MEDS: Building Models and Tools in a Reproducible Health AI Ecosystem

Matthew B. A. McDermott
mattmcdermott8@gmail.com
Harvard Medical School
Boston, Massachusetts, USA

Justin Xu
University of Oxford
Oxford, United Kingdom

Teya S. Bergamaschi
MIT
Cambridge, Massachusetts, USA

Hyewon Jeong
MIT
Cambridge, Massachusetts, USA

Simon A. Lee
UCLA
Los Angeles, California, USA

Nassim Oufattole
MIT
Cambridge, Massachusetts, USA

Patrick Rockenschaub
Medical University of Innsbruck
Innsbruck, Austria

Kamilė Stankevičiūtė
University of Cambridge
Cambridge, United Kingdom

Ethan Steinberg
Stanford University
Stanford, California, USA

Jimeng Sun
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

Robin P. van de Water
Hasso Plattner Institute, University of Potsdam
Potsdam, Germany

Michael Wornow
Stanford University
Stanford, California, USA

John Wu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

Zhenbang Wu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

Abstract

Health AI suffers from a systemic reproducibility crisis that irreparably hinders research across both academia and industry [4, 5]. One key tool poised to solve this crisis is the Medical Event Data Standard (MEDS), a comprehensive data format and open-source ecosystem designed to enhance reproducibility and interoperability of AI research using longitudinal Electronic Health Records (EHR) [6]. Currently adopted by over 15 institutions globally, MEDS encompasses various open-source tools, published models, and data processing pipelines, enabling streamlined model development and robust benchmarking.

In this tutorial, participants will gain key hands-on experience in working with the MEDS format to perform efficient, reproducible, state-of-the-art AI research over real health data. Participants will transform data into the MEDS format, preprocess data, build predictive models, and contribute to the decentralized MEDS-DEV benchmarking platform. Interactive exercises using Jupyter notebooks will provide hands-on experience and practical skills for reproducible health AI research. Attendees will leave equipped with practical knowledge to use the MEDS schema and ecosystem of open source tools to build reproducible, state-of-the-art AI models.

KDD '25, Toronto, ON, Canada

ACM ISBN 979-8-4007-1454-2/2025/08
https://doi.org/10.1145/3711896.3737608

ACM Reference Format:
Matthew B. A. McDermott, Justin Xu, Teya S. Bergamaschi, Hyewon Jeong, Simon A. Lee, Nassim Oufattole, Patrick Rockenschaub, Kamilė Stankevičiūtė, Ethan Steinberg, Jimeng Sun, Robin P. van de Water, Michael Wornow, John Wu, and Zhenbang Wu. 2025. MEDS: Building Models and Tools in a Reproducible Health AI Ecosystem. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3–7, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3711896.3737608

1 Target Audience and Prerequisites

This tutorial is designed for researchers from academia and industry working with longitudinal EHR data. Participants should have basic to intermediate familiarity with Python, Jupyter notebooks, EHR data structures, Python data science libraries (e.g., PyTorch, pandas, polars), and both foundational and more modern machine learning concepts such as binary classification, the area under the receiver operating characteristic (AUROC), and pre-training/fine-tuning/"foundation model" learning paradigms.

2 Tutorial Goals

Attendees will learn to:

- Understand MEDS data standards and interoperability.
- Convert datasets to the MEDS format.

- Utilize MEDS-compatible tools, such as ACES [14], MEDS-Transforms, MEDS TorchData, MEDS-Evaluation, and MEDS-Tab [9].
- Develop, train, and evaluate state-of-the-art AI models.
- Contribute to and leverage the MEDS-DEV platform for reproducible benchmarking.

3 Tutorial Structure

The tutorial opens with a brief introduction and interactive participant survey to identify attendees' backgrounds and interests. Next, we explore the reproducibility crisis in health AI and demonstrate how MEDS addresses these issues by standardizing data practices and promoting interoperability.

Participants will gain a deep understanding of the MEDS schema, its design principles, and its compatibility with existing data formats through interactive discussions and Q&A sessions.

Attendees will engage in hands-on activities transforming raw datasets into MEDS format, specifically utilizing the MIMIC-IV demo dataset [3]. They will interact with MEDS ETL templates and receive references to similar resources for additional datasets like eICU [11], AUMCdb [1], and NWICU [7].

The ACES configuration language [14] will be introduced to extract predictive tasks, accompanied by interactive exercises. Building on these steps, participants will build predictive models, encompassing data preprocessing (normalization, tokenization, tensorization), model planning, and implementation in PyTorch. They will gain practical experience by training demo versions of their models over real MEDS demo datasets and see how to evaluate models using standardized methodologies.

