Machine Learning for Health (ML4H) 2024

Helen Zhou[1] HLZHOU@CS.CMU.EDU
Stefan Hegselmann[1] STEFAN.HEGSELMANN@CHARITE.DE
Elizabeth Healey[2] EHEALEY@MIT.EDU
Trenton Chang[2] CTRENTON@UMICH.EDU
Caleb Ellington[2] CELLINGT@CS.CMU.EDU
Michael Leone[3] MLEONE@ANDREW.CMU.EDU
Vishwali Mhasawade[4] VISHWALI.MHASAWADE@NYULANGONE.ORG
Sana Tonekaboni[4] STONEKAB@BROADINSTITUTE.ORG
Winston Chen[5] CHENWT@UMICH.EDU
Hyewon Jeong[6] HYEWONJ@MIT.EDU
Xiaoxiao Li[6] XIAOXIAO.LI@ECE.UBC.CA
Juyeon Heo[7] JH2324@CAM.AC.UK
Payal Chandak[8] CHANDAK@MIT.EDU
Ayush Noori[8] ANOORI@COLLEGE.HARVARD.EDU
Sarah Jabbour[9] SJABBOUR@UMICH.EDU
Jessica Dafflon[9] JESSICA@VALENCELABS.COM
Jerry Ji[10] JERRYJI@CS.TORONTO.EDU
Jivat Neet Kaur[10] JNKAUR@BERKELEY.EDU
Amin Adibi[10] AMIN.ADIBI@UBC.CA
Xu Cao[10] XUCAO2@ILLINOIS.EDU
Meera Krishnamoorthy[11] MEERAK@UMICH.EDU
Yidi Huang[11] YIDIH@UPENN.EDU
Fabian Gröger[12] FABIAN.GROEGER@UNIBAS.CH
Aishwarya Mandyam[13] AM2@STANFORD.EDU
Niloufar Saharhkhiz[13] NILOUFAR.SAHARKHIZ@FDA.HHS.GOV
Teya Bergamaschi[14] TEYA@MIT.EDU
William Boag[15] WBOAG@MIT.EDU
Jeroen Berrevoets[15] JEROEN.BERREVOETS@MATHS.CAM.AC.UK
Matthew Lee[16] MATTHEW.LEE1@PENNMEDICINE.UPENN.EDU
Kyle Heuton[16] KYLE.HEUTON@TUFTS.EDU
Peniel Argaw[17] PENIELARGAW@MICROSOFT.COM
Haoran Zhang[17] HAORANZ@MIT.EDU

Organizing committee for ML4H 2024
[1] *Publications Chair,* [2] *Workflow Chair,* [3] *Workflow Subchair,* [4] *Program Chair,* [5] *Program Subchair,* [6] *Logistics Chair,* [7] *Logistics Subchair,* [8] *Demos & Perspectives Chair,* [9] *Outreach Chair,* [10] *Research Roundtable Subchair,* [11] *Career Mentorship Subchair,* [12] *Reviewer Mentorship Subchair,* [13] *Author Mentorship Subchair,* [14] *Demos & Perspectives Subchair,* [15] *Communications Chair,* [16] *Finance Chair,* [17] *General Chair*

1. Introduction

The fourth Machine Learning for Health (ML4H) symposium was held on December 15-16, 2024, in Vancouver, BC, Canada. Similar to the last three years (Roy et al., 2021; Parziale et al., 2022; Hegselmann et al., 2023b), the symposium was a standalone event co-located with the Neural Information Processing Systems (NeurIPS) conference.

ML4H 2024 invited high-quality submissions describing innovative research in a variety of health-related disciplines including healthcare, biomedicine, and public health. Works could be submitted to either the archival *Proceedings* track, or the non-

archival *Findings* track. The Proceedings track targeted mature, cohesive works with technical sophistication and high-impact relevance to health. The Findings track promoted works that would spark new insights, collaborations, and discussions at ML4H. Both tracks were given the opportunity to share their work through the in-person poster session. This year, ML4H received 155 submissions to the Proceedings track, and 70 submissions to the Findings track. Of these, 62 Proceedings and 33 Findings track papers were ultimately accepted, with 36 additional Proceedings track papers accepted as Findings track papers. In addition to the typical submission tracks, ML4H offered a *Demos* track as well as a new *Perspectives* track (see Section 2.5).

To encourage new connections and bring together community members from all levels of experience, ML4H offered three types of mentorship programs. Author mentorship was intended to help mentees improve the quality of their submissions before they are reviewed, reviewer mentorship to improve the quality of reviews, and career mentorship to provide advice on career paths. To help foster further connection with like-minded individuals on the day of the event, participants were given the opportunity to participate in a meetup during ML4H. In addition, we hosted research roundtables to discuss important and timely topics related to machine learning for health.

In this front matter, we provide an overview of the ML4H 2024 Symposium (Section 2), including the paper review process, summary statistics about the submissions, and the program. We analyze trends in the accepted works versus previous ML4H workshops and symposiums (Section 3). We summarize characteristics of the ML4H community (Section 4), reflect and present conclusions (Section 5), and conclude with acknowledgements (Section 6).

2. Symposium

2.1. Program

Our program at ML4H 2024 included seven invited talks from experts in fields related to our two thematic sessions. Each session began with two 30-minute invited keynotes, followed by a 15-minute perspectives talk and a 45-minute panel discussion, including the keynote and perspective speakers. The second session also ended with a 30-minute closing keynote.

The first thematic session was on *foundation models in healthcare*. It is a pivotal topic because foundation models provide a unified, scalable approach to analyzing diverse and complex health data, such as imaging, electronic health records, and genomics. These models leverage vast pretraining to generalize across tasks, enabling breakthroughs in diagnostics, personalized medicine, and disease prediction. By reducing reliance on task-specific data and fine-tuning, foundation models accelerate innovation and democratize access to AI capabilities in healthcare, ultimately improving patient outcomes and supporting evidence-based care. We invited papers focusing on this area as well as expert speakers leading this emerging area. This session began with keynote talks by Jason Fries and Yukun Zhou and a perspective talk by Matthew McDermott.

The second thematic session was on the Challenges of *Deploying AI models in healthcare*. Deploying AI models in healthcare is important for improving diagnosis, treatment personalization, and operational efficiency, potentially saving lives and reducing costs. However, it is challenging due to strict regulatory requirements, data privacy concerns, and the need for models to be interpretable and trustworthy. Ensuring robust performance across diverse patient populations and seamless integration into clinical workflows also presents hurdles. Addressing these challenges requires a focus on ethical AI development, rigorous testing, and collaboration between technologists, healthcare providers, and regulators. Hence, we invited papers focusing on this area and speakers with experience working on deploying AI for healthcare in general. It began with keynote talks by Adarsh Subbaswamy and Shannon McWeeney and a perspective talk by Isaac Kohane. Finally, the session ended with a closing keynote talk by Su-in Lee.

