakage.

For integration strategies, progression should follow the principle of "late fusion first, then mid-level fusion." When sample size and modality accessibility are limited, late fusion should be adopted first. As multi-center samples and annotation standards are gradually refined, transition to mid-level fusion: jointly characterize pathway-level omics, imaging, and other data within the same time window, and use multi-modal Transformers (with cross-modal attention and mask training to enhance robustness to missing modalities) or heterogeneous graph neural networks (constructing graphs of "gene/metabolite/microbiota—pathway—phenotype" and performing message passing for importance attribution) to improve mechanistic interpretability (47, 48). Existing small-sample prototypes have shown that combined metabolomic-microbiomic multi-omics exhibit cross-modal consistent signals in inflammation/oxidative stress pathways (49). However, regardless of the approach, multi-omics integration can only be transformed into interpretable, verifiable, and implementable decision support in NICU settings if discrimination, calibration, net benefit, and robustness all meet the required standards.

## 4 Analysis of the actual obstacles in the clinical application of AI prediction models

### 4.1 Data-related limitations

The implementation of DL models within clinical settings necessitates seamless integration with healthcare information systems, including EHRs and imaging databases. However, this process often encounters challenges such as high technical barriers, long implementation periods, large capital investment, and complex compliance requirements. Especially in developing regions with lower levels of informatization, the data foundation is weak, and obtaining high-quality samples is limited. Moreover, the training data has insufficient representativeness for some populations, all of which affect the generalization performance of the model (50). Furthermore, medical records often suffer from missing information and incorrect input. The data structures from different sources such as EHR, wearable devices, and medical images vary greatly, making it difficult to integrate across institutions, which undermines the stability and accuracy of the model (51–54). Currently, the related research on BPD lacks open and shared datasets and code libraries, which limits the verification of results and the re-development of models, and hinders its generalizability across different institutions.

## 4.2 Model-related limitations

1. **Lack of interpretability ("Black Box" issue):** Deep learning-based AI models often function as "black boxes" due to their inherent complexity and opaque decision-making mechanisms. The lack of transparency may undermine clinician trust, particularly in critical clinical decisions such as treatment discontinuation or experimental interventions, where comprehensible rationale is essential for adoption (55).
2. **Insufficient generalization ability and external validation:** AI models are susceptible to overfitting, wherein they capture noise or irrelevant patterns specific to the training dataset. While such models may exhibit high accuracy in controlled experiments, their application is ineffective in different populations or medical systems. The prevalence, manifestations and data sources of BPD are greatly influenced by factors such as race, gestational age, and socioeconomic status, which limits the universality of the model (56). At present, most models lack systematic external validation studies, especially cross-institutional and multi-center data validation is rare, and their applicability in real-world environments cannot be confirmed.
3. **The model update mechanism is not perfect:** Clinical data and disease spectra have dynamic characteristics. If the model lacks a real-time monitoring and update mechanism, it is prone to "model drift", that is, changes in the distribution of input data lead to performance degradation (57). Ideally, an AI model should have the capabilities of continuous performance monitoring, drift detection, and automated re-training, forming a closed-loop feedback system. Currently, most BPD prediction models lack drift monitoring mechanisms, as well as no standardized procedures for model re-training or version iteration, making it difficult to ensure their long-term stability and clinical applicability (58).

## 4.3 Clinical integration and ethical supervision aspects

1. **The integration and transformation path of clinical applications are unclear:** Most BPD AI models are limited to research or tool levels and have not been deeply coupled with treatment paths and intervention processes, making it difficult to guide actual clinical behaviors. Currently, there is a lack of standardized prediction, stratification, intervention processes, and operational guidelines, which limits their practical application in the clinical setting.
2. **Ethical compliance and regulatory mechanisms are lagging behind:** When AI is applied to the prediction of neonatal diseases, it involves sensitive populations, and its data usage and model deployment must strictly comply with ethical norms. Currently, there is a lack of specialized ethical guidelines and data regulatory mechanisms for pediatric AI models. Issues such as informed consent, data anonymization, algorithm bias correction, and liability attribution for model failures all urgently require clear legal and institutional definitions (59).

# 5 From prediction to intervention: building an AI-driven clinical intervention path for BPD

## 5.1 Establishing a "time window–PALM" translation framework for NICUs

Currently, the research on AI in BPD management mainly focuses on risk prediction, but there is less discussion on how to convert the prediction results into specific intervention measures (60). Based on existing systematic evidence, this review proposes a Time Window–PALM (Predict–Act–Learn–Monitor) framework, which takes $\lesssim$24 h, 3–7 days, and $\approx$28 days after birth as key nodes in the disease course, requires the model to output calibrated individualized risk at each node and use Decision Curve Analysis (DCA) to provide net benefits under acceptable thresholds, and maps risk thresholds to specific intervention measures (e.g., ventilator fine-tuning, oxygen therapy targets, medication/fluid management, imaging re-evaluation); within fixed time windows, the following should be routinely reported: calibration and net benefit, alert rate, and robustness across subgroups (e.g., by gestational age or center), which are used to trigger or delay interventions, and subsequently, real-world outcomes are fed back to recalibrate the model and conduct incremental update training while continuous monitoring of model performance and drift is required to ensure traceability and auditability.

## 5.2 Individualized treatment intervention strategies at different stages

The clinical value of AI lies in achieving refined, phased, and individualized treatment guidance (6). Within the first 24 h after birth, the model provides calibrated risk, net benefits at selected thresholds, and simultaneously presents PPV, NPV, and alert rate; if the risk exceeds the center's threshold, parameter optimization for lung-protective ventilation is first implemented (limiting tidal volume within the allowable oxygenation/carbon dioxide targets, setting appropriate PEEP, and reducing high-pressure impact), and lung recruitment strategies and oxygen therapy targets are reviewed based on CXR/LUS, followed by re-evaluation of risk trajectory and adverse events at 24–48 h to verify the immediate effect.

From 3 to 7 days after birth, calibrated individualized risk assessment is updated under the same methodological framework and interpreted in conjunction with inflammatory profiles/cardiopulmonary load signals (e.g., CRP/IL-6/IL-8, NT-proBNP, LUS scores); when both risk and net benefits meet the standards, stratified medication and fluid management can be initiated: for infants with severe inflammatory phenotypes and weaning difficulties, glucocorticoid regimens are selected based on evidence, while infants with suspected increased fluid/

pulmonary vascular load receive fluid restriction/diuresis and supportive therapy, with re-testing at 48–72 h to evaluate efficacy and safety.

At approximately 28 days after birth, if infants still have persistent oxygen dependence, longitudinal oxygen demand curves, weight gain/nutrient intake, KL-6/NT-proBNP, and imaging features are comprehensively interpreted to distinguish phenotypes dominated by interstitial damage/alveolar simplification or accompanied by significant pulmonary vascular load: the former prioritizes nutritional and antioxidant support, rehabilitation, and weaning assessment, while the latter undergoes multi-disciplinary cardiopulmonary evaluation and necessary advanced imaging/functional tests, with the establishment of a long-term follow-up plan.

To facilitate understanding of practical implementation, two contextual examples are provided below: A very preterm infant was identified as high-risk by the model 6 h after admission, with positive threshold net benefits shown by DCA; the team accordingly adjusted down ventilator parameters and calibrated oxygen therapy targets, and re-evaluation at 48 h showed decreased OI/OSI and controllable alert rate, so the intervention was retained and documented. On day 5, NT-proBNP increased and LUS score worsened, with risk exceeding the threshold again; short-course low-dose glucocorticoids and fluid restriction were administered in accordance with prescription guidelines, and re-evaluation at 72 h determined whether to upgrade imaging or discontinue pharmacological intervention.

## 5.3 Operational key points and required resources for the PALM closed-loop

Predict: Use features accessible at the time of prediction, and routinely report calibration and threshold ranges indicated by DCA; it is recommended to centrally establish time-windowed risk control tables (including: risk, threshold, net benefit, preset actions, and re-evaluation time points) (61).

Act: Solidify "threshold-action" into executable medical orders/pathways (ventilator parameter targets, oxygen therapy ranges, recruitment strategies, inclusion/exclusion criteria for glucocorticoid and fluid management); run in parallel (shadow mode) in the initial launch phase to prevent premature automation (62).

Learn: Conduct incremental/periodic retraining and threshold recalibration based on real-world outcomes; supplement with feature ablation and key subgroup reviews to maintain stability (62).

Monitor: Continuously monitor performance, fairness, and drift; retain audit logs and version control; align with incident reporting-rollback processes, as well as information notification and appeal channels for family members and medical staff to ensure traceability and accountability (63).

# 6 Future research prospects

Although the application of artificial intelligence (AI) in the risk prediction and individualized management of bronchopulmonary dysplasia (BPD) has achieved initial results (6, 12). However, its clinical application still faces multiple challenges such as data, models, ethics, and clinical pathways. Future research can focus on the following key directions:

First, a multi-center, high-quality clinical database should be established. Currently, most models are based on single-center data, which is limited by sample size and population heterogeneity, resulting in weak generalization ability. Efforts should be made to establish structured clinical databases covering multiple regions, multiple ethnic groups, and multiple NICUs. The data format and label system should be unified to provide a data foundation for model development, external validation, and multimodal integration, and to promote algorithm openness and research reproducibility.

Second, enhance model transparency and interpretability. Black-box models limit doctors' trust and willingness to adopt. In the future, interpretable artificial intelligence (XAI) technologies should be introduced (64) to provide visualizations of the contributions of clinical variables and tracking of decision paths, enhancing the clinical reviewability and adoption intention of the model.

Third, promote the deep integration of the model with the clinical system. The AI model should seamlessly integrate with clinical information system (HIS/EHR) through standard interfaces, automatically extracting key variables (such as gestational age, ventilator parameters, oxygen saturation, blood gas indicators, etc.) to enable real-time prediction and risk alerts (65–68). At the same time, a dynamic update mechanism should be established to adapt to data drift and changes in the clinical environment, ensuring the long-term stable operation of the model. For example, a periodic re-training mechanism and an automatic warning system should be introduced to build an "adaptive model update platform".

Fourth, develop individualized prediction and mechanism-driven models. The development process of BPD is time-dependent, and it is suitable to adopt time series models to integrate continuous physiological signals (69–71). This enables dynamic identification of the disease progression trend. To reduce the impact of model drift in the BPD intervention path, it is necessary to regularly collect the latest data of BPD children for re-training (72). A real-time data monitoring and performance warning mechanism can be established, and the deviation between the input data and the initial training set can be analyzed regularly. To improve the model's adaptability, XAI tools such as SHAP can be introduced to trace the reasons for drift, clarify the reasons for changes in feature contributions, and optimize the individualized intervention strategies of the clinical path in a targeted manner. In addition, integrating multi-omics data (such as transcriptome, proteome, microbiome, etc.) will help identify the biological subtypes of BPD (73), it helps to identify biological subtypes and promotes precise prediction and the formulation of treatment paths.

Fifth, improve ethical governance and regulatory mechanisms. The clinical application of AI models must comply with ethical and legal norms. Data security can be protected through technologies such as differential privacy and federated learning

(74), to clarify the ownership and responsibility of data usage. At the same time, establish an AI ethics approval and supervision mechanism for newborns to ensure their compliant and safe application in clinical settings.

# 7 Discussion

The advancement of AI within the medical domain has highlighted its increasing potential in the prediction and management of BPD among preterm infants. The present review has provided a comprehensive evaluation of AI applications in the development of BPD prediction models, including aspects such as multimodal data integration, model architecture and refinement, performance assessments of existing models, and limitations related to clinical implementation. Collectively, AI offers innovative tools to facilitate the early detection and individualized management of BPD. However, the practical application of these tools remains constrained by several unresolved challenges, necessitating ongoing interdisciplinary collaboration and model refinement.

Currently, a substantial proportion of AI-based BPD prediction models are limited to internal validation, with minimal utilization of external datasets for independent evaluation. Given the considerable variability in neonatal care practices, data acquisition protocols, and infant characteristics across NICUs, model generalizability and applicability require verification through external validation. Consequently, the establishment of a broad, multi-center collaborative framework is imperative. Such a network would support the harmonization of data standards and facilitate data sharing, thereby enhancing the diversity and representativeness of training datasets and strengthening the reliability of predictive models in diverse clinical contexts.

Ethical and legal issues must be considered in the clinical promotion of AI models. AI models typically rely on personal data, such as health records, hematological biomarkers, imaging data, and multimodal data, including genomics. As the volume of these data continues to increase, particularly when integrating multiple data sources and developing new ones, privacy concerns become more prominent (75). However, overly stringent privacy protections during data collection, use, and sharing may impede the potential benefits that patients could derive from using this data to advance AI-driven solutions, such as new diagnostic or treatment options (75, 76). Emerging methods, such as generating synthetic data through AI-generated samples, offer a way to protect privacy while fostering innovation. This synthetic data can be shared securely and used to develop new models (77). Additionally, legal disputes persist regarding the ownership and usage rights of health data, with current regulations failing to clearly define the ownership of AI training data. Therefore, prior to the clinical application of AI models, it is essential to establish comprehensive safety assessments and accountability mechanisms.

The Good Machine Learning Practice (GMLP) proposed by the FDA/IMDRF outlines 10 full-lifecycle principles. The EU's AI Act classifies AI for medical use as a high-risk system, requiring the establishment of lifecycle-wide risk management, high-quality data governance, transparency and user information, human oversight, as well as post-market monitoring and traceable documentation—all of which align with the NICU's goal of protecting minors (78). For pediatric scenarios, the ACCEPT-AI framework and UNICEF's AI Policy Guidelines for Children can be integrated to implement child-centered rights protection.

Under this framework, this study proposes three categories of actionable solutions:

Informed consent adopts a "stratified + dynamic" model (signed by parents/guardians as proxies, with reviews at admission, discharge, and follow-up nodes). The consent document must define the purpose and secondary use of data, withdrawal procedures and timelines, and implement "minimum necessary collection" and de-identification.

Bias auditing runs through both data and models: assess representativeness by gestational age, weight, gender, and region/center; report calibration and subgroup performance simultaneously during internal and external validation, and use Decision Curve Analysis (DCA) to present net benefits at different thresholds; develop plans for regular fairness reviews and bias correction.

For accountability and operational governance, a three-line responsibility system consisting of "clinical lead + data protector + engineering lead" is implemented: complete model and Data Protection Impact Assessment (DPIA) before launch; first run in shadow mode (model operates in parallel with manual processes without directly issuing medical orders) and gradually integrate with "gating thresholds"; during operation, conduct drift monitoring, retain audit logs and version management, and establish incident reporting-rollback channels as well as explanation and appeal channels for family members and medical staff.

In recent years, various AI models related to BPD have emerged continuously, but the ones that have truly achieved clinical routine application are still very limited. Most AI medical models are still at the development and preliminary validation stage, and only a very few can pass strict external validation. Moreover, the models that can be deployed in clinical settings or undergo prospective clinical trials are even more limited (79–81). For instance, a systematic review included 65 studies, covering 158 developed models and 108 externally validated models. Although some models have good discrimination, they generally have high bias risks, and most lack external validation and calibration assessment, which limits their clinical application (81). Although the French EPIPAGE-2 cohort has undergone cross-border validation, its predictive ability has reached a relatively high level (c-statistic 0.80–0.81), yet due to regional calibration deviations, it is still difficult to be generalized to real clinical settings (82).

To promote the effective transformation of AI models in the clinical management of BPD, it is suggested that the clinical needs be incorporated into the model development process to construct an interpretable and integrable model architecture. In terms of data, a multi-center shared database should be

established to promote the standardization of data collection, storage and exchange, and enhance the generalization and reproducibility of the model (83). After deployment, a model performance monitoring mechanism should be set up and a dynamic update strategy should be adopted to deal with "model drift". In terms of technical implementation, privacy-preserving computing (such as federated learning) can be introduced to facilitate cross-institutional collaboration. In terms of regulation, a unified AI medical regulation system has not yet been established globally, and only some countries have explored preliminary paths. It is recommended to draw on the management ideas of the FDA regarding AI medical software (74, 84). The supervision should cover model approval review, performance assessment, ethical review, continuous monitoring and exit mechanism, to ensure the safety, effectiveness and fairness of the model at all stages of clinical application. It is suggested that the national health administrative department take the lead in supervision, and collaborate with interdisciplinary experts from ethics, law, information technology, etc., to establish a "Medical AI Safety Assessment Platform", formulate unified supervision norms, and achieve cross-institutional and cross-regional collaborative supervision.

The application of AI in BPD prediction has demonstrated considerable potential; however, its clinical translation still faces several challenges, including poor model interpretability, insufficient data quality, and inadequate ethical regulations. Moving forward, AI technology can effectively assist physicians in identifying high-risk infants at an early stage and optimizing intervention strategies, thereby improving both the quality of survival and long-term prognosis for premature infants, provided there is collaborative effort across technical, ethical, and institutional domains.

## 8 Conclusion

BPD is a prevalent and serious respiratory disease among premature infants, artificial intelligence technology provides a brand-new perspective for its early identification and precise intervention. This review systematically summarizes the research progress of BPD prediction models from aspects such as AI modeling principles, representative tools, clinical transformation obstacles and future directions. It focuses on the critical comprehensive evaluation of AI methods and clinical tools, highlighting core challenges such as model transparency, data standardization, system integration and ethical supervision, at the same time, this article discussed the PALM translation framework (Predict–Act–Learn–Monitor), which is organized around key clinical nodes. Future research needs to focus more on the deployment adaptability of AI models in real clinical environments, and build a complete transformation chain covering model design, verification, integration, regulation and feedback. To this end, priority should be given to promoting the construction of multi-center standardized databases, developing AI tools with interpretability and dynamic update capabilities, exploring privacy protection mechanisms such as federated learning, and developing closed-loop intelligent systems that can be embedded in clinical processes. This will truly achieve the transformation from "risk prediction" to "precise intervention", and ultimately improve the long-term health outcomes of premature infants with BPD.

## Author contributions

XZ: Writing – original draft, Formal analysis, Data curation. AW: Writing – review & editing, Conceptualization, Formal analysis. RX: Writing – original draft, Formal analysis. DL: Writing – original draft, Writing – review & editing, Funding acquisition.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by a grant from Shandong Province Medical and Health Science and Technology Project (No. 202406031054).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

# References


1. Bell EF, Hintz SR, Hansen NI, Bann CM, Wyckoff MH, DeMauro SB, et al. Mortality, in-hospital morbidity, care practices, and 2-year outcomes for extremely preterm infants in the US, 2013–2018. *JAMA*. (2022) 327(3):248–63. doi: 10.1001/jama.2021.23580

2. Voynow JA. "New" bronchopulmonary dysplasia and chronic lung disease. *Paediatr Respir Rev*. (2017) 24:17–8. doi: 10.1016/j.prrv.2017.06.006

3. Van Marter LJ, Kuban KC, Allred E, Bose C, Dammann O, O'Shea M, et al. Does bronchopulmonary dysplasia contribute to the occurrence of cerebral palsy among infants born before 28 weeks of gestation? *Arch Dis Child Fetal Neonatal Ed*. (2011) 96(1):F20–9. doi: 10.1136/adc.2010.183012

4. Jeon GW, Oh M, Chang YS. Definitions of bronchopulmonary dysplasia and long-term outcomes of extremely preterm infants in Korean neonatal network. *Sci Rep*. (2021) 11(1):24349. doi: 10.1038/s41598-021-03644-7

5. Keles E, Bagci U. The past, current, and future of neonatal intensive care units with artificial intelligence: a systematic review. *NPJ Digit Med*. (2023) 6(1):220. doi: 10.1038/s41746-023-00941-5

6. Jha T, Suhail S, Northcote J, Moreira AG. Artificial intelligence in bronchopulmonary dysplasia: a review of the literature. *Information*. (2025) 16(4):262. doi: 10.3390/info16040262

7. Rusin CG, Acosta SI, Vu EL, Ahmed M, Brady KM, Penny DJ. Automated prediction of cardiorespiratory deterioration in patients with single ventricle. *J Am Coll Cardiol*. (2021) 77(25):3184–92. doi: 10.1016/j.jacc.2021.04.072

8. Miller C, Manious M, Portnoy J. Artificial intelligence and machine learning for anaphylaxis algorithms. *Curr Opin Allergy Clin Immunol*. (2024) 24(5):305–12. doi: 10.1097/ACI.0000000000001015

9. Saria S, Rajani AK, Gould J, Koller D, Penn AA. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci Transl Med*. (2010) 2(48):48ra65. doi: 10.1126/scitranslmed.3001304

10. Moreira AG, Husain A, Knake LA, Aziz K, Simek K, Valadie CT, et al. A clinical informatics approach to bronchopulmonary dysplasia: current barriers and future possibilities. *Front Pediatr*. (2024) 12:1221863. doi: 10.3389/fped.2024.1221863

11. Pammi M, Aghaeepour N, Neu J. Multiomics, artificial intelligence, and precision medicine in perinatology. *Pediatr Res*. (2023) 93(2):308–15. doi: 10.1038/s41390-022-02181-x

12. Verder H, Heiring C, Ramanathan R, Scoutaris N, Verder P, Jessen TE, et al. Bronchopulmonary dysplasia predicted at birth by artificial intelligence. *Acta Paediatr*. (2021) 110(2):503–9. doi: 10.1111/apa.15438

13. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci*. (2021) 2(6):420. doi: 10.1007/s42979-021-00815-1

14. Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare. *Injury*. (2023) 54(Suppl 3):S69–73. doi: 10.1016/j.injury.2022.01.046

15. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electron Markets*. (2021) 31:685–95. doi: 10.1007/s12525-021-00475-2

16. Martin B, Bennett TD, DeWitt PE, Russell S, Sanchez-Pinto LN. Use of the area under the precision-recall curve to evaluate prediction models of rare critical illness events. *Pediatr Crit Care Med*. (2025) 26(6):e855–9. doi: 10.1097/PCC.0000000000003752

17. Nevin L. Advancing the beneficial use of machine learning in health care and medicine: toward a community understanding. *PLoS Med*. (2018) 15(11):e1002708. doi: 10.1371/journal.pmed.1002708

18. Laughon MM, Langer JC, Bose CL, Smith PB, Ambalavanan N, Kennedy KA, et al. Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants. *Am J Respir Crit Care Med*. (2011) 183(12):1715–22. doi: 10.1164/rccm.201101-0055OC

19. Lei J, Sun T, Jiang Y, Wu P, Fu J, Zhang T, et al. Risk identification of bronchopulmonary dysplasia in premature infants based on machine learning. *Front Pediatr*. (2021) 9:719352. doi: 10.3389/fped.2021.719352

20. Dai D, Chen H, Dong X, Chen J, Mei M, Lu Y, et al. Bronchopulmonary dysplasia predicted by developing a machine learning model of genetic and clinical information. *Front Genet*. (2021) 12:689071. doi: 10.3389/fgene.2021.689071

21. Chou HY, Lin YC, Hsieh SY, Chou HH, Lai CS, Wang B, et al. Deep learning model for prediction of bronchopulmonary dysplasia in preterm infants using chest radiographs. *J Imaging Inform Med*. (2024) 37(5):2063–73. doi: 10.1007/s10278-024-01050-9

22. Moreira A, Tovar M, Smith AM, Lee GC, Meunier JA, Cheema Z, et al. Development of a peripheral blood transcriptomic gene signature to predict bronchopulmonary dysplasia. *Am J Physiol Lung Cell Mol Physiol*. (2023) 324(1):L76–87. doi: 10.1152/ajplung.00250.2022

23. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns (N Y)*. (2023) 4(9):100804. doi: 10.1016/j.patter.2023.100804

24. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic group "evaluating diagnostic tests and prediction models" of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. (2019) 17(1):230. doi: 10.1186/s12916-019-1466-7

25. Ambalavanan N, Weese-Mayer DE, Hibbs AM, Claure N, Carroll JL, Moorman JR, et al. Cardiorespiratory monitoring data to predict respiratory outcomes in extremely preterm infants. *Am J Respir Crit Care Med*. (2023) 208(1):79–97. doi: 10.1164/rccm.202210-1971OC

26. Gursoy T, Hayran M, Derin H, Ovali F. A clinical scoring system to predict the development of bronchopulmonary dysplasia. *Am J Perinatol*. (2015) 32(7):659–66. doi: 10.1055/s-0034-1393935

27. Xie F, Yuan H, Ning Y, Ong MEH, Feng M, Hsu W, et al. Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies. *J Biomed Inform*. (2022) 126:103980. doi: 10.1016/j.jbi.2021.103980

28. Corsini I, Parri N, Ficial B, Ciarcià M, Migliaro F, Capasso L, et al. Lung ultrasound in Italian neonatal intensive care units: a national survey. *Pediatr Pulmonol*. (2022) 57(9):2199–206. doi: 10.1002/ppul.26025

29. Vanhaverbeke K, Van Eyck A, Van Hoorenbeeck K, De Winter B, Snoeckx A, Mulder T, et al. Lung imaging in bronchopulmonary dysplasia: a systematic review. *Respir Med*. (2020) 171:106101. doi: 10.1016/j.rmed.2020.106101

30. Xiao SZ, Zhou LW, Wang LL, Zheng Q, Yang J. Clinical applicatton value of imaging-AI in bronchopulmonary dysplasia of prematurity. *J Clin Images Med Case Rep*. (2025) 6(2):3459. doi: 10.52768/2766-7820/3459

31. Draker N, Torry DS, Torry RJ. Placenta growth factor and sFlt-1 as biomarkers in ischemic heart disease and heart failure: a review. *Biomark Med*. (2019) 13(9):785–99. doi: 10.2217/bmm-2018-0492

32. Procianoy RS, Hentges CR, Silveira RC. Vascular endothelial growth factor/placental growth factor heterodimer levels in preterm infants with bronchopulmonary dysplasia. *Am J Perinatol*. (2016) 33(5):480–5. doi: 10.1055/s-0035-1566294

33. Zhang Z, Wu W, Hou L, Jiang J, Wan W, Li Z. Cytokines and exhaled nitric oxide are risk factors in preterm infants for bronchopulmonary dysplasia. *Biomed Res Int*. (2021) 2021:6648208. doi: 10.1155/2021/6648208

34. Cakir U, Tayman C, Yucel C. A novel diagnostic marker for the severity of bronchopulmonary dysplasia in very low birth weight infants: interleukin-33. *Pediatr Allergy Immunol Pulmonol*. (2019) 32(1):12–7. doi: 10.1089/ped.2019.0994

35. Jiang J, Mao Y, Zhou Q, Wu J. Neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio are early predictors of bronchopulmonary dysplasia. *Medicine (Baltimore)*. (2023) 102(35):e34987. doi: 10.1097/MD.0000000000034987

36. Cakir U, Tayman C, Tugcu AU, Yildiz D. Role of systemic inflammatory indices in the prediction of moderate to severe bronchopulmonary dysplasia in preterm infants. *Arch Bronconeumol*. (2023) 59(4):216–22. doi: 10.1016/j.arbres.2023.01.003

37. He Q, Tang Y, Huang J, Rao Y, Lu Y. The value of KL-6 in the diagnosis and assessment of interstitial lung disease. *Am J Transl Res*. (2021) 13(8):9216–23.

38. Dilli D, Özyazici A, Dursun A, Beken S. Predictive values of plasma KL-6 in bronchopulmonary dysplasia in preterm infants. *Turk J Med Sci*. (2017) 47(2):621–6. doi: 10.3906/sag-1512-78

39. Khan S, Concina VA, Schneider D, Westgate P, Arriagada S, Bada H. Role of NT-proBNP in the prediction of moderate to severe bronchopulmonary dysplasia in preterm infants. *Pediatr Pulmonol*. (2020) 55(2):376–82. doi: 10.1002/ppul.24610

40. Méndez-Abad P, Zafra-Rodríguez P, Lubián-López S, Benavente-Fernández I. NTproBNP is a useful early biomarker of bronchopulmonary dysplasia in very low birth weight infants. *Eur J Pediatr*. (2019) 178(5):755–61. doi: 10.1007/s00431-019-03347-2

41. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Brief Bioinform*. (2022) 23(2):bbab569. doi: 10.1093/bib/bbab569

42. Piersigilli F, Lam TT, Vernocchi P, Quagliariello A, Putignani L, Aghai ZH, et al. Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. *Metabolomics*. (2019) 15(2):20. doi: 10.1007/s11306-019-1482-9

43. Pintus MC, Lussu M, Dessì A, Pintus R, Noto A, Masile V, et al. Urinary 1H-NMR metabolomics in the first week of life can anticipate BPD diagnosis. *Oxid Med Cell Longev*. (2018) 2018:7620671. doi: 10.1155/2018/7620671

44. Baraldi E, Giordano G, Stocchero M, Moschino L, Zaramella P, Tran MR, et al. Untargeted metabolomic analysis of amniotic fluid in the prediction of preterm delivery and bronchopulmonary dysplasia. *PLoS One*. (2016) 11(10):e0164211. doi: 10.1371/journal.pone.0164211

45. Lohmann P, Luna RA, Hollister EB, Devaraj S, Mistretta TA, Welty SE, et al. The airway microbiome of intubated premature infants: characteristics and changes that predict the development of bronchopulmonary dysplasia. *Pediatr Res*. (2014) 76(3):294–301. doi: 10.1038/pr.2014.85

46. Lal CV, Olave N, Travers C, Rezonzew G, Dolma K, Simpson A, et al. Exosomal microRNA predicts and protects against severe bronchopulmonary dysplasia in extremely premature infants. *JCI Insight*. (2018) 3(5):e93994. doi: 10.1172/jci.insight.93994

47. Im JE, Yoon SA, Shin YM, Park S. Real-time prediction for neonatal endotracheal intubation using multimodal transformer network. *IEEE J Biomed Health Inform*. (2023) 27(6):2625–34. doi: 10.1109/JBHI.2023.3267521

48. Qu Z, Chen T, Zhou X, Zhu F, Guo W, Xu Y, et al. Multi-modal disease prediction with hierarchical self-supervised learning. *IEEE J Biomed Health Inform*. (2025) 29(10):7057–67. doi: 10.1109/JBHI.2025.3561546

49. El Saie A, Fu C, Grimm SL, Robertson MJ, Hoffman K, Putluri V, et al. Metabolome and microbiome multi-omics integration from a murine lung inflammation model of bronchopulmonary dysplasia. *Pediatr Res*. (2022) 92(6):1580–9. doi: 10.1038/s41390-022-02002-1

50. Franklin G, Stephens R, Piracha M, Tiosano S, Lehouillier F, Koppel R, et al. The sociodemographic biases in machine learning algorithms: a biomedical informatics perspective. *Life (Basel)*. (2024) 14(6):652. doi: 10.3390/life14060652

51. Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J*. (2021) 19:3735–46. doi: 10.1016/j.csbj.2021.06.030

52. Baena-Miret S, Reverter F, Vegas E. A framework for block-wise missing data in multi-omics. *PLoS One*. (2024) 19(7):e0307482. doi: 10.1371/journal.pone.0307482

53. Tang AS, Woldemariam SR, Miramontes S, Norgeot B, Oskotsky TT, Sirota M. Harnessing EHR data for health research. *Nat Med*. (2024) 30(7):1847–55. doi: 10.1038/s41591-024-03074-8

54. Kim MK, Rouphael C, McMichael J, Welch N, Dasarathy S. Challenges in and opportunities for electronic health record-based data analysis and interpretation. *Gut Liver*. (2024) 18(2):201–8. doi: 10.5009/gnl230272

55. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform*. (2021) 113:103655. doi: 10.1016/j.jbi.2020.103655

56. Lewis TR, Kielt MJ, Walker VP, Levin JC, Guaman MC, Panitch HB, et al. Association of racial disparities with in-hospital outcomes in severe bronchopulmonary dysplasia. *JAMA Pediatr*. (2022) 176(9):852–9. doi: 10.1001/jamapediatrics.2022.2663

57. Cabanillas Silva P, Sun H, Rezk M, Roccaro-Waldmeyer DM, Fliegenschmidt J, Hulde N, et al. Longitudinal model shifts of machine learning-based clinical risk prediction models: evaluation study of multiple use cases across different hospitals. *J Med Internet Res*. (2024) 26:e51409. doi: 10.2196/51409

58. Feng J, Gossmann A, Sahiner B, Pirracchio R. Bayesian logistic regression for online recalibration and revision of risk prediction models with performance guarantees. *J Am Med Inform Assoc*. (2022) 29(5):841–52. doi: 10.1093/jamia/ocab280

59. Chng SY, Tern MJW, Lee YS, Cheng LT, Kapur J, Eriksson JG, et al. Ethical considerations in AI for child health and recommendations for child-centered medical AI. *NPJ Digit Med*. (2025) 8(1):152. doi: 10.1038/s41746-025-01541-1

60. Hwang JK, Kim DH, Na JY, Son J, Oh YJ, Jung D, et al. Two-stage learning-based prediction of bronchopulmonary dysplasia in very low birth weight infants: a nationwide cohort study. *Front Pediatr*. (2023) 11:1155921. doi: 10.3389/fped.2023.1155921

61. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *Br Med J*. (2024) 385:e078378. doi: 10.1136/bmj-2023-078378

62. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. *Br Med J*. (2024) 386:e078276. doi: 10.1136/bmj-2023-078276

63. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *Br Med J*. (2025) 388:e082505. doi: 10.1136/bmj-2024-082505

64. Hassan SU, Abdulkadir SJ, Zahid MSM, Al-Selwi SM. Local interpretable model-agnostic explanation approach for medical imaging analysis: a systematic literature review. *Comput Biol Med*. (2025) 185:109569. doi: 10.1016/j.compbiomed.2024.109569

65. Taylor JA, Stout JW, de Greef L, Goel M, Patel S, Chung EK, et al. Use of a smartphone app to assess neonatal jaundice. *Pediatrics*. (2017) 140(3):e20170312. doi: 10.1542/peds.2017-0312

66. Ji J, Ling XB, Zhao Y, Hu Z, Zheng X, Xu Z, et al. A data-driven algorithm integrating clinical and laboratory features for the diagnosis and prognosis of necrotizing enterocolitis. *PLoS One*. (2014) 9(2):e89860. doi: 10.1371/journal.pone.0089860

67. Reed RA, Morgan AS, Zeitlin J, Jarreau PH, Torchin H, Pierrat V, et al. Machine-learning vs. expert-opinion driven logistic regression modelling for predicting 30-day unplanned rehospitalisation in preterm babies: a prospective, population-based study (EPIPAGE 2). *Front Pediatr*. (2021) 8:585868. doi: 10.3389/fped.2020.585868

68. Cabrera-Quiros L, Kommers D, Wolvers MK, Oosterwijk L, Arents N, van der Sluijs-Bens J, et al. Prediction of late-onset sepsis in preterm infants using monitoring signals and machine learning. *Crit Care Explor*. (2021) 3(1):e0302. doi: 10.1097/CCE.0000000000000302

69. Sullivan BA, Fairchild KD. Vital signs as physiomarkers of neonatal sepsis. *Pediatr Res*. (2022) 91(2):273–82. doi: 10.1038/s41390-021-01709-x

70. Kumar N, Akangire G, Sullivan B, Fairchild K, Sampath V. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. *Pediatr Res*. (2020) 87(2):210–20. doi: 10.1038/s41390-019-0527-0

71. Fairchild KD, Srinivasan V, Moorman JR, Gaykema RP, Goehler LE. Pathogen-induced heart rate changes associated with cholinergic nervous system activation. *Am J Physiol Regul Integr Comp Physiol*. (2011) 300(2):R330–9. doi: 10.1152/ajpregu.00487.2010

72. Jenkins DA, Martin GP, Sperrin M, Riley RD, Debray TPA, Collins GS, et al. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? *Diagn Progn Res*. (2021) 5(1):1. doi: 10.1186/s41512-020-00090-3

73. Lal CV, Bhandari V, Ambalavanan N. Genomics, microbiomics, proteomics, and metabolomics in bronchopulmonary dysplasia. *Semin Perinatol*. (2018) 42(7):425–31. doi: 10.1053/j.semperi.2018.09.004

74. Hanif A, Lu C, Chang K, Singh P, Coyner AS, Brown JM, et al. Federated learning for multicenter collaboration in ophthalmology: implications for clinical diagnosis and disease epidemiology. *Ophthalmol Retina*. (2022) 6(8):650–6. doi: 10.1016/j.oret.2022.03.005

75. Price WN 2nd, Cohen IG. Privacy in the age of medical big data. *Nat Med*. (2019) 25(1):37–43. doi: 10.1038/s41591-018-0272-7

76. Seastedt KP, Schwab P, O'Brien Z, Wakida E, Herrera K, Marcelo PGF, et al. Global healthcare fairness: we should be sharing more, not less, data. *PLoS Digit Health*. (2022) 1(10):e0000102. doi: 10.1371/journal.pdig.0000102

77. Torkzadehmahani R, Nasirigerdeh R, Blumenthal DB, Kacprowski T, List M, Matschinske J, et al. Privacy-preserving artificial intelligence techniques in biomedicine. *Methods Inf Med*. (2022) 61(S 01):e12–27. doi: 10.1055/s-0041-1740630

78. Aboy M, Minssen T, Vayena E. Navigating the EU AI act: implications for regulated digital medical products. *NPJ Digit Med*. (2024) 7(1):237. doi: 10.1038/s41746-024-01232-3

79. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. (2019) 25(1):44–56. doi: 10.1038/s41591-018-0300-7

80. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *Br Med J*. (2020) 368:m689. doi: 10.1136/bmj.m689

81. Romijn M, Dhiman P, Finken MJJ, van Kaam AH, Katz TA, Rotteveel J, et al. Prediction models for bronchopulmonary dysplasia in preterm infants: a systematic review and meta-analysis. *J Pediatr*. (2023) 258:113370. doi: 10.1016/j.jpeds.2023.01.024

82. Torchin H, Dhiman P, Ancel PY, Durrmeyer X, Jarreau PH, Nuytten A, et al. Early prediction of bronchopulmonary dysplasia: comparison of modelling methods, development and validation studies. *Pediatr Res*. (2025). doi: 10.1038/s41390-025-04170-2

83. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART On FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc*. (2016) 23(5):899–908. doi: 10.1093/jamia/ocv189

84. Elmas G, Dar SUH, Korkmaz Y, Ceyani E, Susam B, Ozbey M, et al. Federated learning of generative image priors for MRI reconstruction. *IEEE Trans Med Imaging*. (2023) 42(7):1996–2009. doi: 10.1109/TMI.2022.3220757

frontiers | Frontiers in Pediatrics

TYPE Original Research
PUBLISHED 08 December 2025
DOI 10.3389/fped.2025.1677292

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Xu Liu,
The Chinese University of Hong Kong, China
Om Prakash,
Council of Scientific and Industrial Research (CSIR), India

*CORRESPONDENCE
Jian-xin Sun
✉ nbfe2025@163.com

RECEIVED 31 July 2025
REVISED 21 November 2025
ACCEPTED 24 November 2025
PUBLISHED 08 December 2025

CITATION
Ye L-z, Sun J-x, Chen J, Cen K-k, Bi Y and Lu Y-c (2025) Machine learning model for predicting urinary tract infection risk in febrile children under 3 years of age.
Front. Pediatr. 13:1677292.
doi: 10.3389/fped.2025.1677292

COPYRIGHT
© 2025 Ye, Sun, Chen, Cen, Bi and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Machine learning model for predicting urinary tract infection risk in febrile children under 3 years of age

Le-zhen Ye, Jian-xin Sun*, Jing Chen, Kuan-kuan Cen, Ye Bi and Yun-cong Lu

Department of Paediatrician, Women's and Children's Hospital of Ningbo University, Ningbo, China

**Objective:** Urinary tract infection (UTI) is a common childhood infectious disease. Accurate prediction of UTI risk in febrile children enables timely intervention and helps avoid long-term complications such as renal scarring.
**Methods:** 1,556 cases of febrile children under 3 years of age were retrospectively analyzed, and feature variables were screened using LASSO regression. Seven machine learning (ML) algorithms, including Random Forest, were used to construct the UTI prediction model. The model performance was evaluated based on comprehensive indices, including area under the curve (AUC), calibration curve, and decision curve analysis, from which the optimal prediction model was selected. The SHAP method was applied to analyze the decision-making mechanism of the model.
**Results:** Among the seven ML models, Random Forest performed best, achieving an AUC of 0.88 in the test set, an AUPRC of 0.824, optimal calibration (ICI = 0.12), and decision curve analysis showed superior performance compared to other ML algorithms. Through LASSO regression screening and SHAP analysis, seven core predictors were established: age, WBC count, previous UTI episodes, PLT, fever peak, CRP, prenatally detected renal abnormalities. These key indicators helped to construct an accurate prediction system for UTI risk in febrile children.
**Conclusions:** The ML model constructed in this study can accurately predict UTI risk in febrile children under 3 years of age. The visual decision interpretation achieved through the SHAP framework can assist clinicians in quickly identifying high-risk children.