The tutorial concludes with exploring MEDS-DEV's decentralized benchmarking capabilities. Attendees will also explore how to run other MEDS-ready models, such as MEDS-Tab [9], CEHR-BERT [10], GenHPF [2], MOTOR [13], CORE-BEHRT [8], and ETHOS [12]. The session ends with group discussions on predictive task definitions, benchmarking validity, and contributions to the MEDS ecosystem.

4 Societal Impacts

This tutorial directly addresses reproducibility challenges in health AI, promoting transparency, interoperability, and reliability. By adopting MEDS standards and accessible, robust tools, researchers can accelerate health AI advancements, improve patient outcomes, and facilitate equitable healthcare solutions globally.

References

[1] P W G Elbers. 2019. AmsterdamUMCdb v1.0.2.
[2] Kyunghoon Hur, Jungwoo Oh, Junu Kim, Jiyoun Kim, Min Jae Lee, Eunbyeol Cho, Seong-Eun Moon, Young-Hak Kim, Louis Atallah, and Edward Choi. 2023. GenHPF: General Healthcare Predictive Framework for multi-task multi-source learning. *IEEE J. Biomed. Health Inform.* PP, 1 (Oct. 2023), 1–12.
[3] Alistair E W Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Benjamin Moody, Brian Gow, Li-Wei H Lehman, Leo A Celi, and Roger G Mark. 2023. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* 10, 1 (Jan. 2023), 1.
[4] Matthew McDermott. 2025. The (lack of?) Science of Health AI. In *Proceedings of the 3rd Machine Learning for Health Symposium*. PMLR.
[5] Matthew B A McDermott, Shirly Wang, Nikki Marinsek, Rajesh Ranganath, Luca Foschini, and Marzyeh Ghassemi. 2021. Reproducibility in machine learning for health research: Still a ways to go. *Sci. Transl. Med.* 13, 586 (March 2021), eabb1655.
[6] MEDS Working Group: Bert Arnrich, Edward Choi, Jason A. Fries, Matthew B. A. McDermott, Jungwoo Oh, Tom J. Pollard, Nigam Shah, Ethan Steinberg, Michael Wornow, Robin van de Water. 2024. Medical Event Data Standard (MEDS): Facilitating Machine Learning for Health. In *ICLR 2024 Workshop TS4H*.
[7] Dana Moukheiber, William Temps, Bhadrappa Molgi, Yikuan Li, Alice Lu, Prasanth Nannapaneni, Abdulrahman Chahin, Sicheng Hao, Felipe Torres Fabregas, Leo Anthony Celi, Adrian Wong, Maxwell Lloyd, Xavier Borrat Frigola, Hyung-Chul Lee, Daniel Schneider, Tom Pollard, Yuan Luo, Abel Kho, and Roger Mark. 2024. Northwestern ICU (NWICU) database.
[8] Mikkel Odgaard, K Klein, Sanne Moller Thysen, E Jimenez-Solem, Martin Sillesen, and Mads Nielsen. 2024. CORE-BEHRT: A carefully optimized and rigorously evaluated BEHRT. In *Machine Learning for Healthcare Conference*, Kaivalya Deshpande, Madalina Fiterau, Shalmali Joshi, Zachary Lipton, Rajesh Ranganath, and Iñigo Urteaga (Eds.), Vol. abs/2404.15201. PMLR.
[9] Nassim Oufattole, Teya Bergamaschi, Aleksia Kolo, Hyewon Jeong, Hanna Gaggin, Collin M Stultz, and Matthew B A McDermott. 2024. MEDS-Tab: Automated tabularization and baseline methods for MEDS datasets. *arXiv [cs.LG]* (Oct. 2024).
[10] Chao Pang, Xinzhuo Jiang, Krishna S Kalluri, M Spotnitz, Ruijun Chen, A Perotte, and K Natarajan. 2021. CEHR-BERT: Incorporating temporal information from structured EHR data to improve prediction tasks. 158 (Nov. 2021), 239–260.
[11] Tom J Pollard, Alistair E W Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. 2018. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data* 5 (Sept. 2018), 180178.
[12] Pawel Renc, Yugang Jia, Anthony E Samir, Jaroslaw Was, Quanzheng Li, David W Bates, and Arkadiusz Sitek. 2024. Zero shot health trajectory prediction using transformer. *NPJ Digit. Med.* 7, 1 (Sept. 2024), 256.
[13] Ethan Steinberg, Jason Alan Fries, Yizhe Xu, and Nigam Shah. 2023. MOTOR: A Time-to-Event Foundation Model For Structured Medical Records. In *The Twelfth International Conference on Learning Representations*.
[14] Justin Xu, Jack Gallifant, ALISTAIR JOHNSON, and Matthew B.A. McDermott. 2025. ACES: Automatic Cohort Extraction System for Event-Stream Datasets. In *The Thirteenth International Conference on Learning Representations*. https://openreview.net/forum?id=P4XmKjXTrM