In contrast to prior years, the program was split across two days instead of one, with the first thematic session on day one and the second thematic session on day two.

On day one, in addition to the invited talks and panels, we held research roundtables and mentorship roundtables, a spotlight talk session for authors of selected papers to present their work, and a poster session for all authors to present and attendees to ask questions. Day one concluded with an evening social at Pinnacle Vistas 360°, where ML4H 2024 participants could network, continue their discussions, and enjoy a social atmosphere.

Day two started similarly with invited talks and panels, followed by research roundtables and mentorship roundtables. Before the spotlight talks and poster session, however, there was additionally a spotlight demonstrations session. The general demonstrations sessions were held concurrently with both poster sessions. Day two concluded with a closing keynote and remarks.

Overall, we structured the program to be heavier on presentations in the morning. The afternoon was followed by more open and mobile sessions (poster session, research roundtables, and evening social), so that the interesting content presented earlier in the day could open up fruitful discussions that flowed into the evening and hopefully into lasting conversations and even collaborations beyond the symposium.

2.2. Paper Selection

Submission Statistics This year, we received 225 full submissions (Proceedings and Findings), constituting a 13.6% increase over the previous year (Hegselmann et al., 2023b). This was comprised of 155 submissions to the Proceedings track and 70 submissions to the Findings track.

The program committee consisted of 277 reviewers and 26 area chairs. Our review process was double-blind and we ensured each submission received at least 3 quality reviews.

This year, we accepted 62 papers to the Proceedings track for an acceptance rate of 40%. Consistent with previous years, for a subset of submissions we gave authors the option to transfer their work from the Proceedings track to the Findings track. This year, we accepted 36 Proceedings submissions as Findings papers. Of the Findings submissions, 33 papers were accepted (47% acceptance rate).

2.3. Paper Awards

This year, best paper awards were presented to the top proceedings paper from each of the three submission areas, as well as the best findings paper. This corresponded to the following awards:

- Best Proceedings Paper: Models and Methods
- Best Proceedings Paper: Applications and Practice
- Best Proceedings Paper: Impact and Society
- Best Findings Paper

To decide on the awards, the General Chairs started by gathering a subset of papers that were accepted in the Proceedings and Findings track, corresponding to those with top average reviewer scores. For each track, we manually examined the reviews, the authors' rebuttals to them, and the metareview. Based on this, we created a smaller subset of papers where the authors have addressed all major weaknesses raised by the reviewers, and the area chair strongly recommends the paper. We read each of the papers in this subset, and chose to give the award to the paper in each category with the most significance, *i.e.*, one that has the most potential impact, and will generate the most fruitful discussion.

Winners of the awards were announced during the closing remarks on the last day of the symposium. Authors were given a cash prize and a commemorative certificate.

2.4. Outreach Programs

This year, our outreach efforts consisted of five main programs: author mentorship, reviewer mentorship, career mentorship, and research roundtables, and attendee interest matching. The author mentorship program ran up to the paper submission deadline, followed by the reviewer mentorship program, which continued until the end of the review period. The career mentorship program took place between the review period and the symposium date. All four programs ended with in-person mentorship roundtables to foster smaller group discussions each topic. In addition to the mentorship roundtables, we also held research roundtables on specialized research topics of interest to the ML4H community. Finally, the new attendee matching program consisted of matching conference attendees based on common interests to foster more in-person connections.

Author Mentorship The Author Mentorship Program focuses on pairing less experienced authors with senior researchers to provide feedback on their paper submission, with the overall goal of improving both submission quality and fostering future collaboration. Mentors were expected to meet with mentees on a bi-weekly basis for two months prior to the paper submission deadline. The goal of these interactions is to provide a space for the mentee to describe ongoing research and receive advice about presentation and experimental development from the mentors. Mentors ranged in experience from senior graduate students to professors. The mentors also ranged in pro-

fession, with several coming from industry research backgrounds.

This year, we simultaneously recruited 29 mentors who were matched based on research expertise and interests with 58 mentees. Each mentor was paired with anywhere between one and four mentees depending on their time constraints. We used a matching algorithm that considered factors such as field(s) of expertise, mentor time capacity, and willingness to provide advice in fields related to expertise. This matching algorithm was used several times because we kept the signup forms for both mentees and mentors open until 1 month prior to the submission deadline. To ensure the program was proceeding smoothly, both mentors and mentees received check-in emails twice. The goal of these check-in emails was to identify if the mentor-mentee matches had met, and whether any re-matching was necessary. In the vast majority of cases, there were at most minor issues, and in a few cases, we re-matched the mentees or assigned new mentees to existing mentors.

Upon completion of the program, we sent out feedback forms to both the mentors and mentees. Responses noted that even if the mentee did not submit a paper to the ML4H main conference, they were able to benefit from the mentorship program and build connections in similar research fields. The most prevalent complaint was a lack of time for the mentorship program. We attribute this in part to the fact that we left the signup forms open for a long time, and were matching mentor-mentee pairs until one month prior to the submission deadline. In future iterations of this program, we imagine that the best option is to close the signup forms 2 months prior to the submission deadline.

Reviewer Mentorship The Reviewer Mentorship Program is designed to elevate the capabilities of junior reviewers by matching them with senior reviewers, with the ultimate goal of refining the review process. The program's foundation is the professional development of junior reviewers, typically graduate students, through a systematic approach that includes evaluating academic papers, in-depth feedback discussions, and guidance in responding to authors. The purpose of the reviewer mentorship program also extends beyond skill development; it is about building new connections within the ML4H community and elevating the overall quality of the review process. The expectation of this program is twofold: mentors will provide constructive feedback on review drafts, and mentees will refine their reviews accordingly. These interactions improve the immediate quality of reviews and foster a culture of continuous learning and collaboration.

This year, 73 individuals applied for the program, of which all confirmed their participation through OpenReview. We succeeded in pairing 26 mentors with 47 mentees. The mentors, who had at least a doctoral degree or strong industry experience with significant review and publication experience, voluntarily provided guidance. Their expertise ensured that the match with mentees was based on shared research interests and appropriate experience levels. They also had input on the number of mentees they preferred to mentor.

Participants represented a diverse cross-section of the ML4H community, with their ethnic and race backgrounds, including categories such as Asian (58%), Black or African American (10%), White (8%), Middle Eastern (5%), and 19% of others and individuals not preferring to disclose this information. Gender-wise, 53% of the participants are male, 29% female, and 18% preferred not to say. The technical and academic background of the participants include senior PhD students (26%), junior PhD students (21%), industry experts (14%), postdocs (14%), master students (11%), professors (8%), industry PhD holders (5%), and others (1%). This diversity contributed to the program's richness, enabling varied insights and experiences. The clarity of the program's expectations was well-acknowledged, with the majority finding the matching process between participants effective, reflecting the program's success in fostering compatible and beneficial pairings. Mentorship connections formed a core part of the program's success, with many participants reporting establishing meaningful relationships.