KEYWORDS

machine learning, prediction model, SHAP, urinary tract infection, UTI

## 1 Introduction

Urinary tract infection (UTI) is a common childhood infectious disease (1). The clinical presentation varies significantly according to age: older children (typically aged ≥3 years) tend to present with typical symptoms of urinary tract irritation, such as urinary frequency and dysuria, which are seldom missed. In contrast, infants and young children (aged <3 years) usually lack specific manifestations and may present with fever as the primary symptom, accompanied by atypical signs, such as crying, lethargy, feeding difficulties, and growth retardation, which are frequently overlooked (2). Studies have shown that even in healthcare settings with high clinical vigilance and adequate diagnostic resources, the missed rate of febrile UTI in infants and young

children remains at as high as 50%–70% (3), representing a significant diagnostic challenge. More critically, delayed treatment of UTI is associated with permanent renal scarring (4–7), particularly in febrile UTI, where the incidence of renal scarring ranges from 10% to 30% (2, 8–10). These pathological changes may lead to a significant increase in the risk of long-term complications, including hypertension and chronic kidney disease (11). Despite the clinical importance of early recognition and standardized treatment of UTI, factors such as insidious symptoms and difficulty in obtaining urine specimens in infants and children make early and accurate recognition of urinary tract infections a challenging task (12).

Previous studies have shown associations between obesity, bladder and bowel dysfunction (BBD), vesicoureteral reflux, age, vitamin D deficiency, fever (temperature ≥39°C), and UTI occurrence in children (13–18). However, these studies have been limited in two critical aspects. First, most have focused on exploring the diagnostic value of single or few clinical factors rather than comprehensive multivariable assessment. Second, many studies have relied on traditional statistical methods that typically assume linear relationships in the log-odds scale, which may not adequately capture the non-linear patterns and complex variable interactions present in clinical data without extensive manual feature engineering.

With the rapid development of artificial intelligence technology, machine learning (ML) algorithms capable of modeling non-linear relationships and complex interactions, combined with explainable AI frameworks such as SHAP, have demonstrated advantages in disease risk prediction by constructing multidimensional models while maintaining clinical transparency (19–21). In this study, we retrospectively analyzed the clinical data of 1,556 febrile children under 3 years of age and constructed a UTI risk prediction model using multiple ML algorithms. We compared the diagnostic performance of different algorithms and established an early warning system for UTI with clinical application value. This approach provides an evidence-based foundation for optimizing the diagnostic approach for fever in children.

# 2 Materials and methods

## 2.1 Data sources

A total of 4,971 febrile children admitted to the Women's and Children's Hospital of Ningbo University from January 1, 2020, to December 31, 2024, were identified by reviewing outpatient and inpatient electronic medical records. After applying predefined inclusion and exclusion criteria, 1,556 patients were included in the final analysis.

Inclusion Criteria: (1) Fever was defined as core temperature of ≥38.0°C. (2) Age 28 days to 3 years. (3) Hospitalization duration >24 h. (4) Complete clinical data.

Exclusion Criteria:

1. Cases presenting with predominant symptoms strongly suggesting alternative diagnoses, including:
    1) Respiratory diseases: persistent cough requiring antitussive therapy, tachypnea, wheezing, or respiratory distress
    2) Gastrointestinal diseases: vomiting (>3 episodes per day) or diarrhea (≥3 loose stools per day), with symptoms lasting >24 h, as the primary complaint
    3) Neurological disorders: seizures, altered consciousness, focal neurological deficits, or meningeal signs
    4) Rheumatological/autoimmune disorders: arthritis, characteristic rashes (e.g., malar rash, photosensitivity), or documented autoimmune disease history
2. Cases lacking both urinalysis and urine culture within 24 h of presentation.

This study was carried out in accordance with the Declaration of Helsinki and approved by the Medical Ethics Committee of the Women's and Children's Hospital of Ningbo University (EC2023-011).

Figure 1 presents the flowchart of the methodological process.

## 2.2 Methodology

This study used the R software (version 4.4.1), Python (version 3.11.9), and related extension packages for predictive model construction and evaluation.

### 2.2.1 Study variables

Twelve variables were included: (1) Demographic and clinical characteristics, including age, sex, weight status, prenatally detected renal abnormalities, previous UTI episodes, and feeding mode. (2) Clinical symptoms, including fever peak. (3) Laboratory findings, including white blood cell count (WBC), neutrophil percentage (*N*%), hemoglobin (Hb), platelet count (PLT), C-reactive protein (CRP).

The variables were defined as follows:

1. Diagnosis of urinary tract infection: UTI was diagnosed by positive urine culture (excluding obvious contaminants) and/or ≥5 leukocytes per high-power field in centrifuged urine sediment. Positive culture was defined as ≥100,000 CFU/mL of a single pathogen for midstream specimens or $\geq 5 \times 10^4$ CFU/mL for catheterized specimens.
2. Weight status: According to the 2006 WHO standard for evaluation of physical development of children, the interval from −1SD to +1SD was defined as normal, <−1SD as underweight, and >+1SD as overweight.
3. Prenatally detected renal abnormalities: prenatal ultrasound demonstrating abnormalities of the urinary tract including: (1) unilateral/bilateral hydronephrosis or pelvic dilatation; (2) duplicate collecting system or horseshoe kidney; (3) renal agenesis or dysplasia; (4) multicystic dysplastic kidney; and (5) ureterocele, ureteral stenosis, or ectopic ureteral insertion.
4. Fever peak: the highest value of rectal or tympanic temperature prior to admission, as reported by caregivers or documented in medical records.

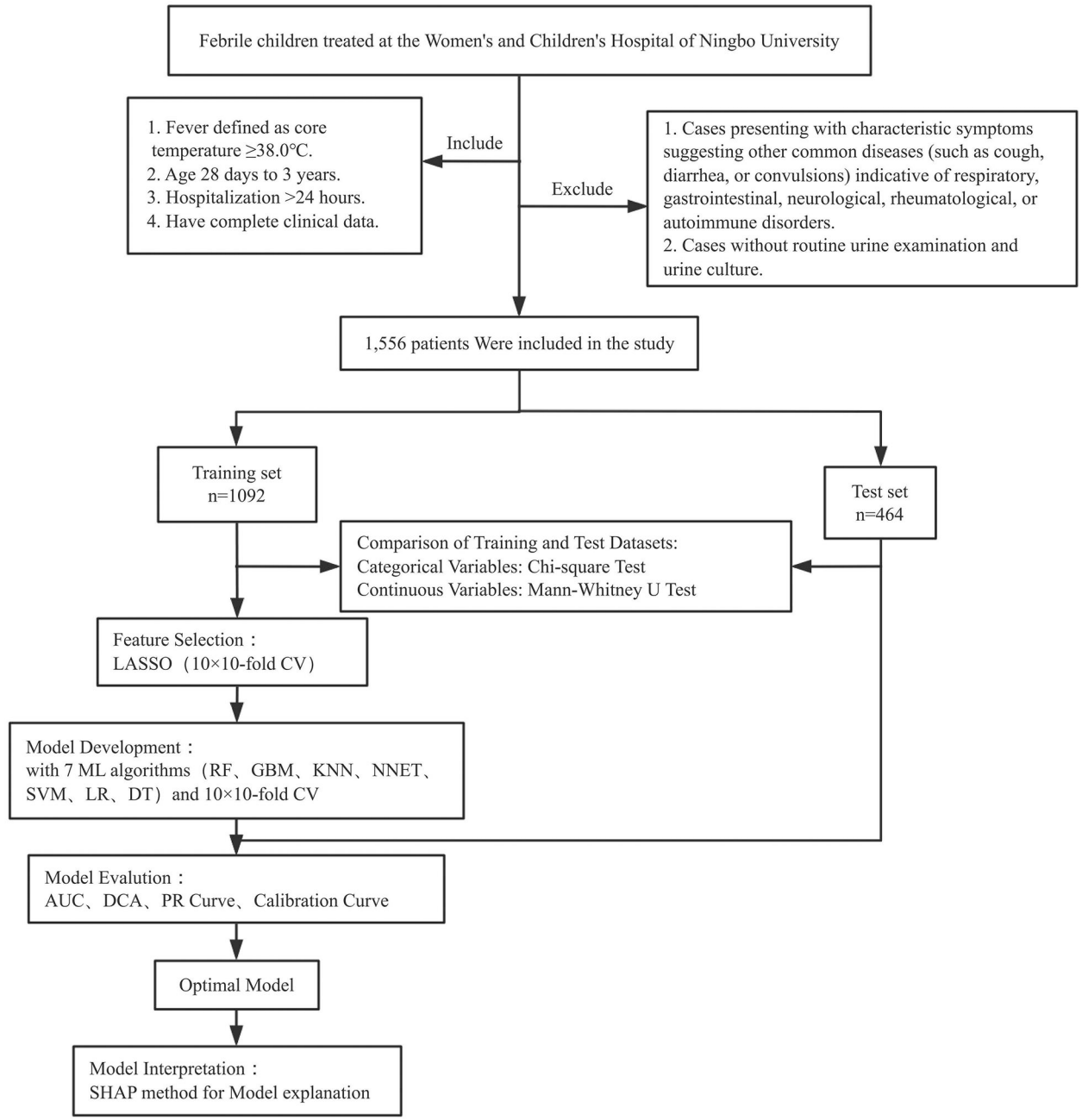


FIGURE 1
Flowchart of the methodological process for the prediction model.

5. Feeding mode: categorized by the main feeding mode within the first six months of life, including breastfeeding, mixed feeding, and artificial feeding.
6. Blood tests: Complete blood count (including WBC, *N*%, Hb, PLT) and CRP were performed at the time of presentation after the onset of fever symptoms.
7. Urinalysis and urine culture: Urine specimens were collected as early as possible after presentation, ideally before antibiotic administration. Collection methods included clean-catch midstream collection for cooperative older children, sterile urethral catheterization when clinically indicated, and bag collection for younger infants when other methods were not feasible. Urinalysis included: (1) microscopic examination for white blood cells, red blood cells, and bacteria; (2) dipstick testing for leukocyte esterase, nitrites, and protein. Urine culture referred to quantitative bacterial culture with antimicrobial susceptibility testing when indicated. When multiple urine cultures were available during the same febrile episode, the first culture result obtained within 24 h of presentation was used for analysis. Cases with discordant culture results were reviewed by two independent pediatric infectious disease specialists, with

consensus determination based on clinical context, specimen quality, and colony counts.

### 2.2.2 Sample size determination

Our sample size is rigorously justified based on both the traditional 10 EPV rule and contemporary best-practice guidelines by Riley et al. (22). Table 1 presents the calculation process, formula, and results of sample size determination.

### 2.2.3 Grouping methods

The caret package (version 7.0.1) was used to randomly divide all patients into training and test sets in a ratio of 7:3, where the training set contained 1,092 samples and the test set contained 464 samples. Based on the diagnostic criteria, 513 children were diagnosed with UTI (294 males, 57.3%) and 1,043 children were classified as non-UTI cases.

For comparisons between the training and test sets, continuous variables were expressed as medians and interquartile ranges and were compared using the Mann–Whitney *U* test. Categorical variables were expressed as numbers and percentages and compared using the chi-square test. Statistical significance was set at $p < 0.05$. Statistical analyses were performed using Statistical Package for the SPSS (version 29.0). Comparison of characteristics between the groups showed well-balanced baseline characteristics, with no statistically significant differences (all $p > 0.05$). Table 2 provides a comprehensive overview of the baseline characteristics between the two groups.

### 2.2.4 Construction and evaluation of prediction models

Data screening: LASSO regression analysis was performed using the glmnet package (version 4.1.8). Through 10-fold cross-validation, the optimal $\lambda$ value was selected based on the lambda.1se criterion, and feature variables with non-zero coefficients were screened to extract key features.

Comprehensive multi-model analysis: The caret package (version 7.0.1) was used to construct seven ML models: Logistic Regression (glm), Random Forest (rf), Gradient Boosted Tree (gbm), Neural Network (nnet), Decision Tree (rpart), Support

TABLE 2 Summary of baseline data analysis results.

| Variable | Training set (*n* = 1,092) | Test set (*n* = 464) | *Z* | *p* |
|---|---|---|---|---|
| UTI, *n* (%) | | | | |
| Yes | 358 (32.8) | 155 (33.4) | 0.057 | 0.812 |
| No | 734 (67.2) | 309 (66.6) | | |
| Sex, *n* (%) | | | | |
| Male | 579 (53.0) | 265 (57.1) | 2.195 | 0.138 |
| Female | 513 (47.0) | 199 (42.9) | | |
| Age, *n* (%) | | | | |
| <3 months | 245 (22.4) | 113 (24.4) | 1.195 | 0.550 |
| 3 months to 1 year | 456 (41.8) | 197 (42.5) | | |
| 1–3 years | 391 (35.8) | 154 (33.2) | | |
| Weight status, *n* (%) | | | | |
| >+1SD | 395 (36.2) | 148 (31.9) | 2.825 | 0.244 |
| <−1SD | 84 (7.7) | 35 (7.5) | | |
| −1SD to +1SD | 613 (56.1) | 281 (60.6) | | |
| Fever peak, *n* (%) | | | | |
| ≤39°C | 474 (43.4) | 205 (44.2) | 0.079 | 0.778 |
| >39°C | 618 (56.6) | 259 (55.8) | | |
| Prenatally detected renal abnormalities, *n* (%) | | | | |
| Abnormal | 29 (2.7) | 11 (2.4) | 0.106 | 0.745 |
| Normal | 1,063 (97.3) | 453 (97.6) | | |
| Previous UTI episodes, *n* (%) | | | | |
| ≥2 times | 19 (1.7) | 8 (1.7) | 0.165 | 0.921 |
| 1 time | 67 (6.1) | 26 (5.6) | | |
| 0 time | 1,006 (92.1) | 430 (92.7) | | |
| Feeding mode | | | | |
| Artificial feeding | 216 (19.8) | 73 (15.7) | 4.505 | 0.105 |
| Mixed feeding | 278 (25.5) | 135 (29.1) | | |
| Breastfeeding | 598 (54.8) | 256 (55.2) | | |
| Anemia | | | | |
| Yes | 153 (14.0) | 66 (14.2) | 0.012 | 0.912 |
| No | 939 (86.0) | 398 (85.8) | | |
| WBC ($\times 10^9$/L), median | 13,600 (7,500, 18,900) | 13,900 (7,833, 19,400) | −0.367 | 0.714 |
| *N* (%), median | 55,000 (44,000, 64,075) | 56,000 (42,000, 64,000) | −0.700 | 0.484 |
| PLT ($\times 10^9$/L), median | 290,000 (213,000, 378,750) | 282,000 (209,250, 384,000) | −0.466 | 0.641 |
| CRP (mg/L), median | 27,000 (6,400, 69,500) | 33,800 (7,525, 73,475) | −1.660 | 0.097 |

TABLE 1 Sample size requirements for prediction model development.

| Method | Sample size (*n*) | Formula |
|---|---|---|
| 10 Events Per Candidate Predictor Parameter | 365 | At least 10 events per predictor parameter (EPV ≥ 10) |
| Four-Step Procedure | | |
| Precise Outcome Risk Estimation | 340 | $n = \left(\frac{1.96}{\delta}\right)^2 \phi(1-\phi)$ |
| Minimizing Prediction Error | 943 | $n = \exp\left(\frac{-0.508+0.259\ln(\phi)+0.504\ln(P)-\ln(\mathrm{MAPE})}{0.544}\right)$ |
| Reducing Predictor Effect Shrinkage | 549 | $n = \frac{P}{(S-1)\ln\left(1-\frac{R^2_{cs}}{S}\right)}$ |
| Minimizing Optimism in Model Fit | 181 | $R^2_{\mathrm{Nagelkerke}} = \frac{R^2_{cs}}{\max(R^2_{cs})}, \max(R^2_{cs}) = 1-\exp\left(\frac{2\ln L_{\mathrm{null}}}{n}\right)$ $\ln L_{\mathrm{null}} = E\ln\left(\frac{E}{n}\right)+(n-E)\ln\left(1-\frac{E}{n}\right), S = \frac{R^2_{cs}}{R^2_{cs}+\delta\max(R^2_{cs})}$ $n = \frac{P}{(S-1)\ln\left(1-\frac{R^2_{cs}}{S}\right)}$ |

Vector Machine (svmRadial), and K-Nearest Neighbors (knn). For model evaluation, the pROC package (version 1.18.5) was used to plot the ROC curves, the PRROC package (version 1.4) was used to analyze precision-recall curves, and decision curve analysis was used to calculate the net gain of each model under different threshold probabilities. Model calibration was evaluated using the integrated calibration index (ICI), with lower values indicating better calibration performance. These models were tested, and the performance metrics of the training and test sets were compared and analyzed to select the optimal model.

Model Interpretation: To enhance model interpretability, Python (version 3.11.9) was used to calculate the SHAP values, which were combined with feature importance ranking, swarm plots, and dependency plots to facilitate model interpretation at both global and individual levels.

# 3 Results

## 3.1 Screening of factors characterizing urinary tract infections in febrile children

UTI diagnosis was the dependent variable, and 12 independent factors were subjected to the LASSO regression analysis. LASSO reduces overfitting by compressing variable coefficients through L1 regularization and addresses multicollinearity issues. Ten-fold cross-validation for the regularization parameter ($\lambda$) selection (Figure 2A) showed that the optimal $\lambda$ value (lambda.1se) was 0.0327. Seven variables were ultimately selected from the 12 independent variables, including age, WBC count, previous UTI episodes, PLT, fever peak, CRP, and prenatally detected renal abnormalities. The distribution patterns of these seven LASSO-selected variables in the training dataset are shown in Figure 2C, with all features demonstrating statistically significant differences between UTI and non-UTI groups (all $p < 0.001$).

## 3.2 Analysis of machine learning models

We used seven models, namely Random Forest (RF), Gradient Boosting Tree (GBM), K-Nearest Neighbors (KNN), Neural Network (NNET), Support Vector Machine (SVM), Decision Tree (DT), and Logistic Regression (LR), for training and 10-fold cross-validation. The models were evaluated using metrics including AUC.

ROC curve evaluation: In the ROC curve comparison of the training set (Figure 3A), RF (AUC = 0.91) and GBM (AUC = 0.873) achieved relatively high AUC values. In the ROC curve comparison of the test set (Figure 3B), RF (AUC = 0.88) and GBM (AUC = 0.859) performed optimally.

Calibration curve assessment: Calibration analysis revealed that all models demonstrated suboptimal calibration, with ICI values ranging from 0.120 to 0.228. Among these, the RF (ICI = 0.120) and NNET (ICI = 0.160) models showed relatively better calibration performance compared to other algorithms (Figure 3C). The calibration curves revealed systematic underestimation of UTI risk, particularly in the moderate probability range (0.4–0.6), with predicted probabilities lower than observed frequencies.

Decision curve analysis (DCA): The net benefit performance showed that the Random Forest model performed better overall. Random Forest consistently outperformed other models including Logistic Regression, KNN, SVM, and GBM in the moderate to high probability threshold range (0.2–0.8), with its curve remaining above the "Treat all" reference line in the clinically optimal range (approximately 0.1–0.6), indicating its advantages in balancing overtreatment and underdiagnosis risks.

Precision-recall curve evaluation: In the test set precision-recall curve comparison (Figure 3E), both Random Forest (AUPRC = 0.824) and GBM (AUPRC = 0.8) performed well.

The Random Forest model was identified as the optimal model based on a comprehensive evaluation of multiple metrics. Table 3 lists the detailed performance metrics of the seven models.

## 3.3 Interpretation of the model by SHAP

To visually explain the key predictive features in the random forest model (the final selected model based on optimal performance), we used SHAP values to quantify feature contributions, as illustrated in Figures 4A–C. In the swarm plot, each point represents the feature contribution value for a patient sample, with red points indicating high feature values and blue points indicating low feature values. The horizontal positions of the points reflect the positive or negative impact of the SHAP value on the prediction. In the bar chart, features are sorted from high to low based on their mean absolute SHAP values, and the length of the bars visually indicates the feature's importance (the longer the bar, the stronger the impact on the prediction). These two charts not only show the global ranking of feature importance, but also reveals the direction and degree of influence of each feature value on individual predictions. Figure 4C shows the dependency plots for each predictor, which demonstrate that children with younger age (especially those ≤1 year old), elevated white blood cell count, previous UTI episodes, elevated platelet count and CRP, prenatally detected renal abnormalities, and fever peak ≤39°C were more likely to be diagnosed with urinary tract infection.

## 3.4 Clinical application

Based on the above SHAP analysis results, we have developed a web-based UTI prediction system for febrile children under 3 years of age, available at https://uti-prediction.yezhiqiu.cn. By visiting this website and entering clinical indicators such as age, white blood cell count, and platelet count, users can quickly obtain UTI risk probability predictions. Furthermore, the system provides targeted medical recommendations based on the

10.3389/fped.2025.1677292

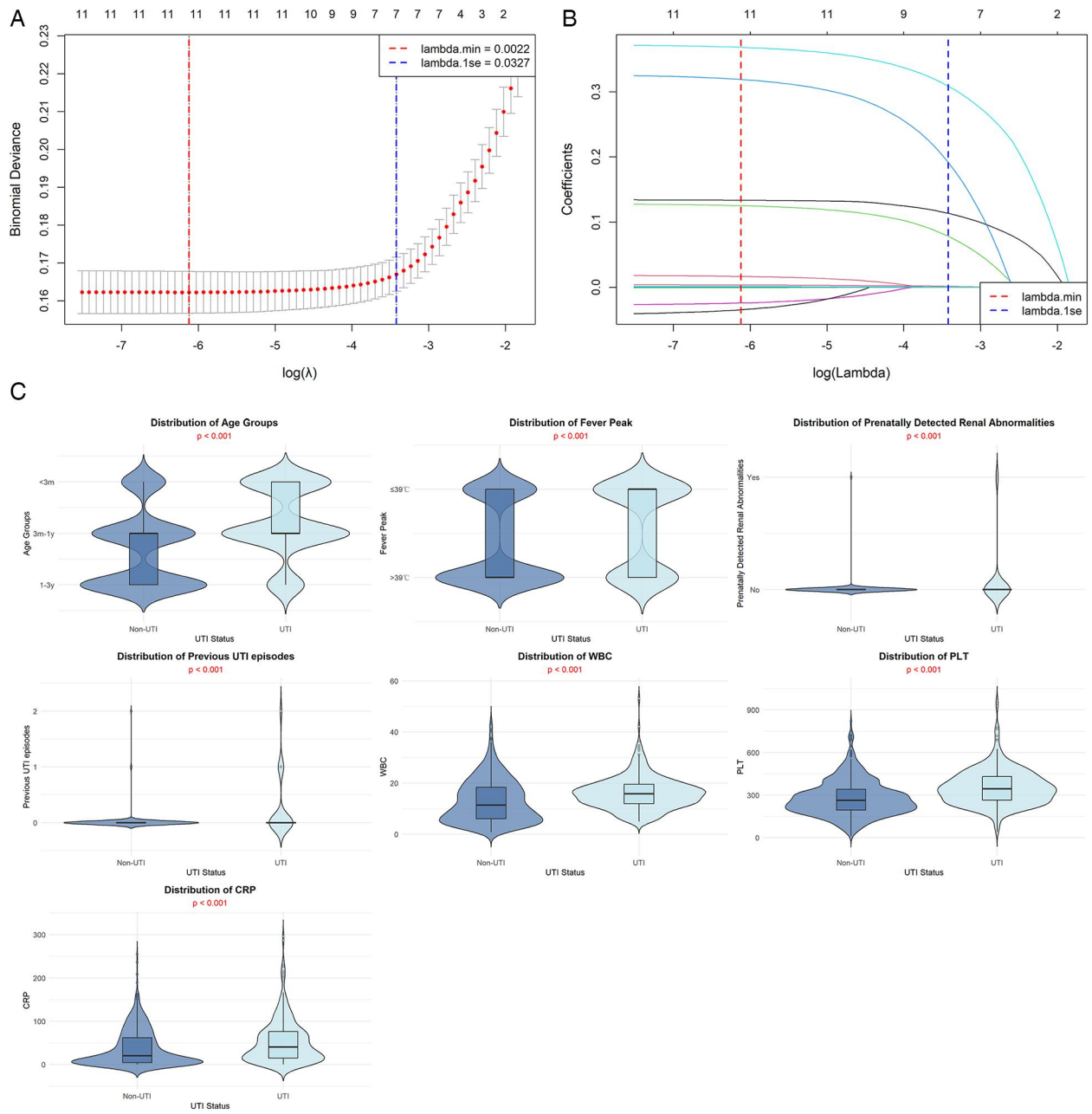


FIGURE 2
**(A)** Illustrates the variation of binomial deviance with log($\lambda$), with lambda.1se chosen to balance model simplicity and prediction accuracy, ultimately retaining seven variables with non-zero coefficients. The coefficient path diagram in **(B)** shows how variable coefficients gradually shrink toward zero as log($\lambda$) increases from −7 to −2 (corresponding to increasing $\lambda$ values). This process demonstrates that: (1) high $\lambda$ values [e.g., log($\lambda$) = −2] apply strong regularization, compressing most coefficients to zero and reducing model complexity; (2) low $\lambda$ values [e.g., log($\lambda$) = −7] retain more variables but increase overfitting risk; therefore, the optimal $\lambda$ = 0.0327 was chosen to balance prediction accuracy with clinical interpretability. **(C)** Shows the clinical features distribution comparison between UTI and non-UTI groups in the training dataset.

model's analysis to assist clinicians in initial screening and clinical decision-making.

## 4 Discussion

In this study, seven key predictors (age, WBC count, previous UTI episodes, PLT, fever peak, CRP, prenatally detected renal abnormalities) were screened from 12 clinical indicators by LASSO regression, and an optimal ML algorithm was identified based on multi-model comparisons for UTI prediction in febrile children. Accurate prediction of UTI risk in febrile children under 3 years of age is of great importance in clinical practice. Consequently, previous studies have examined the risk factors for urinary tract infection in children. However, no suitable ML prediction model has been developed based on these findings.

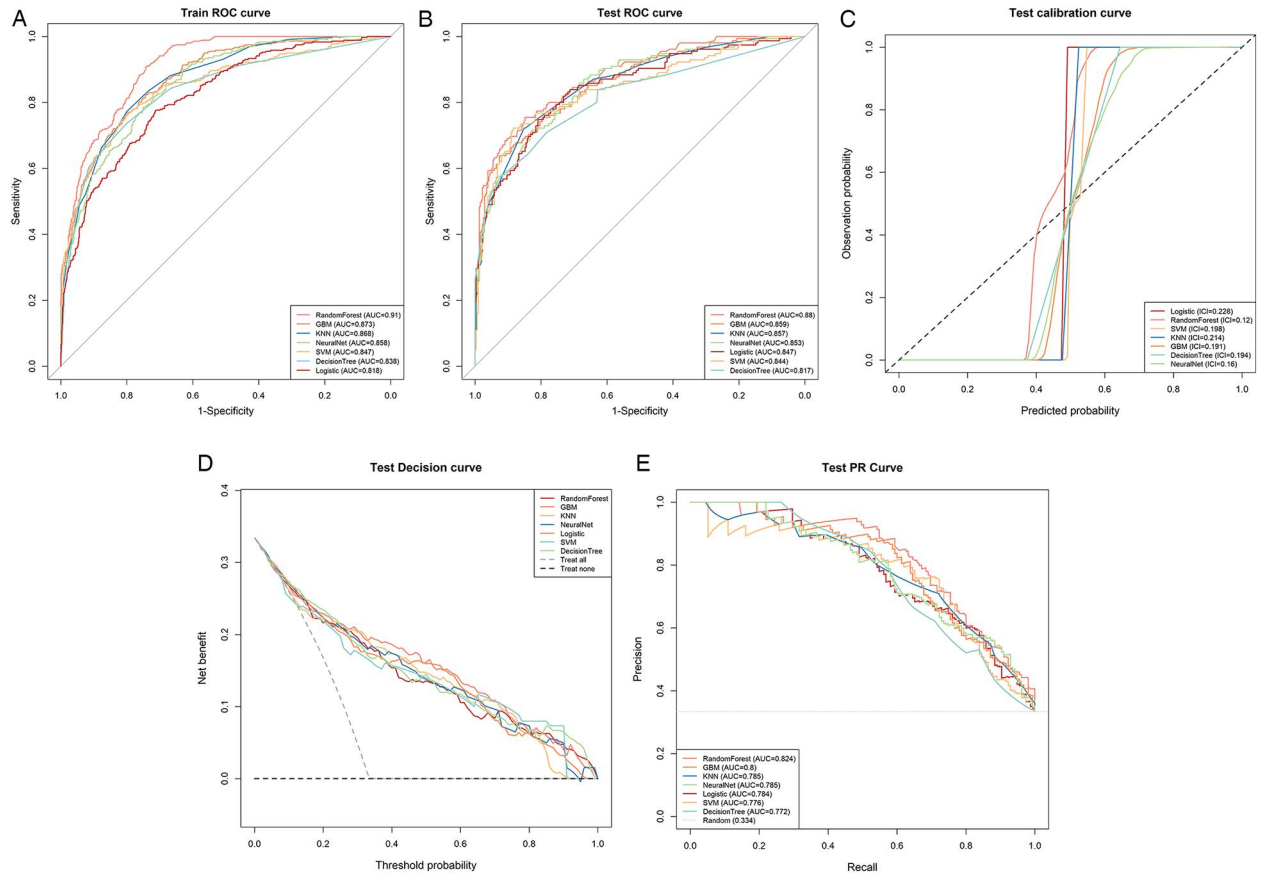


FIGURE 3
Comprehensive analysis of ML models: **(A)** training set ROC and AUC with ten-fold cross-validation. **(B)** Test set ROC and AUC. **(C)** Calibration curves for the test set: the *y*-axis represents the observed probability, the *x*-axis represents the predicted probability, and the dashed diagonal line represents the reference line. The closer the fitted line is to the reference line, the more accurate the model's predictions are. **(D)** DCA shows that RF performs better than the other ML algorithms. **(E)** Test set precision-recall (PR) curve and Average Precision (AP). The *y*-axis represents precision and the *x*-axis represents recall, and higher AP values indicate better model performance. Different colors represent different models, and the values are shown as averages.

TABLE 3 Performance of ML models for predicting UTI.

| Model | Train-AUC | Test-AUC | Test-AUPRC | Test-sensitivity | Test-specificity | Test-accuracy | ICI |
|---|---|---|---|---|---|---|---|
| RF | 0.910 | 0.880 | 0.824 | 0.755 | 0.848 | 0.832 | 0.120 |
| GBM | 0.873 | 0.859 | 0.800 | 0.639 | 0.926 | 0.825 | 0.191 |
| KNN | 0.868 | 0.857 | 0.740 | 0.723 | 0.851 | 0.804 | 0.214 |
| LR | 0.818 | 0.847 | 0.784 | 0.748 | 0.809 | 0.802 | 0.228 |
| SVM | 0.847 | 0.844 | 0.776 | 0.723 | 0.883 | 0.812 | 0.198 |
| NNET | 0.858 | 0.853 | 0.785 | 0.716 | 0.828 | 0.810 | 0.160 |
| DT | 0.838 | 0.817 | 0.777 | 0.574 | 0.922 | 0.806 | 0.194 |

A meta-analysis by Marjo Renko indicated correlations between urinary tract infections and factors including obesity, insufficient fluid intake, breastfeeding, and circumcision, but did not quantify the predictive weight of each factor (23). The ML model developed by Sriram Ramgopal's team achieved risk stratification of febrile infants and was suitable for identifying severe bacterial infections in infants up to 2 months of age but could not be used for specific prediction of urinary tract infections in children (24). Shang-Chien Li et al. constructed a UTI prediction model for febrile children under 3 years of age using traditional logistic regression analysis and developed a nomogram (25). However, this approach failed to fully exploit the advantages of ML algorithms in modeling complex feature relationships.