By analyzing ratings from area chairs for each reviewer, we found no significant difference in the scores assigned to mentees and mentors. Since no difference was found, it shows that even inexperienced reviewers were able to provide high-quality reviews with proper guidance and help.

The feedback sessions between the mentors and mentees were a critical component, with many participants finding them instrumental in enhancing their review skills and providing constructive guidance. However, challenges in handling the workload of the reviews for mentees were reported, indicating a need for a possible reduced load of reviews for mentees. Finally, participants were highly willing to re-engage

with the program, with approximately 85% expressing a desire to participate again next year.

Career Mentorship The goal of the career mentorship program is to match mentees with mentors capable of offering guidance on various career-related subjects, such as formulating a comprehensive research plan, engaging in healthcare research within industry, and maintaining a healthy work-life balance. This year, the program included three components: 30-minute individual mentoring sessions, one 1-hour virtual panel discussion, and one 1-hour in-person roundtable discussion. While the group sessions provided mentees with more general career tips, where selected mentors shared their career experiences and lessons learned, the individual mentor-mentee sessions provided a more informal environment for the mentee to ask specific and individualized questions. The in-person roundtable was facilitated by a junior chair by Naveen Raman, a PhD student in the Machine Learning Department at Carnegie Mellon University, and two senior chairs, one from industry and one from academia. The online event consisted of two sessions held at different times of day to accommodate different time zones. Each session featured four mentors representing diverse career backgrounds.

In this year's program, a total of 101 mentees and 46 mentors enrolled and were matched based on their experience and interest. The mentees' backgrounds ranged from high-school seniors to junior faculty, with the greatest proportion of mentees being undergraduates and masters students (34%). Most mentors were industry researchers with PhDs (28%) or academic faulty (26%). While the greatest proportion of participants were based in North America (50 % overall), we also recruited participants from Africa (8%), Western Europe (12%), Asia (20%), Middle East (1%), as well as Central and South America (5%).

Research Roundtables The goal of the research roundtables is to foster smaller group discussions on specialized topics of interest to the ML4H community. In light of the successful reception of the one-day in-person roundtables during ML4H 2023 (Hegselmann et al., 2023a), we introduced new topics and held these sessions along with mentorship roundtables over two days in ML4H 2024. The structure of each roundtable session followed that of ML4H 2022 and ML4H 2023 (Hegselmann et al., 2023a; Parziale et al., 2022), including one invited senior and two invited junior chairs for each topic. Senior chairs were the invited experts in each topic domain who led the sessions, while junior chairs were students who were responsible for moderating them. This year, we hosted 13 roundtables across two separate sessions (8 on the first day, 6 on the second day) with detailed lists of topics and their chairs provided below.

Roundtable Topics and Chairs

1. **Author Mentorship:** What publication venues should I look for when I want to write about interdisciplinary research? What are ways to facilitate research collaborations across people with a medical background and people with a stats/computer science background? What did you learn from the mentorship program? How did it vary across mentors and mentees?
    - Senior Chair: Charles Delahunt
    - Junior Chair: Elizabeth Remfry
2. **Career Mentorship:** How did you decide between academia and industry? What was an important inflection point in your trajectory and how did you navigate it? Day in the life? How did you pick your research direction? What do you think are the ripe topics for early-career researchers to get into?
    - Academia Chair: Samaneh Nasiri
    - Industry Chair: Elise Jortberg
    - Junior Chair: Naveen Raman
3. **Reviewer Mentorship:** What makes a good review and a good reviewer? How do we, as conferences, incentivize reviewers to write quality reviews? What kind of countermeasures are there against low-quality or generated reviews? How does reviewing for a journal and conference differ? What are the differences between reviewing for a specialized ML conference and reviewing for a general ML conference?
    - Senior Chair: Stephen Pfohl
    - Junior Chair: Fabian Gröger
4. **Integrating AI into Clinical Workflows:** What level of evidence is required decide to deploy or retire a model? How do emerging governance structure for AI deployment in hospitals (e.g., Chief AI officers) help these efforts?

- Senior Chair: Adarsh Subbaswamy
- Junior Chairs: Elizabeth Healey, Ayush Noori

5. **Foundation Models and Multimodal AI:** How do we effectively integrate multiple data sources (e.g., Electronic Health Records (EHRs), mHealth devices, images, genomics) for ML applications in healthcare? How does this work in real-time in a hospital? How to assess safety, effectiveness, and trustworthiness of foundation models for health applications? What are some low-hanging fruit opportunities to use large language models in healthcare?

   - Senior Chair: Jason Fries
   - Junior Chairs: Maxwell Xu, Hejie Cui

6. **Causality:** How can recent advances in AI/ML help discover causal relations using clinical data? To what extent can we use observational data to emulate randomized trials, to evaluate the causal effect of any treatment?

   - Senior Chair: Rahul Krishnan
   - Junior Chairs: Jonny Xi, Trenton Chang

7. **Bias/Fairness:** Do clinicians and computer scientists have the same understanding of algorithmic bias and fairness? What are the best practices for continuous evaluation of deployed models? Do foundation models pose a challenge to the assessment of fairness and bias?

   - Senior Chair: Leo Celi
   - Junior Chairs: Alina Peluso, Amin Adibi

8. **Social AI and Healthcare:** Advances in Computer Vision and NLP have led to the development of models capable of describing various aspects of human social communication and behavior. These aspects play a critical role in understanding developmental disorders such as autism and mental health symptoms like suicidal ideation. How can we harness these technological advancements to devise innovative methods for diagnosing and treating mental and developmental conditions, ultimately improving outcomes?

   - Senior Chair: James M. Rehg
   - Junior Chairs: Yurui Cao, Yuwei Zhang

9. **Population Health and Survival Analysis:** Where do we stand with ML's role in population health? How can ML be applied for time-to-event survival analysis?

   - Senior Chair: Ehsan Karim
   - Junior Chairs: Belal Hossain, Hanna Frank

10. **Drug Discovery and Development:** How can Machine learning/LLMs/Foundation models be leveraged to accelerate the identification of potential drug candidates, and what are the challenges in optimizing the drug development pipeline using deep learning models?

    - Senior Chair: Michael Craig
    - Junior Chairs: Zuheng (David) Xu, Geofferey Woollard

11. **Challenges of Interdisciplinary Research:** Doctors and AI experts need to work together to develop Health AI. How can such collaborations form? What is expected from collaborations to make them successful?

    - Senior Chair: Dennis Shung
    - Junior Chair: Vasiliki Bikia, Xi (Nicole) Zhang

12. **Health Economics, Policy, and Reimbursement:** How do we decide whether deploying and maintaining an AI or ML model is cost-effective? Who should pay for AI models?

    - Senior Chair: Ian Cromwel
    - Junior Chair: Shuvom Sadhuka, Ross Duncan

13. **Clinician-AI Interaction:** How can we understand how models are affecting clinical decisions, or clinical processes? What is the current state of clinician-AI interaction research in health AI?

    - Senior Chair: Shannon McWeeney
    - Junior Chair: Yuan Pu, Shannon Zejiang Shen

14. **Personalization and Heterogeneity in Medicine:** How can machine learning models effectively capture patient heterogeneity while ensuring personalized treatment recommendations? What are the key challenges in balancing personalization with generalizability of solutions?

- Senior Chair: Mohsen Sadatsafavi
- Junior Chair: Yuan Xia, Sazan Mahbub

15. **Public Datasets and Benchmarks:** How can we create incentives for institutions to release healthcare datasets for public use? What obstacles are there for releasing health datasets for public use, and how can we mitigate potential obstacles?

    - Senior Chair: Rahmat Beheshti
    - Junior Chair: Adibvafa Fallahpour, John Wu

16. **Health AI in Low- and Middle-Income Countries:** What are the unique challenges and opportunities of AI in low- and middle-income countries? How can we help these countries preparing and preventing epidemics and pandemics? How can AI be adapted to meet such challenges?

    - Senior Chair: Megan Coffee
    - Junior Chair: Wenqian Ye, Brighton Nuwagira

Attendee Matching This year we had a new event during the main conference where we matched people based on their similar research or career interests. During the registration people were asked if they wanted to take part or not of this event. The goal of the match ups was to connect people who otherwise might not have connected based on their interests, due to the large scale of the symposium.

In this year's program, a total of 154 attendees registered. Participants' top three research interests were Natural Language Processing & LLMs, Computer Vision, and Causal Inference & Discovery. 87 participants were students, and 67 participants were not. Participants were notified of their pairing prior to the event and decided on their own how and when to meet each other throughout the event.

2.5. Demonstrations and Perspectives Tracks

For the second year in a row, ML4H offered a Demonstrations (Demos) track (Hegselmann et al., 2023b). With an increasing number of machine learning-based software products receiving regulatory approval as medical devices from authorities such as the U.S. FDA, Chinese NMPA, and the European Commission, there is a critical need for a forum to evaluate these tools in real-world environments. The ML4H Demos track provided a platform to explore the challenges, assumptions, and practical requirements associated with deploying ML4H technologies in clinical and operational settings. Designed to bridge the gap between theoretical proof-of-concept research and tangible, real-world utility, the Demos track showcased practical applications of machine learning technologies in healthcare.

This year, we received 20 submissions to the Demos track. Submissions consisted of a two-page "spec sheet," *i.e.*, a short write-up describing the ML4H tool, technology, and application, as well as a video demonstration of the tool in-use at most two-minutes in length. Submissions were evaluated for relevance to the ML4H field, maturity of the tool or project, significance and impact of the tool, quality and clarity of the submission, and discussion of challenges and lessons learned.

Our review process was single-blind and each submission received two reviews, along with a metareview. In addition to 7 volunteer reviewers from the ML4H organizing committee, we invited authors who submitted to the Demos Track to participate in the review process as well. Eight authors volunteered to provide additional reviews for the Demos Track. All volunteers confirmed that, if assigned a paper for which they have a conflict of interest, they would recuse themselves from reviewing that submission and immediately inform the Demos & Perspectives Committee Chairs. No conflicts of interest were reported.

We accepted 10 submissions to the Demos track, for an acceptance rate of 50%. Importantly, some impressive submissions with high reviewer ratings were rejected because the Demos & Perspectives Chairs, in consultation with the General Chairs, determined that they did not meet the criteria for real-world deployment and evaluation. Accepted submissions were non-archival, and authors were provided the opportunity to present their live demo on the day of the symposium alongside the main poster sessions. We also invited three submissions (15%) to deliver a six-minute spotlight talk at the symposium.

In addition to the Demos track, we also offered a new Perspectives track this year. Submissions for the Perspectives track were accepted by invitation only, with 18 individuals being invited to submit manuscripts. Of these, 3 perspectives were submitted, and 2 were ultimately accepted for publication, resulting in an acceptance rate of 66%.

Submissions were 2-4 pages in length and focused on deployment and foundation models. Each submission underwent a single-blind review process, receiving 2–3 reviews from members of the AHLI board. Submissions were evaluated for relevance, significance, and impact to the ML4H field. Reviewers were instructed to disclose any potential conflicts of interest; one reviewer identified a conflict and was reassigned to another submission.

Perspective authors delivered 15-minute talks at the symposium and participated in a panel discussion on the same day as their talk.

3. Analysis of Accepted Works

3.1. Structured Data Analysis

During submission, authors were asked to categorize their submission into one of three general areas: *Applications and Practice*, *Impact and Society*, and *Models and Methods*. Authors were also asked to further describe their work by selecting at least one specific subject area. Reviewers and area chairs were asked for their expertise in the same areas in a registration survey. Reviewers were allowed to select more than one general area. Note that not all reviewers and area chairs completed the registration survey, with 192 of 277 reviewers (69.3%) and 12 of 26 area chairs (46.2%) completing the survey.

First, we analyze the proportion of submissions by general and specific areas (Figure 1). The vast majority of accepted works self-selected *Applications and Practice* (49.2%, n=65) or *Models and Methods* (43.2%, n=57), with a small fraction of papers self-selecting the *Impact and Society* area (7.6%, n=10). The distribution of accepted papers by area was similar in both Findings and Proceedings track papers.

For specific subject areas, trends varied slightly across Findings and Proceedings track papers (Figure 2). The top areas in the Proceedings track were *Explainability & Interpretability* (11.6%, n=8), followed by a three-way tie between *Generative Models* and *Representation Learning*, and *Other* (8.7%, n=6). In the Findings track, *Supervised Learning* accounted for the most papers (38.7%, n=24), followed by *Representation Learning* (37.1%, n=23) *Generative Models* (33.9%, n=21), *Other* (29.0%, n=18), and *Public & Social Health* (25.8%, n=16). While supervised learning, representation learning, and generative models were popular topics among accepted papers at ML4H 2023, we note an uptick in submissions that marked the *Other* category (Hegselmann et al., 2023b).

As a secondary analysis, we also show the percentages of accepted submissions, reviewers, and area chairs (conditioned on survey completion) in Figure 4, stratified by specific subject areas. Reviewers and area chairs indicated areas of expertise that were broadly well-correlated the distribution of general and specific subject areas in accepted papers.