In this study, the SHAP method was used to provide an interpretable analysis of the ML model, demonstrating the advantages of modeling complex features. The results showed that age was an important factor in determining UTI risk. Infants and young children, especially those younger than 3 months of age, have a significantly higher incidence of UTI than

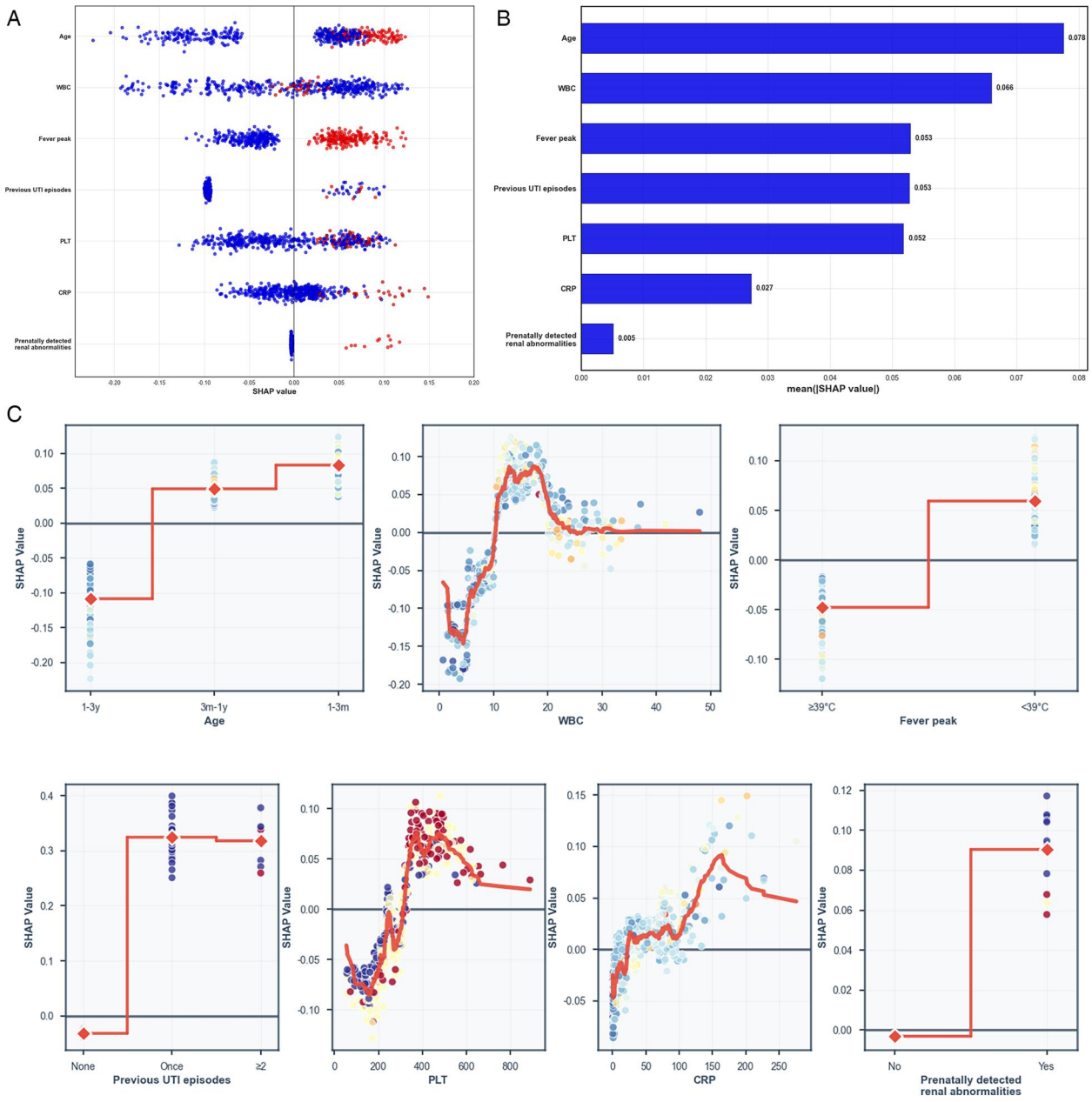


FIGURE 4
SHAP value analysis of feature importance in the random forest model. **(A)** Swarm plot showing SHAP value distribution for each feature across patient samples (red: high feature values; blue: low feature values). **(B)** Bar chart of mean absolute SHAP values ranked by feature importance. **(C)** Dependency plots demonstrating that younger age (≤1 year), elevated white blood cell count, previous UTI episodes, elevated platelet count and CRP, prenatally detected renal abnormalities, and fever peak ≤39°C were associated with increased UTI probability.

children of other ages (26), which is consistent with the results of the present study.

However, our analysis revealed no significant associations between sex or weight and UTI risk, findings that contrast with established literature. Previous studies, including Tej K et al. (4), have demonstrated that females exhibit significantly higher UTI risk than males, except during early infancy. Regarding the relationship between weight status and febrile UTI, existing evidence remains controversial. While Hyung Eun Yim et al. (27) suggested that weight abnormalities (including underweight, overweight, and obesity) may increase UTI susceptibility, other studies have reported no significant association between obesity and febrile UTI risk in hospitalized young children (28). This discrepancy may be related to the fact that sex-age interactions were not analyzed in this study. Additionally, the analysis of individual indicators may have been confounded, considering that children with abnormal weight often have other comorbidities.

Elevated white blood cell counts and platelet counts are important features in predicting UTI in febrile children.

Previous studies have also found that children with UTI who have elevated leukocyte and platelet counts are more likely to have renal involvement (29, 30). Leukocytosis, a common marker of inflammatory response, is associated with the risk of UTI (31). The SHAP analysis in this study not only confirmed the importance of these markers in predicting UTI occurrence but also quantified their relative contribution to the predictive model.

Previous studies have demonstrated that children with urinary tract infections under 3 months of age are more susceptible to high fever (17), which contrasts with our findings. Our study revealed that children with UTI exhibited lower peak fever temperatures compared to children with fever from other etiologies. This finding may be attributed to the earlier healthcare-seeking behavior and prompt treatment initiation in younger children, who represent the high-risk population for UTI.

In prenatal urological ultrasound examination, urinary tract dilation (UTD) is the most common abnormal finding (32). In most cases, such dilation resolves spontaneously (33); however, approximately one-third of UTD cases persist after birth or are diagnosed as congenital anomalies of the kidney and urinary tract (CAKUT) (34). In the early postnatal period, the incidence of UTI in children with UTD is estimated to be 8% to 22%. Additionally, other urinary system abnormalities, such as solitary kidney, are also considered risk factors for increased occurrence of urinary tract infections (35).

UTI should be considered in every child with fever without a source (36). Accurately identifying UTI in febrile children is a challenging clinical task, especially for younger children, whose nonspecific symptoms, insidious signs, and difficulties in urine sample collection significantly increase diagnostic difficulty (12). By developing a ML-based UTI risk prediction model, this study not only breaks through the limitations of traditional diagnostic methods but also enhances clinical utility through model interpretability analysis. Specifically, feature contribution analysis under the SHAP framework can visualize the decision weights of key predictors (e.g., age, WBC, PLT), transforming the model from a "black box" into a clinically understandable decision support tool.

This dual advantage of high-precision prediction and transparent interpretation provides clinicians with a multi-dimensional risk assessment system that helps achieve multiple clinical goals. By accurately predicting UTI risk in high-risk children, it can guide individualized antibiotic use to avoid over- or under-treatment. The system enables combining risk stratification results with the selective use of urine culture and other tests, ensuring diagnostic accuracy while reducing unnecessary testing. Additionally, clinicians can develop preventive intervention strategies based on interpretable characteristics, such as initiating early antibiotic treatment for those with elevated inflammatory markers. The integrated application of these strategies will promote pediatric UTI management from "passive treatment" to "active prevention and control".

This study has several limitations that require attention. First, the urine collection method poses challenges, particularly for younger children under 3 months of age where bag collection is commonly used in clinical practice. While this approach is practical, it may increase the risk of sample contamination. Although catheterization significantly reduces contamination rates, its invasive nature and limited parental acceptance make routine clinical implementation difficult. Second, the retrospective single-center design presents additional constraints, as data completeness was limited by electronic medical record quality, and we did not control for potential confounders such as prior antibiotic use, which may affect the results. Multicenter prospective cohort studies would provide better validation of our findings. Third, the Random Forest model showed suboptimal calibration (ICI = 0.120) with systematic underestimation in the moderate probability range (0.4–0.6), a known limitation of tree-based ensemble methods. While this does not affect the model's discriminative ability (AUC = 0.88), absolute probability estimates should be interpreted cautiously in clinical practice. Clinicians should use model predictions as relative risk indicators to guide comprehensive assessment rather than as definitive probability estimates for individual patients.

## 5 Conclusion

This study successfully addresses a critical gap in pediatric UTI management by developing a clinically interpretable ML model for febrile children under 3 years of age. Through analysis of 1,556 cases, we identified seven key predictors using LASSO regression and demonstrated that the Random Forest model achieved superior performance (AUC = 0.88, AUPRC = 0.824, ICI = 0.12) compared to six other ML algorithms. This study introduces three principal innovations: overcoming traditional linear modeling limitations by capturing complex non-linear interactions among clinical variables; integrating the SHAP framework to transform the model from a "black box" into a transparent clinical decision support tool; and developing a web-based system that enables real-time risk assessment at the point of initial fever evaluation. These advances represent a paradigm shift from passive treatment to proactive risk-based management in pediatric UTI care, facilitating individualized antibiotic stewardship, risk-stratified diagnostic testing, and timely interventions.

While our study provides valuable insights into UTI risk assessment in febrile children younger than 3 years of age, several limitations warrant acknowledgment, including the retrospective single-center design, urine collection methodology challenges, and suboptimal calibration in moderate probability ranges. Multicenter prospective validation studies are therefore needed to refine model performance and assess the impact on clinical outcomes and cost-effectiveness. Nevertheless, this study establishes a robust foundation for ML-driven clinical decision support in pediatric infectious diseases and demonstrates the feasibility of combining accuracy with clinical interpretability to advance pediatric healthcare delivery.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Medical Ethics Committee of the Women's and Children's Hospital of Ningbo University. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by-product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

L-zY: Software, Writing – original draft. J-xS: Methodology, Writing – review & editing. JC: Writing – review & editing. K-kC: Data curation, Writing – review & editing. YB: Data curation, Writing – review & editing. Y-cL: Data curation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by Medical and Health Science and Technology Programme of Zhejiang Province (No. 2023KY1114); Ningbo Top Medical and Health Research Program (No. 2022020405); Ningbo Medical Key Disciplines (No. 2022-B17); Ningbo Medical Clinical Research Centre (No. 2019A21002).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Tullus K, Shaikh N. Urinary tract infections in children. *Lancet*. (2020) 395:1659–68. doi: 10.1016/S0140-6736(20)30676-0

2. Brandström P, Hansson S. Urinary tract infection in children. *Pediatr Clin N Am*. (2022) 69:1099–114. doi: 10.1016/j.pcl.2022.07.003

3. Bunting-Early TE, Shaikh N, Woo L, Cooper CS, Figueroa TE. The need for improved detection of urinary tract infections in young children. *Front Pediatr*. (2017) 5:24. doi: 10.3389/fped.2017.00024

4. Mattoo TK, Shaikh N, Nelson CP. Contemporary management of urinary tract infection in children. *Pediatrics*. (2021) 147(2):e2020012138. doi: 10.1542/peds.2020-012138

5. Esposito S, Biasucci G, Pasini A, Predieri B, Vergine G, Crisafi A, et al. Antibiotic resistance in paediatric febrile urinary tract infections. *J Glob Antimicrob Resist*. (2022) 29:499–506. doi: 10.1016/j.jgar.2021.11.003

6. Kane MMD. Diagnosing and treating urinary tract infections in the outpatient setting. *Pediatr Ann*. (2022) 51:e175–7. doi: 10.3928/19382359-20220314-01

7. Chandra T, Bajaj M, Iyer RS, Chan SS, Bardo DME, Chen J, et al. ACR appropriateness criteria® urinary tract infection-child: 2023 update. *J Am Coll Radiol*. (2024) 21:S326–42. doi: 10.1016/j.jacr.2024.02.025

8. Shaikh N, Haralam MA, Kurs-Lasky M, Hoberman A. Association of renal scarring with number of febrile urinary tract infections in children. *JAMA Pediatr*. (2019) 173:949–52. doi: 10.1001/jamapediatrics.2019.2504

9. Gkiourtzis N, Stoimeni A, Glava A, Chantavaridou S, Michou P, Cheirakis K. Prophylaxis options in children with a history of recurrent urinary tract infections: a systematic review. *Pediatrics*. (2024) 154:e2024066758. doi: 10.1542/peds.2024-066758

10. Uslu Gökceoğlu A, Taş N. Renal scarring in children with febrile urinary tract infection. *J Pediatr (Rio J)*. (2025) 101:370–4. doi: 10.1016/j.jped.2024.10.011

11. Yang SS, Tsai JD, Kanematsu A, Han C-H. Asian guidelines for urinary tract infection in children. *J Infect Chemother*. (2021) 27:1543–54. doi: 10.1016/j.jiac.2021.07.014

12. Marsh MC, Yepes Junquera G, Stonebrook E, Spencer JD, Watson JR. Urinary tract infections in children. *Pediatr Rev*. (2024) 45:260–70. doi: 10.1542/pir.2023-006017

13. Shaikh N, Hoberman A, Keren R, Ivanova A, Gotman N, Chesney RW, et al. Predictors of antimicrobial resistance among pathogens causing urinary tract infection in children. *J Pediatr*. (2016) 171:116–21. doi: 10.1016/j.jpeds.2015.12.044

14. Zaffanello M, Banzato C, Piacentini G. Management of constipation in preventing urinary tract infections in children: a concise review. *Eur Res J*. (2019) 5:236–43. doi: 10.18621/eurj.412280

15. Grier WR, Kratimenos P, Singh S, Guaghan JP, Koutroulis I. Obesity as a risk factor for urinary tract infection in children. *Clin Pediatr (Phila)*. (2016) 55:952–6. doi: 10.1177/0009922815617974

16. Chidambaram S, Pasupathy U, Geminiganesan S, Divya R. The association between vitamin D and urinary tract infection in children: a case-control study. *Cureus*. (2022) 14:e25291. doi: 10.7759/cureus.25291

17. Lejarzegi A, Fernandez-Uria A, Gomez B, Velasco R, Benito J, Mintegi S. Febrile urinary tract infection in infants less than 3 months of age. *Pediatr Infect Dis J*. (2023) 42:e278. doi: 10.1097/INF.0000000000003947

18. Balighian E, Burke M. Urinary tract infections in children. *Pediatr Rev*. (2018) 39:3–12. doi: 10.1542/pir.2017-0007

19. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. (2015) 349:255–60. doi: 10.1126/science.aaa8415

20. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med*. (2018) 284:603–19. doi: 10.1111/joim.12822

21. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine. *N Engl J Med*. (2023) 388:1201–8. doi: 10.1056/NEJMra2302038

22. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *Br Med J*. (2020) 368:m441. doi: 10.1136/bmj.m441

23. Renko M, Salo J, Ekstrand M, Pokka T, Pieviläinen O, Uhari M, et al. Meta-analysis of the risk factors for urinary tract infection in children. *Pediatr Infect Dis J*. (2022) 41:787–92. doi: 10.1097/INF.0000000000003628

24. Ramgopal S, Horvat CM, Yanamala N, Alpern ER. Machine learning to predict serious bacterial infections in young febrile infants. *Pediatrics*. (2020) 146:e20194096. doi: 10.1542/peds.2019-4096

25. Li S-C, Chi H, Huang F-Y, Chiu N-C, Huang C-Y, Chang L, et al. Building nomogram plots for predicting urinary tract infections in children less than three years of age. *J Microbiol Immunol Infect*. (2023) 56:111–9. doi: 10.1016/j.jmii.2022.08.006

26. Veauthier B, Miller MV. Urinary tract infections in young children and infants: common questions and answers. *Am Fam Physician*. (2020) 102:278–85.

27. Yim HE, Han KD, Kim B, Yoo KH. Impact of early-life weight status on urinary tract infection in children: a nationwide population-based study in Korea. *Epidemiol Health*. (2021) 43:e2021005. doi: 10.4178/epih.e2021005

28. Okada M, Kijima E, Yamamura H, Nakatani H, Yokoyama H, Imai M, et al. Obesity and febrile urinary tract infection in young children. *Pediatr Int*. (2022) 64(1):e14686. doi: 10.1111/ped.14686

29. Kocaaslan R, Dilli D, Çitli R. Diagnostic value of the systemic immune-inflammation Index in newborns with urinary tract infection. *Am J Perinatol*. (2024) 41:e719–27. doi: 10.1055/s-0042-1757353

30. Daniel M, Szymanik-Grzelak H, Sierdziński J, Podsiadły E, Kowalewska-Młot M, Pańczyk-Tomaszewska M. Epidemiology and risk factors of UTIs in children—a single-center observation. *J Pers Med*. (2023) 13:138. doi: 10.3390/jpm13010138

31. Fahimi D, Khedmat L, Afshin A, Noparast Z, Jafaripor M, Beigi EH, et al. Clinical manifestations, laboratory markers, and renal ultrasonographic examinations in 1-month to 12-year-old Iranian children with pyelonephritis: a six-year cross-sectional retrospective study. *BMC Infect Dis*. (2021) 21:189. doi: 10.1186/s12879-021-05887-1

32. Çaltek HÖ, Çaltek NC, Aras D, Çolak TNC, Okşen E, Yavuz S, et al. Prenatal diagnosis and postnatal outcomes of congenital kidney and urinary tract anomalies: results from a tertiary center. *BMC Pregnancy Childbirth*. (2025) 25(1):598. doi: 10.1186/s12884-025-07723-9

33. Chiodini B, Ghassemi M, Khelif K, Ismaili K. Clinical outcome of children with antenatally diagnosed hydronephrosis. *Front Pediatr*. (2019) 7:103. doi: 10.3389/fped.2019.00103

34. Herthelius M. Antenatally detected urinary tract dilatation: long-term outcome. *Pediatr Nephrol*. (2023) 38(10):3221–7. doi: 10.1007/s00467-023-05907-z

35. Herndon CDA, Otero HJ, Hains D, Sweeney RM, Lockwood GM. Perinatal urinary tract dilation: recommendations on Pre-/Postnatal imaging, prophylactic antibiotics, and follow-up: clinical report. *Pediatrics*. (2025) 156(1):e2025071814. doi: 10.1542/peds.2025-071814

36. Buettcher M, Trueck J, Niederer-Loher A, Heininger U, Agyeman P, Asner S, et al. Swiss Consensus recommendations on urinary tract infections in children. *Eur J Pediatr*. (2021) 180(3):663–74. doi: 10.1007/s00431-020-03714-4

frontiers | Frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 11 December 2025
DOI 10.3389/frai.2025.1692207

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire, United Kingdom

REVIEWED BY
John Pearson,
University of Utah Hospital, United States
Elham Albaroudi,
University of Salford, United Kingdom

*CORRESPONDENCE
Haoqi Cai
✉ chq_jtum@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 25 August 2025
REVISED 23 November 2025
ACCEPTED 24 November 2025
PUBLISHED 11 December 2025

CITATION
Wang Y, Yuan Y, Liu Y and Cai H (2025) Association of Shanghai air pollution with postoperative infection in adolescent orthopedic patients: a study using a deep learning-based evolutionary model. *Front. Artif. Intell.* 8:1692207. doi: 10.3389/frai.2025.1692207

COPYRIGHT
© 2025 Wang, Yuan, Liu and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Association of Shanghai air pollution with postoperative infection in adolescent orthopedic patients: a study using a deep learning-based evolutionary model

Yufeng Wang†, Yang Yuan†, Yang Liu and Haoqi Cai*

Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

**Background:** Surgical site infections (SSI) represent severe complications in adolescent orthopedic surgery. Shanghai's complex air pollution profile creates a critical context to investigate multi-pollutant impacts on SSI risk in this vulnerable population.
**Methods:** We analyzed 32,261 adolescent SSI cases from Shanghai (2019–2024) alongside high-resolution pollution/meteorological data. An evolutionary deep learning model (CNN-BiGRU-Attention optimized by Improved StarFish Algorithm) and generalized additive models (GAMs) assessed lagged effects, age/gender stratification, and concentration-response relationships.
**Results:** $NO_2$ and $SO_2$ showed significantly positive associations with SSI risk at lag0 (concurrent day); $O_3$ exhibited protective effects (strongest at lag05: −2.396% [95% CI: −3.349% to −1.443%] per 10 μg/$m^3$ increase); Age stratification: 7–14 ages groups demonstrated heightened sensitivity to $NO_2$/$SO_2$. $O_3$ effects varied across age groups; Gender differences: $O_3$'s negative association was stronger in males; Dose–response: $NO_2$/$SO_2$ showed monotonic increases with no safety thresholds; $O_3$ displayed a straight line curve.
**Conclusion:** Multi-pollutant exposure modulates SSI risk in adolescents, with $NO_2$/$SO_2$ as risk factors and $O_3$ showing context-dependent protection. Deep learning identified $SO_2$/$NO_2$/$O_3$ as dominant predictors, supporting perioperative air-quality interventions.

KEYWORDS

**surgical site infection, air pollution, adolescent orthopedics, deep learning, time-series analysis**

## 1 Introduction

Surgical site infections (SSI) are catastrophic complications following orthopedic procedures such as joint replacements and fracture fixation (Kolasiński, 2018; Tan et al., 2020; Tang et al., 2024). They prolong hospitalization, increase healthcare burdens, and may cause permanent dysfunction via deep infections like osteomyelitis (Tucci et al., 2019; Bath et al., 2022; Hammoor et al., 2021; Salášek et al., 2023). Despite advances in aseptic techniques, global SSI rates remain high, particularly among adolescents—a group with open growth

plates, rich vascularity, and distinct immune responses that may heighten vulnerability to exogenous pathogens (Rudic et al., 2023; Mallet et al., 2022).

Recent environmental epidemiology reveals that air pollutants can breach traditional infection control barriers: $PM_{2.5}$ transports endotoxins across alveolar barriers, activating systemic inflammation via Toll-like receptor 4 (TLR4) and impairing fibroblast migration/angiogenesis; $NO_2$ triggers excessive neutrophil extracellular trap (NET) release, creating DNA scaffolds for bacterial biofilm formation; and while ozone ($O_3$) has broad-spectrum antimicrobial properties, chronic exposure depletes surfactant protein A (SP-A), compromising macrophage clearance of *Staphylococcus aureus*. These findings collectively suggest a causal chain linking environmental exposure to SSI pathogenesis (Chehrassan et al., 2024; Liu et al., 2025).

As the core megalopolis of China's Yangtze River Delta, Shanghai exhibits a complex air pollution profile ideal for such investigations (Liu et al., 2024). Monitoring data reveal a seasonal $PM_{2.5}$-$O_3$ antagonism: coal-driven emissions elevate $PM_{2.5}/SO_2$ in winter, whereas intense photochemistry increases $O_3$ in summer (Wang et al., 2025). This "multi-pollutant alternation" dynamically influences surgical wound microenvironments—animal studies confirm $PM_{2.5}$ suppresses TGF-β1 (a key healing factor) by >50%, while clinical cohorts report 12.7% increased deep infection risk per 10 μg/m$^3$ $SO_2$ rise (Chehrassan et al., 2024; Li et al., 2025). Adolescents face amplified risks due to pollutant bioaccumulation in metabolically active growth plates and Th1/Th2 immune imbalance during puberty (Chen et al., 2019; Matia-Garcia et al., 2021). However, existing studies focus on adults or single pollutants, neglecting multi-pollutant synergies and adolescent susceptibility.

This study leverages Shanghai's 2019–2024 SSI data to address three questions: (1) Do $PM_{2.5}$, $NO_2$, $SO_2$, and $O_3$ exhibit pollutant-specific exposure windows? (2) Are adolescents hypersensitive to certain pollutants? (3) Do concentration-response curves show safety thresholds? By integrating environmental monitoring, deep learning, and epidemiological frameworks, we aim to build a "pollutant exposure–immune dysregulation–infection risk" cascade model to inform pediatric orthopedic center siting and perioperative air-quality management. We focused exclusively on Shanghai for several reasons: First, as a coastal megacity with a stable, large population and a well-defined urban core, it provides a contained environment to study urban pollution effects with minimized population mobility confounding factors. Second, its distinct seasonal pollution pattern ($PM_{2.5}/SO_2$-dominated in winter vs. $O_3$-dominated in summer) offers a natural experiment to investigate multi-pollutant impacts. While this design limits immediate generalizability to rural or industrially distinct regions, it allows for a rigorous, high-resolution analysis within a critical and representative urban context.

# 2 Methods

## 2.1 Data collection

Shanghai, situated on the estuary of the Yangtze River along China's east coast, is a major port city with a predominantly flat topography. Administratively, it is divided into 16 municipal districts, including the bustling central districts like Huangpu, Xuhui, and Jing'an, the expansive Pudong New Area (east of the Huangpu River, home to the financial hub Lujiazui), and outer districts such as Minhang, Baoshan, Jiading, and Songjiang. Shanghai experiences a humid subtropical monsoon climate, with a permanent resident population exceeding 24 million people, Shanghai stands as one of the most populous cities in the world and is a global center for finance, trade, and transportation. Information regarding daily outpatient visits for SSI between January 1, 2019, and December 31, 2024, was obtained from Zhongshan Hospital and Ruijin Hospital. Clinicians in the outpatient departments entered medical details into the hospital's computerized information system. Diagnoses of SSI were verified using the International Classification of Diseases, Tenth Revision (ICD-10) (Yanosky et al., 2024), along with clinical diagnostic criteria. The data were organized by diagnosis and personal attributes, including gender, age, and place of residence. To more precisely assess the influence of air pollution exposure in Shanghai's primary urban zone on SSI outpatient numbers, the data were refined to exclude patients whose permanent residences were outside the city's main urban area. The records were maintained daily, with no dates missing from the dataset. The study also incorporated environmental factors by leveraging information from the ShangHai Environmental Monitoring Center. This information included the mean daily levels of multiple pollutants from 2019 to 2024, such as particulate matter ($PM_{10}$ and $PM_{2.5}$), ozone ($O_3$), carbon monoxide (CO), nitrogen dioxide ($NO_2$), and sulfur dioxide ($SO_2$), Data was collected from 31 monitoring stations distributed across ShangHai's urban areas. Based on the statistical analysis of outpatient data, we excluded individuals who permanently resided in four areas of Shanghai known for higher pollution levels: Kangqiao Industrial Park in Pudong New Area, the Nanda Area in Baoshan District, the Taopu Intelligence & Innovation Community in Putuo District, and the Wujing Area in Minhang District. The pollution levels in these four areas are comparatively severe, and their inclusion in the study could have introduced significant bias into the results. Consequently, patients from these areas were excluded from our study population, accounting for 2.64, 1.98, 0.73, and 0.56% of the total outpatient visits, respectively. Additionally, the research integrated meteorological variables, like mean daily relative humidity and temperature, sourced from the ShangHai Meteorological Bureau. Data quality control was rigorously performed. For clinical SSI data, cases with missing critical fields (e.g., diagnosis date, age) were excluded (<0.5% of total records). For air pollution and meteorological data, any station reporting consecutive missing values for over 24 h was flagged. Isolated missing values (constituting <3% of the total dataset) were imputed using a linear interpolation method. Pollutant concentrations falling outside physiologically plausible ranges were treated as erroneous and removed prior to analysis. The final dataset was complete for all analysis variables.

To address potential confounding from surgical volume variations, we obtained annual orthopedic surgery records from participating hospitals (Zhongshan/Ruijin, 2019–2024). Total adolescent orthopedic procedures remained stable across seasons (Cool season: 12,384 ± 1,203; Warm: 13,097 ± 1,415; $p = 0.22$, t-test), with no significant temporal trend (ß = −0.17%, $p = 0.55$, linear regression). Additionally, surgical case-mix was homogeneous: >92% were fracture repairs (ORIF/external fixation), and geographic distributions showed >85% patients resided within 5 km of their operating hospital

(Supplementary Table S1). This consistency mitigates concerns that SSI outpatient fluctuations stem from procedural variability.

## 2.2 Evolutionary deep learning model

The proposed evolutionary deep learning model integrates a CNN-BiGRU-Attention architecture with swarm intelligence algorithms to optimize time-series data analysis. The CNN component extracts local spatiotemporal features, the BiGRU module captures long-short term dependencies, and the attention mechanism dynamically weights critical temporal segments. Through swarm intelligence-based hyperparameter optimization, key parameters including learning rate and convolutional kernel size are automatically tuned, thereby enhancing model generalizability and predictive performance. Figure 1 illustrates the architectural schematic.

### 2.2.1 CNN-BiGRU-attention architecture

This hybrid architecture combines convolutional neural networks (CNN), bidirectional gated recurrent units (BiGRU), and attention mechanisms through four operational phases: ① *Feature extraction:* The CNN layer employs 1D convolutions with multiple filter banks to capture localized patterns in meteorological and pollution time-series data; ② *Sequence modeling*: Bidirectional GRU layers process the extracted features in forward and reverse temporal directions, learning hierarchical representations of long-range dependencies; ③ *Attention weighting:* A self-attention layer computes normalized importance scores (0–1 range) for each time step through scaled dot-product operations, focusing on diagnostically relevant exposure windows; ④*Predictive output:* Weighted features are concatenated and processed through fully connected layers with sigmoid activation to generate final predictions.

### 2.2.2 Swarm intelligence optimization

While the CNN-BiGRU-Attention architecture effectively extracts deep temporal features from meteorological and environmental pollution data, the predictive accuracy of the model remains critically dependent on hyperparameter configuration. Conventional hyperparameter optimization methods, such as grid search, often exhibit suboptimal precision and computational inefficiency. To address these limitations, this study leverages the global optimization capabilities of swarm intelligence algorithms, integrating them with deep learning to optimize three pivotal hyperparameters: learning rate, BiGRU convolutional kernel size, and neuron count.

Building upon these foundations, we propose an enhanced swarm intelligence algorithm - the Improved StarFish Optimization Algorithm (ISFOA) - through strategic modifications to the baseline StarFish Optimization Algorithm (SFOA) (Zhong et al., 2025). Two key algorithmic advancements are implemented:

① *Chaotic population initialization*

The Tent mapping equation governs initial population distribution to ensure uniform coverage of the hyperparameter search space:

$$Z_{t+1} = \begin{cases} \dfrac{Z_t}{\alpha}, 0 \le Z_t \le \alpha \\ \dfrac{1-Z_t}{1-\alpha}, \alpha < Z_t < 1 \end{cases}$$

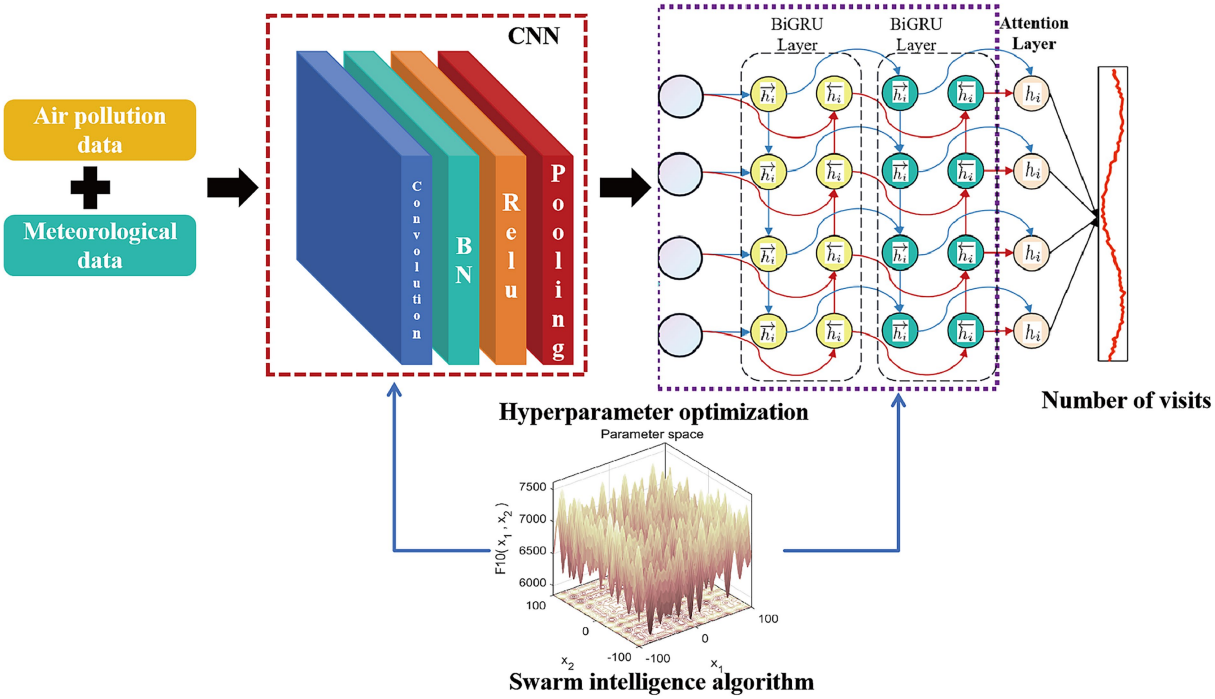


FIGURE 1
Schematic diagram of the deep learning evolutionary model structure.

The chaotic sequence is subsequently mapped to the solution space through linear transformation:

$$x_{id} = x_L + \left(x_U - x_L\right) \cdot z_{id}$$

where $x_{id}$ denotes the position of the $i^{th}$ starfish individual in the $d^{th}$ dimension, $x_U$和$x_L$ represent the upper/lower bounds of the search space, and $z_{id}$ corresponds to the chaotic sequence value.

② *Adaptive lens imaging opposition learning*

To counteract population diversity loss during later iterations, a dynamic opposition strategy is implemented:

$$x_j^* = \frac{a_j + b_j}{2} + \frac{a_j + b_j}{2k} - \frac{x_j}{k}$$

where k, the regulating factor, adaptively evolves as:

$$k = \left(1 + \left(\frac{t}{T}\right)^{\frac{1}{2}}\right)^{10}$$

### 2.2.3 Model validation

To ensure model robustness and prevent overfitting, we employed a 5-fold time-series cross-validation strategy. The model's performance was monitored using the loss on a held-out validation set (20% of training data), and early stopping was implemented with a patience of 10 epochs. We compared our proposed evolutionary deep learning model against several benchmarks, including Generalized Additive Models (GAM), Random Forest, and standalone CNN and GRU models. Performance was evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The superior performance of our model is detailed in Supplementary Table S2.

## 2.3 Deep learning interpretability analysis

The interpretability of the constructed machine learning model was rigorously investigated through deep learning feature layer visualization and Shapley value analysis. Rooted in cooperative game theory, the Shapley value method offers a theoretically equitable mechanism to quantify individual feature contributions to model predictions by calculating their marginal explanatory impacts. Two specialized visualization tools were employed to decode the model's decision logic: ① *Shapley summary plot:* Visualizes the distributional characteristics of feature contributions through scatter point density, where horizontal axis positioning reflects effect direction (positive/negative) and color intensity indicates covariate value magnitudes; ② *Shapley feature importance plot:* Ranks all input variables by their mean absolute Shapley values, with bar lengths proportional to their overall influence weights. This hierarchy explicitly reveals dominant environmental determinants (e.g., PM2.5 lag effects) and secondary meteorological factors in SSI risk prediction.

The synergistic application of these techniques enabled comprehensive interrogation of the model's spatiotemporal reasoning patterns, particularly in identifying critical pollution exposure windows preceding SSI onset.

## 2.4 Traditional time-series analysis

The examination of SSI outpatient visit trends generally adheres to a Poisson distribution, often exhibiting a non-linear association with explanatory factors. Consequently, our study employed a generalized additive model (GAM) (Islam et al., 2024), supplemented by time-series analysis, to establish relationships between SSI outpatient visits, air pollution concentrations, and meteorological data organized chronologically.

During the construction of the Generalized Additive Model (GAM), various variables were integrated based on insights from prior research. The natural spline (ns) function was applied to calendar time, and the model's performance was evaluated using the Akaike Information Criterion (AIC), where smaller AIC values indicated a more optimal fit. Furthermore, we assessed residual autocorrelations through partial autocorrelation functions (PACF) (Chen et al., 2025). Considering the AIC and PACF results alongside existing studies, our model allocated three degrees of freedom (df) per year to account for long-term trends and seasonal variations in SSI outpatient visits (Meng et al., 2023).

Additionally, to account for potential non-linear influences from weather conditions, we assigned four degrees of freedom (df) to relative humidity and eight degrees of freedom to average temperature. Finally, the variable 'day of the week' was incorporated as an additional categorical factor in the model (Tsai and Yang, 2024). The structure of the independent model is outlined as follows:

$$\begin{aligned}\log\mathrm{E}\left(Y\beta\right) = \beta Zt + \mathrm{DOW} + \mathrm{ns}\left(\mathrm{time},\mathrm{df}\right) + \mathrm{ns}\left(\mathrm{temperature},6\right) + \\ \mathrm{ns}\left(\mathrm{humidity},3\right) + \mathrm{Holiday} + \alpha\end{aligned}$$

In our framework, $E(Yt)$ signifies the anticipated count of outpatient visits on day $t$, where $\beta$ represents the logarithmic relative rate of visits associated with each unit rise in pollutant concentration. $Zt$ denotes the pollutant levels on day t$t$, while $DOW$ accounts for variations due to weekdays. The term $ns$(time, $df$) refers to the natural spline function applied to calendar time. Likewise, $ns$(temperature, $df$) and $ns$(humidity, $df$) are natural spline functions used to model temperature and humidity, with degrees of freedom set to 9 and 3, respectively. The 'Holiday' variable is included to account for fluctuations during holidays. The model's intercept is represented by $\alpha\alpha$. For the two-pollutant model, a second pollutant is introduced into $Zt$ to assess the combined impact of multiple pollutants.

To ensure the robustness of our findings, an extensive series of sensitivity analyses was performed. Initially, a single-pollutant model was employed to assess the relationship between air pollutants and SSI cases. Subsequently, we investigated various lag intervals, including individual day lags (1 to 5 days) and cumulative lag periods (lag 01, 03, 05), to analyze both short-term and extended effects of pollutant exposure. Additionally, dual-pollutant models were introduced to account for interactions between multiple pollutants. Age-specific and gender-specific analyses were also conducted to evaluate how pollution impacts vary across different age groups. Using the GAM approach, we plotted dose–response curves to illustrate the association between SSI outpatient visits and pollutant concentrations.

All analyses were conducted using R software (version 2.15.1) with the 'mgcv' package, and statistical significance was defined as a

*p*-value below 0.05. The study evaluated variations in SSI outpatient visits corresponding to every 10 μg/m$^3$ rise in daily air pollutant levels.

# 3 Results

## 3.1 Basic descriptive statistics

Table 1 provides an overview of the descriptive statistics. Between January 1, 2019, and December 31, 2024 (a total of 1,095 days), 32,261 SSI cases were recorded. The mean daily outpatient admissions amounted to 12.66. When categorized by age, the distribution was 10,46 (32.41%) for individuals under 6, 16,24 (50.35%) for those aged 7–14, and 5,56 (17.24%) for those 15–18. The average daily concentrations of air pollutants were 52.97 μg/m$^3$ for $SO_2$, 46.18 μg/m$^3$ for $PM_{2.5}$, 63.28 μg/m$^3$ for $NO_2$, 95.64 μg/m$^3$ for $PM_{10}$, 1.38 mg/m$^3$ for CO, and 67.76 μg/m$^3$ for $O_3$ (8-h maximum). The $PM_{2.5}/PM_{10}$ ratio averaged 0.48, highlighting the prevalence of $PM_{2.5}$. The mean daily relative humidity was 75.40%, while the average temperature was 18.21 °C. Annual orthopedic surgery volume exhibited no significant correlation with pollutant levels ($r = -0.21$ to 0.18, $p > 0.05$) and stable seasonal distribution (Figure 2). Case-mix consistency was confirmed: fracture repairs accounted for 92.8 ± 1.6% of all procedures across hospitals.

## 3.2 Construction of evolutionary deep learning model

The ISFOA demonstrated superior optimization performance (Supplementary Figures S1, S2), and its application to train the CNN-BiGRU-Attention model yielded the training dynamics shown in Supplementary Figure S3, feature visualization of the flatten layer output (Figure 3) revealed that $SO_2$, $NO_2$, and $O_3$ concentrations exhibited the highest variation rates and magnitude in feature activation maps, indicating their dominant informational contribution to prediction outputs. Interpretability analysis further corroborated these findings, with SHAP value rankings identifying $SO_2$, $NO_2$, and $O_3$ as the top three predictive determinants (Figure 4).

## 3.3 Spearman correlation analysis

Table 2 displays the Spearman correlation coefficients between meteorological factors and air pollutants. $PM_{2.5}$, $PM_{10}$, $NO_2$, $SO_2$, and CO generally demonstrated moderate to strong correlations with one another. With the exception of $PM_{2.5}$, most pollutants exhibited negative associations with humidity or temperature. $O_3$ displayed inverse correlations with $PM_{2.5}$, $PM_{10}$, $NO_2$, $SO_2$ and NO.

TABLE 1 Summary statistics for daily outpatient visits, air pollutant levels, and meteorological data.

| Feature | Mean | Min | Max | P25 | P50 | P75 | SD |
|---|---|---|---|---|---|---|---|
| SSI | 12.66 | 0.0 | 45.0 | 7.0 | 12.0 | 17.0 | 6.90 |
| Sex | | | | | | | |
| Male | 6.97 | 0.0 | 27.0 | 4.0 | 6.0 | 10.0 | 4.20 |
| Female | 5.68 | 0.0 | 21.0 | 3.0 | 5.0 | 8.0 | 3.65 |
| Age | | | | | | | |
| 0–6 | 4.10 | 0.0 | 16.0 | 2.0 | 4.0 | 6.0 | 2.83 |
| 7–14 | 6.37 | 0.0 | 25.0 | 3.0 | 6.0 | 9.0 | 3.95 |
| 15–18 | 2.18 | 0.0 | 14.0 | 1.0 | 2.0 | 3.0 | 1.86 |
| Season | | | | | | | |
| Cool | 12.49 | 0.0 | 45.0 | 7.0 | 12.0 | 17.0 | 7.15 |
| Warm | 12.82 | 1.0 | 37.0 | 8.0 | 12.0 | 17.0 | 6.64 |
| Air pollutants concentrations (24-h average) | | | | | | | |
| $PM_{2.5}$ (μg/m$^3$) | 46.18 | 3.0 | 427.0 | 21.0 | 42.0 | 67.0 | 38.91 |
| $PM_{10}$ (μg/m$^3$) | 95.64 | 3.0 | 643.0 | 52.0 | 83.0 | 120.0 | 67.36 |
| $O_3$ (μg/m$^3$) 8-h maximum | 67.76 | 4.0 | 198.0 | 38.0 | 56.0 | 83.7 | 38.26 |
| CO (mg/m$^3$) | 1.38 | 0.3 | 5.76 | 0.9 | 1.28 | 1.76 | 0.67 |
| $NO_2$ (μg/m$^3$) | 63.28 | 12.0 | 164.0 | 48.0 | 60.0 | 75.0 | 24.48 |
| SO2 (μg/m$^3$) | 52.97 | 10.0 | 234.0 | 35.0 | 56.0 | 74.0 | 35.14 |
| Meteorological measure (24-h average) | | | | | | | |
| Temperature (°C) | 18.21 | −2.5 | 36.5 | 10.4 | 19.1 | 26.2 | 8.98 |
| Relative humidity (%) | 75.40 | 32.0 | 71.0 | 66.0 | 75.0 | 83.0 | 12.71 |

$SO_2$, sulfur dioxide; $NO_2$, nitrogen dioxide; CO, carbon monoxide; $O_3$, ozone; $PM_{2.5}$, particulate matter with aerodynamic diameter less than 2.5 μm; $PM_{10}$, particulate matter with aerodynamic diameter less than 10 μm; PMc, particulate matter with aerodynamic with diameter 2.5–10 um; SD, standard error; P25, P50, and P75: the 25th, 50th, and 75th percentiles.

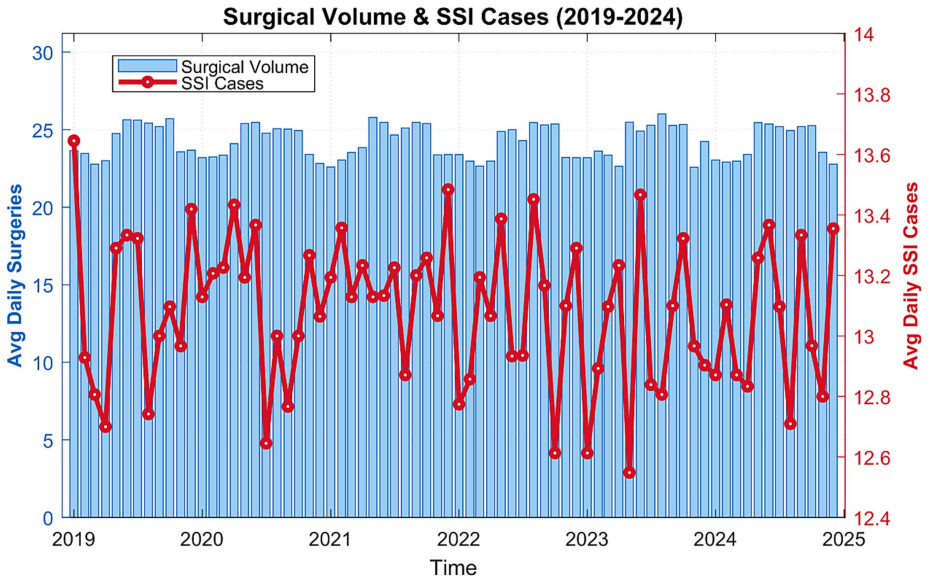


FIGURE 2
Temporal trends in surgical volume vs. SSI cases.

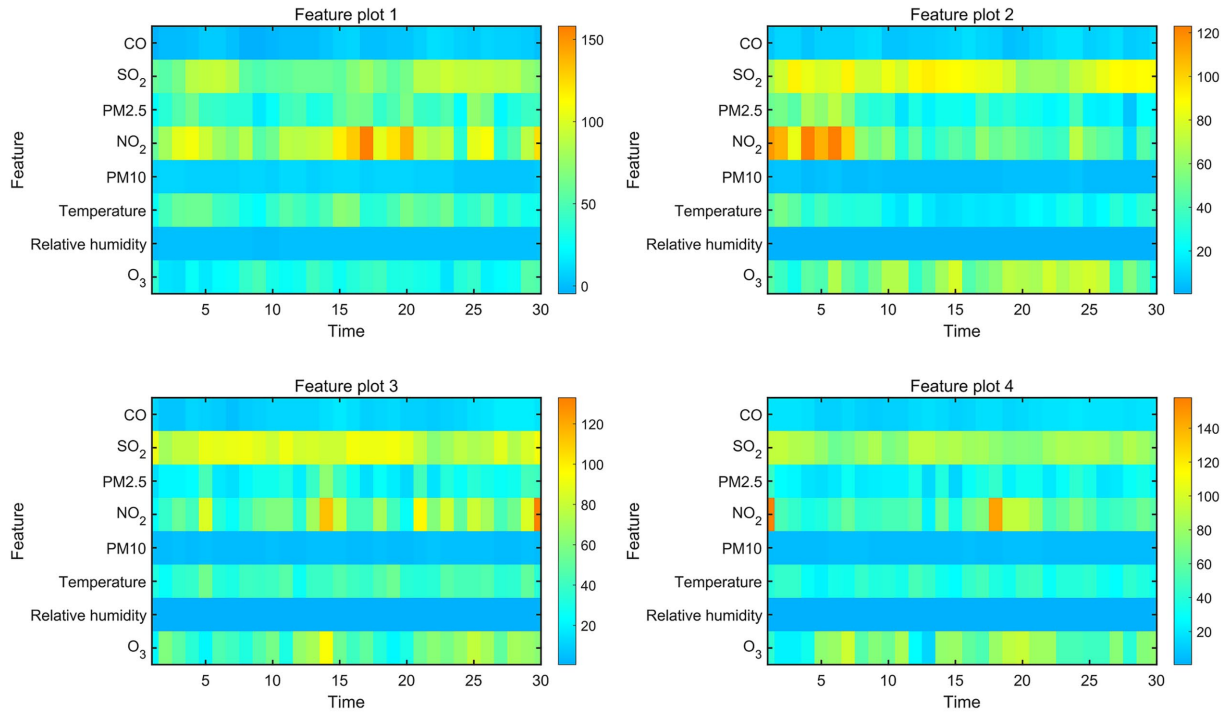


FIGURE 3
Deep learning feature visualization.

## 3.4 Single pollutants models

Table 3 outlines the relationships between air pollutants and daily SSI outpatient visits across single-lag days (lag0–lag5) and cumulative exposure periods (lag 01, 03, 05). The effects are expressed as percentage changes (mean and 95% confidence intervals) in daily visits per 10 μg/m$^3$ increase in $PM_{2.5}$, $PM_{10}$, $NO_2$, $SO_2$, and $O_3$, or per 0.1 mg/m$^3$ increase in CO. With the exception of $O_3$, the most pronounced and statistically significant associations between air pollutants and SSI were observed at lag01. For $NO_2$ and $SO_2$, positive and statistically significant correlations were identified in both moving average exposure periods (lag01, lag03, and lag05) and single-lag days

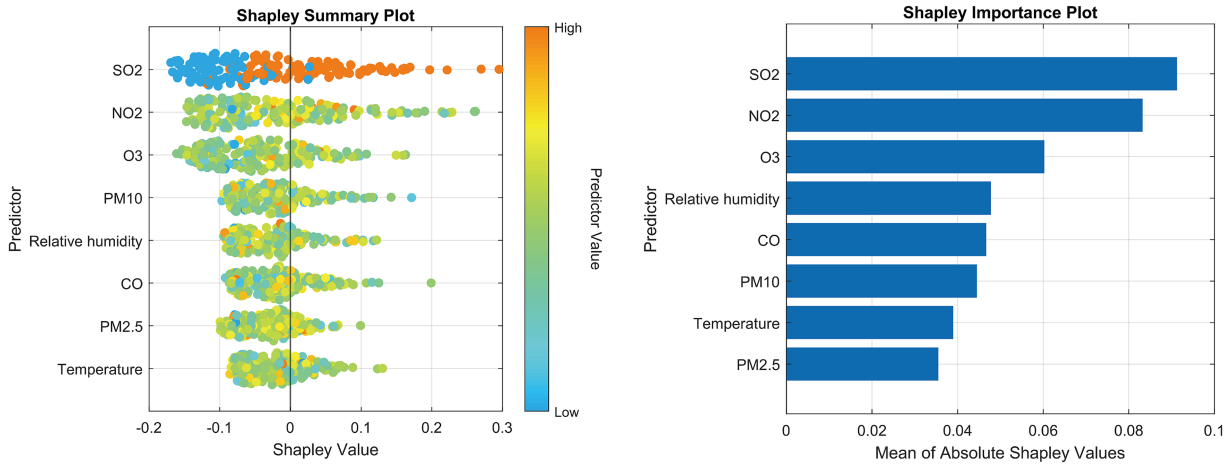


FIGURE 4
SHAP analytical results. Left panel displays Shapley summary plot illustrating feature effect directions; Right panel shows Shapley importance ranking by absolute contribution values.

TABLE 2 Spearman correlation coefficients between daily air pollutant concentrations and weather conditions.

| Variable | $PM_{2.5}$ | $SO_2$ | $NO_2$ | $O_3$ | CO | Temperature | Humidity |
|---|---|---|---|---|---|---|---|
| $PM_{10}$ | 0.825** | 0.615** | 0.719** | −0.199** | 0.688** | −0.419** | −0.159** |
| $PM_{2.5}$ | | 0.724** | 0.751** | −0.216** | 0.716** | −0.473** | 0.029 |
| $SO_2$ | | | 0.722** | −0.385** | 0.782** | −0.618** | −0.194** |
| $NO_2$ | | | | −0.205** | 0.701** | −0.398** | −0.162** |
| $O_3$ | | | | | −0.484** | −0.692** | −0.157** |
| CO | | | | | | −0.799** | −0.043 |
| Temperature | | | | | | | −0.001 |

**$p < 0.01$.

(lag0, lag1). $O_3$ exhibited a negative association with SSI in cumulative exposure periods (lag 01, 03, 05) and single-lag days (lag1–lag5), with the most notable effect at lag05: a 10 μg/m$^3$ rise in $O_3$ was linked to a 2.396% (95% CI: 3.349, 1.443%) reduction in outpatient visits. Notably, $PM_{2.5}$ and CO, despite showing significant Spearman correlations with other pollutants (Table 2), did not demonstrate statistically significant independent associations with SSI risk in the single-pollutant lag models, and thus were not the primary focus of subsequent stratified and interpretability analyses.

## 3.5 Age models

Figure 5 illustrates the outcomes of age-specific analyses across various lag models. Three pollutants ($NO_2$, $SO_2$, and $O_3$) exhibited distinct relationships with SSI outpatient visits depending on the lag model. Notably, the positive association of $SO_2$ was only significant for the 7–14 age group at lag0-4 and lag01-03, while the positive association of $NO_2$ was significant at lag 0, lag1, and lag 01 for both the <6 and 7–14 age groups. Neither $SO_2$ nor $NO_2$ demonstrated a significant relationship in the 15–18 age group at lag 01 to lag 05. In contrast, the most pronounced negative association of $O_3$ was particularly significant for the 7–14 age group in the age-specific analysis, with stronger effects observed in this group compared to the overall population. Overall, more significant positive associations of $SO_2$ and $NO_2$ were identified for adults aged 7–14, whereas $O_3$ showed a negative association with SSI in the 7–14 age group.

## 3.6 Gender models

Figure 6 displays the gender-specific model, revealing that $SO_2$ and $NO_2$ exhibit positive correlations with SSI outpatient visits in both male and female groups, with no notable differences between the two. In contrast, the negative correlation effect of $O_3$ is more pronounced in the male group than in the female group.

## 3.7 Two-pollutant models

Table 4 presents the outcomes of two-pollutant models at lag 0 (with $O_3$ at lag 4). We noted that the effects of $NO_2$ and $SO_2$ became stronger statistically significant when adjusted for other pollutants. However, the associations of $SO_2$ and $NO_2$ lost statistical significance after controlling for $NO_2$ and $SO_2$, respectively. Regarding $O_3$, the relationship showed no significant change after adjusting for other pollutants.

TABLE 3 The lag effects for single pollutants.

| Lag | $PM_{10}$ | $PM_{2.5}$ | $SO_2$ | $NO_2$ | $O_3$ | CO |
|---|---|---|---|---|---|---|
| 0 | 0.027 (−0.183, 0.238) | 0.151 (−0.194, 0.496) | 2.536 (1.131, 3.941)*** | 2.853 (1.393, 4.313)*** | −0.209 (−0.860, 0.442) | 0.221 (−0.142, 0.584) |
| 1 | −0.022 (−0.227, 0.183) | 0.029 (−0.304, 0.362) | 1.478 (0.091, 2.865)* | 2.216 (0.767, 3.666)** | −0.739 (−1.331, −0.147)* | −0.001 (−0.355, 0.354) |
| 2 | −0.076 (−0.274, 0.122) | 0.003 (−0.318, 0.324) | 0.791 (−0.580, 2.162) | 1.003 (−0.391, 2.397) | −0.809 (−1.341, −0.278)** | −0.167 (−0.507, 0.172) |
| 3 | −0.125 (−0.329, 0.079) | 0.053 (−0.272, 0.378) | 0.764 (−0.580, 2.108) | 0.074 (−1.288, 1.437) | −0.787 (−1.305, −0.268)** | −0.283 (−0.626, 0.061) |
| 4 | −0.185 (−0.397, 0.027) | −0.029 (−0.362, 0.304) | 0.677 (−0.652, 2.007) | −0.322 (−1.689, 1.044) | −1.037 (−1.559, −0.516)*** | −0.330 (−0.679, 0.019) |
| 5 | −0.108 (−0.313, 0.098) | 0.047 (−0.276, 0.369) | −0.886 (−2.237, 0.465) | −0.524 (−1.891, 0.844) | −0.946 (−1.457, −0.435)*** | −0.268 (−0.611, 0.075) |
| 01 | −0.002 (−0.233, 0.228) | 0.086 (−0.284, 0.456) | 2.575 (0.955, 4.194)** | 3.157 (1.532, 4.783)*** | −0.859 (−1.671, −0.048)* | 0.104 (−0.284, 0.492) |
| 03 | −0.104 (−0.368, 0.160) | 0.033 (−0.386, 0.452) | 2.505 (−0.535, 4.475)* | 2.511 (0.633, 4.389)** | −1.661 (−2.546, −0.776)*** | −0.144 (−0.574, 0.285) |
| 05 | −0.180 (−0.480, 0.120) | 0.057 (−0.413, 0.528) | 2.431 (0.150, 4.712)* | 1.839 (−0.260, 3.938) | −2.396 (−3.349, −1.443)*** | −0.262 (−0.733, 0.208) |

Percentage change (mean and 95% confidence interval) in daily SSI outpatient visits per 10 μg/m³ increase ($PM_{10}$, $PM_{2.5}$, $SO_2$, $NO_2$, $O_3$) or per 0.1 mg/m³ increase (CO) in pollutant concentrations across various lag days. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

## 3.8 Exposure-response curves

Figure 7 depicts the exposure-response curves for $SO_2$ and $NO_2$ (lag 0) and $O_3$ (lag 4) in relation to SSI outpatient visits. The curves indicate that the associations of $NO_2$ and $SO_2$ were positive and exhibited no apparent thresholds. Conversely, $O_3$ showed a negative association with SSI outpatient visits. The slope remained relatively flat as concentrations rose from 0 μg/m³ to 150 μg/m³ but became steeper at concentrations exceeding 150 μg/m³.

# 4 Discussion

Our hybrid methodological framework reveals complex links between multi-pollutant exposure and SSI risk in the Yangtze River Delta. Both SHAP interpretability and GAM analyses confirm the risk-enhancing effects of $NO_2$/$SO_2$. Pathologically, these gases may: (1) induce systemic inflammation that impairs macrophage phagocytosis and epithelial integrity; and (2) form $NO_2$/$SO_2$-bound complexes that deliver endotoxins to surgical sites, disrupting collagen remodeling. Paradoxically, $O_3$ showed protective associations—a phenomenon contextualized by Shanghai's photochemical pollution (Liu et al., 2025; Supphapipat et al., 2024; Montagna et al., 2019; He et al., 2023). Summer-derived $O_3$ exerts antimicrobial effects but concurrently damages pulmonary surfactants, inducing compensatory immunosuppression (Agostini et al., 2021). Two-pollutant models further revealed effect modification between $NO_2$ and $SO_2$, suggesting synergistic exposure from traffic/industrial sources. Our deep learning interpretability and GAM analyses consistently identified $SO_2$, $NO_2$, and $O_3$ as the dominant predictors, while $PM_{2.5}$ and CO, though correlated, showed weaker direct associations. This suggests that the effects of $PM_{2.5}$ and CO on SSI risk in this cohort may be largely mediated or confounded by their strong co-linearity with the gaseous pollutants ($SO_2$ and $NO_2$), or that their biological pathways to impacting surgical site immunity are less direct in the context of adolescent orthopedic surgery. Despite lacking granular surgical metadata, three evidence lines confirm pollution-SSI associations are not confounded by procedural factors: (1) Surgical volume showed minimal seasonal/correlative fluctuation with pollutants; (2) Case-mix homogeneity (>92% fracture repairs) reduces bias from infection-prone procedures; (3) Pollutant effects peaked at lag0/lag1 – biologically implausible if driven by delayed surgical volume changes.

Age stratification highlighted adolescents' unique vulnerability. The heightened vulnerability in 7–14-year-olds likely stems from accelerated bone remodeling and immature immunity during active growth phases. $SO_2$/$NO_2$ exposure may disrupt osteoblast–osteoclast balance, impairing surgical repair, while younger children's developing immune systems amplify inflammatory responses. By ages 15–18, skeletal maturation and stabilized immunity confer resilience against pollution-driven SSI risk. Notably, $O_3$'s protective effect was more pronounced in males, aligning with their larger lung surface area and higher $O_3$ uptake (Liu et al., 2025; Supphapipat et al., 2024). Exposure-response curves confirmed no safe thresholds for $NO_2$/$SO_2$, while $O_3$ displayed a plateau effect at low concentrations and a steep slope >150 μg/m³—indicating a potential "harvest effect" where short-term disinfection masks long-term immune damage. In adolescents, whose immune systems are in a state of dynamic development and whose bones undergo active remodeling, the pro-inflammatory and tissue-disruptive effects of $NO_2$ and $SO_2$ may be particularly detrimental. The heightened inflammatory response could dysregulate the delicate balance of osteoblast and osteoclast activity at the surgical site, impairing early bone healing and creating a niche more susceptible to colonization by pathogens. The observed protective association of $O_3$ presents a fascinating paradox. Its well-documented antimicrobial properties in the gaseous phase may theoretically reduce bacterial load in the ambient air and on superficial wounds, potentially lowering the initial inoculum for infection. This acute, disinfectant effect might dominate in the short-term perioperative period we analyzed. However, this must be contextualized against extensive literature on $O_3$'s toxicity. Chronic or high-dose exposure damages respiratory epithelium, depletes protective surfactant proteins, and can induce a state of compensatory immunosuppression. We hypothesize that our study captured the net short-term effect, where the antimicrobial action outweighed the sub-acute immunosuppressive effects in this specific post-surgical context. The straight-line dose–response curve at higher concentrations, however, hints at a potential threshold beyond which its harmful effects might prevail, a critical area for future investigation.

The ISFOA-optimized hybrid model uncovered spatiotemporal exposure patterns overlooked by conventional approaches, particularly gaseous pollutants' critical lag0 windows. SHAP interpretability transcended statistical associations by hierarchically ranking $SO_2$/$NO_2$/$O_3$ as dominant features. This synergy informs actionable strategies: (1) deploy real-time pollution monitors in

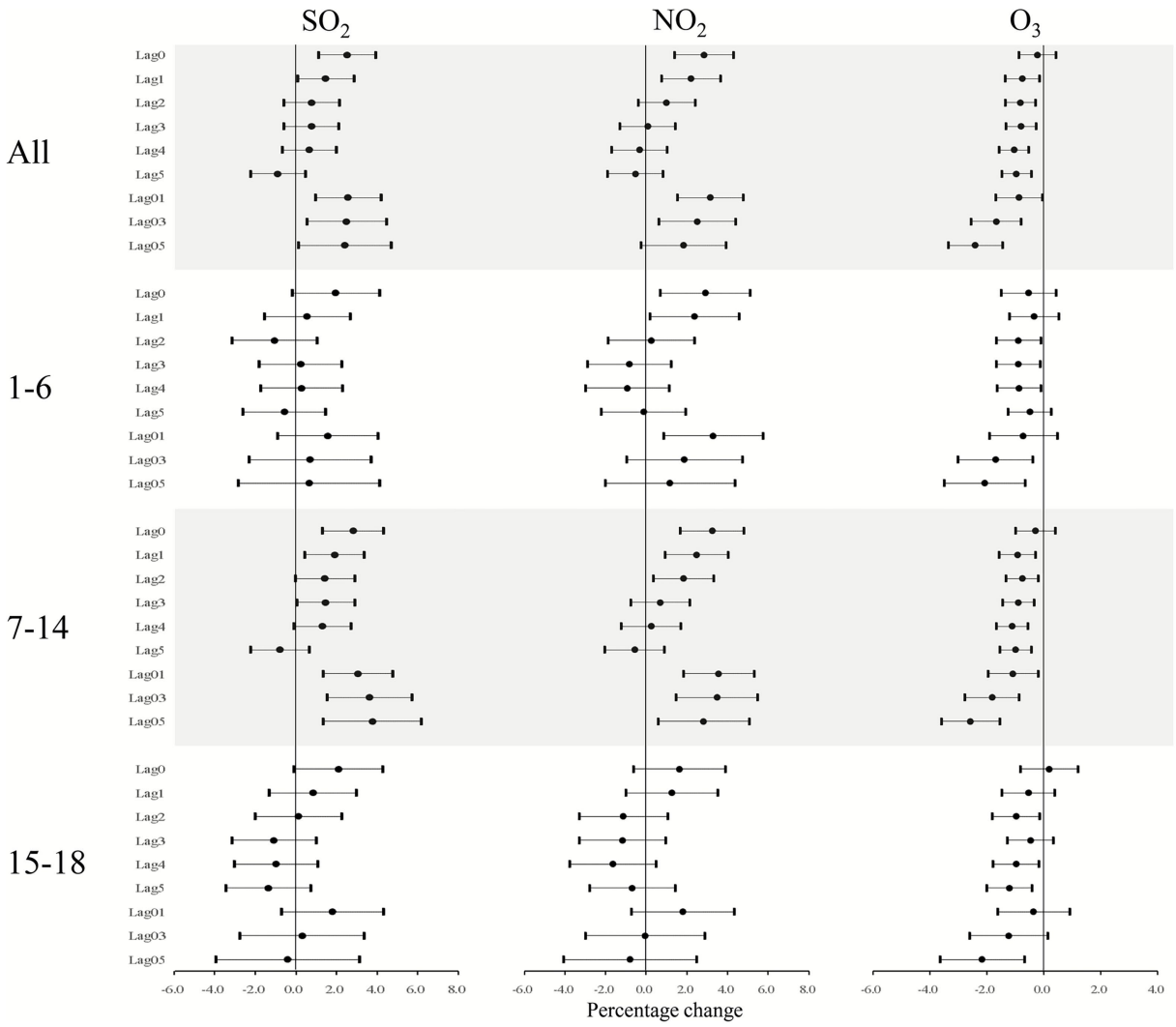


FIGURE 5
Percentage increase in daily SSI outpatient visits linked to a 10 μg/m$^3$ rise in pollutant concentrations ($NO_2$, $SO_2$, and $O_3$) across various lag days in age-specific analysis models. Results are presented as means with 95% confidence intervals.

orthopedic wards, activating air purification during industrial/traffic peaks; (2) schedule elective pediatric surgeries in high-$O_3$ seasons while avoiding high-$NO_2$/$SO_2$ periods. Future studies should dissect pollutant-bone immunology interactions to enable precision environmental interventions.

Our findings have tangible implications for clinical practice and health policy. Firstly, hospitals, particularly those specializing in pediatric orthopedics, could invest in real-time indoor air quality monitoring and high-efficiency particulate air (HEPA) filtration systems with activated carbon filters to mitigate $NO_2$ and $SO_2$ ingress. Secondly, for elective surgeries, our data suggest a potential benefit in scheduling procedures during seasons or days with lower ambient $NO_2$/$SO_2$ and moderate $O_3$ levels, as forecasted by local environmental agencies. While implementing such a 'pollution-aware' surgical schedule requires logistical planning and validation in prospective studies, it represents a low-cost, high-potential intervention. Finally, these results strengthen the case for urban public health policies aimed at reducing traffic and industrial emissions, framing them not just as environmental issues but as direct contributors to surgical patient safety and healthcare costs. In addition, The observed statistical association is unlikely to represent direct causation from a single day's exposure. A more plausible explanation is that the pollution level on the visit day acts as a proxy for recent cumulative exposure, which may have pre-conditioned the patient's immune system. The surgery then served as a "second hit," leading to a dysregulated inflammatory response that increased infection susceptibility.

This study has several limitations. First, while we employed spatial optimization, exposure misclassification remains possible due to the use of ambient monitoring data rather than personal measurements. Second, despite adjusting for major confounders, residual confounding from unmeasured factors (e.g., environmental tobacco smoke, occupational dust exposure, or detailed individual socioeconomic status) cannot be fully ruled out. Third, the observational design precludes definitive causal inferences, which is particularly relevant for the intriguing protective association of $O_3$; this finding may reflect unmeasured confounding or complex effect modification rather than a direct biological effect; Fourth, the ecological study design, while useful

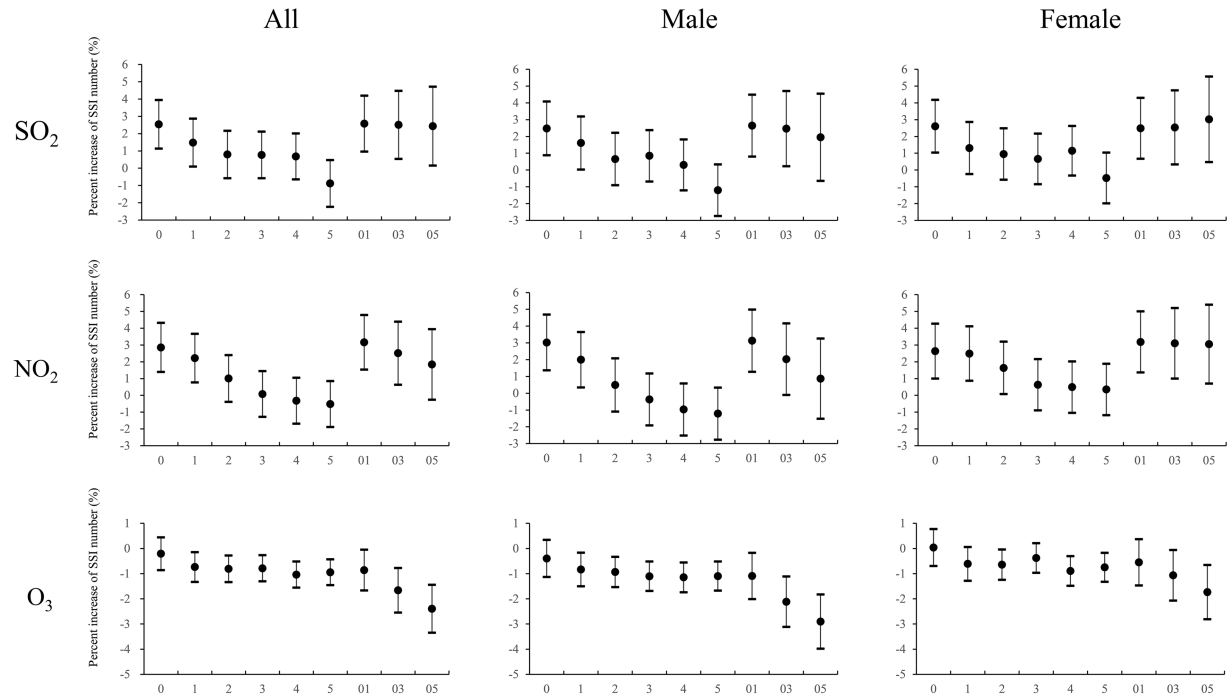


FIGURE 6
Percentage increase in daily SSI outpatient visits associated with a 10 μg/m³ increase in pollutant concentrations ($NO_2$, $SO_2$, and $O_3$) at different lag days in sex-specific analysis models. Values are shown as means and 95% confidence intervals.

TABLE 4 **Percentage increase in daily SSI outpatient visits linked to a 10 μg/m³ rise ($NO_2$, $SO_2$, $O_3$) in pollutant concentrations in two-pollutant models.**

| Primary pollutant | Two-pollutant models | Estimates |
|---|---|---|
| $SO_2$ | – | 2.536 (1.131, 3.941)*** |
| | +$PM_{10}$ | 3.599 (1.930, 5.268)*** |
| | +$PM_{2.5}$ | 3.906 (2.148, 5.663)*** |
| | +$NO_2$ | 1.392 (−0.367, 3.150) |
| | +CO | 2.516 (1.107, 3.926)*** |
| | +$O_3$ | 2.915 (1.246, 4.585)*** |
| $NO_2$ | – | 2.853 (1.393, 4.313)*** |
| | +$PM_{10}$ | 4.696 (2.799, 6.593)*** |
| | +$PM_{2.5}$ | 5.023 (3.044, 7.002)*** |
| | +$SO_2$ | 1.818 (−0.178, 3.814) |
| | +CO | 2.951 (1.479, 4.423)*** |
| | +$O_3$ | 3.916 (1.990, 5.843)*** |
| $O_3$ | – | −1.037 (−1.559, −0.516)*** |
| | +$PM_{10}$ | −1.036 (−1.558, −0.514)*** |
| | +$PM_{2.5}$ | −1.037 (−1.559, −0.515)*** |
| | +$SO_2$ | −1.008 (−1.530, −0.486)*** |
| | +$NO_2$ | −0.980 (−1.503, −0.458)*** |
| | +CO | −1.033 (−1.555, −0.511)*** |

***$P < 0.001$.

for generating hypotheses, had limited capacity for causal inference. By analyzing associations at the group level, it could suggest but not establish causality, as controlling for confounders is difficult. A further limitation of our outpatient dataset was the absence of individual-level details on medications, infection sites, and antibiotic use, a challenge common to ecological studies. Finally, the generalizability of our findings may be constrained as they are derived from a single megacity with a unique pollution profile and demographic context. Future research should integrate personal monitoring, incorporate more granular clinical and social covariates, and employ multi-center designs across diverse geographic settings to validate our results and explore the underlying mechanisms, such as through the analysis of wound microbiome virulence factors or host immune epigenetic modifications.