3.2. Topic Modeling

Similar to previous years (Hegselmann et al., 2023b; Parziale et al., 2022; Roy et al., 2021; Sarkar et al., 2020), we performed topic modeling over the text of the accepted ML4H proceedings and findings papers. The top topics were *Medical Imaging, Pre-trained LLMs on Medical Data, Clinical Datasets,* and *Representation Learning*. Compared to the previous year (Hegselmann et al., 2023b), medical imaging continued to be a prominent topic, and there was a greater proportion of topics which were at the intersection of previous topics and large language models. There was again a decline in explainability research, but an increase in learning from human feedback. These topics are in turn different from previous years (Parziale et al., 2022; Roy et al., 2021; Sarkar et al., 2020), indicating evolving research interests of the ML4H community over the years.

4. The ML4H Community

The community found at the intersection of machine learning and health continues to expand in a number of metrics, including the number of research publications (ML4H submissions increased by 13.6% this year; Section 2.2), interdisciplinary conferences and their attendance, and collaborative initiatives between academia, industry and healthcare institutions. This growth is also reflected in the increasing availability of open-source medical datasets, the rise in funding and investments directed toward AI-driven healthcare solutions, and the adoption of advanced algorithms by major health systems. In particular, the profound impact of machine learning on healthcare has been especially evident in its transformative contributions to protein structure prediction and drug discovery, as highlighted by the achievements of this year's winners of the Nobel Prize in Chemistry (Callaway, 2024).

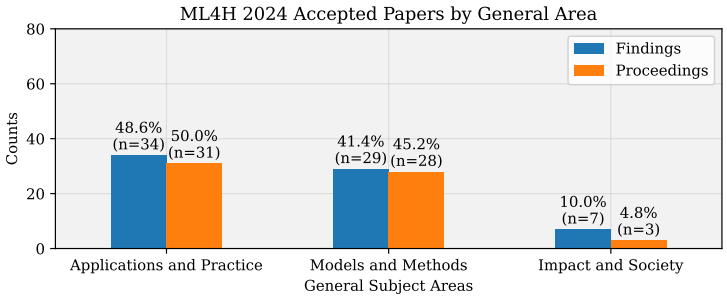

Figure 1: ML4H 2024 accepted papers by self-reported general subject area, split by track (Findings and Proceedings). Percentages computed with respect to the total # of accepted papers in each track.

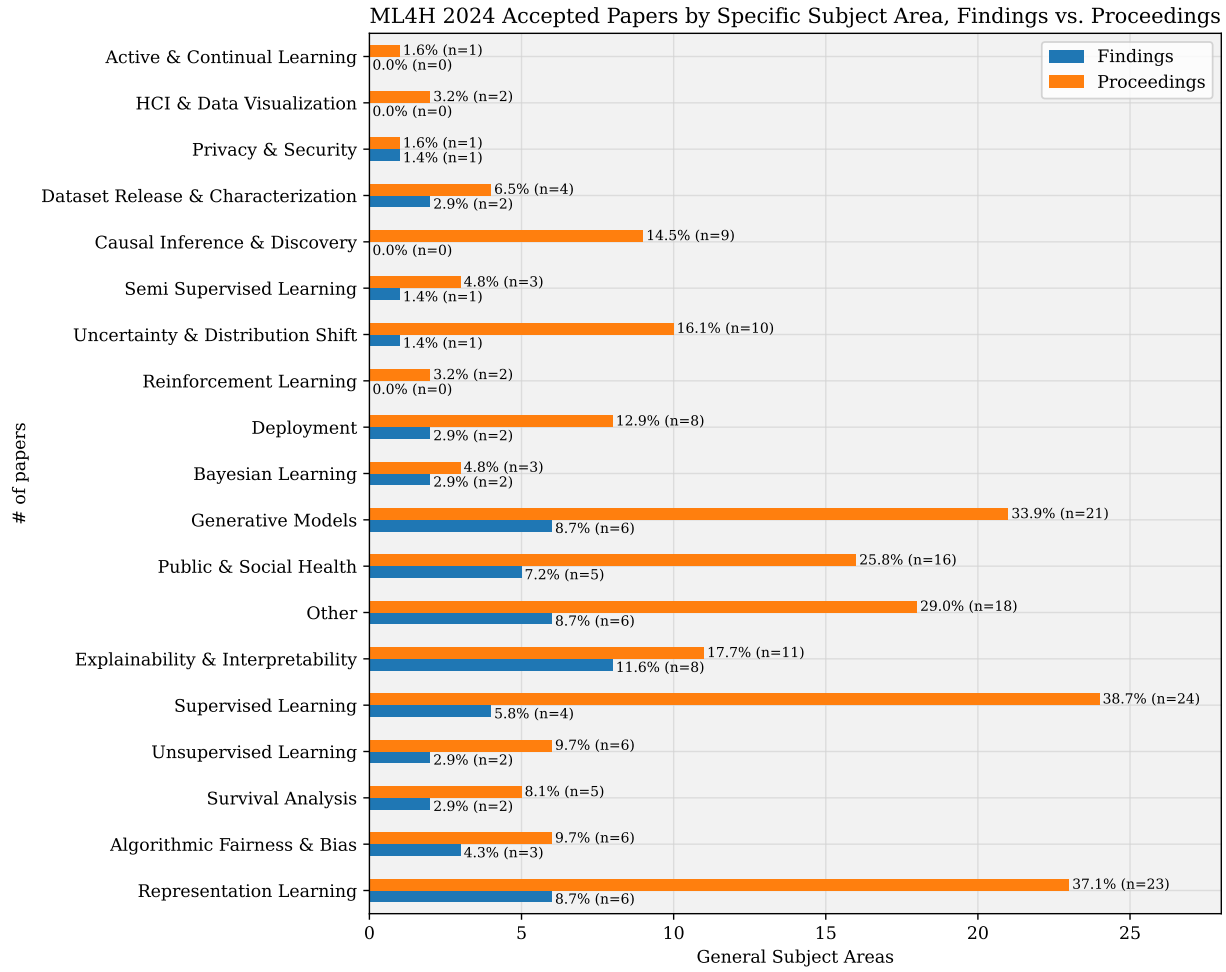

Figure 2: ML4H 2024 accepted papers by self-selected specific subject areas, split by track (Findings, blue vs. Proceedings, orange). Percentages computed with respect to the total # of accepted papers in each track. Total percentages may not sum to 100.0% because authors are permitted to select multiple values.

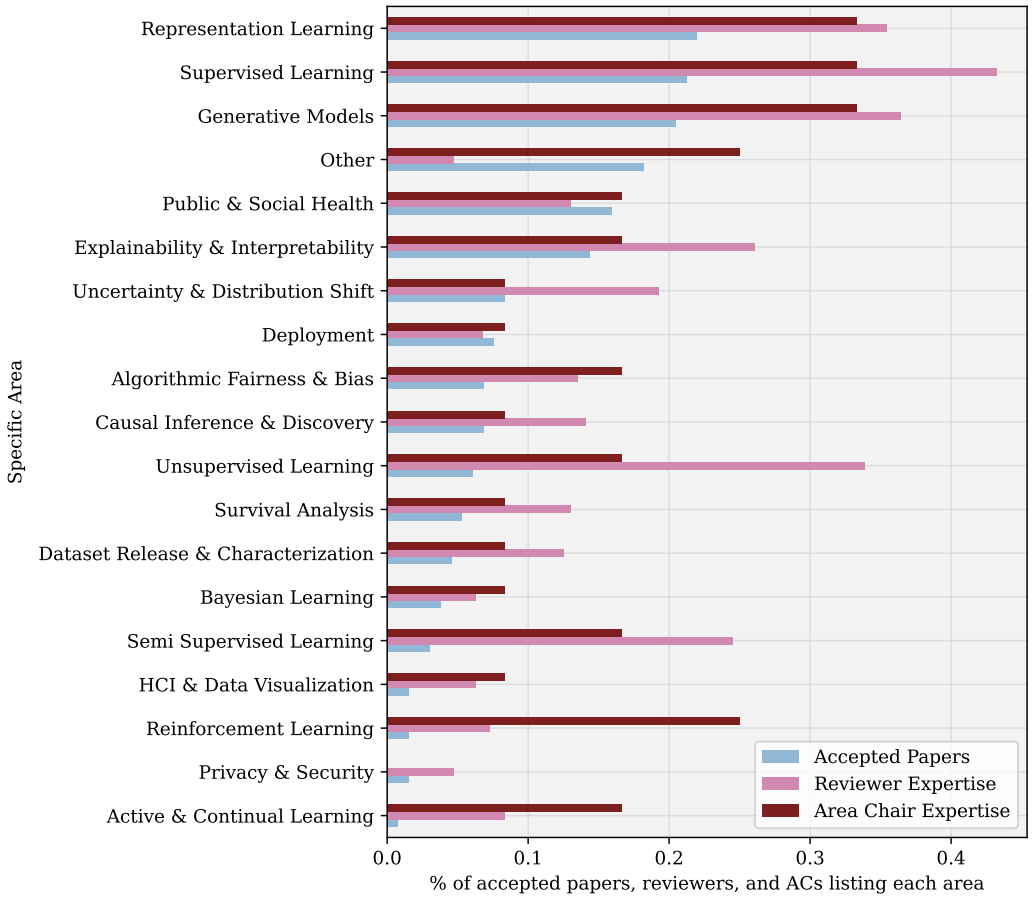

Figure 3: ML4H 2024 accepted papers, reviewer, and area chair expertise by self-selected specific area.

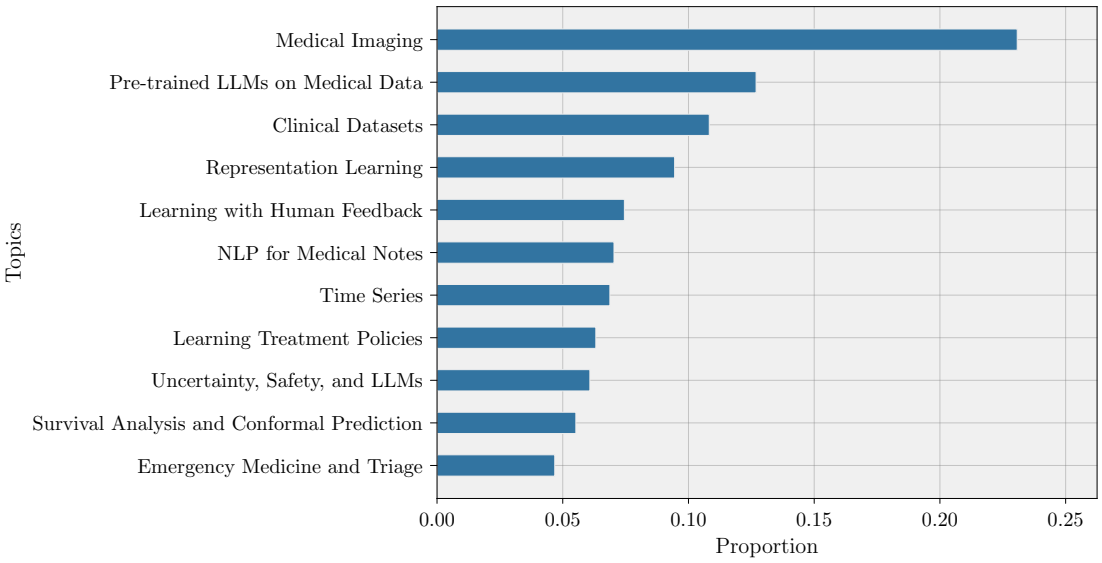

Figure 4: LDA marginal topic distribution of accepted works in ML4H 2024.

This year, following last year's approach, ML4H was an in-person only conference. ML4H expanded for the first time ever, into a 1.5 day conference. As of December 2nd, 2024, 232 attendees have registered for the event. One of the goals for the conference was to build a broad, accessible, inclusive, and engaged community. In order to evaluate this, we asked several optional questions on the registration form.

4.1. Demographic Distribution

Of the 83.2% of registrants who indicated their country of residence, 72% reside in North America, 17.1% reside in Europe, 7.2% reside in Asia, and the remainder from Africa, Oceania and South America. The country with the highest representation is the United States of America, followed by Canada, the United Kingdom, and Switzerland.

Of the 76.7% of registrants who provided their age group, 47.1% are between the ages of 21 and 30, 35.1% are between 31 and 40, and the remainder are older than 40 and younger than 20. Of those who specified their gender, 60.3% identified as Male and 35.9% identified as Female. Of those who provided their race, 43.4% identified as Asian or Asian American, 34.1% identified White (non-Hispanic), and the remainder identified as Middle Eastern or North African, Black or African American, and Hispanic or Latino/a/x.

4.2. Background & Experience

The ML4H symposium aims to attract participants from varying educational and professional backgrounds. Similar to previous years, we found that the majority of participants (79.2%) identified *Machine Learning / Computer Science* as their primary community. 16% selected *Health / Medicine*, 3.3% selected *Informatics*, and 1.5% selected *Other*. From previous years, we saw a 30% increase in registrants who stated *Health / Medicine* as their primary community. This year, 54.1% of attendees are students. The background of the participants are 79.2% in academia, 15.6% in industry, 2.4% in clinical practice, and 1.4% in government.

5. Conclusions

The field of machine learning for healthcare continues to grow and mature. The fourth Machine Learning for Health Symposium, ML4H 2024, served as an important hub for collaboration and innovation in this interdisciplinary field. This year, the symposium expanded its scope with new tracks and programs, including a broadened mentorship initiative and the introduction of a Perspectives Track, showcasing critical insights into foundational and applied research. With a 13.6% increase in submissions and a diverse range of accepted works, ML4H 2024 highlighted the field's sustained growth and relevance. Keynote talks and thematic sessions tackled pressing challenges, such as the deployment of medical AI and the integration of foundation models in healthcare. The symposium also fostered active engagement through poster sessions, research roundtables, and live demonstrations of real-world ML applications. ML4H's ongoing commitment to mentorship and inclusivity is reflected in the participation of over 300 mentees and mentors across various programs, underscoring its role in cultivating the next generation of leaders in machine learning for health. As the field matures, ML4H continues to provide a critical forum for cutting-edge research, collaboration, and discourse at the intersection of machine learning and healthcare.

6. Acknowledgements

6.1. AHLI Board

The organizers thank the Board of the Association for Health Learning and Inference (AHLI) for its support in the organization of the symposium. A special thanks goes to Tasmie Sarker, who helped in organization as well as coordinated between the organizing committee and the AHLI Board.

6.2. Area Chairs

The organizers thank all area chairs for their expert guidance in reviewing submissions to ML4H: Adji Bousso Dieng, Ayah Zirikly, Brett Beaulieu-Jones, Chandan Singh, Charles B. Delahunt, Collin Stultz, Edward Choi, Haoqi Sun, Irene Y. Chen, Jack Gallifant, Jimeng Sun, Masoud Rouhizadeh, Monica Agrawal, Neel Dey, Prince Ebenezer Adjei, Reza Sadeghi, Samantha Kleinberg, Sana Tonekaboni, Shengpu Tang, Stephen R Pfohl, Thomas Hartvigsen, Tristan Naumann, Vivek Kumar Singh, Weimin Zhou, Yogatheesan Varatharajah, and Zepeng Frazier Huo.

6.3. Reviewers

The proceedings and discussions would not have been possible without the generous reviewers who provided meaningful feedback and evaluation of all submissions. We are grateful to all reviewers: Aahlad Manas Puli, Abdullah Mamun, Abhijeet Parida, Adrien Coulet, Aishwarya Mandyam, Ajinkya K Mulay, Akshata Kishore Moharir, Akshay Goel, Alan Kaplan, Alejandro Lozano, Alex Fedorov, Alexej Gossmann, Ameya Wagh, Amin Adibi, Amy X. Lu, Ananya Joshi, Andrew Julio Barros, Anna Zapaishchykova, Antonio Parziale, Anurag Jayant Vaidya, Ari Ercole, Arinbjörn Kolbeinsson, Arjun Sondhi, Arsh Verma, Aryan Jadon, AS Soares, Atif Khan, Benjamin C Warner, Bhawesh Kumar, Çağlar Hızlı, Cécile Trottet, Cédric Wemmert, Chang Hu, Chen Zhang, Chris Brown, Christina X Ji, Christopher YK Williams, Daeun Kyung, Dang Nguyen, Daniel Capellán-Martín, David A Bloore, David Remy Bellamy, David Zimmerer, Dhruv Naik, Dibyendu Roy Chowdhury, Divya D. Kulkarni, Dmytro Lopushanskyy, Dokyoon Kim, Dominik Meier, DongHyeon Seo, Donna Tjandra, Ece Ozkan, Eike Petersen, Elise Jortberg, Elliot Creager, Elliot Schumacher, Eloise Berson, Emmanuel Klu, Ethan Harvey, Fabian Gröger, Fang Wu, Fanyang Yu, Felipe Parodi, Frank Rudzicz, Giovanni Cinà, Guanjie Huang, Gyubok Lee, Hamed Fayyaz, Hammaad Adam, Han Yu, Hao He, Haocheng Bi, Haolin Yang, Haoran Zhang, Hardy, Hariharan Ragothaman, Harri Lähdesmäki, Heike Leutheuser, Hiba Ahsan, Houliang Zhou, Hugo Figueiras, Hyewon Jeong, HyunGeun Lee, Hyungyung Lee, Ilker Demirel, Irene Y. Chen, Ivana K. Parker, Jarett Dewbury, Jason Poulos, Jennifer Williams, Jennifer Yu, Jia Guo, Jiaming Zeng, Jie Lian, Jiechao Gao, Jieshi Chen, Jing Su, Jinpei Han, Joel Persson, John Michael Giorgi, Jong Hak Moon, Jongoh Jeong, José D. Posada, Jun-En Ding, Jungwoo Oh, Junyi Gao, Justin Cosentino, Juyong Kim, Kaiwen Zha, Karim Kadry, Kathryn Wantlin, Katie Matton, Kazuma Kobayashi, Kenza Bouzid, Kuancheng Wang, Kumail Alhamoud, Kwan Ho Ryan Chan, Laya Rafiee Sevyeri, Leopoldo Julian Lechuga Lopez, Lorenzo A. Rossi, Lovedeep Gondara, Lucía Prieto Santamaría, Luna Zhang, Macton Mgonzo, Malvern Madondo, Mantek Singh, Manuel Burger, Marie-Laure Charpignon, Marta Avalos, Martin Meinel, Maryam Daniali, Marzia Hoque Tania, Matthew Eric Lee, Matthew M. Engelhard, Matthew Peroni, Max Homilius, Md Osman Gani, Meera Krishnamoorthy, Megan Coffee, Meghasyam Tummalacherla, Mehul Motani, Miao Zhang, Mikio Tada, Milos Vukadinovic, Min Hun Lee, Minjae Lee, MohammadAli Shaeri, Mollie Marian Mckillop, Monica Munnangi, Mozhgan Saeidi, Muhammad Ali, Muhammad Muneeb Afzal, Munib Mesinovic, Nandita Damaraju, Nassim Oufattole, Natalie R Davidson, Neeraj Wagh, Nikita Mehandru, Nimshi Venkat Meripo, Nivetha Jayakumar, Onintze Zaballa, Onur Poyraz, Pankaj Pandey, Pankhi Kashyap, Peiqi Wang, Peng Cao, Phyllis M Thangaraj, Prajwal Kailas, Preetish Rath, Prince Mireku, Priyanka Mary Mammen, Purity Mugambi, Qinfeng Zhu, Qixuan Jin, Rafal Dariusz Kocielnik, Rahmatollah Beheshti, Ričards Marcinkevičs, Ridwan Alam, Robin Khatri, Roozbeh Yousefzadeh, Roy Adams, Rudraksh Tuwani, Ryan Andrew Chi, Ryan King, Ryan Schuerkamp, Saeed Alahmari, Saksham Bhutani, Salvatore Tedesco, Sam Freesun Friedman, Samyak Jain, Sandeep Angara, Sandhya Tripathi, Sarah Brueningk, Sasikanth Kotti, Saurav Sengupta, Seongsu Bae, Serina Chang, Seung Eun Yi, Shahriar Noroozizadeh, Shalini Saini, Shayan Fazeli, Shenao Yan, Shuo Shuo Liu, Sicong Huang, Siddhant shah, Sidhika Balachandar, Siun Kim, Siyi Tang, Soumyadeep Roy, Srikar Katta, Stefan Groha, Stefan Hegselmann, Stephanie Hyland, Sumedha Rai, Sumit Mukherjee, Sunny Gupta, Supriya Nagesh, Tamilarasan Kannadasan, Tessa Han, Tiffany Callahan, Tomas Bosschieter, Tong Xia, Tonmoy Hossain, Toon Vanderschueren, Tzu-Ting Wei, Utsha Saha, Vaibhav Ganatra, Vajratiya Vajrobol, Vasco Prudente, Vasundhara Agrawal, Vin Bhaskara, Vincent Jeanselme, Vishwas Suryanarayanan, Vivek Shankar, Walter Gerych, Walter Nelson, Wangzhi Dai, Wenbin Zhang, Wenfei Fang, Willa Potosnak, William Boag, William La Cava, Wouter A.C. van Amsterdam, Xiao Fan, Xiaobin Shen, Xiaochen Zheng, Xiaomeng Dong, Xueqiao Peng, Yan Gao, Yasar Mehmood, Ye Jin, Ye Zhang, Yefan Tao, Yeonie Heo, Yihang Tao, Yike Zhang, Yiwen Chen, Yoav Wald, Yogesh Kumar, You Cheng, Yousef Nademi, Yuan Zhao, Yurui Cao, Yuta Kobayashi, Yuwen Chen, Yuxin Qiao, Zhe Huang, Zhenbang Wu, Zhengjian Kang, Zhi Yu, Zhichao Yang, Zhiwei Zhang, Zhongyuan Liang, Zijian Zhang, Zixuan Wang, Ziyang Song, and Zongliang Ji.

References

Ewen Callaway. Chemistry nobel goes to developers of alphafold ai that predicts protein structures. *Na-*

ture, 634:525–526, 10 2024. ISSN 0028-0836. doi: 10.1038/d41586-024-03214-7.

Stefan Hegselmann, Antonio Parziale, Divya Shanmugam, Shengpu Tang, Kristen Severson, Mercy Nyamewaa Asiedu, Serina Chang, Bonaventure F. P. Dossou, Qian Huang, Fahad Kamran, Haoran Zhang, Sujay Nagaraj, Luis Oala, Shan Xu, Chinasa T. Okolo, Helen Zhou, Jessica Dafflon, Caleb Ellington, Sarah Jabbour, Hyewon Jeong, Harry Reyes Nieva, Yuzhe Yang, Ghada Zamzmi, Vishwali Mhasawade, Van Truong, Payal Chandak, Matthew Lee, Peniel Argaw, Kyle Heuton, Harvineet Singh, and Thomas Hartvigsen. Machine learning for health (ml4h) 2023. In Stefan Hegselmann, Antonio Parziale, Divya Shanmugam, Shengpu Tang, Mercy Nyamewaa Asiedu, Serina Chang, Tom Hartvigsen, and Harvineet Singh, editors, *Proceedings of the 3rd Machine Learning for Health Symposium*, volume 225 of *Proceedings of Machine Learning Research*, pages 1–12. PMLR, 10 Dec 2023a. URL https://proceedings.mlr.press/v225/hegselmann23a.html.

Stefan Hegselmann, Antonio Parziale, Divya Shanmugam, Shengpu Tang, Kristen Severson, Mercy Nyamewaa Asiedu, Serina Chang, Bonaventure FP Dossou, Qian Huang, Fahad Kamran, et al. Machine learning for health (ml4h) 2023. In *Machine Learning for Health (ML4H)*, pages 1–12. PMLR, 2023b.

Antonio Parziale, Monica Agrawal, Shengpu Tang, Kristen Severson, Luis Oala, Adarsh Subbaswamy, Sayantan Kumar, Elora Schoerverth, Stefan Hegselmann, Helen Zhou, Ghada Zamzmi, Purity Mugambi, Elena Sizikova, Girmaw Abebe Tadesse, Yuyin Zhou, Taylor Killian, Haoran Zhang, Fahad Kamran, Andrea Hobby, Mars Huang, Ahmed Alaa, Harvineet Singh, Irene Y. Chen, and Shalmali Joshi. Machine learning for health (ml4h) 2022. In Antonio Parziale, Monica Agrawal, Shalmali Joshi, Irene Y. Chen, Shengpu Tang, Luis Oala, and Adarsh Subbaswamy, editors, *Proceedings of the 2nd Machine Learning for Health symposium*, volume 193 of *Proceedings of Machine Learning Research*, pages 1–11. PMLR, 28 Nov 2022. URL https://proceedings.mlr.press/v193/parziale22a.html.

Subhrajit Roy, Stephen Pfohl, Girmaw Abebe Tadesse, Luis Oala, Fabian Falck, Yuyin Zhou, Liyue Shen, Ghada Zamzmi, Purity Mugambi, Ayah Zirikly, Matthew B. A. McDermott, and Emily Alsentzer. Machine learning for health (ml4h) 2021. In Subhrajit Roy, Stephen Pfohl, Emma Rocheteau, Girmaw Abebe Tadesse, Luis Oala, Fabian Falck, Yuyin Zhou, Liyue Shen, Ghada Zamzmi, Purity Mugambi, Ayah Zirikly, Matthew B. A. McDermott, and Emily Alsentzer, editors, *Proceedings of Machine Learning for Health*, volume 158 of *Proceedings of Machine Learning Research*, pages 1–12. PMLR, 04 Dec 2021. URL https://proceedings.mlr.press/v158/roy21a.html.

Suproteem K. Sarkar, Subhrajit Roy, Emily Alsentzer, Matthew B. A. McDermott, Fabian Falck, Ioana Bica, Griffin Adams, Stephen Pfohl, and Stephanie L. Hyland. Machine learning for health (ml4h) 2020: Advancing healthcare for all. In Emily Alsentzer, Matthew B. A. McDermott, Fabian Falck, Suproteem K. Sarkar, Subhrajit Roy, and Stephanie L. Hyland, editors, *Proceedings of the Machine Learning for Health NeurIPS Workshop*, volume 136 of *Proceedings of Machine Learning Research*, pages 1–11. PMLR, 11 Dec 2020. URL https://proceedings.mlr.press/v136/sarkar20a.html.