## 5 Conclusion

This study establishes significant associations between multi-pollutant exposure and SSI risk in Shanghai's adolescent orthopedic patients. $NO_2$ and $SO_2$ emerged as consistent risk factors, with peak effects at lag01 and heightened susceptibility among older adolescents (15–18 years). Conversely, $O_3$ demonstrated protective effects, potentially attributable to its antimicrobial properties, though its straight line dose–response curve warrants caution at high concentrations. The evolutionary deep learning model (optimized via ISFOA) outperformed traditional analyses in identifying critical

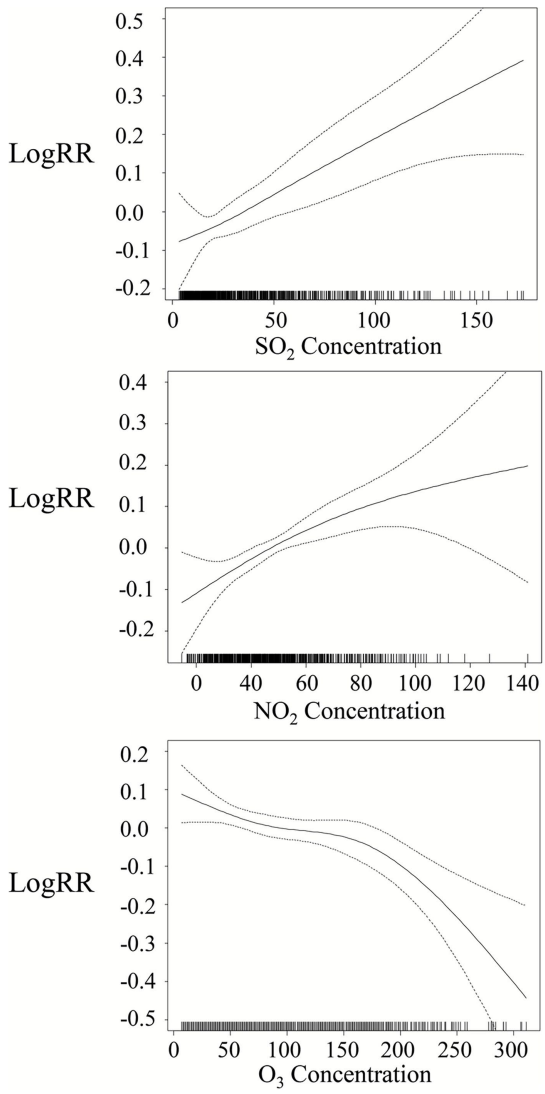


FIGURE 7
Smoothing curves illustrating air pollution effects. The x-axis represents pollutant concentrations (μg/m³ for $NO_2$, $SO_2$, and $O_3$) in different models (lag 0 for $NO_2$ and $SO_2$, lag 4 for $O_3$). The solid line indicates the estimated mean percentage change in daily SSI outpatient visits, while the dotted lines denote twice the point-wise standard error.

exposure windows and pollutant hierarchies, with SHAP interpretability confirming $SO_2$, $NO_2$, and $O_3$ as dominant predictors. These findings advocate for real-time air-quality monitoring in orthopedic wards, strategic scheduling of elective surgeries during high-$O_3$/low-$NO_2$-$SO_2$ periods, and integrated environmental health policies in surgical practice. Future research should address individual-level exposure misclassification and elucidate $O_3$'s immunomodulatory mechanisms via epigenetic approaches.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Ethics Committee of Shanghai Jiao Tong University School of Medicine. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

YW: Software, Investigation, Writing – original draft, Data curation, Funding acquisition. YY: Resources, Supervision, Validation, Writing – review & editing. YL: Methodology, Writing – original draft, Supervision, Resources, Data curation, Project administration. HC: Conceptualization, Writing – review & editing, Methodology, Formal Analysis.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by National Natural Science Foundation of China (No. 81802186).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2025.1692207/full#supplementary-material

# References


Agostini, F., Faccini, M., Fitarelli, F., Ortiz, M. A. L., Salmeron, S., Oliveira, R. C. G., et al. (2021). In vitro comparison of antibacterial effect of ozonated water and ozonated gas. *Ozone Sci. Eng.* 43, 394–400. doi: 10.1080/01919512.2020.1811636

Bath, M. F., Davies, J., Suresh, R., and Machesney, M. R. (2022). Surgical site infections: a scoping review on current intraoperative prevention measures. *Ann. R. Coll. Surg. Engl.* 104, 571–576. doi: 10.1308/rcsann.2022.0075

Chehrassan, M., Nikouei, F., Shakeri, M., Behnamnia, A., Mahabadi, E. A., and Ghandhari, H. (2024). The role of environmental and seasonal factors in spine deep surgical site infection: the air pollution, a factor that may be underestimated. *Eur. Spine J.* 33, 3148–3153. doi: 10.1007/s00586-024-08183-z

Chen, C. S., Kuo, T., Kuo, H., Kuo, T. C., Kuo, H. C., Tseng, Y. J., et al. (2019). Metabolomics of children and adolescents exposed to industrial carcinogenic pollutants. *Environ. Sci. Technol.* 53, 5454–5465. doi: 10.1021/acs.est.9b00392

Chen, C. C., Tsai, S. S., Yeh, C. N., and Yang, C. Y. (2025). Health benefits of a reduction in ambient fine particulate matter levels for post-neonatal infant survival in Taiwan. *J. Toxicol. Environ. Health A* 88, 536–545. doi: 10.1080/15287394.2025.2469079

Hammoor, B., Matsumoto, H., Marciano, G., Dziesinski, L., Wang, K., Roye, B. D., et al. (2021). Surgical site infections in pediatric spinal surgery after implementation of a quality assurance program. *Spine Deform* 9, 125–133. doi: 10.1007/s43390-020-00192-4

He, T., Jin, L., and Li, X. (2023). On the triad of air PM pollution, pathogenic bioaerosols, and lower respiratory infection. *Environ. Geochem. Health* 45, 1067–1077. doi: 10.1007/s10653-021-01025-7

Islam, M. T., Kamal, A. S. M. M., Islam, M. M., and Hossain, S. (2024). Impact of climate change on dengue incidence in Singapore: time-series seasonal analysis. *Int. J. Environ. Health Res.* 34, 3988–3998. doi: 10.1080/09603123.2024.2337827

Kolasiński, W. (2018). Surgical site infections - review of current knowledge, methods of prevention. *Pol. Przegl. Chir.* 91, 41–47. doi: 10.5604/01.3001.0012.7253

Li, S., Zhou, R., Wu, W., Luan, M., Li, X., Liu, C., et al. (2025). Fine particulate matter induces cardiac fibrosis via the CHOP/TXNIP/NLRP3 pathway in C57 BL/6 mice. *Int. Immunopharmacol.* 161:115073. doi: 10.1016/j.intimp.2025.115073

Liu, L., Liu, C., Chen, R., Feng, R., Zhou, Y., Wang, L., et al. (2024). Associations of ambient air pollution and daily outpatient visits for pediatric atopic dermatitis in Shanghai, China. *Ecotoxicol. Environ. Saf.* 286:117231. doi: 10.1016/j.ecoenv.2024.117231

Liu, S., Zhou, S., Li, Y., Cao, L., Lv, G., Peng, L., et al. (2025). Lag analysis of the effect of air pollution on orthopedic postoperative infection in Hebei Province and Xinjiang Uygur autonomous region. *Sci. Rep.* 15:12919. doi: 10.1038/s41598-025-95550-5

Mallet, C., Meissburger, V., Caseris, M., Happiette, A., Chinnappa, J., Bonacorsi, S., et al. (2022). Does the use of intrawound povidone-iodine irrigation and local vancomycin powder impact surgical site infection rate in adolescent idiopathic scoliosis surgery? *Eur. Spine J.* 31, 3020–3028. doi: 10.1007/s00586-022-07340-6

Matia-Garcia, I., Vadillo, E., Pelayo, R., Muñoz-Valle, J. F., García-Chagollán, M., Loaeza-Loaeza, J., et al. (2021). Th1/Th2 balance in young subjects: relationship with cytokine levels and metabolic profile. *J. Inflamm. Res.* 14, 6587–6600. doi: 10.2147/JIR.S342545

Meng, Y., Liu, Z., Hao, J., Tao, F., Zhang, H., Liu, Y., et al. (2023). Association between ambient air pollution and daily hospital visits for cardiovascular diseases in Wuhan, China: a time-series analysis based on medical insurance data. *Int. J. Environ. Health Res.* 33, 452–463. doi: 10.1080/09603123.2022.2035323

Montagna, M. T., Rutigliano, S., Trerotoli, P., Napoli, C., Apollonio, F., D'Amico, A., et al. (2019). Evaluation of air contamination in Orthopaedic operating theatres in hospitals in southern Italy: the IMPACT project. *Int. J. Environ. Res. Public Health* 16:581. doi: 10.3390/ijerph16193581

Rudic, T. N., Althoff, A. D., Kamalapathy, P., and Bachmann, K. R. (2023). Surgical site infection after primary spinal fusion surgery for adolescent idiopathic scoliosis: an analysis of risk factors from a Nationwide insurance database. *Spine (Phila Pa 1976)* 48, E101–E106. doi: 10.1097/BRS.0000000000004591

Salášek, M., český, R., Whitley, A., Šídlo, K., Klézl, P., and Džupa, V. (2023). Surgical site infections after stabilization of pelvic ring injuries: a retrospective analysis of risk factors and a meta-analysis of similar studies. *Int. Orthop.* 47, 1331–1344. doi: 10.1007/s00264-023-05719-8

Supphapipat, K., Leurcharusmee, P., Chattipakorn, N., and Chattipakorn, S. C. (2024). Impact of air pollution on postoperative outcomes following organ transplantation: evidence from clinical investigations. *Clin. Transpl.* 38:e15180. doi: 10.1111/ctr.15180

Tan, T., Lee, H., Huang, M. S., Rutges, J., Marion, T. E., Mathew, J., et al. (2020). Prophylactic postoperative measures to minimize surgical site infections in spine surgery: systematic review and evidence summary. *Spine J.* 20, 435–447. doi: 10.1016/j.spinee.2019.09.013

Tang, X. F., Bin, X., Qu, K. Y., Liu, H. J., Lei, H., Li, W. F., et al. (2024). Antibiotic prophylaxis for surgical wound infections in clean and clean-contaminated surgery: an updated systematic review and meta-analysis. *Int. J. Surg.* 110, 5818–5832. doi: 10.1097/JS9.0000000000001882

Tsai, S. S., and Yang, C. Y. (2024). The impacts of reduction in ambient fine particulate air pollution on natural-cause mortality in Taiwan. *J. Toxicol. Environ. Health A* 87, 855–862. doi: 10.1080/15287394.2024.2384396

Tucci, G., Romanini, E., Zanoli, G., Pavan, L., Fantoni, M., and Venditti, M. (2019). Prevention of surgical site infections in orthopaedic surgery: a synthesis of current recommendations. *Eur. Rev. Med. Pharmacol. Sci.* 23, 224–239. doi: 10.26355/eurrev_201904_17497

Wang, Y., Li, Z., Qian, F., Huang, C., and Hu, J. (2025). Characterization of multi-pollutant air quality in China using health-centric air pollution indices. *Environ. Pollut.* 381:126565. doi: 10.1016/j.envpol.2025.126565

Yanosky, J. D., Washington, A., Foulke, G. T., Guck, D., Butt, M., and Helm, M. F. (2024). Air pollution and incident sarcoidosis in Central Pennsylvania. *J. Toxicol. Environ. Health A* 87, 763–772. doi: 10.1080/15287394.2024.2369255

Zhong, C., Li, G., Meng, Z., Li, H., Yildiz, A. R., and Mirjalili, S. (2025). Starfish optimization algorithm (SFOA): a bio-inspired metaheuristic algorithm for global optimization compared with 100 optimizers. *Neural Comput. & Applic.* 37, 3641–3683. doi: 10.1007/s00521-024-10694-1

frontiers | Frontiers in Pediatrics

TYPE Original Research
PUBLISHED 15 December 2025
DOI 10.3389/fped.2025.1659812

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Kirsten Tulchin-Francis,
Nationwide Children's Hospital, United States
Felice Curcio,
University of Sassari, Italy

*CORRESPONDENCE
Klemens Vertesich
✉ klemens.vertesich@meduniwien.ac.at

RECEIVED 04 July 2025
REVISED 04 November 2025
ACCEPTED 27 November 2025
PUBLISHED 15 December 2025

CITATION
Vertesich K, Ortmayr J, Windhager R and Willegger M (2025) Reliability of ChatGPT answers to common questions on developmental dysplasia of the hip as an information source for parents.
Front. Pediatr. 13:1659812.
doi: 10.3389/fped.2025.1659812

COPYRIGHT
© 2025 Vertesich, Ortmayr, Windhager and Willegger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Reliability of ChatGPT answers to common questions on developmental dysplasia of the hip as an information source for parents

Klemens Vertesich*, Joachim Ortmayr, Reinhard Windhager and Madeleine Willegger

Department of Orthopedics and Trauma Surgery, Medical University of Vienna, Vienna, Austria

**Background:** Artificial intelligence (AI), particularly AI-based large language models (LLM) like ChatGPT, is increasingly shaping how information is accessed, offering patients a new source for understanding complex medical conditions. Given the physical, emotional, and logistical challenges that parents are faced when their baby is diagnosed with developmental dysplasia of the hip (DDH), the demand for clear and accessible educational resources is high. This study aimed to evaluate the quality and reliability of ChatGPT's responses to frequently asked questions about DDH.
**Methods:** This study assessed the quality of responses generated by the AI chatbot ChatGPT 4o to eight frequently asked questions about DDH, derived from real consultations in a pediatric orthopedic clinic Responses were generated during one interaction per question using a ChatGPT account not previously exposed to medical information. Responses were evaluated by two individual readers using a standardized rating system, comparing them to current literature, patient education resources, and consensus guidelines. Each response was categorized by its level of informational accuracy and completeness, and descriptive statistics were calculated to quantify performance.
**Results:** ChatGPT 4o was able to generate structured responses to all eight parental questions. The responses were rated in 12.5% excellent, 25.0% satisfactory with minimal clarification, 50.0% satisfactory with moderate clarification, and 12.5% unsatisfactory due to missing or inaccurate information.
**Conclusion:** ChatGPT provided satisfactory answers to questions about DDH and may serve as a useful supplementary information resource for parents. However, due to limitations in presenting detailed diagnostic and treatment pathways, it should be viewed as an adjunct to, not a replacement for, specialist medical consultation.

KEYWORDS

**developmental dysplasia of the hip, artificial intelligence, large language model, infant hip, hip screening**

# 1 Introduction

Artificial intelligence (AI) has rapidly gained prominence in the medical field, continuously evolving and reshaping daily clinical practice (1). The development of AI-based large language models (LLM) such as ChatGPT has significantly transformed how information is accessed and utilized.

The internet has become a major source of medical information, with over 70% of patients routinely using it for health-related queries, and more than 80% of parents searching online for medical information concerning their children (2, 3). As the internet continues to serve as a primary platform for health education and decision-making, it is crucial for healthcare professionals, including surgeons, to remain up to date on emerging technologies that are influencing how patients obtain medical advice (4). In pediatric orthopedics, certain conditions can significantly affect the daily lives of patients and their families. Impairments such as reduced mobility, chronic pain, and limited joint function may not only disrupt physical activities but also contribute to psychological and emotional comorbidities (5).

Developmental dysplasia of the hip (DDH) represents one of the most common pediatric orthopedic conditions with an incidence of 1%–3% of infants (5), characterized by a broad clinical spectrum, ranging from acetabular dysplasia to complete dislocation of the femoral head. Typically diagnosed in the neonatal period through screening by clinical examinations, DDH can have lasting implications on mobility, function, and pain throughout childhood, adolescence, and even into adulthood if not appropriately managed (6). Management strategies for DDH, including the use of orthotic devices such as the Pavlik harness or spica casts, often place a substantial burden on caregivers, particularly during the early neonatal phase. These interventions may disrupt daily routines, require extensive compliance, and pose emotional and logistical challenges for families. Furthermore, the potential need for surgical correction in more severe or treatment-resistant cases can generate considerable anxiety and uncertainty among parents (6, 7). In such contexts, families frequently seek not only comprehensive medical guidance from healthcare professionals but also accessible, accurate, and comprehensible educational resources. The availability of reliable information becomes critical in helping parents understand the condition, adhere to treatment protocols, and participate actively in shared decision-making.

As AI tools become more integrated into healthcare communication, it is essential to assess their performance in delivering reliable, comprehensible, and patient-centered information. Several studies have assessed the accuracy, completeness, and clinical relevance of chatbot-generated responses to patient questions concerning common orthopedic procedures and conditions, including total hip arthroplasty (THA), slipped capital femoral epiphysis, and hip arthroscopy. These investigations have generally demonstrated that AI-generated content can provide satisfactory and informative answers, potentially supporting patient education and pre-consultation awareness (4, 8, 9). Further, the use of LLMs as assisting tools for investigation of pediatric diseases were described. Although results may look promising defining LLMs as additive tool, authors generally urge to caution quality inconsistencies and misinformation risks (10–12).

The aim of this study was to evaluate the quality, accuracy, and completeness of ChatGPT-generated responses to a set of frequently asked questions (FAQ) by parents regarding DDH. By comparing these responses with established patient education materials and current evidence-based literature, we aimed to assess the potential utility and limitations of ChatGPT as a supplementary tool for parents.

# 2 Methods

This study was designed as a quality assessment of responses generated by an AI chatbot. No patient or caregiver data were used, and ethical approval was not required due to the absence of human subjects. The study design was adapted from the methodology described by Mika et al. (8).

FAQs related to developmental DDH were identified and collected during consultations at our pediatric orthopedic outpatient clinic, which specializes in infant hip dysplasia and serves as a referral center for ultrasound-based screening of the infant hip. Questions were monitored over a time period of 6 months, from November 2024 to April 2025. Based on these observations, eight representative questions were formulated. These questions were formulated to represent a heterogeneity of clinically relevant issues in DDH. The questions were evaluated by all authors and revised upon further analysis. Questions were then submitted to ChatGPT 4o (OpenAI, San Francisco, CA, USA), with no follow-up or clarifying prompts provided. The used account was newly registered and was not previously exposed to orthopedic or medical questions. Questions were phrased and presented in English. The AI-generated responses were independently reviewed and evaluated by two pediatric orthopedic surgeons (K.V., consultant, and M.W., fellowship trained consultant). Inter-rater agreement was analyzed by using inter class correlation coefficient (ICC). After individual evaluation all answers were re-evaluated and definitive score was assigned upon agreement of all authors. Each response was assessed in comparison with current evidence-based literature. When applicable, responses were also evaluated against patient education resources and contemporary consensus statements from the Pediatric Orthopaedic Society of North America (POSNA), the European Federation of National Associations of Orthopaedics and Traumatology (EFORT), and peer-reviewed publications indexed in PubMed. Responses were rated according to the AI response rating system established by Mika et al., which categorizes output as follows: “excellent response requiring no further clarification,” “satisfactory response requiring minimal clarification,” “satisfactory response requiring moderate clarification,” and “unsatisfactory response requiring substantial clarification” (Table 1). For further analysis, results were stratified by the degree of informational deficiency: minimal (1–2 items

omitted or requiring clarification), moderate (3–4 items), and unsatisfactory (more than 4 items omitted or requiring clarification, or inclusion of significant inaccuracies or misinformation). Numerical scores were assigned to each rating category. Descriptive statistical analyses and ICC were performed using SPSS version 29.0 (IBM Corp., Armonk, NY, USA).

# 3 Results

ChatGPT 4o responded to all eight questions, generating answers structured in bullet-point format, typically ranging from three to seven items per response. Each response included subheadings with key points highlighted in bold, aiming to enhance readability and information retention for lay users. After analyzing the questions between the two readers there was an inter-rater agreement assessed by ICC of 0.93 ($p < 0.001$). Upon evaluation and agreement of the authors, the responses were classified as follows: one response was rated as excellent, requiring no further clarification; two responses were rated as satisfactory with minimal clarification; four responses were categorized as satisfactory with moderate clarification; and one response was deemed unsatisfactory, requiring substantial clarification due to omission of critical information and inclusion of misleading or inaccurate content. Table 2 provides an overview on classification of each question as well as a structured description and appraisal on readability for the targeted audience of each response.

TABLE 1 Overview on the used classification system published by Mika et al. (8).

| ChatGPT Response Rating System | |
|---|---|
| Score | Response description/classification |
| 1 | Excellent response, requiring no further clarification |
| 2 | Satisfactory response, requiring minimal clarification |
| 3 | Satisfactory response, requiring moderate clarification |
| 4 | Unsatisfactory response, requiring substantial clarification |

In the following section, each question posed to ChatGPT 4o is presented alongside the corresponding AI-generated response and a detailed analysis is provided for each.

1. What is the likelihood of my infant developing hip dysplasia?

- The likelihood depends on several factors:
  - **Family history:** If a close relative has had developmental dysplasia of the hip (DDH), the risk is higher.
  - **Breech presentation:** Babies born in a breech position (buttocks or feet first) are at higher risk.
  - **Gender:** Girls are more likely to develop DDH than boys.
  - **First-born children:** They have a slightly higher risk due to the tighter fit in the womb.
  - **Swaddling practices:** Tight swaddling that restricts leg movement can increase the risk.
  - The incidence of DDH ranges from **1 to 2 per 1,000 live births** for severe cases, though mild instability may be more common.

## 3.1 Answer analysis: satisfactory response, requiring minimal clarification

ChatGPT accurately identified several relevant factors that influence the incidence of DDH. It highlighted the increased risk associated with a positive family history, noting that first-degree relatives may have up to a 12-fold higher risk (13). The chatbot also addressed the mechanical theory related to DDH pathogenesis, correctly identifying breech positioning as a significant risk factor. However, other mechanical risk factors—such as oligohydramnios, macrosomia, and multiple pregnancy—were not mentioned (14, 15). Female sex, a critical risk factor, was correctly identified by the chatbot. Notably, up to 80% of DDH cases requiring treatment occur in females without any other associated risk factors. The chatbot also correctly identified being a first-born child as a risk factor for DDH (16).

TABLE 2 Overview on questions presented to ChatGPT 4o with classification and scoring of their response accuracy.

| Question | Response accuracy scoring | Possible errors of responses | Readability |
|---|---|---|---|
| 1. What is the likelihood of my infant developing hip dysplasia? | 2 | Lack in detailed information | Sufficient readability |
| 2. Are there prenatal screening methods available for detecting developmental dysplasia of the hip? | 1 | No inaccuracy | Sufficient readability |
| 3. How can developmental dysplasia of the hip be identified in my infant? | 3 | Outdated/not relevant information | Reduced readability/use of medical terms |
| 4. What treatment options are available if my infant is diagnosed with developmental dysplasia of the hip? | 3 | Outdated/not relevant information | Reduced readability/use of medical terms |
| 5. Will surgery be required if my infant is diagnosed with developmental hip dysplasia? | 3 | Outdated/not relevant information | Reduced readability/use of medical terms |
| 6. How long does treatment typically last for developmental dysplasia of the hip? | 2 | Lack in detailed information | Reduced readability/use of medical terms |
| 7. Are there potential long-term risks for my child if developmental dysplasia of the hip is diagnosed during infancy? | 4 | Incorrect and missing information | Sufficient readability |
| 8. Is total hip arthroplasty likely to be necessary later in life if my child is diagnosed with developmental dysplasia of the hip as an infant? | 3 | Outdated/not relevant information | Sufficient readability |
| | Mean: 2.6 | | |

Cultural practices such as swaddling, which limit leg mobility and place the hips in extension and adduction, were recognized by the chatbot as significant contributors to DDH development (17, 18). However, associations between DDH and other congenital conditions, such as torticollis or clubfoot deformity, were not detected (14, 19, 20). The chatbot underestimated the incidence of clinically significant DDH requiring treatment, which is reported to range from 5 to 13 per 1,000 live births. Nonetheless, it did acknowledge the higher prevalence of milder forms of hip instability. Literature indicates that general hip instability is present in approximately 1%–1.5% of newborns, with nearly 90% of cases resolving spontaneously by eight weeks of age (21). The readably of the questions was sufficient and can be well interpreted by caregivers. Information were lacking in additional and detailed information and therefore we classified the results as satisfactory response, requiring minimal clarification.

1. Are there prenatal screening methods available for detecting developmental dysplasia of the hip?

   - Currently, there are **no reliable prenatal screening methods** to detect DDH.
   - Some risk factors, such as breech presentation or restricted fetal movement, may be identified via **prenatal ultrasounds**, but these are not definitive indicators of DDH.

## 3.2 Answer analysis: excellent response, requiring no further clarification

The chatbot accurately stated that, at present, no reliable method for prenatal screening of DDH is routinely used in clinical practice. It correctly described all clinically relevant and established diagnostic methods applicable during the neonatal period. However, recent studies have explored the feasibility of prenatal DDH screening, reporting promising results using an ultrasound technique modified after Graf around the 34th week of gestation (22–24). Although not yet integrated into routine clinical practice, these findings offer potential for future research and development in the field. Additionally, the chatbot appropriately noted that certain mechanical risk factors for DDH can be identified during the prenatal period (14, 15). The generated output was clear and easily understandable. Taken together, this response was considered to provide comprehensive and informative content for parents.

2. How can developmental dysplasia of the hip be identified in my infant?

   - DDH can be identified through:
     - **Physical exams:** Pediatricians perform hip stability tests such as the **Ortolani** and **Barlow maneuvers** during routine check-ups.
     - **Imaging:** Ultrasound is used for infants under 6 months, while x-rays may be used for older babies.
     - **Signs in your baby:** Asymmetrical thigh folds, a leg that appears shorter, or difficulty spreading the legs apart during diaper changes may indicate DDH.

## 3.3 Answer analysis: satisfactory response, requiring moderate clarification

ChatGPT correctly identified physical examination as a key component in the assessment of DDH. It referenced the Ortolani and Barlow maneuvers, which are routinely performed during the neonatal period. The Barlow maneuver involves adduction and posterior translation of the hip to provoke dislocation, whereas the Ortolani maneuver attempts to reduce a dislocated hip through abduction and anterior translation (21, 25). Additionally, ChatGPT recognized limited hip abduction as an important clinical sign of DDH. It also described a typical parental observation during diaper changes—difficulty abducting the hip—as a potential indicator of hip dislocation. This clinical finding is indeed characteristic of DDH and warrants further diagnostic evaluation (7, 26, 27). The chatbot also mentioned inguinal fold asymmetry, which, while associated with DDH, has limited diagnostic value, as it may appear in up to 30% of otherwise normal hips (28). Clinical examination alone may not be sufficient for diagnosing DDH. Notably, over 90% of individuals who underwent THA due to undetected DDH had inadequate screening during infancy (29). Regarding imaging, ChatGPT accurately identified it as a critical component of DDH evaluation. During the neonatal period—up to 4–6 months of age—ultrasound is the preferred imaging modality. The widely accepted Graf method assesses the bony and cartilaginous coverage of the femoral head by the acetabulum and is used universally as a screening tool in several countries (30). The method described by Terjesen et al. represents an alternative approach that incorporates dynamic parameters of hip development (31). he universal implementation of ultrasound screening remains a topic of ongoing debate. The European Society of Paediatric Radiology currently recommends ultrasound only for newborns with specific risk factors, such as breech presentation or a positive family history of DDH (15). After six months of age, when ossification of the femoral head has progressed, radiographs become the primary imaging method. However, due to concerns regarding radiation exposure, radiography is typically reserved for cases with a clinical suspicion of DDH rather than used for general screening purposes (32). While ChatGPT identified most key aspects of DDH assessment, several explanations lacked sufficient depth. Therefore, this response was classified as satisfactory, requiring moderate clarification.

3. What treatment options are available if my infant is diagnosed with developmental dysplasia of the hip?

   - Treatment depends on the severity and age of diagnosis:
     - **Pavlik harness:** A soft brace used to keep the hips in a stable position, typically for babies under 6 months old.

- **Closed reduction:** A nonsurgical procedure to reposition the hip, followed by a cast (spica cast) to maintain alignment.
- **Open reduction:** Surgical correction may be needed for more severe cases.
- **Post-treatment monitoring:** Regular follow-ups with imaging to ensure proper hip development.

## 3.4 Answer analysis: satisfactory response, requiring moderate clarification

The primary goal in the treatment of DDH is to achieve a concentric and stable reduction of the femoral head within the acetabulum, while minimizing the risk of complications such as avascular necrosis (AVN) of the femoral head or re-dislocation (15). ChatGPT correctly identified the Pavlik harness as the most commonly used and important orthotic device for early treatment of DDH. It is generally considered the first-line therapy for infants with reducible hips (33). By positioning the hips in flexion and abduction, the harness facilitates spontaneous reduction of a dislocated hip. However, its effectiveness relies heavily on proper application and continuous use, requiring close cooperation from both patients and parents (34). The chatbot recognized the Pavlik harness as the preferred initial treatment method but did not elaborate on its limitations or contraindications. It also appropriately identified closed reduction followed by spica casting as a common next step in cases where conservative treatment fails, and acknowledged open reduction as a necessary intervention in more severe or unresponsive cases. However, it did not clearly explain the clinical indications that determine the transition from closed to open reduction. In practice, both closed and open reductions are performed under general anesthesia in the operating room. If closed reduction is unsuccessful—either due to failure to achieve a stable, concentric position of the femoral head or inconclusive findings on intraoperative arthrography—open reduction is indicated (35, 36). Open reduction is more frequently required in older children or when anatomical blocks to reduction prevent closed reduction (15). Follow-up assessments using plain radiographs, 3D radiographic techniques, or magnetic resonance imaging (MRI) are essential to confirm maintained reduction and to detect early signs of AVN (37, 38). The answers output was comprehensive. Treatment names such as Pavlik harness were not explained making it initially not fully understandable for persons without medical knowledge and definitely requires further follow-up questions and explanations. Taken together the response was classified as satisfactory, requiring moderate clarification.

4. Will surgery be required if my infant is diagnosed with developmental hip dysplasia?
    - Surgery is not always required:
        - Mild cases often resolve with noninvasive treatments like the **Pavlik harness**.
    - Surgery (closed or open reduction) is more likely if:
        - The diagnosis is delayed.
        - Noninvasive treatments are unsuccessful.
        - The hip is severely displaced.

## 3.5 Answer analysis: satisfactory response, requiring moderate clarification

The chatbot correctly stated that surgical intervention is not always necessary in the management of DDH. As outlined in the previous section, the Pavlik harness is considered the first-line treatment in the majority of DDH cases, with many pathological changes resolving within the first six weeks of life (39). Reported success rates for the Pavlik harness range between 85% and 92% (40). When non-surgical management fails, treatment must be escalated to closed or open reduction (15, 41). The chatbot also identified unsuccessful non-operative treatment and delayed or late diagnosis as risk factors for surgical intervention. This is consistent with current literature, which links delayed treatment to an increased likelihood of requiring multiple surgical procedures (42). However, the chatbot did not elaborate on the age-related risks associated with surgical treatment, such as the potential need for acetabular interventions in older children. Studies report that up to 19% of patients who initially respond well to Pavlik harness therapy may later exhibit residual acetabular dysplasia, necessitating additional surgical procedures—most commonly acetabular osteotomies—during childhood (43–45). While the chatbot successfully identified general factors that may lead to surgical intervention, providing a useful overview for parents, it lacked coherence in explaining the clinical decision-making process regarding when surgery becomes necessary. Therefore, this response was classified as satisfactory, requiring moderate clarification.

5. How long does treatment typically last for developmental dysplasia of the hip?
    - Treatment duration varies:
        - **Pavlik harness:** Typically worn for 6–12 weeks, depending on progress.
        - **Spica cast:** May be needed for 3–6 months after closed or open reduction.
        - Regular monitoring may continue for **several years** to ensure normal hip development.

## 3.6 Answer analysis: satisfactory response, requiring minimal clarification

ChatGPT successfully identified variations in treatment duration for DDH. It accurately reported the typical duration of Pavlik harness therapy as ranging from 6 to 12 weeks (7, 46). However, in cases with suboptimal response to treatment, prolonged use may be necessary, and escalation to surgical intervention may be required. The chatbot also correctly stated that the duration of spica casting following closed or open

reduction is generally around three months. However, it did not elaborate on the fact that treatment duration may be extended based on radiographic findings during follow-up (36). Additionally, ChatGPT recognized the importance of continued clinical and radiographic monitoring following initial treatment. Long-term follow-up is essential to ensure proper hip development and to identify residual dysplasia or complications such as AVN. Overall, the response provided by the chatbot was relatively accurate and informative, though it required minimal clarification. The answer was understandably phrased. The use of medical terms for immobilization requires follow-up explanation or further search for caregivers. Therefore, it was classified as satisfactory, requiring minimal clarification.

6. Are there potential long-term risks for my child if developmental dysplasia of the hip is diagnosed during infancy?
   - If treated early, most children develop normal hip function with minimal long-term risks.
   - Without treatment, DDH can lead to:
     - Hip pain or arthritis in adolescence or adulthood.
     - Limping or uneven leg lengths.
     - Decreased range of motion.

## 3.7 Answer analysis: unsatisfactory response, requiring substantial clarification

ChatGPT correctly identified that early and adequate treatment is key for normal hip development and function (6, 47, 48). Under 5% of suspicious findings on ultrasound persist or worsen during treatment (46). In the second part of its response, the chatbot mentioned potential long-term symptoms such as hip pain, leg-length discrepancy, and reduced range of motion. However, it did not clearly specify the timeframe in which these complications may arise. Despite sufficient early treatment, some patients with DDH may continue to experience hip instability, which can impair mobility and cause pain beginning in childhood or adolescence (44, 49). In such cases, surgical intervention, such as femoral osteotomies, may be indicated to reorient the femoral head, improve joint stability, and stimulate proper acetabular development (45, 50). Acetabular osteotomies are required to enhance acetabular coverage (15). Acetabuloplasties reshape the morphology of the cup and are used in younger patients where the acetabular growth plate has not yet ossified. More complex procedures, such as periacetabular osteotomy, may be necessary to achieve adequate acetabular reorientation and joint congruity (51). Age and skeletal maturity are critical factors that influence the choice and timing of surgical interventions. In cases of residual DDH, surgical treatment can significantly improve long-term outcomes. However, the optimal timing of such procedures remains a topic of ongoing debate in the literature (6, 40). While ChatGPT mentioned possible long-term symptoms, it failed to address the potential need for additional surgical treatments over the course of the disease. This represents a significant omission, particularly in the context of providing comprehensive information to parents. Therefore, the response was classified as unsatisfactory, requiring substantial clarification.

7. Is total hip arthroplasty likely to be necessary later in life if my child is diagnosed with developmental dysplasia of the hip as an infant?
   - Early diagnosis and treatment significantly reduce the likelihood of needing a total hip arthroplasty (hip replacement) later in life.
   - If untreated or if treatment is delayed, the risk of early-onset osteoarthritis increases, which may necessitate hip replacement in adulthood.

## 3.8 Answer analysis: satisfactory response requiring moderate clarification

The chatbot clearly recognized the importance of early and adequate diagnosis and treatment of DDH in reducing the risk of THA later in life. It accurately stated that undiagnosed or inadequately treated DDH can result in gait abnormalities, restricted range of motion, and chronic pain due to the development of osteoarthritis (52). By disrupting normal hip biomechanics, untreated hip dysplasia can accelerate the degeneration of both femoral and acetabular cartilage, ultimately leading to secondary osteoarthritis. Literature indicates that approximately 25%–50% of patients with persistent DDH develop radiographic or symptomatic signs of osteoarthritis by the age of 50, often requiring THA due to pain and functional impairment (53, 54). While the chatbot identified the key consequences of untreated DDH, including the long-term risk of osteoarthritis and subsequent surgical intervention, the response lacked depth and omitted additional context that would be valuable for informing and reassuring parents. Despite that the phrased answer was easily readable and understandable. Therefore, although the main points were addressed, the response could have been strengthened with more comprehensive detail.

# 4 Discussion

The LLM ChatGPT 4o demonstrated the ability to generate accurate and relevant responses to critical parental questions concerning DDH. Most of the responses required only minimal or moderate clarification. However, when compared with expert medical advice and critical appraisal of current literature, gaps in the depth and specificity of information became evident. Given the complexity and evolving nature of DDH diagnosis and treatment algorithms, fully capturing the nuance of clinical decision-making remains a challenge for AI systems. ChatGPT showed potential in delivering useful information to parents. It may provide an overview on complex pathologies like DDH to initially comprehend the disease and the necessity of treatment, while it cannot replace professional consultation. Further, the

ability of parents to interpret AI-generated information varies widely and may be further impaired in emotionally stressful situations such as a DDH diagnosis. This creates clinical and ethical concerns while incomplete or overly reassuring responses could delay medical consultation, reinforce misconceptions, or undermine trust in healthcare professionals. Final treatment indication and detailed explanation in each individual case cannot be replaced by LLMs. ChatGPT may serve as a valuable initial source of information, helping to prepare patients and caregivers for more informed discussions with healthcare providers.

The use of AI-generated medical content, particularly through chatbots such as ChatGPT, has garnered growing interest in the medical, and especially orthopedic community. As AI tools rapidly evolve, they are transforming how patients and caregivers access health information, which in turn may influence healthcare-seeking behavior and treatment decisions (55). The ability of LLMs to process and communicate medical information has been supported by recent publications across a range of orthopedic conditions (4, 8, 9, 56–59). In this study, ChatGPT provided generally sufficient and helpful answers to FAQs about DDH. However, given the continuous development of AI models and the potential for variability in responses across different interactions, it is essential that AI-generated medical information be reviewed and verified by qualified healthcare professionals (4, 8).

Several limitations of this study must be acknowledged. The study design incorporated inherently subjective elements, including the selection and formulation of FAQs from parents, as well as the observation and documentation of the AI-driven conversation process. Additionally, the evaluation and classification of the chatbot responses were based on the authors' interpretation, which may introduce bias despite efforts to maintain objectivity. To enhance transparency and reproducibility, all AI-generated outputs were documented and shared. Furthermore, in order to minimize variability and improve internal validity, each question was submitted to the chatbot a single time. While this approach reduces the impact of response variation, it must be noted that repeated queries can yield slightly different responses, even if the core content and intent remain consistent. Additionally, responses of LLMs may vary in terms of reproducibility and generalizability limiting the results regarding these parameters. The study was conducted using ChatGPT-4o, which, at the time of inquiry, represented the most advanced version of the model available. Currently there is a more recent version, ChatGPT 5, available. However, ChatGPT does not possess continuous access to the internet and is limited by a training data cutoff of April 2024. Consequently, the chatbot may not reflect the most recent literature or clinical guidelines published thereafter. Importantly, ChatGPT was not assigned a specific role, for example physician or parent, during the interaction, nor was it provided with contextual information about the target audience. This approach was chosen to simulate a typical, unscripted interaction between users and the AI, as might occur in real-world usage. Nevertheless, the absence of such framing could have influenced the style and depth of responses, potentially affecting their classification.

## 5 Conclusion

ChatGPT frequently provided satisfactory responses to parental questions related to DDH. Based on the findings of this study, it may serve as a helpful supplementary source of information for parents confronted with a DDH diagnosis in their newborn child. However, the chatbot did not consistently present structured diagnostic pathways or detailed treatment algorithms, and most responses required some degree of clarification or supplementation. As such, AI-based chatbots should be regarded as adjunct tools rather than standalone sources of medical advice. In clinical practice, these tools may complement—but not replace—comprehensive consultation with a pediatric orthopedic specialist. Rapid developments of LLMs however may quickly lead to an implication and use of these models in clinical practice. Therefore continued development, validation, and oversight will be essential to ensure the safe and effective integration of AI-driven technologies into clinical practice as well as parent and patient education. LLMs will be useful and essential tool in future clinical practice and ongoing research, improvement and validation of these models is essential to further improve patient specific treatments.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

KV: Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. JO: Data curation, Formal analysis, Writing – original draft, Writing – review & editing. RW: Methodology, Supervision, Writing – review & editing, Writing – original draft. MW: Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing, Data curation, Project administration.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Conflict of interest

RW reports consulting fees from Johnson & Johnson, Medical Limited, Stryker European, and Operations Limited,

and institutional agreements between the Medical University of Vienna and De Puy Synthes, Johnson & Johnson, all of which are not related to this study. MW reports honoraria for lectures and course instruction from Arthrex Inc., all of which are not related to this study.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References


1. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metab Clin Exp*. (2017) 69S:S36. doi: 10.1016/j.metabol.2017.01.011

2. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online health information seeking among US adults: measuring progress toward a healthy people 2020 objective. *Public Health Rep*. (2019) 134(6):617. doi: 10.1177/0033354919874074

3. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are parents getting it right? A survey of Parents' internet use for children's health care information. *Interact J Med Res*. (2015) 4(2):e12. doi: 10.2196/ijmr.3790

4. Adelstein JM, Sinkler MA, Li LT, Mistovich RJ. ChatGPT responses to common questions about slipped capital femoral epiphysis: a reliable resource for parents? *J Pediatr Orthop*. (2024) 44(6):353. doi: 10.1097/BPO.0000000000002681

5. Fuglkjaer S, Dissing KB, Hestbaek L. Prevalence and incidence of musculoskeletal extremity complaints in children and adolescents. A systematic review. *BMC Musculoskelet Disord*. (2017) 18(1):418. doi: 10.1186/s12891-017-1771-2

6. Young JR, Anderson MJ, O'Connor CM, Kazley JM, Mantica AL, Dutt V. Team approach: developmental dysplasia of the hip. *JBJS Rev*. (2020) 8(9):e20.00030. doi: 10.2106/JBJS.RVW.20.00030

7. Alves C, Truong WH, Thompson MV, Suryavanshi JR, Penny CL, Do HT, et al. Diagnostic and treatment preferences for developmental dysplasia of the hip: a survey of EPOS and POSNA members. *J Child Orthop*. (2018) 12(3):236. doi: 10.1302/1863-2548.12.180034

8. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. *J Bone Joint Surg Am*. (2023) 105(19):1519. doi: 10.2106/JBJS.23.00209

9. Ozbek EA, Ertan MB, Kindan P, Karaca MO, Gursoy S, Chahla J. ChatGPT can offer at least satisfactory responses to common patient questions regarding hip arthroscopy. *Arthroscopy*. (2024) 41(6):1806–27. doi: 10.1016/j.arthro.2024.08.036

10. Young CC, Enichen E, Rivera C, Auger CA, Grant N, Rao A, et al. Diagnostic accuracy of a custom large language model on rare pediatric disease case reports. *Am J Med Genet A*. (2025) 197(2):e63878. doi: 10.1002/ajmg.a.63878

11. Yan C, Li Z, Liang Y, Shao S, Ma F, Zhang N, et al. Assessing large language models as assistive tools in medical consultations for Kawasaki disease. *Front Artif Intell*. (2025) 8:1571503. doi: 10.3389/frai.2025.1571503

12. Temsah MH, Abouammoh N, Alsatrawi M, Al-Jelaify M, Altamimi I, ChatGPT Research Consortium P, et al. Exploring early perceptions and experiences of ChatGPT in pediatric critical care: a qualitative study among health care professionals. *Clin Pediatr (Phila)*. (2025) 99228251370556. doi: 10.1177/00099228251370556

13. Stevenson DA, Mineau G, Kerber RA, Viskochil DH, Schaefer C, Roach JW. Familial predisposition to developmental dysplasia of the hip. *J Pediatr Orthop*. (2009) 29(5):463. doi: 10.1097/BPO.0b013e3181aa586b

14. Dunn PM. Perinatal observations on the etiology of congenital dislocation of the hip. *Clin Orthop Relat Res*. (1976) 119:11.

15. Vaquero-Picado A, Gonzalez-Moran G, Garay EG, Moraleda L. Developmental dysplasia of the hip: update of management. *EFORT Open Rev*. (2019) 4(9):548. doi: 10.1302/2058-5241.4.180019

16. Dawod MS, Alswerki MN, Abuqudiri AZ, Albadaineh AA, Mahmoud LM, Altarawneh DJ, et al. The impact of maternal age and educational level on developmental dysplasia of the hip diagnosis and screening: a descriptive comparative study. *Adv Orthop*. (2023) 2023:6793645. doi: 10.1155/2023/6793645

17. Loder RT, Skopelja EN. The epidemiology and demographics of hip dysplasia. *ISRN Orthop*. (2011) 2011:238607. doi: 10.5402/2011/238607

18. Schwend RM, Shaw BA, Segal LS. Evaluation and treatment of developmental hip dysplasia in the newborn and infant. *Pediatr Clin North Am*. (2014) 61(6):1095. doi: 10.1016/j.pcl.2014.08.008

19. Perry DC, Tawfiq SM, Roche A, Shariff R, Garg NK, James LA, et al. The association between clubfoot and developmental dysplasia of the hip. *J Bone Joint Surg Br*. (2010) 92(11):1586. doi: 10.1302/0301-620X.92B11.24719

20. Paton RW, Choudry QA, Jugdey R, Hughes S. Is congenital talipes equinovarus a risk factor for pathological dysplasia of the hip?: a 21-year prospective, longitudinal observational study. *Bone Joint J*. (2014) 96-b(11):1553. doi: 10.1302/0301-620X.96B11.34130

21. Barlow TG. Early diagnosis and treatment of congenital dislocation of the hip. *Proc R Soc Med*. (1963) 56(9):804. doi: 10.1177/003591576305600920

22. Contro E, Larcher L, Lenzi J, Benfenati A, Massinissa Magini G, Galeati G, et al. Prenatal screening for developmental displacement of the hip: the BUDDHA (Pre-birth ultrasound for developmental displacement of the hip assessment) study. *Diagnostics (Basel)*. (2021) 11(5):868. doi: 10.3390/diagnostics11050868

23. Bevilacqua C, Boscarato V, Pieroni G, Fraternali E, Lattanzi G, Aspriello SD, et al. The feasibility of prenatal ultrasonic screening for developmental dysplasia of the hip. *Diseases*. (2024) 12(8):188. doi: 10.3390/diseases12080188

24. Komut E, Zehir S. Can developmental dysplasia of the hip be identified in the prenatal period? A pilot study: ultrasonographic evaluation and postnatal follow-up results of fetal hips in the third trimester. *Acta Orthop Traumatol Turc*. (2021) 55(3):196. doi: 10.5152/j.aott.2021.20143

25. Ortolani M. Congenital hip dysplasia in the light of early and very early diagnosis. *Clin Orthop Relat Res*. (1976) 6(119):6–10.

26. Sankar WN, Weiss J, Skaggs DL. Orthopaedic conditions in the newborn. *J Am Acad Orthop Surg*. (2009) 17(2):112. doi: 10.5435/00124635-200902000-00007

27. Choudry Q, Goyal R, Paton RW. Is limitation of hip abduction a useful clinical sign in the diagnosis of developmental dysplasia of the hip? *Arch Dis Child*. (2013) 98(11):862. doi: 10.1136/archdischild-2012-303121

28. Dezateux C, Rosendahl K. Developmental dysplasia of the hip. *Lancet*. (2007) 369(9572):1541. doi: 10.1016/S0140-6736(07)60710-7

29. Engesaeter I, Lie SA, Lehmann TG, Furnes O, Vollset SE, Engesaeter LB. Neonatal hip instability and risk of total hip replacement in young adulthood: follow-up of 2,218,596 newborns from the medical birth registry of Norway in the Norwegian arthroplasty register. *Acta Orthop*. (2008) 79(3):321. doi: 10.1080/17453670710015201

30. Graf R. Fundamentals of sonographic diagnosis of infant hip dysplasia. *J Pediatr Orthop*. (1984) 4(6):735. doi: 10.1097/01241398-198411000-00015

31. Terjesen T, Holen KJ, Tegnander A. Hip abnormalities detected by ultrasound in clinically normal newborn infants. *J Bone Joint Surg Br*. (1996) 78(4):636. doi: 10.1302/0301-620X.78B4.0780636

32. Kane TP, Harvey JR, Richards RH, Burby NG, Clarke NM. Radiological outcome of innocent infant hip clicks. *J Pediatr Orthop B*. (2003) 12(4):259. doi: 10.1097/01.bpb.0000060292.16932.05

33. Pavlik A. The functional method of treatment using a harness with stirrups as the primary method of conservative therapy for infants with congenital dislocation of the hip. 1957. *Clin Orthop Relat Res*. (1992) (281):4.

34. Thacker MM. CORR insights (®): higher pavlik harness treatment failure is seen in graf type IV ortolani-positive hips in males. *Clin Orthop Relat Res*. (2016) 474(8):1855. doi: 10.1007/s11999-016-4865-5

35. Moraleda L, Albiñana J, Salcedo M, Gonzalez-Moran G. Dysplasia in the development of the hip. *Rev Esp Cir Ortop Traumatol*. (2013) 57(1):67. doi: 10.1016/j.recot.2012.10.005

36. Studer K, Williams N, Studer P, Baker M, Glynn A, Foster BK, et al. Obstacles to reduction in infantile developmental dysplasia of the hip. *J Child Orthop*. (2017) 11(5):358. doi: 10.1302/1863-2548.11.170031

37. Baljoun A, Bradley CS, Verma Y, Wedge JH, Gargan MF, Kelley SP. Intraoperative 3D fluoroscopy during open reduction for DDH: an effective alternative to postoperative CT/MRI. *J Pediatr Orthop*. (2024) 44(8):462. doi: 10.1097/BPO.0000000000002728

38. Sachleben B, Perry DC, Wedge J, Kelley SP. Intraoperative assessment of closed reduction for developmental dislocation of the hip using 3-dimensional fluoroscopy. *J Pediatr Orthop*. (2015) 35(3):246. doi: 10.1097/BPO.0000000000000245

39. Dunn PM. The anatomy and pathology of congenital dislocation of the hip. *Clin Orthop Relat Res*. (1976) (119):23.

40. Harris IE, Dickens R, Menelaus MB. Use of the Pavlik harness for hip displacements. When to abandon treatment. *Clin Orthop Relat Res*. (1992) (281):29.

41. Khoshhal KI, Kremli MK, Zamzam MM, Akod OM, Elofi OA. The role of arthrography-guided closed reduction in minimizing the incidence of avascular necrosis in developmental dysplasia of the hip. *J Pediatr Orthop B*. (2005) 14(4):256. doi: 10.1097/01202412-200507000-00004

42. Luhmann SJ, Bassett GS, Gordon JE, Schootman M, Schoenecker PL. Reduction of a dislocation of the hip due to developmental dysplasia. Implications for the need for future surgery. *J Bone Joint Surg Am*. (2003) 85(2):239. doi: 10.2106/00004623-200302000-00009

43. Nakamura J, Kamegaya M, Saisu T, Someya M, Koizumi W, Moriya H. Treatment for developmental dysplasia of the hip using the Pavlik harness: long-term results. *J Bone Joint Surg Br*. (2007) 89(2):230. doi: 10.1302/0301-620X.89B2.18057

44. Malvitz TA, Weinstein SL. Closed reduction for congenital dysplasia of the hip. Functional and radiographic results after an average of thirty years. *J Bone Joint Surg Am*. (1994) 76(12):1777. doi: 10.2106/00004623-199412000-00004

45. Moraleda L, Bravo C, Forriol F, Albiñana J. Does orientation of the femoral head affect acetabular development? An experimental study in lamb. *J Pediatr Orthop*. (2019) 39(8):416. doi: 10.1097/BPO.0000000000000974

46. Bache CE, Clegg J, Herron M. Risk factors for developmental dysplasia of the hip: ultrasonographic findings in the neonatal period. *J Pediatr Orthop B*. (2002) 11(3):212. doi: 10.1097/00009957-200207000-00004

47. Lankinen V, Helminen M, Bakti K, Valipakka J, Laivuori H, Hyvarinen A. Known risk factors of the developmental dysplasia of the hip predicting more severe clinical presentation and failure of Pavlik harness treatment. *BMC Pediatr*. (2023) 23(1):148. doi: 10.1186/s12887-023-03935-0

48. Aarvold A, Perry DC, Mavrotas J, Theologis T, Katchburian M, Group BDC. The management of developmental dysplasia of the hip in children aged under three months: a consensus study from the British society for children's orthopaedic surgery. *Bone Joint J*. (2023) 105-B(2):209. doi: 10.1302/0301-620X.105B2.BJJ-2022-0893.R1

49. Morcuende JA, Meyer MD, Dolan LA, Weinstein SL. Long-term outcome after open reduction through an anteromedial approach for congenital dislocation of the hip. *J Bone Joint Surg Am*. (1997) 79(6):810. doi: 10.2106/00004623-199706000-00002

50. Blockey NJ. Derotation osteotomy in the management of congenital dislocation of the hip. *J Bone Joint Surg Br*. (1984) 66(4):485. doi: 10.1302/0301-620X.66B4.6746678

51. López-Carreño E, Carillo H, Gutiérrez M. Dega versus salter osteotomy for the treatment of developmental dysplasia of the hip. *J Pediatr Orthop B*. (2008) 17(5):213. doi: 10.1097/BPB.0b013e32830850eb

52. Stein-Zamir C, Volovik I, Rishpon S, Sabi R. Developmental dysplasia of the hip: risk markers, clinical screening and outcome. *Pediatr Int*. (2008) 50(3):341. doi: 10.1111/j.1442-200X.2008.02575.x

53. Vitale MG, Skaggs DL. Developmental dysplasia of the hip from six months to four years of age. *J Am Acad Orthop Surg*. (2001) 9(6):401. doi: 10.5435/00124635-200111000-00005

54. Gala L, Clohisy JC, Beaulé PE. Hip dysplasia in the young adult. *J Bone Joint Surg Am*. (2016) 98(1):63. doi: 10.2106/JBJS.O.00109

55. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. (2021) 8(2):e188. doi: 10.7861/fhj.2021-0095

56. Taylor WL, Cheng R, Weinblatt AI, Bergstein V, Long WJ. An artificial intelligence Chatbot is an accurate and useful online patient resource prior to total knee arthroplasty. *J Arthroplasty*. (2024) 39(8S1):S358. doi: 10.1016/j.arth.2024.02.005

57. Varady NH, Lu AZ, Mazzucco M, Dines JS, Altchek DW, Williams RJ 3rd, et al. Understanding how ChatGPT may become a clinical administrative tool through an investigation on the ability to answer common patient questions concerning ulnar collateral ligament injuries. *Orthop J Sports Med*. (2024) 12(7):23259671241257516. doi: 10.1177/23259671241257516

58. Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. *Front Med (Lausanne)*. (2024) 11:1477898. doi: 10.3389/fmed.2024.1477898

59. Johns WL, Kellish A, Farronato D, Ciccotti MG, Hammoud S. ChatGPT can offer satisfactory responses to common patient questions regarding elbow ulnar collateral ligament reconstruction. *Arthrosc Sports Med Rehabil*. (2024) 6(2):100893. doi: 10.1016/j.asmr.2024.100893

frontiers | Frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 16 January 2026
DOI 10.3389/frai.2025.1702087

Check for updates

OPEN ACCESS

EDITED BY
Kaya Kuru,
University of Central Lancashire,
United Kingdom

REVIEWED BY
Vinothkumar Kolluru,
Stevens Institute of Technology, United States
Liu Bo,
Shanghai University of Engineering Sciences, China

*CORRESPONDENCE
Agnese Graziosi
✉ agnese.graziosi@unimi.it

†These authors have contributed equally to this work

RECEIVED 09 September 2025
REVISED 19 December 2025
ACCEPTED 24 December 2025
PUBLISHED 16 January 2026

CITATION
Frasca M, Gazzaniga G, Graziosi A, De Nicolo V, De Giacomo C, Martinelli S, Senatore M, Romandini A, Moretti C, Pattarino GAC, Proto A, Danesi R, Scaglione F, Vago G, La Torre D and Pani A (2026) Artificial intelligence and precision medicine: a pilot study predicting optimal ceftaroline dosage for pediatric patients.
*Front. Artif. Intell.* 8:1702087.
doi: 10.3389/frai.2025.1702087

COPYRIGHT
© 2026 Frasca, Gazzaniga, Graziosi, De Nicolo, De Giacomo, Martinelli, Senatore, Romandini, Moretti, Pattarino, Proto, Danesi, Scaglione, Vago, La Torre and Pani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Artificial intelligence and precision medicine: a pilot study predicting optimal ceftaroline dosage for pediatric patients

Maria Frasca[1†], Gianluca Gazzaniga[2†], Agnese Graziosi[1,3*†], Valentina De Nicolo[1], Costantino De Giacomo[4], Stefano Martinelli[5], Michele Senatore[6], Alessandra Romandini[6], Chiara Moretti[4], Giulia Angela Carla Pattarino[4], Alice Proto[5], Romano Danesi[1,6], Francesco Scaglione[1], Gianluca Vago[1], Davide La Torre[1,7,8] and Arianna Pani[1]

[1]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy, [2]Department of General Surgery and Surgical Specialty Paride Stefanini, Sapienza University of Rome, Rome, Italy, [3]Department of Medical Biotechnology and Translational Medicine, Postgraduate School of Clinical Pharmacology and Toxicology, University of Milan, Milan, Italy, [4]Pediatric Unit, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy, [5]Neonatal Intensive Care Unit, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy, [6]Chemical-Clinical Analyses Unit, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy, [7]SKEMA Business School, Université Côte d'Azur, Nice, France, [8]Department of Mathematics, University of Bologna, Bologna, Italy

**Background:** Accurate drug dosing in pediatrics is complicated by age-related physiological variability. Standard weight-based dosing may result in either subtherapeutic exposure or toxicity. Machine learning (ML) models can capture complex relationships among clinical variables and support individualized therapy.

**Methods:** We analyzed clinical and pharmacokinetic data from 20 pediatric patients enrolled in the PUERI study (January 2020–November 2021, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy). Eight ML models-including linear regression (LR), ridge regression (RR), lasso regression (LaR), Huber regression (HR), random forest (RF), XGBoost, LightGBM, and a neural network (MLP)-were trained to predict ceftaroline doses that would achieve plasma concentrations close to the therapeutic target of 10 mg/L. Model performance was evaluated using mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination ($R^2$). To ensure interpretability, we applied local interpretable model-agnostic explanations (LIME) to identify the most influential predictors of dose.

**Results:** MLP (MAE 1.53 mg, $R^2$ 0.94) and XGBoost (MAE 2.04 mg, $R^2$ 0.89) outperformed linear models. Predicted doses were more frequently aligned with therapeutic concentrations than those clinically administered. Model-based simulated concentrations fell within the therapeutic range in approximately 85% of cases, and the best ML models showed over 90% patient-level clinical. RF, LightGBM and XGBoost achieved the highest clinical alignment, with 94.2, 92.4 and 91.5% of patients reaching therapeutic levels. Renal function markers, such as serum creatinine and azotemia, together with anthropometric parameters including weight, height, and body mass index, were consistently the most influential features.

**Conclusion:** Advanced ML models can optimize ceftaroline dosing in pediatric patients and outperform traditional dosing strategies. Combining predictive

accuracy with interpretability (via LIME) supports clinical trust and may enhance precision antibiotic therapy while reducing the risks of antimicrobial resistance and toxicity.

KEYWORDS

**artificial intelligence, ceftaroline, drug dosing, explainable AI, machine learning, pediatric pharmacokinetics, personalized medicine**

# 1 Introduction

The correct administration of antibiotics in pediatrics represents a complex challenge for modern pharmacotherapy. Unlike adults, children exhibit significant physiological variability related to age, weight, renal function, and liver maturation, all of which critically influence drug metabolism and elimination (Willems et al., 2021). Standardized dosing based on simple proportional formulas (e.g., mg/kg) is often insufficient to ensure effective therapy personalization, exposing patients to the risk of overdosing, which can lead to adverse effects, or underdosing, with potential therapeutic failure (Anderson and Holford, 2008; Kearns et al., 2003). This challenge is compounded by the growing threat of antimicrobial resistance (AMR). Inappropriate dosing not only risks therapeutic failure but also accelerates the selection and proliferation of drug-resistant bacterial strains, rendering treatments less effective and contributing to one of the most critical global public health threats of this century, with an estimated 10 million deaths per year by 2050 (Murray et al., 2022; Ahmed et al., 2024). In this context, Artificial Intelligence (AI) and Machine Learning (ML) techniques are emerging as innovative tools to address the complexities of personalized medicine, with a particular focus on optimizing drug dosages (Arnold et al., 2025; Smith et al., 2020). By analyzing vast amounts of clinical data and identifying complex patterns among physiological variables, AI models offer a novel approach to individualized pharmacotherapy. This capability enables clinicians to predict optimal antibiotic dosages based on specific patient characteristics, thereby improving treatment efficacy and reducing toxicity risks (Poweleit et al., 2023). Ceftaroline, a fifth-generation cephalosporin, has gained attention for its broad spectrum of action, which includes resistant pathogens like Methicillin-Resistant *Staphylococcus Aureus* (MRSA; Welte et al., 2019). However, its optimal dosage in pediatric patients remains a challenge due to the scarcity of clinical data and significant interindividual variability. While standard doses are outlined in ceftaroline's summary of product characteristics, ensuring therapeutic drug concentrations while avoiding toxicity is crucial. This critical balance underscores the urgent need for individualized dosing strategies (Ruggiero et al., 2019). This study leverages multiple regression models to predict the optimal ceftaroline dose required to achieve a target therapeutic plasma concentration of 10 mg/L. We compare these AI-predicted doses with those actually administered to identify potential instances of under- or overdosing. We also employ Explainable AI (XAI) technologies, such as Local Interpretable Model-agnostic Explanations (LIME), to enhance model interpretability by highlighting the most influential patient characteristics in dose prediction. This research aims to demonstrate how AI can increase the accuracy of ceftaroline administration, paving the way for safer and more effective treatments for pediatric patients.

# 2 Materials and methods

## 2.1 Study design and data source

This retrospective analysis utilized data from the "Observational, single-center, open-label, sequential, pharmacokinetic and tolerability study of ceftaroline in pediatric patients from 2 to 24 months of age, with suspected or confirmed infection (PUERI)." The PUERI study was conducted between January 2020 and November 2021 at the Pediatric Department of the ASST Grande Ospedale Metropolitano Niguarda (Milan, Italy). The study adhered to the Declaration of Helsinki and Good Clinical Practice guidelines and was approved by the Niguarda Hospital Ethics Committee (protocol no. 189–042019, April 15, 2019). Written informed consent was obtained from parents or legal guardians.

## 2.2 Patient population and data collection

The dataset comprised pharmacological, clinical, and demographic information from 21 pediatric patients; one was excluded due to insufficient data, yielding a final cohort of 20. Ceftaroline therapy was prescribed at physician discretion, with documentation of infusion parameters, dosage, and administration time. Opportunistic plasma sampling used residual specimens from routine clinical procedures. Variables collected are detailed in Supplementary Table S1.

## 2.3 Data preprocessing and feature selection

Prior to model development, the dataset underwent a systematic preprocessing pipeline. First, missing data were handled by mean imputation for each numerical feature. Categorical variables were then converted into a numerical format. To ensure uniform scaling and improve model stability, all selected features were standardized to a mean of 0 and a standard deviation of 1. The features used for training, including anthropometric, biochemical, and derived variables, e.g., BMI and the ratio between target and real plasma concentrations (DV ratio), were selected based on their clinical relevance and availability (Table 1). Time elapsed since the first drug administration (TIME) and Infusion rate (mg/h; RATE) were included as retrospective descriptors of the pharmacokinetic context at sampling and were not used for prospective pre-dose prediction. DV ratio was used only as an exploratory feature during training and had no role in dose optimisation. For patients with multiple treatment cycles, each cycle was treated as a distinct event to ensure accurate representation.

TABLE 1 Features used in the ML models for dose prediction.

| Feature | Description |
|---|---|
| TIME | Time since the start of the infusion (minutes) |
| RATE | Infusion rate (mL/min) |
| HT | Patient's height (cm) |
| WT | Patient's body weight (kg) |
| SCR | Serum creatinine level (mg/dl) |
| AZOTEMIA | Blood urea nitrogen concentration (mg/dl) |
| AST | Aspartate Aminotransferase level (U/L) |
| ALT | Alanine Aminotransferase level (U/L) |
| Albumin | Blood albumin concentration (g/dl) |
| Schwartz formula | eGFR estimated using the Schwartz formula |
| Revised Schwartz equation | eGFR using the revised Schwartz formula |
| BMI | Body Mass Index calculated from height and weight |
| DV ratio | Ratio between target and real plasma concentration |

## 2.4 Model development and evaluation

The dataset included 20 pediatric patients, contributing a total of 58 pharmacokinetic events. Since several patients had more than one therapeutic cycle, events were treated as independent units. The 80/20 split was performed at the event level (46 training events and 12 independent test events) using a fixed random seed of 42. The independent test set was not used during hyperparameter tuning or model selection. This approach was used to evaluate model generalization and prevent overfitting. We utilized a range of eight ML regression models to predict the personalized optimized ceftaroline dose, including linear regression (LR), ridge regression (RR), Lasso Regression (LaR), Huber regression (HR), Random Forest (RF), XGBoost, LightGBM, and a Neural Network (MLP). Each model was trained with the optimized dose (Optimized AMT) as the dependent variable. Model performance was evaluated on the test set using standard metrics for regression tasks: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination ($R^2$), and Mean Absolute Percentage Error (MAPE; Supplementary file 2).

## 2.5 Model validation and overfitting control

To ensure robustness given the limited sample size, all models were internally validated using a leave-one-out cross-validation (LOOCV) strategy, which maximizes data utilization and provides an unbiased estimate of generalization performance. Additionally, bootstrap resampling (1,000 iterations) was performed to derive 95% confidence intervals for MAE, RMSE and $R^2$ and to assess model stability. Regularization techniques (L1/L2 penalties) were applied to linear models, while early stopping and dropout were used for tree-based and neural architectures to minimize overfitting and improve generalization. Hyperparameter tuning for RF, XGBoost, LightGBM and MLP was performed outside the LOOCV framework, using predefined configurations based on best practices in the literature.

## 2.6 Explainable AI analysis and dose comparison

After model training, we calculated a personalized optimized dose for each patient by minimizing the absolute deviation between the predicted drug concentration and the target value of 10 mg/L. This Optimized AMT was then compared to the dose actually administered clinically (AMT) to identify potential under- or overdosing. Finally, to enhance model interpretability, the best-performing models were analyzed using Local Interpretable Model-agnostic Explanations (LIME). This technique was used to identify the most significant clinical variables contributing to the dose predictions, providing insight into the key features driving the model's recommendations. LIME was implemented using 5,000 perturbed samples, a kernel width of 0.75, quantile-based discretization, and a fixed random seed of 42.

# 3 Results

## 3.1 Model performance and comparison

The ML models demonstrated varying capabilities in predicting the Optimized AMT. A key metric was the $R^2$ score, which represents the proportion of variability in the dependent variable explained by the model. The Neural Network showed the highest $R^2$ score at 0.94, followed by Random Forest (0.93), XGBoost (0.89), and LightGBM (0.84). The performance of linear models was generally lower, with the Linear Regression achieving an $R^2$ of 0.74 and the Ridge, Lasso, and Huber models ranging from 0.35 to 0.54. Prediction accuracy was further assessed using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The Neural Network achieved the lowest MAE (1.53 mg) and RMSE (3.38 mg), indicating high accuracy. XGBoost also performed exceptionally well, with a low MAE of 2.04 mg and RMSE of 4.70 mg. In contrast, linear models showed higher errors, with Ridge Regression having the largest MAE (7.92 mg) and RMSE (9.94 mg). A summary of the performance metrics for all models is presented in Table 2, the whole results are uploaded in Supplementary file 2.

## 3.2 Clinical alignment of predicted doses

The clinical utility of the models was evaluated by comparing simulated plasma concentrations derived from predicted doses with the therapeutic range. Clinical alignment was defined as the proportion of patients whose model-predicted dose resulted in simulated plasma concentrations falling within the predefined therapeutic window of 1–10 mg/L. RF achieved the highest clinical alignment (94.2% within range). XGBoost and LGBM performed well (91.5 and 92.4% within range, respectively). In contrast, the MLP and LR showed lower alignment (~49.9%), underscoring that models with slightly higher error may still yield more clinically relevant recommendations. Administered versus predicted doses for all models are provided in Table 3.

TABLE 2 Performance comparison of all ML models used in the study.

| Model | MAE (mg) | RMSE (mg) | R2 |
|---|---|---|---|
| Linear Regression | 4.34 | 7.19 | 0.74 |
| Ridge Regression | 7.92 | 9.94 | 0.51 |
| Lasso Regression | 7.61 | 9.58 | 0.54 |
| Huber Regression | 7.20 | 11.41 | 0.35 |
| Random Forest | 4.21 | 6.10 | 0.93 |
| XGBoost | 2.04 | 4.70 | 0.89 |
| LightGBM | 2.92 | 5.59 | 0.84 |
| Neural Network | 1.53 | 3.38 | 0.94 |

## 3.3 Visual representation of model performance

To clearly assess the accuracy and clinical coherence of the best-performing model (XGBoost), two graphical analyses were generated and are presented in Figure 1. Figure 1A shows the relationship between the predicted and observed ceftaroline doses, highlighting a strong correspondence between the two values across patients. This result confirms the model's ability to capture individual variability in dosing requirements. Figure 1B presents the Visual Predictive Check (VPC), which compares the observed plasma concentrations with those simulated using the optimized doses predicted by the XGBoost model (Supplementary file 2; Ratain et al., 2003). Approximately 85% of the simulated plasma concentrations fall within the therapeutic range of 1–10 mg/L. Although some variability is still present, which is expected in a small pediatric cohort, the overall consistency of the results supports the validity of XGBoost as a robust proof-of-concept tool for individualized dose optimization.

To further validate the model's predictive consistency, an additional VPC was performed by comparing the observed plasma concentrations with those simulated using the optimized doses predicted by the XGBoost model. As shown in Figure 2, the observed and simulated concentration curves exhibited a consistent overall trend across patients, confirming that the model accurately captured the shape and magnitude of the plasma concentration distribution. Approximately 85% of simulated concentrations and 37.5% of observed concentrations were within the therapeutic range of 1–10 mg/L, supporting the model's ability to enhance dosing precision and minimize the risk of sub- or supra-therapeutic exposure. It is important to distinguish patient-level clinical alignment (91–94% depending on the model) from the Visual Predictive Check, where approximately 85% of simulated concentrations fell within the therapeutic window. Moreover, the VPCs demonstrated that the model preserved patient-specific exposure patterns while recentering the simulated concentrations around the therapeutic target (≈10 mg/L), indicating that the XGBoost-based optimization improved the dose homogeneity without diminishing interindividual variability.

## 3.4 Correlation between clinical variables and dose

A correlation matrix was used to analyze the linear relationships between the clinical variables and the target dose. This analysis was crucial for identifying the variables with the greatest impact on dose prediction and for understanding the underlying relationships in the data. As expected, physiological parameters like WT (r = 0.82), HT (r = 0.76), and SCR (r = 0.68) showed the strongest positive correlations with the dose. This reflects their direct influence on drug pharmacokinetics. Other variables, such as albumin and liver function parameters (AST, ALT), showed weaker correlations, suggesting a less significant direct impact on dose prediction. Figure 3 displays the complete correlation matrix, highlighting the relationships among all variables.

## 3.5 Explainable AI with LIME

In this study, one of the main objectives was to ensure the interpretability of the predictions made by the ML models. LIME creates a simpler model to explain the behaviour of a complex model for a single prediction, highlighting the features that contributed most to a specific outcome. This approach proved invaluable for local interpretation at an individual patient level. LIME analyses consistently identified WT, HT, SCR, and azotemia as the most influential features across models, with additional contributions from albumin and liver enzymes in some cases. The five most important predictors across models are summarized in Table 4. An illustrative LIME example for MLP is shown in Figure 4, which highlights how specific clinical variables influenced the Optimized AMT for an individual patient. The example shown corresponds to the best-performing model, the Neural Network, selected for its superior predictive accuracy and robustness.

# 4 Discussion

This study presents a comprehensive evaluation of various ML models for predicting an optimized Ceftaroline dose, focusing on both predictive accuracy and clinical utility. Our results establish a clear hierarchy of performance, with advanced models like the MLP and XGBoost demonstrating excellent predictive capability. Specifically, the MLP achieved a MAE of 1.53 mg and an $R^2$ of 0.94, while XGBoost performed similarly with an MAE of 2.04 mg and an $R^2$ of 0.89. The performance of these models significantly surpassed that of traditional linear methods. RR and LaR, for instance, yielded notably lower $R^2$ scores (0.51 and 0.54, respectively) and higher MAEs (7.92 mg and 7.61 mg), confirming their limitations in capturing the complex, non-linear relationships inherent in pharmacokinetic data. Similarly, HR, while more stable with outliers, did not match the performance of the tree-based and neural models, achieving a MAE of 7.20 mg and an $R^2$ of 0.35. The efficacy of gradient boosting was further underscored by LightGBM, which delivered performance metrics close to those of XGBoost (MAE of 2.92 mg and $R^2$ of 0.84). Beyond mathematical accuracy, our analysis revealed a crucial finding regarding the clinical utility of the models. While the MLP provided the most precise predictions from a statistical standpoint, ensemble models like RF and XGBoost showed superior performance in producing clinically meaningful recommendations. Specifically, XGBoost, RF and LGBM predicted doses that resulted in therapeutic plasma concentrations for 91.5, 94.2 and 92.4% of patients, respectively, while the MLP and LR achieved lower

10.3389/frai.2025.1702087

TABLE 3 Comparison between clinically administered doses and model-predicted optimized doses (mg) for representative patients.

| Dose Admin (mg). | LR | RF | XGBoost | MLP | RR | LaR | HR | LGBM |
|---|---|---|---|---|---|---|---|---|
| 102.0 | 102.4 | 97.4 | 102.0 | 102.3 | 101.9 | 99.7 | 100.4 | 101.6 |
| 42.0 | 46.9 | 42.2 | 42.0 | 41.5 | 44.2 | 45.3 | 43.9 | 41.9 |
| 88.2 | 89.7 | 88.3 | 88.2 | 87.5 | 88.1 | 87.4 | 87.9 | 88.1 |
| 37.2 | 38.3 | 39.3 | 37.2 | 37.4 | 38.0 | 38.6 | 37.9 | 37.3 |
| 60.0 | 67.5 | 60.4 | 60.0 | 60.5 | 63.2 | 64.3 | 63.7 | 60.4 |
| 66.0 | 63.1 | 67.1 | 67.0 | 65.9 | 65.8 | 67.3 | 65.9 | 66.7 |
| 96.0 | 83.5 | 91.4 | 96.0 | 95.3 | 91.9 | 89.4 | 90.6 | 95.7 |
| 105.0 | 106.7 | 102.6 | 105.0 | 106.8 | 105.5 | 103.9 | 104.1 | 105.6 |
| 60.0 | 67.2 | 58.1 | 60.0 | 60.7 | 63.4 | 62.1 | 63.0 | 59.8 |
| 49.8 | 56.6 | 51.7 | 49.8 | 47.1 | 52.3 | 54.1 | 53.6 | 50.9 |
| 100.2 | 93.6 | 96.2 | 100.2 | 100.5 | 97.8 | 98.1 | 99.3 | 100.7 |
| 49.8 | 43.9 | 45.0 | 49.8 | 49.8 | 46.7 | 47.3 | 47.0 | 49.5 |
| 36.92 | 36.5 | 38.1 | 36.9 | 37.1 | 36.8 | 36.9 | 37.2 | 37.0 |
| 39.8 | 35.7 | 39.0 | 39.8 | 40.0 | 38.5 | 37.9 | 38.1 | 39.7 |
| 75.0 | 72.4 | 79.5 | 75.0 | 75.1 | 75.6 | 77.8 | 76.2 | 75.8 |
| 26.96 | 28.4 | 32.1 | 27.0 | 27.3 | 29.1 | 29.8 | 29.3 | 27.8 |
| 49.8 | 48.5 | 49.2 | 49.8 | 49.8 | 49.1 | 48.9 | 49.3 | 49.9 |
| 75.0 | 74.0 | 80.4 | 75.0 | 75.5 | 76.0 | 78.3 | 76.8 | 75.3 |
| 30.0 | 29.6 | 33.0 | 30.0 | 30.2 | 31.4 | 32.0 | 31.5 | 30.6 |
| 102.0 | 107.2 | 101.7 | 102.0 | 100.2 | 104.5 | 102.1 | 102.7 | 101.9 |

Predictions from each model are shown alongside the corresponding administered dose. LR, Linear Regression; RF, Random Forest; XGBoost, Extreme Gradient Boosting; MLP, Neural Network; RR, Ridge Regression; LaR, Lasso Regression; HR, Huber Regression; LGBM, Light Gradient Boosting Machine.

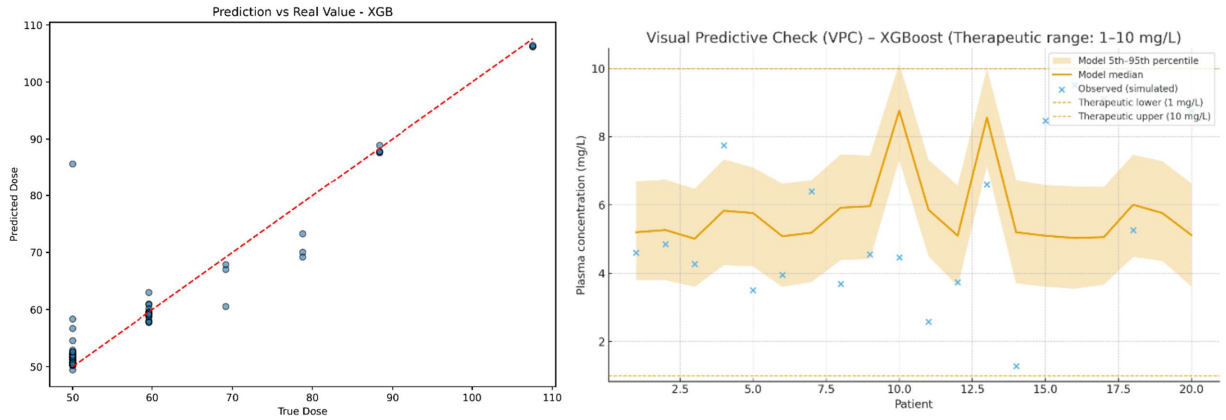


FIGURE 1
Visual representation of model performance (XGBoost). Left Observed vs. predicted ceftaroline doses. The dashed red line represents the identity line ($y = x$); blue points correspond to individual patients. Right Distributional visual predictive check (VPC). The shaded yellow area represents the 5th–95th percentile prediction interval derived from 1,000 bootstrap simulations, and the orange line indicates the model median. Blue crosses show the observed plasma concentrations. The dashed red lines delimit the therapeutic range (1–10 mg/L). This VPC illustrates the agreement between the observed data and the predictive distribution generated by the model, rather than a point-by-point fit.

alignment (≈49.9%). This discrepancy highlights a critical point for clinical decision support systems: a low mathematical error does not always equate to high clinical relevance. The high accuracy of the MLP may stem from its ability to model complex, non-linear patterns that lead to highly precise yet potentially out-of-range dose predictions. Conversely, the ensemble models, despite slightly lower accuracy, consistently recommended doses that fall within a clinically safe and effective therapeutic window, making them more valuable in a real-world setting. This suggests that, for dose optimization, models that prioritize a high percentage of clinically aligned predictions may be more valuable than those focused solely on minimizing mathematical error. The interpretability of these models was a

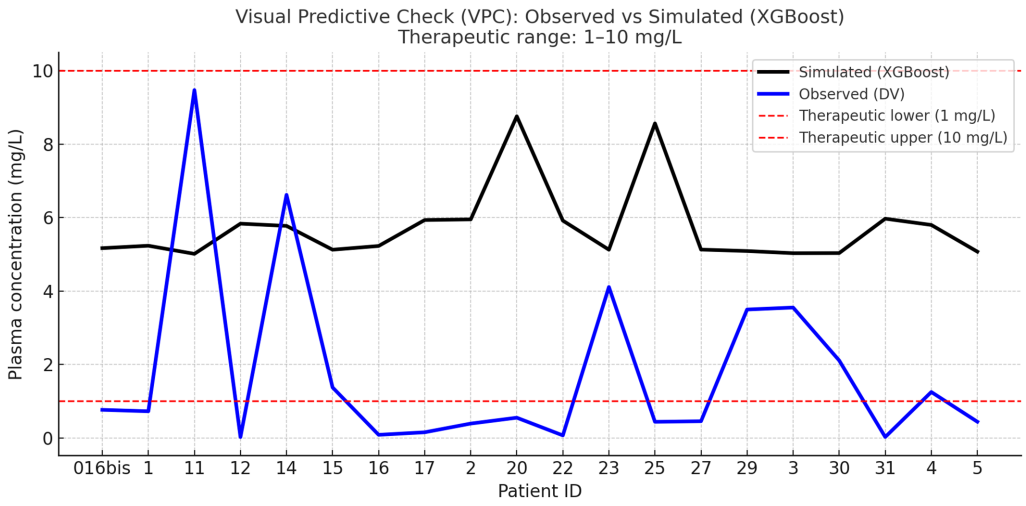


FIGURE 2
Combined visual predictive check (XGBoost). Blue line: median observed plasma concentrations (DV) per patient. Black line: median simulated concentrations obtained from model-predicted optimized doses. Red dashed lines: therapeutic range limits (1–10 mg/L). This combined VPC compares the observed and model-simulated concentration trends across patients. The divergence between curves reflects the difference between empirical dosing and model-optimized dosing, confirming the model's ability to reproduce the overall distributional pattern while reducing inter-patient variability.

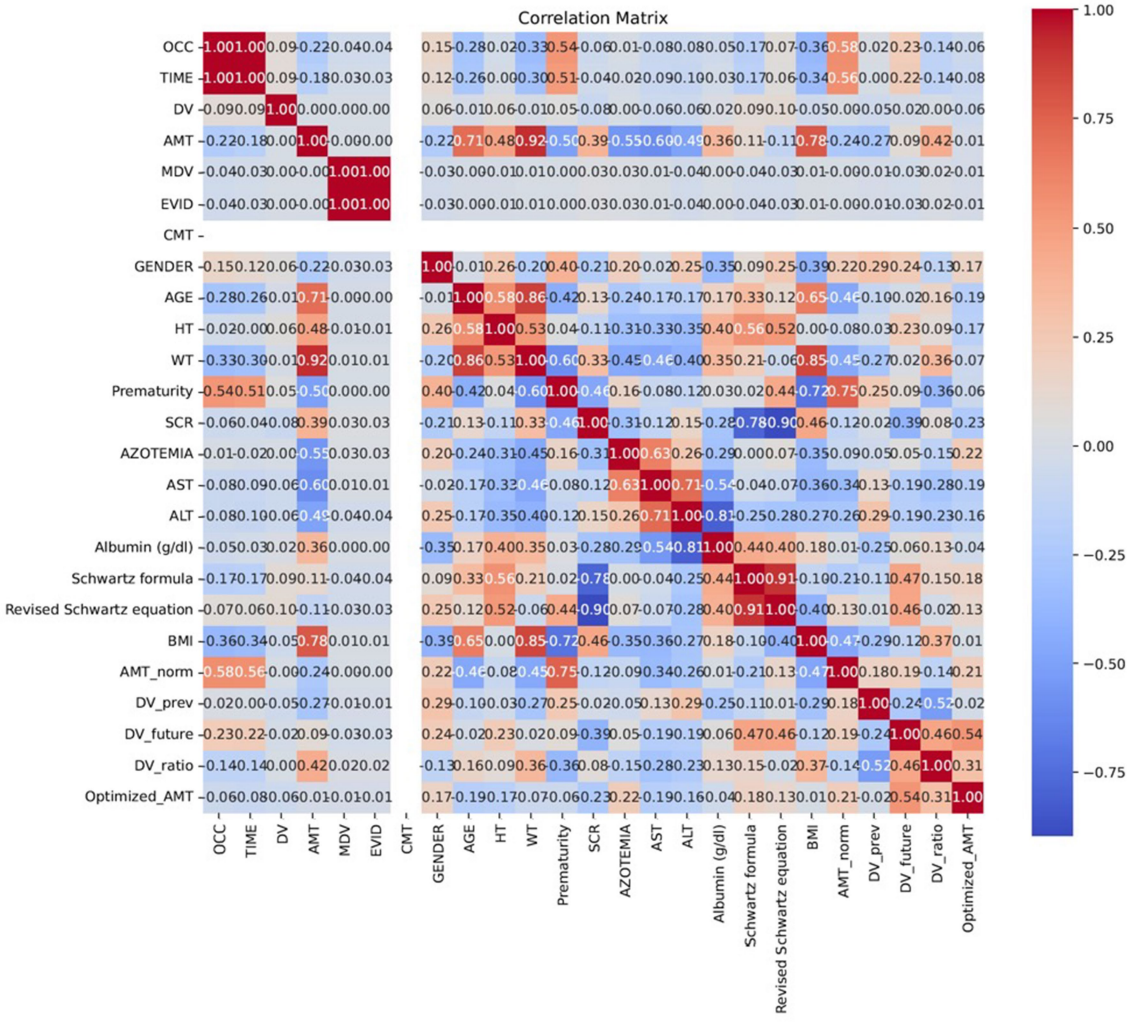


FIGURE 3
Correlation matrix of clinical variables. The figure highlights the linear relationships between key predictors, such as WT, HT, SCR, and the AMT. Variables like albumin, AST, and ALT show weaker correlations, indicating a less direct influence on the target variable.

TABLE 4 The top five features contributing to dose predictions across ML models.

| Model | Top 5 Features |
|---|---|
| Linear Regression | Weight (WT), Height (HT), SCR, Azotemia, Age |
| Ridge Regression | Weight (WT), Height (HT), SCR, Azotemia, Age |
| Lasso Regression | Weight (WT), Height (HT), SCR, Azotemia, Albumin |
| Huber Regression | Weight (WT), Height (HT), SCR, Azotemia, Age |
| Random Forest | Weight (WT), Height (HT), SCR, Azotemia, Albumin |
| XGBoost | Weight (WT), Height (HT), SCR, Azotemia, AST |
| LightGBM | Weight (WT), Height (HT), SCR, Azotemia, AST |
| Neural Network | Weight (WT), Height (HT), SCR, Azotemia, AST |

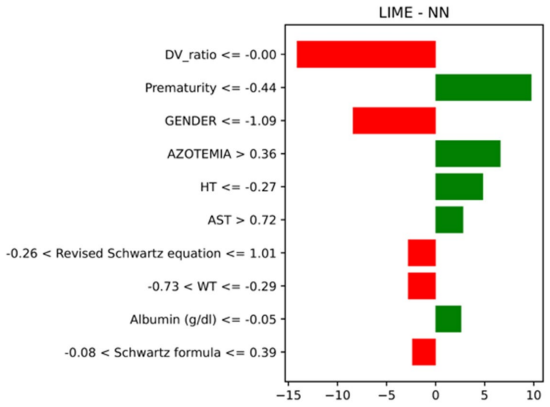


FIGURE 4
LIME explanation for a representative patient generated by MLP. The graph illustrates the contribution of individual clinical variables to the predicted dose. Positive contributions indicate variables that increased the predicted dose, while negative contributions represent variables that reduced the prediction.

cornerstone of our study, achieved through both correlation analysis and the application of LIME. Correlation analysis validated the strong influence of key physiological parameters, such as WT, HT, and SCR, which aligns with established pharmacokinetic principles for Ceftaroline. Notably, we observed a strong correlation between WT and HT ($r = 0.89$) and between SCR and blood urea nitrogen ($r = 0.72$). The application of LIME to "black-box" models, such as the MLP, provided critical, per-patient explanations for predictions. This is vital for clinical adoption as it allows physicians to trust and understand the reasoning behind a recommended dose, ensuring consistency with their expertise. Our analysis of feature importance confirmed the central role of renal parameters like SCR and azotemia in predicting dosage, which is consistent with the primary route of Ceftaroline elimination (Zhou et al., 2021). However, given the pharmacokinetic vulnerability of neonatal and especially preterm patients, all model-derived optimized doses in this study were considered strictly exploratory and were not used for clinical decision-making. Moreover, the uneven distribution of renal function markers in this small cohort may bias model behaviour and dose suggestions, and this potential source of error cannot be robustly quantified without larger prospective datasets. Conversely, the appearance of features such as AST and ALT among locally important variables in LIME explanations is likely influenced by multicollinearity and small-sample variability, and was therefore not interpreted as indicating a direct mechanistic role in dose–concentration prediction. Interestingly, our analysis also uncovered subtle sex-based differences, with males tending to require slightly higher doses. While modest, this finding suggests that sex-specific physiological factors, such as differences in body composition or renal function, may warrant further investigation to enhance the precision and equity of future dosing models. Traditional population pharmacokinetic (popPK) models remain the cornerstone of dose optimization, offering valuable mechanistic insights into how drugs are absorbed, distributed, metabolized, and eliminated. However, these models often require predefined structural assumptions and may not easily capture the complex, non-linear relationships that exist among clinical variables in real-world settings. In this work, we propose a complementary, data-driven perspective: the use of ML to model such variability directly from clinical data, without relying on strict compartmental structures. Rather than replacing popPK approaches, our method can enhance them, helping clinicians explore patient-specific factors and identify subtle interactions that traditional frameworks might overlook especially for peculiar population such as infants. The inclusion of explainability tools such as LIME further bridges the gap between computational modeling and clinical reasoning, allowing predictions to be interpreted in a way that aligns with medical decision-making.

## 5 Significance of the study and limitations

This study, although based on a limited pediatric cohort ($n = 20$), serves as a proof of concept for the application of ML in personalized dose optimization. The small sample size reflects the well-known challenges of conducting pharmacokinetic research in neonates and young infants but nonetheless provides valuable insights in this difficult-to-study population. The strong performance of advanced models warrants validation in larger, multicenter cohorts. Despite the limited sample size, the risk of model overfitting was carefully mitigated through the use of LOOCV and bootstrap-based validation, combined with regularization and early stopping strategies. The convergence and performance metrics remained stable across validation iterations, supporting the reliability of the findings within the intended proof-of-concept framework. Nevertheless, larger multicentric datasets and external validation will be essential to confirm model generalizability and clinical applicability. Future research will also explore the use of synthetic data generation to expand training variability and further enhance model robustness. Formal hypothesis testing for pairwise model comparison was not performed because paired statistical tests are underpowered and unstable in datasets of this size; therefore, model

performance is reported descriptively using LOOCV metrics and bootstrap confidence intervals.

## 6 Conclusion

Advanced ML models, particularly XGBoost, RF, and MLP, improve the prediction of ceftaroline dosing in pediatric patients. While MLP achieved the greatest statistical accuracy, ensemble models demonstrated superior clinical alignment, underscoring the importance of balancing precision with therapeutic applicability. The successful application of LIME and correlation analysis further validates that ML models can be both powerful and interpretable, a prerequisite for their seamless integration into clinical practice. The analysis of clinical parameters reaffirmed the predominant role of renal function and anthropometric characteristics in dose determination, which is consistent with known pharmacological principles. These findings advocate for AI-driven dose optimization to advance personalized pediatric pharmacotherapy and contribute to mitigating AMR (Branda and Scarpa, 2024).

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Niguarda Hospital Ethics Committee (protocol no. 189–042019, April 15, 2019). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

MF: Formal analysis, Investigation, Software, Writing – original draft, Writing – review & editing. GG: Data curation, Writing – review & editing. AG: Data curation, Writing – original draft, Writing – review & editing. VD: Writing – review & editing. CD: Writing – review & editing. SM: Writing – review & editing. MS: Writing – review & editing. AR: Writing – review & editing. CM: Writing – review & editing. GP: Writing – review & editing. APr: Writing – review & editing. RD: Writing – review & editing. FS: Writing – review & editing. GV: Writing – review & editing. DL: Writing – review & editing. APe: Writing – review & editing.

## Funding

The author(s) declared that financial support was not received for this work and/or its publication.

## Acknowledgments

We thank the Pediatric Unit and Neonatal Intensive Care Unit at ASST Grande Ospedale Metropolitano Niguarda for their collaboration.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2025.1702087/full#supplementary-material

## References

Ahmed, S. K., Hussein, S., Qurbani, K., Ibrahim, R. H., Fareeq, A., Mahmood, K. A., et al. (2024). Antimicrobial resistance: impacts, challenges, and future prospects. *J. Med. Surg. Public Health* 2:100081. doi: 10.1016/j.glmedi.2024.100081

Anderson, B. J., and Holford, N. H. G. (2008). Mechanism-based concepts of size and maturity in pharmacokinetics. *Annu. Rev. Pharmacol. Toxicol.* 48, 303–332. doi: 10.1146/annurev.pharmtox.48.113006.094708

Arnold, A., McLellan, S., and Stokes, J. M. (2025). How AI can help us beat AMR. *NPJ Antimicrob. Resist.* 3:18. doi: 10.1038/s44259-025-00085-4

Branda, F., and Scarpa, F. (2024). Implications of artificial intelligence in addressing antimicrobial resistance: innovations, global challenges, and healthcare's future. *Antibiotics* 13:502. doi: 10.3390/antibiotics13060502

Kearns, G. L., Abdel-Rahman, S. M., Alander, S. W., Blowey, D. L., Leeder, J. S., and Kauffman, R. E. (2003). Developmental pharmacology — drug disposition, action, and therapy in infants and children. *N. Engl. J. Med.* 349, 1157–1167. doi: 10.1056/NEJMra035092

Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., et al. (2022). Global burden of bacterial antimicrobial resistance

in 2019: a systematic analysis. *Lancet* 399, 629–655. doi: 10.1016/S0140-6736(21)02724-0

Poweleit, E. A., Vinks, A. A., and Mizuno, T. (2023). Artificial intelligence and machine learning approaches to facilitate therapeutic drug management and model-informed precision dosing. *Ther. Drug Monit.* 45, 143–150. doi: 10.1097/FTD.0000000000001078

Ratain, MJ, William, K., and Plunkett, J. (2003). Principles of pharmacokinetics. Journal of Medicine, surgery, and Public Health, Elsevier. Available online at: https://www.sciencedirect.com/science/article/pii/S2949916X24000343

Ruggiero, A., Ariano, A., Triarico, S., Capozza, M. A., Ferrara, P., and Attinà, G. (2019). Neonatal pharmacology and clinical implications. *Drugs Context* 8:212608. doi: 10.7573/dic.212608

Smith, N. M., Lenhard, J. R., Boissonneault, K. R., Landersdorfer, C. B., Bulitta, J. B., Holden, P. N., et al. (2020). Using machine learning to optimize antibiotic combinations: dosing strategies for meropenem and polymyxin B against carbapenem-resistant *Acinetobacter baumannii*. *Clin. Microbiol. Infect.* 26, 1207–1213. doi: 10.1016/j.cmi.2020.02.004

Welte, T., Kantecki, M., Stone, G. G., and Hammond, J. (2019). Ceftaroline fosamil as a potential treatment option for *Staphylococcus aureus* community-acquired pneumonia in adults. *Int. J. Antimicrob. Agents* 54, 410–422. doi: 10.1016/j.ijantimicag.2019.08.012

Willems, J., Hermans, E., Schelstraete, P., Depuydt, P., and De Cock, P. (2021). Optimizing the use of antibiotic agents in the pediatric intensive care unit: a narrative review. *Pediatr. Drugs* 23, 39–53. doi: 10.1007/s40272-020-00426-y

Zhou, J., You, X., Guo, G., Ke, M., Xu, J., Ye, L., et al. (2021). Ceftaroline dosage optimized for pediatric patients with renal impairment using physiologically based pharmacokinetic modeling. *J. Clin. Pharmacol.* 61, 1646–1656. doi: 10.1002/jcph.1944

## Glossary

**AI** - Artificial Intelligence

**AMR** - Antimicrobial Resistance

**AMT** - Amount administered

**ALT** - Alanine Aminotransferase

**AST** - Aspartate Aminotransferase

**BMI** - Body Mass Index

**DV ratio** - Ratio between target and real plasma concentration

**eGFR** - Estimated Glomerular Filtration Rate

**HR** - Huber Regression

**LaR** - Lasso Regression

**LGBM** - Light Gradient Boosting Machine

**LIME** - Local Interpretable Model-agnostic Explanations

**LR** - Linear Regression

**MAE** - Mean Absolute Error

**MAPE** - Mean Absolute Percentage Error

**MLP** - Neural Network

**PK** - Pharmacokinetics

**RF** - Random Forest

**RMSE** - Root Mean Squared Error

**RR** - Ridge Regression

**SCR** - Serum Creatinine

**WT** - Weight

**HT** - Height

**XAI** - Explainable AI

**XGBoost** - Extreme Gradient Boosting

frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 18 February 2026
DOI 10.3389/frai.2026.1662264

Check for updates

OPEN ACCESS

EDITED BY
Gregory R. Hart,
Institute for Disease Modeling (IDM),
United States

REVIEWED BY
Soomin Shin,
Korea National University of Transportation, Republic of Korea
Kyungwon Kim,
Incheon National University, Republic of Korea

*CORRESPONDENCE
Debora Marques de Miranda
debora.m.miranda@gmail.com

RECEIVED 08 July 2025
REVISED 03 January 2026
ACCEPTED 16 January 2026
PUBLISHED 18 February 2026

CITATION
Carvalho IF, Couto da Silva AP, Lacerda AM, Meira W Jr., Bastos Bispo Ferreira D, Malloy Diniz L, Serpa ALdO, Machado MCL, Romano-Silva MA, Miranda DM and Pappa GL (2026) Machine learning strategies for predicting pediatric suicidal behaviors in a Brazilian emergency setting. *Front. Artif. Intell.* 9:1662264. doi: 10.3389/frai.2026.1662264

COPYRIGHT
© 2026 Carvalho, Couto da Silva, Lacerda, Meira., Bastos Bispo Ferreira, Malloy Diniz, Serpa, Machado, Romano-Silva, Miranda and Pappa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Machine learning strategies for predicting pediatric suicidal behaviors in a Brazilian emergency setting

Isis F. Carvalho[1], Ana Paula Couto da Silva[1], Anisio M. Lacerda[1], Wagner Meira Jr.[1], Danilo Bastos Bispo Ferreira[2], Lys Malloy Diniz[2], Alexandre Luiz de Oliveira Serpa[3], Maria Carolina Lobato Machado[2], Marco A. Romano-Silva[2,4,5], Debora Marques de Miranda[5]* and Gisele Lobo Pappa[1]

[1]Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, [2]Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, [3]PPG em Ciências do Desenvolvimento Humano, Universidade Presbiteriana Mackenzie, São Paulo, SP, Brazil, [4]Shanghai Mental Health Centre, Shanghai, China, [5]Centro de Tecnologia e Medicina Molecular, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

**Background:** Suicide is a leading cause of death worldwide, yet its prediction remains a challenge. This difficulty arises not only because suicidal behavior is a rare event in the general population, leading to significant class imbalance in datasets, but also due to its complex, multi-causal nature involving a non-linear interplay of sociodemographic and clinical factors. Furthermore, while the majority of suicides occur in middle-income countries, there is a lack of predictive models tailored to these specific social contexts. This study evaluates machine learning strategies in an enriched clinical setting: a pediatric psychiatric emergency center in Brazil.

**Methods:** We analyzed a comprehensive database of 2,365 youth seeking emergency care. We benchmarked three machine learning algorithms, namely Logistic Regression, Random Forest, and XGBoost, to predict three outcomes: self-harm, suicidal ideation, and suicide attempts. To address class imbalance, we applied oversampling techniques to the training data. We also used SHapley Additive exPlanations (SHAP) values to quantify each feature's contribution to the predictions.

**Findings and interpretation:** In this setting, suicide-related behaviors represented 28.7% of the clinical demand. The Random Forest model combined with oversampling was the most effective strategy, achieving sensitivities of 78.04% for suicidal ideation, 71.18% for suicide attempts, and 69.37% for self-harm. Specificity remained consistently above 75%. SHAP value analysis revealed that social determinants were critical predictors, highlighting that social conditions in middle-income populations introduce unique variables that significantly influence suicidal risk. While accuracy for suicide attempts remained a challenge, SHAP provided clear clinical insights into the drivers of risk.

**Conclusions:** Machine learning, specifically Random Forest models together with oversampling and SHAP, demonstrates strong potential for identifying suicidal risk in pediatric emergency settings. By integrating clinical data with

social determinants, these models provide a transparent and scalable strategy for early identification in regions with limited specialized psychiatric resources.

KEYWORDS

adolescents, children, prediction, risk features, self-harm, suicide attempt, suicide ideation

# 1 Introduction

Suicide claims a life every 40 s globally, and each loss impacts 60–135 people (Knipe et al., 2022). In 2020, a meta-analysis reported a suicide rate of 4.9/100,000 and a pooled rate of 17% for suicidal ideation among adolescents (Bersia et al., 2022). In developed countries, suicide stands as the foremost cause of death among children and adolescents (Knipe et al., 2022). The often underestimated burden of suicide and related behaviors weighs heavily on low- and middle-income countries (LMICs) (Knipe et al., 2022; Naghavi, 2019), which account for 75% of suicides worldwide. Hence, efforts to understand this phenomenon in these nations remain necessary, as few studies have examined them (Klonsky et al., 2016; Robinson et al., 2018). The differences in suicide rates in LMICs are mainly attributed to economic, social and cultural features, such as access to social support systems or mental health care (Miladinov, 2023).

Self-harm occurs at a 20 times higher frequency than suicide, and often precedes suicide attempts (Mars et al., 2019a). Both engaging in self-harm and being exposed to it are prominent predictors of suicide attempts (Mars et al., 2019a). In LMICs, the incidence of suicide planning and attempts among adolescents, especially girls aged 15–17, is approximately 17% (Uddin et al., 2019). Given the profound individual, societal and household impact of child and adolescent deaths, it is imperative to investigate the factors leading to this condition.

Multiple machine learning (ML) models have been proposed to predict suicide (Harris et al., 2019a). However, predicting suicide is a challenge even for proficient psychiatrists and mental health specialists. Understanding the interplay between social and environmental factors and mental health is crucial, as social factors often contribute to risk (Robinson et al., 2018; Pollock, 2019; Navarro et al., 2021). This holds particular significance in countries such as Brazil, where inequality and social shortcomings can exacerbate psychiatric risks (Glenn et al., 2020; Orellana et al., 2020).

Robust predictors for suicide-related behaviors offer a pathway to intervene in individuals' behavior or monitor specific variables. Harris et al. (2019a) conducted a systematic review of risk assessment tools for predicting adolescent suicide (Harris et al., 2019b). They evaluated ten risk assessment tools used in the US and the UK and found that none of them could effectively predict suicidal behaviors. This reveals the complexity of the task, a matter that warrants significant attention given its widespread impact (Harris et al., 2019a,b).

In a population-based longitudinal study of Scottish young adults aged 18–34 (van Mens et al., 2020), the prediction of suicidal behaviors using traditional ML techniques found that algorithms based on decision trees outperformed regular logistic regression, achieving a sensitivity of 0.47 and a specificity of 0.91 for suicide attempt prediction. When comparing short- and long-term predictions of suicide risk using longitudinal data from structured electronic health records at the Connecticut Children's Medical Center, Su et al. (2020) developed a model with sensitivity 0.62 and specificity 0.90; their model exhibited superior performance for shorter prediction windows. A model derived from a Brazilian cohort revealed a heightened risk of depression among individuals who are female, socially isolated, non-white, involved in drug use and conflicts, experiencing academic difficulties, and victims of maltreatment (Orellana et al., 2020).

Here, we propose models to predict self-harm, suicidal ideation and suicidal attempts using an enriched dataset from the regional emergency unit of the Psychiatric Emergency Center for Children and Adolescents (CEPAI-FHEMI), in Brazil. This facility provides urgent psychiatric care, primarily assisting individuals reporting depression, substance use, and symptoms of agitation and aggressiveness, as detailed in Lobato Machado et al. (2022). This sample had more frequent suicidal behaviors than the general population, which may facilitate the identification of risks to the outcome (Robinson et al., 2018).

# 2 Methods

## 2.1 Ethics approval

This study was approved by the institutional review board (IRB) of the participating institution (COEP Ciencias Médicas). All procedures followed the ethical standards for human subjects research, and a waiver of informed consent was granted due to the retrospective design, in accordance with national regulations. The original data follows the rules from the medical records data storage, which must be retrieved only for patient benefit and must be safely stored for 20 years. Data was anonymized by removing names, addresses, document IDs, and other sensitive information. We report only aggregated patient statistics to avoid any risk of reidentification.

## 2.2 Dataset

We conducted a retrospective observational study with children and adolescents receiving emergency psychiatric care. The dataset comprises 2,365 health records of patients admitted to CEPAI-FHEMIG, some with multiple admissions, resulting in 1,720 unique patients (Lobato Machado et al., 2022). The data was collected from June 2017 to May 2018, and the subjects' ages range from

1 to 18 years. The dataset includes 27 sociodemographic features, such as race, gender, place of birth, household geographical location, school situation, living situation, and 123 clinical features, encompassing details such as reasons for seeking psychiatric assistance, family history of mental disorders, substance abuse information, psychiatric diagnoses received after treatment at the facility, neuro-psychomotor development delays, and previous traumatic events, among other relevant aspects.

We excluded 75 admissions where patients: (i) had no information at all (5); (ii) had only personal information (37); (iii) had no information about the motivations for looking for help or diagnoses (31); (iv) had not completed the screening stages at the center (1); or (v) had missing age information (1). Regarding the features, 12 of 110 contained textual information (e.g., a written reason for the patient leaving school). On average, more than 87% of the data in these 12 features were missing, and therefore these features were discarded. The original dataset also contained 21 features related to suicidal behaviors, which were, unfortunately, mostly missing (more than 80% of instances were missing) and were therefore excluded due to the high risk of data imputation in this context.

With the assistance of psychiatry professionals, out of the remaining 77 features, 57 were considered more informative for the tasks at hand. From the selected features, 33 were categorical. These categorical features were binarized, yielding 21 new binary features (e.g., whether the patient had depression, schizophrenia, or learning difficulties). The final dataset had 2,289 admissions for 1,687 unique patients (1,071 male and 616 female), with each admission described by 154 features.

Three target values—self-harm, suicidal ideation, and suicidal attempt—were extracted from the patient's motivation for seeking help, which was the prediction target. In this scenario, we had 337 out of 2,291 cases of self-harm (14.71%), 309 cases of suicide ideation (13.48%) and 323 positive suicide attempts (14.09%).

Self-harm was operationally defined as any non-suicidal manifestation of self-poisoning or self-injury lacking the explicit intent to induce mortality. Suicide ideation was identified in instances where individuals self-reported contemplation or strategic planning associated with suicidal tendencies. Concurrently, a suicide attempt was characterized as a deliberate act of self-poisoning or self-injury with the explicit aim of inducing mortality.

## 2.3 Methodology

To build prediction models for all outcomes, the classic methodology for handling data was followed. The first step was data preparation, as reported in the previous section. Next, the prepared datasets were used to train ML models, and oversampling was used to address class imbalance. As one of our main objectives is to interpret the contribution of each predictor to the model's decisions, we also employed SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) to understand predictions better. Each of these steps is detailed next.

The three ML models used in our experiments were Logistic Regression (LR), Random Forest (RF), and eXtreme Gradient Boosting (XGB) (Zaki and Meira, 2014). LR was chosen for its interpretability, whereas RF and XGB, as tree-based models, excel at capturing non-linear relationships and may facilitate the emergence of interpretable models. The experiments were conducted using the Scikit-learn (Pedregosa et al., 2011) and XGBoost libraries (Chen and Guestrin, 2016). A detailed description of these algorithms can be found in Bishop (2006).

Given the imbalanced nature of the data, instances with suicidal behaviors were oversampled to compose 30% of the training set. Oversampling is a well-known technique that alters the data distribution in the training set to improve predictive performance by increasing the number of instances of the rarer class.

The models were trained on two distinct feature sets. The first set incorporates all available features and is denoted as "All Features" in the results table. The second set, referred to as "Feature Selection," was generated using pre-selected features based on expert knowledge of the specialist who curated the dataset. It comprises features such as motivations for seeking help, diagnoses during medical hospitalization, whether it is the first-time admission to CEPAI-FHEMIG, and personal data, including gender, age, and the number of people in the household. We are aware that this type of selection can bias the model, and, for that reason, experiments with the full set of available features were also performed, and their results were compared with those obtained using only the features selected by the specialist.

During model training, the target variables were excluded from the set of predictors. Considering the relationship between suicide ideation and suicide attempt, ideation was used as a predictor for suicide attempt. Still, the suicide attempt was removed from the data used for suicide ideation prediction. For the self-harm prediction task, both suicidal ideation and suicide attempts were used as predictors, in line with previous research (Mars et al., 2019b).

Each model generated provided feature importance scores for the predictors, enabling us to compare the most significant features across classification tasks. Additionally, SHAP values were generated (Lundberg and Lee, 2017). SHAP is an approach rooted in game theory and frequently used to explain the output of ML models. It reveals the extent to which each feature contributes to the target feature. The interpretation of SHAP is akin to feature importance, but SHAP goes further by indicating whether each feature has a positive or negative relationship with the predicted value. It allows a richer understanding of feature impact, since it allows us to identify the most influential predictors associated with increased risk of suicide-related behaviors.

## 2.4 Experimental setup

To assess model performance, we used a stratified 10-fold cross-validation strategy, following the methodology detailed in Bishop (2006). To prevent data leakage, we employed patient-level grouping: all admissions associated with a single individual were restricted to a single fold, ensuring the model was never

tested on data from a patient it had already seen during training. For each iteration, the dataset was partitioned into training (eight folds), validation (one fold), and independent test sets (one fold). The validation set was used exclusively for hyperparameter optimization, whereas the test set was used only to assess model generalization. It is important to note that, in experiments using oversampling, the test set retained the original data distribution.

The performance of the models was assessed using four metrics:

1. Area under the receiver operating characteristic curve (AUC): assesses the model's performance across all classification thresholds of the true positive (TP) rate and false positive (FP) rate, providing a comprehensive view of the model's ability to discriminate between classes.
2. Positive predictive value (PPV): proportion of correctly predicted positive instances out of all predicted positive instances, offering insights into the accuracy of positive predictions.
3. Sensitivity: proportion of actual positive instances correctly identified by the classifier, indicating how well the model captures TP cases.
4. Specificity: proportion of actual negative instances correctly identified by the classifier, indicating how well the model captures true negative (TN) cases.

These metrics collectively provide a comprehensive evaluation of the model's performance on classification tasks. However, in the context of this paper, the sensitivity is considered more important than the other metrics, because the impact of not identifying a positive case of suicide attempt (or ideation, or self-harm) is considered more serious than misclassifying a negative case. This metric will be used to evaluate the models' performance relative to one another.

Regarding hyperparameter tuning, for XGB, a grid search was performed to identify a near-optimal parameter combination. This search encompassed four parameters: (1) the number of boosting iterations, (2) the ratio of features used to train a tree, (3) the maximum tree depth, and (4) the ratio of training instances in a subsample.

For RF, five parameters were assessed: (1) number of trees in the forest, (2) minimum number of samples required to split an internal node, (3) maximum depth of the tree, (4) minimum number of samples required to be at a leaf node and (5) whether bootstrap samples are used when building trees. Due to the large number of possible parameter combinations, an initial random search was conducted over the hyperparameter space. Subsequently, the best values identified in this search were evaluated using a grid search to determine the best parameter configuration. Table 1 presents the best parameter values found for Random Forest in the three scenarios investigated.

Once the parameters for both RF and XGB models were determined, further tuning involved adjusting their classification thresholds, which determine whether an example is classified as positive or negative (default value is 0.5), using the validation set. This experiment was also performed for LR. In this scenario, we balanced sensitivity and specificity using the F1 score, which is the harmonic mean of PPV and sensitivity. After that, the models were evaluated on the test sets.

TABLE 1 Best parameter configurations returned by the grid-search for the Random Forest algorithm.

| Parameters | Model output | | |
|---|---|---|---|
| | Suicide attempt | Suicide ideation | Self-harm |
| Number of trees | 10 | 10 | 100 |
| Max tree depth | None | None | None |
| Min. number of split samples | 2 | 4 | 1 |
| Min. number of leaf samples | 10 | 2 | 2 |
| Use of bootstrap | True | True | True |

The three models underwent testing with both the complete set of attributes and the reduced set of features ("Feature Selection"). Both experiments were conducted with and without oversampling.

# 3 Results

## 3.1 Descriptive results

Figure 1 shows the basic statistics of the dataset. As observed in Figure 1a, most patients (~62%) were between 13 and 18 years old. Figure 1b shows the number of admissions to CEPAI-FHEMIG. Although most of the patients were admitted only once, 21% of the individuals were admitted at least twice in the studied period, revealing an incidence of repeated suicide behaviors, in line with the literature. Figure 2 illustrates the distribution of diagnoses given by the psychiatric professional (after assessment). Depression and general mood disorders were diagnosed in 21,69% of all admissions, being the most common diagnosis.

Table 2 presents descriptive statistics for a subset of features, including motives that led the family to seek help and other dependent features considered relevant to the model, using suicide attempt as the dependent variable. The most common reasons were related to agitation, aggressiveness, irritability and learning difficulties, which are not traditionally associated with suicidal behaviors. Factors commonly related to suicidal behavior include depression, self-harm, anxiety, and self-aggressiveness, which were present in 19,07%, 14,7%, 14,36% and 5,28% of the admissions registered, respectively, and follow the distribution in the table.

Suicide-related behaviors were among the motivations given by patients or their guardians to seek psychiatric care. As previously reported, the motivations were used to define the model's outcome variables. The dataset had 337 out of 2,291 cases of self-harm (14.71%), 309 cases of suicide ideation (13.48%) and 323 positive suicide attempts (14.09%).

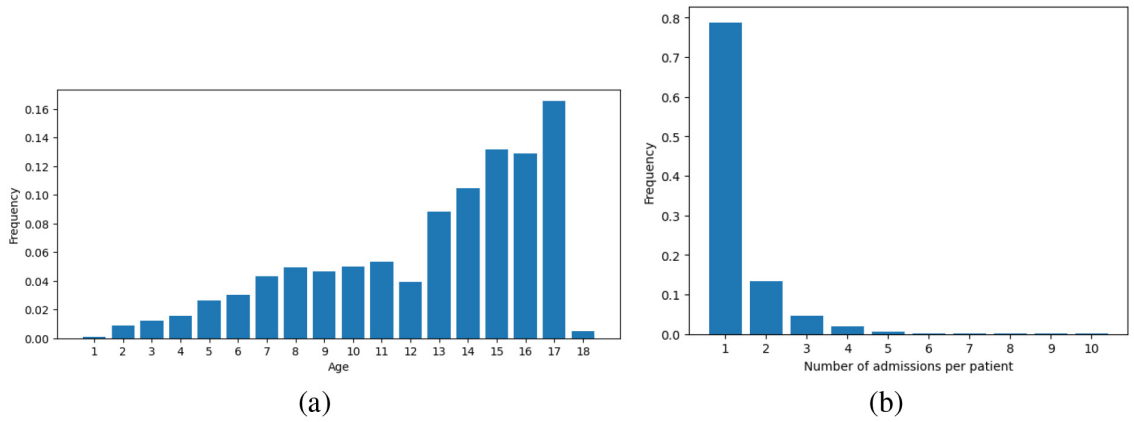


FIGURE 1
Statistics of the dataset. **(a)** Frequency of children searching for care in child psychiatry facilities; **(b)**Frequency of admissions for the same patient.

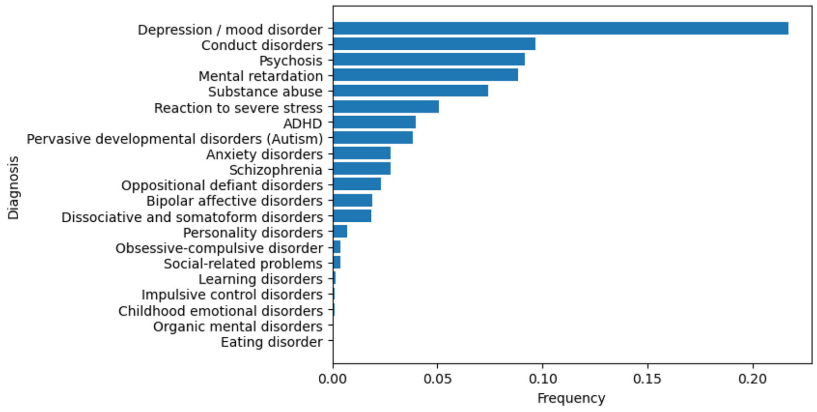


FIGURE 2
Frequency of diagnosis from admission in the psychiatric facility.

## 3.2 Suicide behavior prediction results

Tables 3–5 present the results of the prediction models predicting suicide attempts, suicidal ideation, and self-harm, respectively. In the tables, the acronym OV denotes Oversampling, referring to models in which the data distribution was modified. The ratio of the outcome we want to predict corresponded to 30% of the training dataset. The RF model achieved the best overall performance across the four metrics for all three tasks. Oversampling does not affect the current sample's results, as the method's results with and without oversampling differ only in confidence intervals, which overlap. For suicide attempt and ideation prediction, the best-performing model used only the "Feature Selection" subset of features, achieving a sensitivity of 0.7118, specificity of 0.7592 and PPV of 0.3350 for suicide attempt prediction. For suicide ideation prediction, the method presented a sensitivity of 0.7804, specificity of 0.7763 and PPV of 0.3602. In the case of self-harm, the best model was trained using all available features, resulting in a sensitivity of 0.6937, specificity of 0.8177 and PPV of 0.4086.

In summary, a combination of sensitivity above 0.69 and specificity above 0.75 indicates accurate predictions, correctly identifying more than 69 out of 100 positive cases and more than 75 out of 100 negative cases. However, further studies are still needed to improve PPV. PPV indicates the likelihood that an individual with a positive suicide prediction is truly going to commit suicide. Although in this context false positives are more desirable than false negatives, more accurate models have better economic and social advantages and are more feasible. Improving the current model would require additional data, and we believe longitudinal data are key. We are pursuing efforts to obtain a more comprehensive dataset from CEPAI-FHEMIG. Given that we initially intend to use the system for triage, the burden of low PPV is smaller than that of a fully predictive system in operation. The worst-case scenario would prioritize an individual with low risk over one with higher risk.

Finally, it is important to say that we have analyzed the differences in the metrics regarding the training and test sets, and the model does not present any indication of overfitting. A cross-validation procedure was performed to obtain statistically meaningful results. However, we cannot guarantee the method will generalize to other contexts without further adaptation, and testing it in new scenarios is part of our future work.

TABLE 2 Descriptive statistics of a subset of predictive variables, including motives to seek help, considering suicide attempt as the dependent variable.

| Independent variable | No suicide attempt ($N$ = 1,966) | Suicide attempt ($N$ =323) | $p$-value |
|---|---|---|---|
| Gender (male) | 1,309 (66.6%) | 129 (39.9%) | <0.001 |
| Gender (female) | 657 (33.4%) | 194 (60.1%) | <0.001 |
| Was hospitalized? | 193 (9.8%) | 81 (25.1%) | <0.001 |
| Age | 12.6 ± 3.9 | 15.1 ± 1.9 | <0.001 |
| Depression | 288 (14.6%) | 149 (46.1%) | <0.001 |
| Suicidal ideation | 197 (10.0%) | 112 (34.7%) | <0.001 |
| Self-harm | 230 (11.7%) | 107 (33.1%) | <0.001 |
| Substance abuse | 324 (16.5%) | 84 (26.0%) | – |
| Irritability | 598 (30.4%) | 74 (22.9%) | 0.007 |
| Aggressiveness | 830 (42.2%) | 69 (21.4%) | <0.001 |
| Insomnia | 288 (14.6%) | 63 (19.5%) | 0.031 |
| Isolation | 236 (12.0%) | 59 (18.3%) | 0.003 |
| Agitation | 905 (46.0%) | 49 (15.2%) | <0.001 |
| Anxiety | 279 (14.2%) | 47 (14.6%) | – |
| Hallucination | 271 (13.8%) | 46 (14.2%) | 0.894 |
| School difficulty | 482 (24.5%) | 40 (12.4%) | <0.001 |
| Disobedience | 414 (21.1%) | 38 (11.8%) | <0.001 |
| Mood swings | 156 (7.9%) | 21 (6.5%) | – |
| Attention difficulty | 320 (16.3%) | 15 (4.6%) | <0.001 |
| Delusion | 182 (9.3%) | 15 (4.6%) | 0.008 |
| Self-aggressiveness | 108 (5.5%) | 13 (4.0%) | 0.338 |
| Hypersexuality | 47 (2.4%) | 9 (2.8%) | 0.816 |

$p$-values were obtained with a $\chi^2$ test for categorical attributes and a $t$-test for age, the only numerical attribute.

## 3.3 SHAP values

As we are interested in identifying the factors that most influence the prediction of suicidal behaviors, the impact of each feature on the model's output was assessed using SHAP values. Figures 3–5 depict these values for the three tasks. In these plots, the horizontal axis indicates whether a particular value is associated with a higher or lower prediction value. The colors denote whether that attribute presents a high (red) or low (blue) value. For this analysis, the gender feature is coded as 0 for female and 1 for male.

For example, higher values of "Agitation," "Gender" (which means the patients are male), and "Learning difficulties" tend to produce lower predicted values for a suicide attempt (axis $x$), suggesting that male patients reporting agitation and learning difficulties are less likely to make a suicide attempt than the patients characterized by the opposite values of these binary variables (i.e., female patients, with no signs of agitation or learning difficulties). On the other hand, patients with depression, self-harm and suicidal ideation as motives for looking for help, who were hospitalized in CEPAI-FHEMIG and diagnosed with depression, are more likely to have a suicide attempt.

For suicide ideation (Figure 4), intellectual disability is linked to decreased risk, while isolation and hallucinations are associated with increased risk. As for the self-harm model (Figure 5), the relationships of gender and age with its prediction seem to be more regular, with older female patients being associated with a higher risk, while a patient with no trauma history using only CEPAI-FHEMIG's social services is linked to decreased risk, highlighting the distinct factors contributing to each outcome.

# 4 Discussion

The matter of children and adolescents requiring psychiatric care warrants attention. Roughly 1 in 5 children struggle with a mental, emotional, or behavioral disorder, yet fewer than 20% receive care from a mental health specialist (Vial and Almon, 2023). This lack of access to appropriate treatment is a pressing concern, particularly given that suicidal behavior, the most severe consequence of commonly untreated mental health issues, is a significant risk for this vulnerable population. In LMICs, accessing mental health services is even more difficult. Many individuals do not receive the care they need, and available services are often inefficient.

Suicidal behavior is a serious concern among adolescents, particularly during middle and late adolescence, with consequences that often persist into adulthood, especially for those who have previously attempted suicide. Research indicates that both short-term (6 months) and long-term (5 years) recurrence are common, increasing the risk among individuals who have been hospitalized for suicide attempts (Scoliers et al., 2009; Azcárate-Jiménez et al., 2019). Prevalence rates for suicidal ideation and attempts range from 14.3% to 22.6% and 4.6% to 15.8%, respectively (Van Meter et al., 2023). The transition from adolescence to young adulthood is associated with high healthcare costs and frequently results in interruptions in psychiatric care (Canaway et al., 2023). Late adolescence is the period of greatest risk for suicidal behavior, particularly during stressful times (Whalen et al., 2022; Orri et al., 2020). There is a pressing need to understand the predictors of suicidal behaviors to develop effective preventive interventions and ensure accessible, appropriate care for adolescents and young adults.

While ML excels at general classification and prediction tasks, predicting suicide behaviors remains a complex challenge. Suicide-related behaviors are influenced by numerous, interrelated factors that cannot be addressed with simple solutions. As noted earlier, a suicide attempt prediction model achieved a sensitivity of 0.47 and a specificity of 0.91 (van Mens et al., 2020). Although direct comparison is difficult due to differences in datasets, their study demonstrated higher specificity (0.76) but lower sensitivity (0.71) compared to ours. This underscores the critical importance of calibrating classifier thresholds. Adjusting these thresholds can significantly affect model performance, enabling tailored outcomes aligned with clinical priorities. Given the potential severity of missed cases, we prioritize increased sensitivity, while acknowledging that this often comes at the expense of specificity.

In sum, dealing with prediction tasks of suicidal behaviors is hard because (i) the data has an intrinsic bias—all patients presenting at the emergency had psychiatric symptoms; (ii) the suicide attempt itself is an unbalanced data; (iii) the dataset is limited to a low number of patients; (iv) mispredicting a high-risk patient has more serious consequences, given that it is

TABLE 3 Suicide attempt prediction results for the three classifiers tested with and without oversampling (OV).

| Features | Models | AUC | PPV | Sensitivity | Specificity |
|---|---|---|---|---|---|
| All features | LR | 0.6582 ± 0.0628 | 0.3841 ± 0.0835 | 0.4350 ± 0.1494 | 0.8814 ± 0.0462 |
| | LR OV | 0.6956 ± 0.0678 | 0.3372 ± 0.0542 | 0.5837 ± 0.1614 | 0.8076 ± 0.0577 |
| | RF | 0.7338 ± 0.0659 | 0.3719 ± 0.0829 | 0.6583 ± 0.1067 | 0.8092 ± 0.0547 |
| | RF OV | 0.7379 ± 0.0543 | 0.3461 ± 0.0701 | 0.7047 ± 0.0990 | 0.7710 ± 0.0637 |
| | XGB | 0.6769 ± 0.0489 | 0.3256 ± 0.0565 | 0.5406 ± 0.1007 | 0.8132 ± 0.0394 |
| | XGB OV | 0.6989 ± 0.0786 | 0.3574 ± 0.0734 | 0.5682 ± 0.1610 | 0.8295 ± 0.0434 |
| Feat. selection | LR | 0.6808 ± 0.0523 | 0.4380 ± 0.0951 | 0.4665 ± 0.1135 | 0.8951 ± 0.0376 |
| | LR OV | 0.7264 ± 0.0644 | 0.3686 ± 0.0661 | 0.6428 ± 0.1504 | 0.8100 ± 0.0627 |
| | RF | 0.7225 ± 0.0430 | 0.3151 ± 0.0563 | 0.7055 ± 0.0932 | 0.7395 ± 0.0636 |
| | RF OV | 0.7355 ± 0.0430 | 0.3350 ± 0.0514 | **0.7118 ± 0.1079** | 0.7592 ± 0.0627 |
| | XGB | 0.6812 ± 0.0553 | 0.3155 ± 0.0502 | 0.5628 ± 0.1272 | 0.7995 ± 0.0391 |
| | XGB OV | 0.7132 ± 0.0406 | 0.3012 ± 0.0241 | 0.6899 ± 0.0989 | 0.7364 ± 0.0366 |

In bold, there is the model and strategy with the best sensitivity.

TABLE 4 Suicide ideation prediction results for the three classifiers tested with and without oversampling (OV).

| Features | Models | AUC | PPV | Sensitivity | Specificity |
|---|---|---|---|---|---|
| All features | LR | 0.7053 ± 0.0406 | 0.4640 ± 0.0765 | 0.5075 ± 0.1040 | 0.9031 ± 0.0334 |
| | LR OV | 0.7418 ± 0.0445 | 0.4219 ± 0.0861 | 0.6274 ± 0.1252 | 0.8561 ± 0.0492 |
| | RF | 0.7599 ± 0.0434 | 0.4324 ± 0.0688 | 0.6625 ± 0.1212 | 0.8572 ± 0.0434 |
| | RF OV | 0.7819 ± 0.0540 | 0.4178 ± 0.0751 | 0.7287 ± 0.1373 | 0.8351 ± 0.0467 |
| | XGB | 0.7265 ± 0.0513 | 0.3596 ± 0.0683 | 0.6321 ± 0.1086 | 0.8210 ± 0.0384 |
| | XGB OV | 0.7782 ± 0.0304 | 0.3656 ± 0.0667 | 0.7755 ± 0.0896 | 0.7809 ± 0.0573 |
| Feat. selection | LR | 0.7250 ± 0.0415 | 0.4952 ± 0.1064 | 0.5449 ± 0.1096 | 0.9051 ± 0.0352 |
| | LR OV | 0.7397 ± 0.0364 | 0.4346 ± 0.0696 | 0.6099 ± 0.1046 | 0.8696 ± 0.0397 |
| | RF | 0.7557 ± 0.0340 | 0.3999 ± 0.0730 | 0.6781 ± 0.1020 | 0.8334 ± 0.0471 |
| | RF OV | 0.7783 ± 0.0363 | 0.3601 ± 0.0615 | **0.7804 ± 0.0954** | 0.7763 ± 0.0525 |
| | XGB | 0.6991 ± 0.0466 | 0.3685 ± 0.0615 | 0.5477 ± 0.1001 | 0.8504 ± 0.0362 |
| | XGB OV | 0.7514 ± 0.0505 | 0.3929 ± 0.0643 | 0.6703 ± 0.1304 | 0.8324 ± 0.0475 |

In bold, there is the model and strategy with the best sensitivity.

TABLE 5 Self-harm prediction results for the three classifiers tested with and without oversampling (OV).

| Features | Models | AUC | PPV | Sensitivity | Specificity |
|---|---|---|---|---|---|
| All features | LR | 0.6989 ± 0.0511 | 0.4593 ± 0.0702 | 0.5072 ± 0.1348 | 0.8905 ± 0.0481 |
| | LR OV | 0.7315 ± 0.0582 | 0.4045 ± 0.0831 | 0.6348 ± 0.1394 | 0.8281 ± 0.0628 |
| | RF | 0.7333 ± 0.0367 | 0.4567 ± 0.0678 | 0.5955 ± 0.1039 | 0.8710 ± 0.0490 |
| | RF OV | 0.7557 ± 0.0340 | 0.4086 ± 0.0700 | **0.6937 ± 0.1075** | 0.8177 ± 0.0600 |
| | XGB | 0.7198 ± 0.0288 | 0.3690 ± 0.0438 | 0.6305 ± 0.0873 | 0.8091 ± 0.0492 |
| | XGB OV | 0.7199 ± 0.0571 | 0.4693 ± 0.0692 | 0.5524 ± 0.1425 | 0.8874 ± 0.0447 |
| Feat. selection | LR | 0.7070 ± 0.0492 | 0.4906 ± 0.0809 | 0.5107 ± 0.1155 | 0.9033 ± 0.0391 |
| | LR OV | 0.7398 ± 0.0340 | 0.4112 ± 0.0640 | 0.6469 ± 0.0995 | 0.8327 ± 0.0505 |
| | RF | 0.7233 ± 0.0409 | 0.4594 ± 0.0817 | 0.5695 ± 0.1094 | 0.8771 ± 0.0483 |
| | RF OV | 0.7362 ± 0.0338 | 0.4065 ± 0.0608 | 0.6404 ± 0.0956 | 0.8321 ± 0.0498 |
| | XGB | 0.7155 ± 0.0405 | 0.3947 ± 0.0395 | 0.5902 ± 0.1084 | 0.8408 ± 0.0427 |
| | XGB OV | 0.7049 ± 0.0579 | 0.4877 ± 0.0454 | 0.5025 ± 0.1407 | 0.9073 ± 0.0299 |

In bold, there is the model and strategy with the best sensitivity.

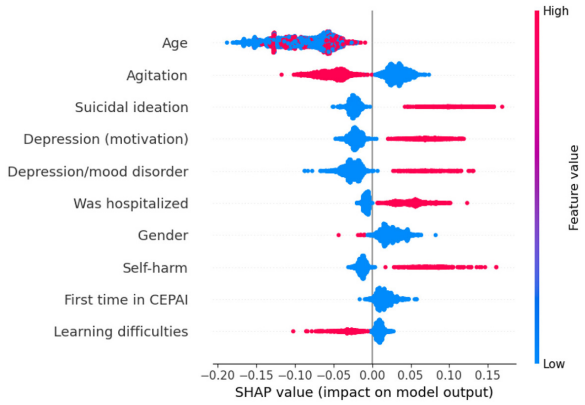


FIGURE 3
SHAP values for suicide attempt prediction. Each line represents a feature in decreasing order of importance, and the violin plot shows the impact of each feature in the model output. Features concentrated on the right side of the graph (positive SHAP) increase the prediction of suicide attempt, while features on the negative side decrease the prediction. Colors indicate whether a high (red) or low (blue) value of the feature is responsible for the increase/decrease in the prediction.

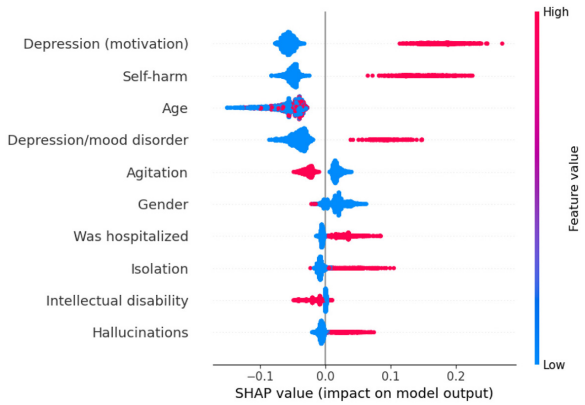


FIGURE 4
SHAP Values for Suicide Ideation Prediction. Each line represents a feature in decreasing order of importance, and the violin plot shows the impact of each feature in the model output. Features concentrated on the right side of the graph (positive SHAP) increase the prediction of suicide ideation, while features on the negative side decrease the prediction. Colors indicate whether a high (red) or low (blue) value of the feature is responsible for the increase/decrease in the prediction.

a cost-sensitive classification task. As most health-related models, the one generated here would need to be adapted to a setting that resembles the prevalence of these events in the general population (Richter et al., 2025). We are currently looking for a new dataset to validate our findings.

Given the numerous predisposing conditions, the relatively low frequency of the tragic event, and the undesirable possibility of error, the models proposed in this study demonstrate acceptable accuracy and can serve as valuable screening tools. These models can aid health professionals by identifying patients at risk of suicidal behavior. They may function effectively as screening tools, prioritizing individuals with a high probability of suicide

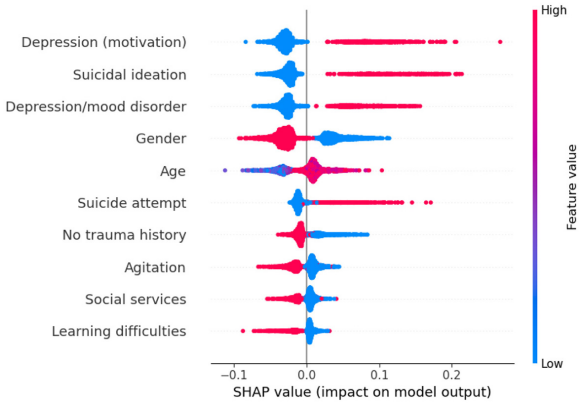


FIGURE 5
SHAP Values for Self-Harm Prediction. Each line represents a feature in decreasing order of importance, and the violin plot shows the impact of each feature in the model output. Features concentrated on the right side of the graph (positive SHAP) increase the prediction of self-harm, while features on the negative side decrease the prediction. Colors indicate whether a high (red) or low (blue) value of the feature is responsible for the increase/decrease in the prediction.

risk for prompt medical attention. Such measures that facilitate access could help minimize adverse outcomes. This approach is particularly crucial in middle-income scenarios where access to psychiatric facilities and treatment is insufficient for the multitude of people in deep need of care (Dalgalarrondo et al., 2023).

The unique context of addressing psychiatric cases in an emergency setting significantly differentiates our work. Patients in crisis alter the population-level expectation of rare events, making them more frequent. The emphasis put on identifying predictors of suicide-related behaviors allowed our models to pinpoint the most relevant factors for self-harm, suicidal ideation, and suicide attempts, permitting timely intervention. The potential of these predictor models lies in prioritizing higher-risk patients for immediate emergency care and identifying the risk factors involved in each case. Given the rarity of the outcome in high-risk populations (lower heterogeneity), there is a clear need to validate our model on data from other psychiatric centers to generalize the results (Varoquaux, 2018). Beyond the difficulties in predicting low-frequency events, another limitation of the study was the associative character of the risk features, which does not add causal knowledge that should be addressed at some point.

The consistent identification of the 10 most relevant features across models for predicting self-harm, suicidal ideation, and suicide attempts highlights the strong interconnection among these outcomes. When assessment resources are limited, key indicators of suicidal behavior become increasingly prominent, especially those linked to lack of treatment, inadequate care, and repeated psychiatric service utilization. Social factors, such as household size, may also shed light on vulnerability, particularly in LMICs, where these dynamics differ. While emergency settings provide a valuable source of cases for model training, they also introduce limitations, as findings may not generalize readily to other clinical contexts.

Depression and its symptoms consistently stand out as major predictors of suicide-related behaviors, as supported by

previous research (Li, 2023; Jiang et al., 2021; Mubasyiroh et al., 2018). Given its substantial predictive value, depression should be prioritized as a key target for intervention. Early detection and effective treatment of depression could substantially reduce the risk of suicide-related outcomes. In addition, self-harm and suicidal ideation are vital warning signs that may indicate the potential for more severe developments. Addressing modifiable social determinants, integrated into risk screening, can further enhance the comprehensiveness of suicide prevention strategies. Limited access to mental health care is also strongly associated with increased risk, as individuals often seek help in emergency settings due to inadequate prior psychiatric support. These findings highlight the urgent need to improve access to mental health services and initiate timely interventions to prevent escalation and severe outcomes.

## 5 Conclusions

This study introduces a machine learning framework designed to predict suicidal behaviors, including suicidal ideation, self-harm, and suicide attempts, within a pediatric psychiatric emergency context. By focusing on a middle-income population in Brazil, this research addresses a significant geographic gap. Although most global suicides occur in LMICs, these settings remain significantly underrepresented in predictive modeling literature.

Our findings lead to several key conclusions regarding the intersection of artificial intelligence and youth mental health: The Random Forest model with oversampling achieved sensitivities ranging from 69% to 78%, demonstrating its clinical utility for identifying at-risk individuals who might otherwise be missed in overstretched emergency services. By using SHAP values, we moved from "black-box" predictions, showing social determinants are primary drivers of risk in middle-income contexts.

Ultimately, this study serves as a proof-of-concept for the deployment of AI-supported screening tools in psychiatric emergency departments. By integrating these models into triage workflows, we can shift from reactive crisis management to proactive, data-driven identification, potentially reducing suicide rates in vulnerable youth populations worldwide.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Comite de Etica em Pesquisa da Ciencias Medicas. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because It consist in a retrospective medical data evaluation.

## Author contributions

IC: Formal analysis, Writing – original draft, Writing – review & editing. AC: Formal analysis, Supervision, Writing – original draft, Writing – review & editing. AL: Formal analysis, Writing – original draft, Writing – review & editing. WM: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. DB: Writing – original draft, Writing – review & editing. LM: Data curation, Formal analysis, Writing – original draft. AS: Conceptualization, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. MM: Data curation, Investigation, Resources, Writing – original draft, Writing – review & editing. MR-S: Conceptualization, Investigation, Resources, Supervision, Funding acquisition, Project administration, Writing – original draft. DM: Conceptualization, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing. GP: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was financed with resources from the National Institute for Science and Technology (INCT NeurotechR), CAPES, and the Centre for Innovation and Artificial Intelligence for Health (CI-IA Saúde), in part with resources from the São Paulo State Research Support Foundation (FAPESP) Process no. 2020/09866-4, from Minas Gerais Research Support Foundation (FAPEMIG) Process No. PPE-00030-21 and UNIMED Belo Horizonte and CNPq.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. We used Chat GPT to improve readability and grammar issues.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Azcárate-Jiménez, L., López-Goñi, J., Goñi-Sarriés, A., Montes-Reula, L., Portilla-Fernández, A., Elorza-Pardo, R., et al. (2019). Repeated suicide attempts: a follow-up study. *Actas Esp. Psiquiatr*. 47, 127–136. Available online at: https://actaspsiquiatria.es/index.php/actas/article/view/245 (Accessed January, 2026).

Bersia, M., Koumantakis, E., Berchialla, P., Charrier, L., Ricotti, A., Grimaldi, P., et al. (2022). Suicide spectrum among young people during the COVID-19 pandemic: a systematic review and meta-analysis. *EClinicalMedicine* 54:101705. doi: 10.1016/j.eclinm.2022.101705

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Canaway, A., Appleton, R., van Bodegom, L., Dieleman, G., Franić, T., Gerritsen, S., et al. (2023). Healthcare costs for young people transitioning the boundary between child/adolescent and adult mental health services in seven European countries: results from the milestone study. *BJPsych Open* 9:e175. doi: 10.1192/bjo.2023.559

Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 785–794. doi: 10.1145/2939672.2939785

Dalgalarrondo, P., Oda, A., Onocko-Campos, R., and Banzato, C. (2023). Challenges facing the psychiatric reform and mental health care in Brazil: critical unmet needs and prospects for better integrating the public and university sectors. *SSM - Ment. Health* 4:100262. doi: 10.1016/j.ssmmh.2023.100262

Glenn, C., Kleiman, E., Kellerman, J., Pollak, O., Cha, C., Esposito, E., et al. (2020). Annual research review: a meta-analytic review of worldwide suicide rates in adolescents. *J. Child Psychol. Psychiatry* 61, 294–308. doi: 10.1111/jcpp.13106

Harris, I., Beese, S., and Moore, D. (2019a). Predicting future self-harm or suicide in adolescents: a systematic review of risk assessment scales/tools. *BMJ Open* 9:e029311. doi: 10.1136/bmjopen-2019-029311

Harris, I., Beese, S., and Moore, D. (2019b). Predicting repeated self-harm or suicide in adolescents and young adults using risk assessment scales/tools: a systematic review protocol. *Syst. Rev*. 8:87. doi: 10.1186/s13643-019-1007-7

Jiang, T., Nagy, D., Rosellini, A., Horváth-Puhó, E., Keyes, K., Lash, T., et al. (2021). Suicide prediction among men and women with depression: a population-based study. *J. Psychiatr. Res*. 142, 275–282. doi: 10.1016/j.jpsychires.2021.08.003

Klonsky, E., May, A., and Saffer, B. (2016). Suicide, suicide attempts, and suicidal ideation. *Annu. Rev. Clin. Psychol*. 12, 307–330. doi: 10.1146/annurev-clinpsy-021815-093204

Knipe, D., Padmanathan, P., Newton-Howes, G., Chan, L., and Kapur, N. (2022). Suicide and self-harm. *Lancet* 399, 1903–1916. doi: 10.1016/S0140-6736(22)00173-8

Li, Y. (2023). Depression and suicide risk prediction based on machine learning models. *J. Educ. Humanit. Soc. Sci*. 15, 302–307. doi: 10.54097/ehss.v15i.9312

Lobato Machado, M., Hibner, M., Nogueira, D., Rezende, M., Bórem, I., da Cunha, L., et al. (2022). Irritability in an open-door pediatric psychiatric emergency service in a middle-income country. *Neuropsychiatr. Enfance Adolesc*. 70, 336–342. doi: 10.1016/j.neurenf.2022.05.007

Lundberg, S., and Lee, S. (2017). "A unified approach to interpreting model predictions" in *Paper Presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017)* (Long Beach, CA).

Mars, B., Heron, J., Klonsky, E., Moran, P., O'Connor, R., Tilling, K., et al. (2019a). Predictors of future suicide attempt among adolescents with suicidal thoughts or non-suicidal self-harm: a population-based birth cohort study. *Lancet Psychiatry* 6, 327–337. doi: 10.1016/S2215-0366(19)30030-6

Mars, B., Heron, J., Klonsky, E., Moran, P., O'Connor, R., Tilling, K., et al. (2019b). What distinguishes adolescents with suicidal thoughts from those who have attempted suicide? A population-based birth cohort study. *J. Child Psychol. Psychiatry* 60, 91–99. doi: 10.1111/jcpp.12878

Miladinov, G. (2023). The economics of suicide: evidence from LMICS and HICS. *J. Res. Soc. Sci. Humanit*. 2, 9–25. doi: 10.56397/JRSSH.2023.06.02

Mubasyiroh, R., Pradono, J., Nurkhotimah, E., Kusumawardani, N., and Idaiani, S. (2018). Depression as a strong prediction of suicide risk. *Glob. J. Health Sci*. 10:52. doi: 10.5539/gjhs.v10n12p52

Naghavi, M. (2019). Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the global burden of disease study 2016. *BMJ* 364:l94. doi: 10.1136/bmj.l94

Navarro, M., Ouellet-Morin, I., Geoffroy, M., Boivin, M., Tremblay, R., Côté, S., et al. (2021). Machine learning assessment of early life factors predicting suicide attempt in adolescence or young adulthood. *JAMA Netw. Open* 4:e211450. doi: 10.1001/jamanetworkopen.2021.1450

Orellana, J., Ribeiro, M., Barbieri, M., Saraiva, M. C., Cardoso, V., Bettiol, H., et al. (2020). Mental disorders in adolescents, youth, and adults in the RPS birth cohort consortium (Ribeirão Preto, Pelotas and São Luís), Brazil. *Cad. Saude Publica* 36:e00154319. doi: 10.1590/0102-311x00154319

Orri, M., Scardera, S., Perret, L., Bolanis, D., Temcheff, C., Séguin, J., et al. (2020). Mental health problems and risk of suicidal ideation and attempts in adolescents. *Pediatrics* 146:e20193823. doi: 10.1542/peds.2019-3823

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res*. 12, 2825–2830. Available online at: http://scikit-learn.sourceforge.net (Accessed January, 2026).

Pollock, N. (2019). Place, the built environment, and means restriction in suicide prevention. *Int. J. Environ. Res. Public Health*, 16. doi: 10.3390/ijerph16224389

Richter, M., Emden, D., and Leenings, R., Winter, N. R., Mikolajczyk, R., Massag, J., et al. (2025). Generalizability of clinical prediction models in mental health. *Mol. Psychiatry* 30, 3632–3639. doi: 10.1038/s41380-025-02950-0

Robinson, J., Bailey, E., Witt, K., Stefanac, N., Milner, A., Currier, D., et al. (2018). What works in youth suicide prevention? A systematic review and meta-analysis. *EClinicalMedicine* 4–5, 52–91. doi: 10.1016/j.eclinm.2018.10.004

Scoliers, G., Portzky, G., van Heeringen, K., and Audenaert, K. (2009). Sociodemographic and psychopathological risk factors for repetition of attempted suicide: a 5-year follow-up study. *Arch. Suicide Res*. 13, 201–213. doi: 10.1080/13811110902835130

Su, C., Aseltine, R., Doshi, R., Chen, K., Rogers, S., Wang, F., et al. (2020). Machine learning for suicide risk prediction in children and adolescents with electronic health records. *Transl. Psychiatry* 10:413. doi: 10.1038/s41398-020-01100-0

Uddin, R., Burton, N., Maple, M., Khan, S., and Khan, A. (2019). Suicidal ideation, suicide planning, and suicide attempts among adolescents in 59 low-income and middle-income countries: a population-based study. *Lancet Child Adolesc Health* 3, 223–233. doi: 10.1016/S2352-4642(18)30403-6

van Mens, K., de Schepper, C., Wijnen, B., Koldijk, S., Schnack, H., de Looff, P., et al. (2020). Predicting future suicidal behaviour in young adults, with different machine learning techniques: a population-based longitudinal study. *J. Affect. Disord*. 271, 169–177. doi: 10.1016/j.jad.2020.03.081

Van Meter, A., Knowles, E., and Mintz, E. (2023). Systematic review and meta-analysis: international prevalence of suicidal ideation and attempt in youth. *J. Am. Acad. Child Adolesc. Psychiatry* 62, 973–986. doi: 10.1016/j.jaac.2022.07.867

Varoquaux, G. (2018). Cross-validation failure: small sample sizes lead to large error bars. *Neuroimage* 180, 68–77. doi: 10.1016/j.neuroimage.2017.06.061

Vial, T., and Almon, A. (2023). Artificial intelligence in mental health therapy for children and adolescents. *JAMA Pediatr*. 177, 1251–1252. doi: 10.1001/jamapediatrics.2023.4212

Whalen, D., Hennefield, L., Elsayed, N., Tillman, R., Barch, D., Luby, J., et al. (2022). Trajectories of suicidal thoughts and behaviors from preschool through late adolescence. *J. Am. Acad. Child Adolesc. Psychiatry* 61, 676–685. doi: 10.1016/j.jaac.2021.08.020

Zaki, M., and Meira, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9780511810114

# Frontiers in
# Artificial Intelligence

**Explores the disruptive technological revolution of AI**

A nexus for research in core and applied AI areas, this journal focuses on the enormous expansion of AI into aspects of modern life such as finance, law, medicine, agriculture, and human learning.

---

## Discover the latest Research Topics

See more →

**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact