Machine Learning for Health (ML4H) 2023

Stefan Hegselmann[1] STEFAN.HEGSELMANN@UNI-MUENSTER.DE
Antonio Parziale[2] ANPARZIALE@UNISA.IT
Divya Shanmugam[3] DIVYAS@MIT.EDU
Shengpu Tang[3] SHENGPU.TANG@GMAIL.COM
Kristen Severson[4] KSEVERSON@MICROSOFT.COM
Mercy Nyamewaa Asiedu[5] MASIEDU@GOOGLE.COM
Serina Chang[5] SERINAC@STANFORD.EDU
Bonaventure F. P. Dossou[6] BONAVENTURE.DOSSOU@MILA.QUEBEC
Qian Huang[6] QHWANG@CS.STANFORD.EDU
Fahad Kamran[7] FHDKMRN@UMICH.EDU
Haoran Zhang[7] HAORANZ@MIT.EDU
Sujay Nagaraj[8] S.NAGARAJ@MAIL.UTORONTO.CA
Luis Oala[9] LUIS.OALA@DOTPHOTON.COM
Shan Xu[9] XUSHAN@CAICT.AC.CN
Chinasa T. Okolo[10] CHINASA@CS.CORNELL.EDU
Helen Zhou[10] HLZHOU@ANDREW.CMU.EDU
Jessica Dafflon[11] JESSICA.DAFFLON@GMAIL.COM
Caleb Ellington[11] CELLINGT@CS.CMU.EDU
Sarah Jabbour[11] SJABBOUR@UMICH.EDU
Hyewon Jeong[11] HYEWONJ@MIT.EDU
Harry Reyes Nieva[11] HARRY.REYES@COLUMBIA.EDU
Yuzhe Yang[11] YUZHE@MIT.EDU
Ghada Zamzmi [11] ALZAMZMIGA@NIH.GOV
Vishwali Mhasawade[12] VISHWALIM@NYU.EDU
Van Truong[12] HEYVAN@UPENN.EDU
Payal Chandak[13] CHANDAK@MIT.EDU
Matthew Lee[13] MATTHEW.LEE1@PENNMEDICINE.UPENN.EDU
Peniel Argaw[14] PENIEL@G.HARVARD.EDU
Kyle Heuton[14] KYLE.HEUTON@TUFTS.EDU
Harvineet Singh[15] FNU.HARVINEETSINGH@UCSF.EDU
Thomas Hartvigsen[15] HARTVIGSEN@VIRGINIA.EDU

Organizing committee for ML4H 2023
[1] Publications Chair, [2] Publications Subchair, [3] Workflow Chair, [4] Workflow Subchair, [5] Program Chair, [6] Program Subchair, [7] Logistics Chair, [8] Logistics Subchair, [9] Demo Chair, [10] Outreach Chair, [11] Outreach Subchair, [12] Communication Chair, [13] Communication Subchair, [14] Finance Chair, [15] General Chair

1. Introduction

The third Machine Learning for Health (ML4H) symposium was held on December 10, 2023, in New Orleans, Louisiana, USA. Following the last two years (Roy et al., 2021; Parziale et al., 2022), the symposium was again a stand-alone event co-located with the Neural Information Processing Systems (NeurIPS) conference.

ML4H 2023 invited high-quality submissions on relevant problems in a variety of health-related disciplines including healthcare, biomedicine, and public health. Two submission tracks were offered: the archival *Proceedings* track, and the non-archival *Findings* track. Proceedings were targeted at mature work with strong technical sophistication and a high impact to health. The Findings track looked for new ideas that could spark insightful discussion, serve as

valuable resources for the community, or could enable new collaborations. Submissions to the Proceedings track, if not accepted, are automatically considered for the Findings track. Accepted publications of both tracks were given a platform for presentation and exchange through an in-person poster session. This year, there were 136 submissions to the Proceedings and 76 submissions to the Findings track. Of these, the organizing committee accepted 39 Proceedings and 71 Findings.

In addition to the submission tracks, ML4H 2023 offered a rich program to connect researchers and stimulate discussions. Mentorship programs for authors, reviewers, and career advice were meant to bring together less experienced members of the community with senior researchers to provide exchange, advice, and feedback. Research roundtables were hosted at the symposium to allow a platform for discussion about timely topics in machine learning for health. Two optional thematic sessions *Machine learning for health equity and global health* and *Generative AI for health: the road ahead* were created to highlight work on these relevant fields. Awards for best Proceedings, Findings, Thematic, and Newcomer papers were nominated to highlight outstanding work submitted to the symposium. Also, a *Demonstrations* track was held during the event to showcase existing tools and projects that translate cutting-edge research into real-world applications.

In this front matter, we provide an overview of the ML4H 2023 symposium including the paper selection process, submission statics, and the entire program (Section 2). Next, we will provide an analysis of the accepted works and provide trends in the research field based on analyses of the previous ML4H workshops and symposiums (Sarkar et al., 2020; Roy et al., 2021; Parziale et al., 2022) (Section 3). In Section 4, we comment on the current statistics of the ML4H community. Lastly, we close with acknowledgments, including a list of area chairs and reviewers for ML4H 2023.

2. Symposium

2.1. Program

Our program at ML4H 2023 included eight invited talks from experts in fields related to our two thematic sessions. The program was divided across the two themes, each session beginning with a 30-minute invited keynote, followed by 20-minute talks by three panelists and a 45-minute panel discussion. The panel discussions included the keynote speaker as well as the panel speakers.

The first thematic session was *Global Health and Health Equity*. As artificial intelligence, and more specifically machine learning (ML), and generative AI continue to gain ground, research studies have found that the models perpetuate unequal treatment, biases, and stereotypes that especially affect groups who are not largely represented in the training data or involved in the ML pipeline. These groups tend to either be from minority and disadvantaged groups in high income countries or from the global south. For ML to be beneficial for everyone, it is important to address these limitations inclusively. Hence we invited papers focusing on these areas as well as speakers with experience working on ML for global health and health equity generally. The session began with a keynote talk by Elaine Nsoesie. It featured panel speakers Emma Pierson, Milind Tambe, and Charles Delahunt. During this session, we discussed machine learning efforts to promote health equity and global health, as well as investigations of bias, fairness, and equity in machine learning models when applied to health, both locally and globally.

The second session was *Generative AI for Health: the Road Ahead*. In recent months, with Open AI's release of GPT 3.5, and Google's release of Med-Palm large language models have gained popularity in health, with applications ranging from question answering tasks, to multimodal tasks such as radiology report interpretation. We invited papers focusing on this area as well as expert speakers on this area. The session began with a keynote talk by Prof. Marinka Zitnik. It featured panel speakers Monica Agrawal, Stefan Harrer, and Tao Tu. This session focused on the use and development of generative AI (such as large language models) for key problems in health, where we discussed opportunities alongside challenges, risks, and ethical questions.

In addition to speaker sessions, we also held two lightning talk sessions for authors of accepted papers to present their work; two poster sessions for authors to present and attendees to ask questions; demonstrations that were concurrent with the poster sessions (Section 2.6); announcement of Best Paper Awards and datathon winners (Section 2.3); and research roundtables near the end of the day (Section 2.5). Finally, the day concluded with an evening social at Napoleon House, where ML4H 2023 partici-

pants could network, continue their discussions, and enjoy a social atmosphere.

We structured the program to be heavier on presentations in the morning and afternoon, followed by more open and mobile sessions (the second poster session, research roundtables, and evening social), so that the interesting content presented earlier in the day could open up into 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 212 abstract submissions and 198 full submissions (Proceedings and Findings track) to the ML4H Symposium, constituting an over 40% increase in submission volume compared to the prior year (Parziale et al., 2022). This underscores the growing respect for ML4H as a venue to discuss state-of-the-art technology at the intersection of machine learning and health.

The program committee consisted of 19 area chairs and 182 reviewers who completed a total of 505 reviews. Each Proceedings submission received at least three high-quality reviews, while each Findings submission received at least two.

Out of the 136 submissions to the Proceedings track, 39 were accepted to appear in the ML4H 2023 proceedings (acceptance rate of 28.7%). We allowed reviewers to recommend transfers of papers from the Proceedings track to the non-archival Findings track. 31 Proceedings submissions were accepted to the Findings track (22.8% transfer rate). Out of the 76 submissions to the Findings track, 40 were accepted (acceptance rate of 52.6%). As a result, there were 71 submissions accepted to the Findings track.

2.3. Paper Awards

The General Chairs and Program Committee (Chairs and Subchairs) worked together to select the paper awards. Our awards consisted of:

- Best Proceedings paper (one winner and one honorable mention): any paper accepted in the Proceedings track was eligible,
- Best Findings paper (one winner and one honorable mention): any paper accepted in the Findings track was eligible,
- Best Newcomer paper: any paper where the first author was submitting to ML4H for the first time was eligible,
- Best Thematic paper(s): any paper that was relevant to the themes was eligible. For the *Generative AI for Health* theme, we included any paper where the authors chose the topic "Generative AI" as one of their paper's related topics when submitting their paper. For the *Global Health and Health Equity* theme, we included any paper where the authors chose the topic "Public & Social Health", but since this topic was less of a perfect fit for the theme, the Workflow Chairs additionally went through the accepted papers and marked those that were relevant to the theme.

We began by gathering a subset of papers that were accepted in the Proceedings track and Findings track. For each track, we ordered the papers by their average score from reviewers (in descending order). Then, we used the score of the 10th paper in this ordering as our cutoff and kept all papers that scored above the cutoff. In the Proceedings track, the cutoff score was 4.00 and 12 papers received an average score of at least 4.00. In the Findings track, the cutoff score was 3.75 and 11 papers received an average score of at least 3.75. Finally, we also included all papers that were marked as "Notable" by area chairs. This resulted in 14 papers in total to review in the Proceedings track and 15 papers in total to review in the Findings track.

For each of these papers, all six committee members—two General Chairs, two Program Chairs, and two Program Subchairs—reviewed the paper and its reviews. The papers were evaluated on the basis of similar criteria to the original reviews, with emphasis placed on significance, i.e., how will this work guide further development in the field of machine learning for health? Based on these evaluations, we chose the Best Proceedings winner and honorable mention from the Proceedings subset of papers, the Best Findings winner and honorable mention from the Findings subset of papers, and the Best Newcomer paper as the best paper across the two tracks that was eligible for the Newcomer award. For the Best Thematic paper awards, we further evaluated each of these papers in terms of their relevance to the theme and significance within the theme's subfield. Based on these criteria, we chose one Best Thematic paper for each of the two themes. We announced the award winners on the day

of the event, in the afternoon of the program before the second poster session.

2.4. Mentorship Programs

This year, we ran three mentorship programs: (1) author mentorship, (2) reviewer mentorship, and (3) career mentorship. The author mentorship program ran until the paper deadline, followed by the reviewer mentorship program, which ran until the review period was over, followed by the career mentorship program, which occurred between the review period and the symposium date. The overarching goal of these programs was to provide hands-on guidance for valuable skills such as paper writing and reviewing, as well as to share the knowledge and experience of members within the ML4H community to make the field of machine learning for healthcare more accessible.

Author Mentorship To foster current and future collaborations while also improving the overall quality of submissions, the ML4H author mentorship program focuses on pairing senior researchers with less experienced authors. For ML4H 2023, only those with existing paper submission ideas and the intention of submitting to the Findings or Proceedings track were eligible to participate as mentees. During the application process, mentors and mentees were asked about their primary and secondary areas of research and then matched based on shared interests. Mentorship occurred throughout various stages of the research process, with mentors providing feedback and suggestions for improvement on aspects such as research design direction, existing and potential models and experiments, analysis of results, and the presentation and organization of the final paper submission.

The Author Mentorship Program began to solicit applications in May 2023 and accepted applications on a rolling basis through August 2023, just prior to the ML4H submission deadline. Participants were encouraged to have bi-weekly one-hour meetings to discuss progress (e.g., mentors and mentees paired during the first round of matches in early June would have approximately four one-hour meetings over the course of eight weeks).

In total, 54 mentees and 23 mentors participated in the program. Among mentees, 50% identified as male, 46% as female, and 4% chose not to disclose their gender. With respect to race or ethnicity, 39% of mentees identified as Asian; 30% as Black, African, or African American; 13% as white; 7% were Middle Eastern or North African; 4% identified with 2 or more races; 0.5% were Hispanic or Latinx; and 4% chose not to disclose their race or ethnicity. With respect to training status, 52% were PhD students, 39% of mentees were undergraduate or master's students, 6% were post-doctoral fellows, and 4% were professors with expertise in fields other than machine learning.

To evaluate the Author Mentorship Program and drive future improvements, participants were asked at the conclusion of the program to complete an online survey to share feedback regarding their experiences. Most mentors and mentees responded to the survey positively. The majority rated their experience in the program as either very good or excellent, with a median score of 4 based on a 5-point Likert scale ranging from 1 (poor) to 5 (excellent). Most respondents also stated that they were either likely or very likely to participate in the program the following year, with a median score of 4 among mentees and 5 among mentors based on a 5-point Likert scale ranging from 1 (very unlikely) to 5 (very likely). Primarily, mentors considered mentor-mentee matching as the main roadblock to a successful experience (69%), while mentees instead most commonly cited time constraints (71%) as the major roadblock. Based on participant feedback, in subsequent years, the Author Mentorship program will seek to start the program earlier in the calendar year as well as recruit mentors with a more diverse set of research expertise in order to meet matching demands and the interests of mentees.

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, who are graduate students, through a systematic approach that includes the evaluation of 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's 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 not only improve the immediate qual-

ity of reviews but also foster a culture of continuous learning and collaboration.

This year, 111 individuals applied for the program, and 97 confirmed their participation through OpenReview. We succeeded in pairing 35 mentors with 62 mentees. The mentors, who had at least a doctoral degree with significant review and publication experience, voluntarily took on the role of providing guidance. Their expertise ensured that the match with mentees was based on shared research interests and appropriate levels of experience, and they also had input on the number of mentees they preferred to mentor. In addition to the post-completion survey for retrospective feedback, we also utilized a check-in form that provided actionable insights throughout the program.

Participants represented a diverse cross section of the ML4H community, with their ethnic and race backgrounds including categories such as Asian (50%), North African and African American (22%), Middle Eastern (5%), Hispanic (1%), and other (4%). Gender-wise, 60% of the participants are male, 37% female, and 3% prefer to not say. The technical and academic background of the participants include senior PhD students (33%), junior PhD students (24%), postdoc (6%), professor (5%), industry PhD holders (9%), industry with 10 years experience in ML (9%), master students (12%), and government PhD holders (2%). This diversity contributed to the richness of the program, enabling a confluence of 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 the establishment of meaningful relationships.

The feedback sessions between the mentors and mentees were a critical component, with a large portion of participants finding them instrumental in enhancing their review skills and providing constructive guidance. However, communication challenges were reported, indicating a need for improved interaction strategies and better communication of expectations. Finally, the willingness of participants to re-engage with the program was high, with approximately 89% 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 the industry, and maintaining a healthy work-life balance. This year, the program included two components: a one-hour group mentoring workshop and a twenty-minute individual mentoring session. While the one-hour group mentoring 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 individual questions. In this year's program, a total of 122 mentees and 42 mentors enrolled and were matched based on their experience and interest. We also honored specific requests from the mentees who asked to be paired with mentors with similar experiences (e.g., how to transition from a medical doctor background to machine learning) or who spoke the same language. The mentees' backgrounds ranged from high-school seniors to professors, with most mentees being Master/Undergrad students (31%) or PhD students 3rd year or higher (27%). Similarly, most mentors were PhD students in their 3rd year or higher (28%) or on industry jobs after having completed a PhD (21%). While the home institution of the mentors was mainly located in North America (66%), most of the mentees were from North America (48%), Africa (22%), and Asia (20%).

2.5. 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 hybrid roundtables during ML4H 2022 (Parziale et al., 2022), we continued these sessions in an in-person format for ML4H 2023. We adopted the topics from the previous year and added descriptions to the topics. The structure of a roundtable session followed that of ML4H 2021 and ML4H 2022 (Roy et al., 2021; Parziale et al., 2022), including invited senior and 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 11 roundtables, with detailed lists of topics and their chairs provided in section 2.5.

Roundtable Topics and Chairs

1. **Health AI Collaborations, Deployment, and Regulation:** One of the barriers to deploying AI models in healthcare is the ability to safely and effectively integrate models into clinical workflows. What are different factors one should consider in presenting AI models to clinicians that result in effective clinician-AI collaborations, and how do we know if these models truly have a significant impact within the healthcare setting? What are the desires of caregivers and clinicians, and what aspects are still lacking? Furthermore, there has been an increased focus on AI regulation by policymakers and industry players in the last few years. How do we ensure all stakeholders are considered in AI policy, and who should be in charge of writing such regulation, if at all?
    - Senior Chair: Jason Fries, Parisa Rashidi
    - Junior Chair: Rahul Thapta, Hussein Mozannar
2. **Integrating AI into Clinical Workflows**: The rate of AI progress in the last few years seems to have major implications for the types of models we train for healthcare purposes. With the types of models we train constantly changing, how can we develop model-agnostic methods to integrate AI into clinical workflows?
    - Senior Chair: Brett Beaulieau-Jones
    - Junior Chair: William Jongwon Han, Nikita Mehandru
3. **Health AI Foundation Models**: Foundation models train on large amounts of data, and there might be benefits to combining the data from multiple sources (hospitals) and over-training separate foundation models for each hospital. What are the pros and cons of doing so?
    - Senior Chair: Matthew McDermott
    - Junior Chairs: Michael Wornow, Vlad Lialin
4. **Large Language Models and Healthcare**: What are some low-hanging fruit opportunities to use large language models in healthcare?
    - Senior Chair: Monica Agrawal
    - Junior Chair: Xin Liu, Alejandro Lozano
5. **Multimodal AI for Health**: How to effectively integrate multiple data sources (e.g., Electronic Health Records (EHRs), images, genomics) for ML applications in healthcare? How does this work in real-time in a hospital?
    - Senior Chair: Marinka Zitnik
    - Junior Chair: Jiacheng Zhu, Rafal Dariusz Kocielnik
6. **Health AI Model Development and Generalizability**: Applying ML models in practice could face multiple challenges including domain shift, annotation quality, and out-of-distribution. How can we ensure the robustness and generalizability of a model?
    - Senior Chair: Berk Ustun
    - Junior Chair: Haoran Zhang, Keith Harrigian
7. **Health AI and Accessibility**: Making AI accessible to all in healthcare is important, but "accessibility" could encompass many things such as infrastructure, compute resources, or access to healthcare in the first place. What are the different components of the healthcare system that could improve patients' accessibility to health AI, and how do these different components play into the development of AI models?
    - Senior Chair: Edward Choi, Kristen Yeom
    - Junior Chair: Edward Lee
8. **Health AI and Patient Privacy**: How can we preserve patient privacy and maintain data security while leveraging machine learning techniques in healthcare?
    - Senior Chair: Gamze Gürsoy
    - Junior Chair: Milos Vukadinovic
9. **Bias/Fairness in Health AI**: Despite its potential, the application of machine learning in healthcare has often resulted in models that reflect and reinforce existing health disparities. How can machine learning promote fairness and enhance global health outcomes?
    - Senior Chair: Marzyeh Ghassemi, Emma Pierson

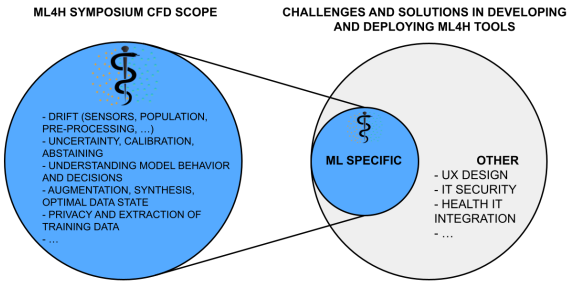

Figure 1: Scope of the new ML4H Demo Track.

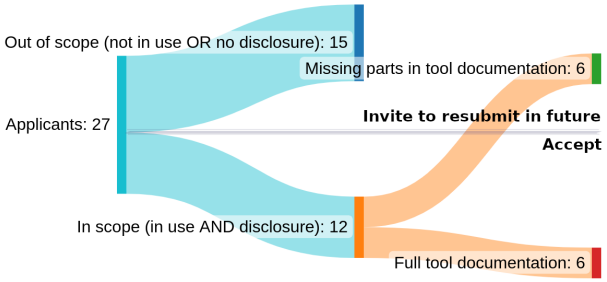

Figure 2: Flow of applications and accepted submissions to the ML4H 2023 Demo Track.

- Junior Chair: Aparna Balagopalan, Sarah Jabbour

10. **ML for Survival Analysis & Epidemiology/Population Health**: Where do we stand with ML's role in population health? How can ML be applied for time-to-event survival analysis? How ML is aiding in preventing and responding to outbreaks of infectious diseases?
    - Senior Chair: George Chen, Sanjat Kanjilal
    - Junior Chair: Vincent Jeanselme
11. **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: Michael Oberst
    - Junior Chair: Katherine Matton, lker Demirel

2.6. Demonstrations Track

This year, ML4H is excited to introduce a new component to the program - the Call for Demonstrations. Increasing numbers of Machine Learning-based Software as Medical Devices are approved by organizations such as US FDA, China NMPA, or EU CE among others. As the ML4H field continues to mature and differentiate, there is a growing need for an interface where assumptions prevalent in ML4H research can be validated against the challenges, solutions, and maturity of real-world ML4H tools. The ML4H Demo track aims at submissions that demonstrate real-world applications of ML4H technologies, bridging the gap from proof-of-concept to practical utility (Figure 1). Submissions will be evaluated according to the following process and criteria.

We received 27 applications for the first demo track. Applicants had to submit a one-page spec sheet and disclose information about the tool and its machine learning components in a form. The submissions were evaluated based on:

- Scope of demo falls within ML4H call
- Maturity of the tool or project (e.g., used but approval not necessary, in approval process, approved by a notified body)
- Quality and clarity of the submission
- Highlighting the role of machine learning methods as a source of solutions or challenges during the development or deployment of the tool.

Six submissions passed the selection criteria (Figure 2) and were invited to present their demos in-person at the symposium. The first batch of ML4H demos, in random order, comprise:

1. *CarDIA-AI Angiogram Triage Tool* by Jeremy Petch, Shuang Di and Walter Nelson from CREATE, Hamilton Health Sciences
2. *Vivalution: AI-Powered Digital Cell Morphology Platform* by S Sree Niranjanaa Bose, Murali Mohan, S Guruprasad, Shefali Hemanth Karve, Anandarama Marvi Hebbar, and Ayan Debnath from Bosch Global Software Technologies, India
3. *COMPASS: Versatile Research Platform for Clinical Studies* by Samuel Kim and Min Sang Kim from Cipherome Inc., San Jose, USA

4. *CheXchoNet: A Tool for Structural Heart Disease Detection* by Shreyas Bhave and Pierre Elias from Columbia University, Department of Biomedical Informatics

5. *Vivaray hb/hb pro: Point-of-Care Hemoglobin Measuring Solution* by S Sree Niranjanaa Bose, Murali Mohan, S Guruprasad, and Shefali Hemanth Karve from Bosch Global Software Technologies, India

6. *A Clinically Actionable Finding on the Path to Identifying and Predicting Pediatric Sepsis* by Shems Saleh, William Ratliff, Marshall Nichols, Mike Revoir, Michael Gao, Mark Sendak, Suresh Balu, and Emily Sterrett from Duke Institute for Health Innovation

3. Analysis of Accepted Works

3.1. Structured Data Analysis

We asked authors to report the subject of their accepted proceedings and findings papers and used data modalities. Accordingly, we also asked reviewers and area chairs during registration for their expertise regarding the same topics and data modalities. Reviewers and area chairs were allowed to select multiple topics and data modalities. Note that due to a technical issues, authors were unable to choose the subjects *Transfer Learning* and *Computer Vision* and the data modality *Graph and Network*. Reviewers and area chairs could not select *Tabular* as a data modality. Figure 3 contains the ratios of the subjects for each subgroup sorted by the accepted submissions. The top five topics are *Supervised Learning*, *Generative Models*, *Representation Learning*, *Public and Social Health*, and *Unsupervised Learning*. In comparison to the topic topics of last year (Parziale et al., 2022), *Generative Models* gained and *Explainability and Interpretabilty* lost relevance. Generally, the overlap between submitted papers and expertise reported by reviewers and area chairs overlapped. Only very few reviewers and area chairs chose the category *Other*

The results for the data modalities are shown in Figure 4. The top three data modalities are *Time Series*, *Tabular*, and *Images*. Again, the overlap between submitted papers and reported expertise is generally high.

3.2. Topic Modeling

Similar to last year (Parziale et al., 2022), we performed topic modeling over the text of accepted proceedings and findings papers using a Latent Dirichlet allocation (Blei et al., 2003; Syed, 2019). The marginal topic distributions are shown in Figure 5. The top topics are *Representation Learning*, *Medical Imaging Analysis*, and *NLP and Dataset*. Compared to the previous edition of the symposium, (Parziale et al., 2022), only *Medical Imaging Analysis*, *Treatment Prediction*, *EHR*, and *ML4H* or similar were also identified last year. The topic distributions of ML4H 2021 and 2020 (Roy et al., 2021; Sarkar et al., 2020) are quite different indicating the change of interest over the years.

4. The ML4H Community

The previous year has brought forth significant advancements in several fields in machine learning with potential major implications for healthcare, from large language models that have improved clinical question answering (Singhal et al., 2023) and protein design (Madani et al., 2023), to increased AI governance as reflected by the US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (White House, 2023). As a result, the size of the machine learning for health community continues to grow, as reflected by both the number of paper submissions (as described in Section 2.2), and the number of attendees in symposiums such as ML4H.

This year, following the lead of many machine learning conferences, ML4H has returned to an in-person only format. As of November 29th, 221 attendees have registered for the event. This represents an increase of 17% compared to the same date the previous year. During registration, we asked several optional questions to attendees, aiming to assess the diversity within the community, as well as to guide us in fostering more equity, diversity, and inclusion in the event program.

4.1. Geographic Distribution

Of the 87.1% of attendees who indicated their country of residence, 69.3% reside in North America, 21.3% reside in Europe, 7.9% reside in Asia, and the remainder from Africa and Oceania. The most represented nation is the United States of America, followed by Switzerland, the United Kingdom, and Canada.

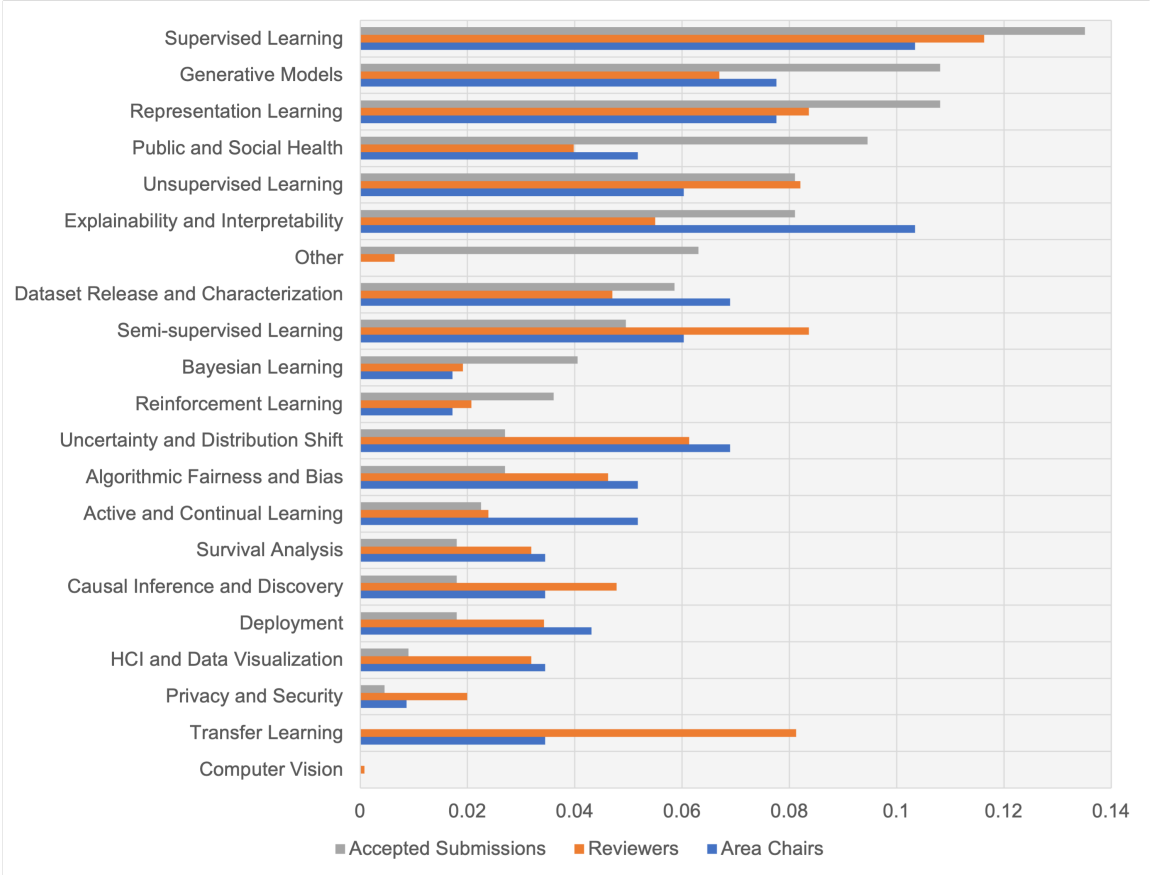

Figure 3: Self-reported topics of accepted submissions and expertise of reviewers and area chairs.

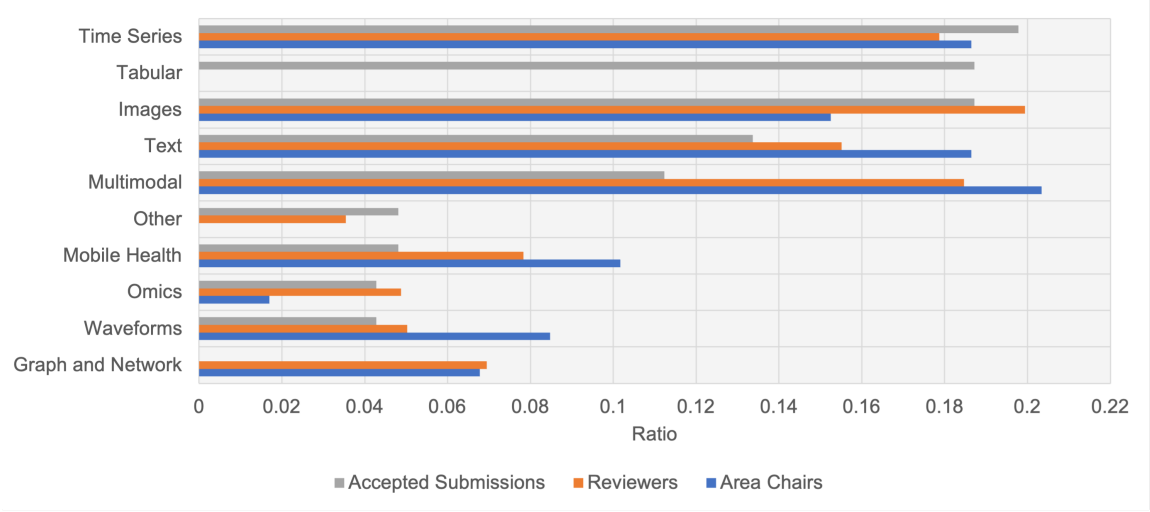

Figure 4: Self-reported data modalities of accepted submissions and expertise of reviewers and area chairs.

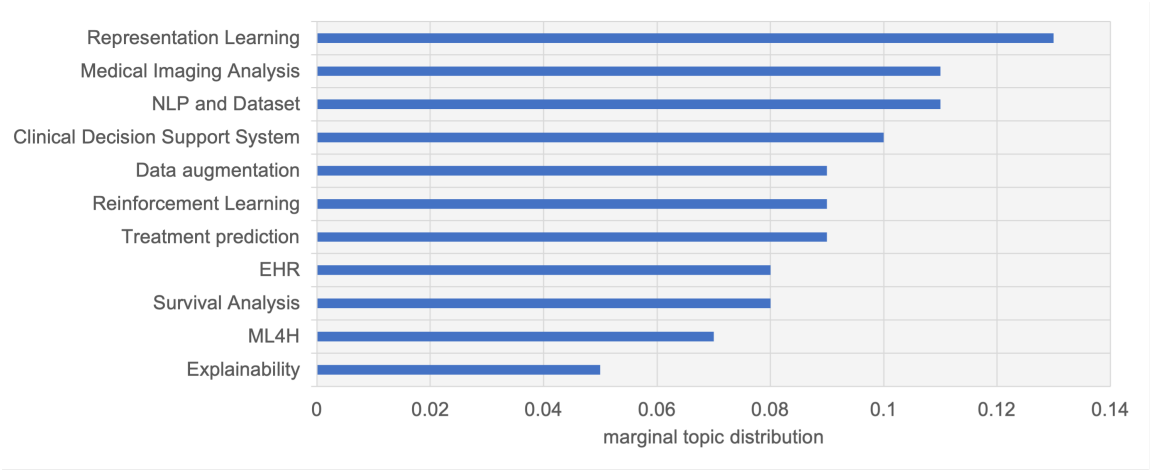

Figure 5: LDA marginal topic distribution of accepted works in ML4H 2023.

4.2. Primary Community

Similar to previous years, we found that the majority of participants (78.1%) identified *Machine Learning / Computer Science* as their primary community. 12.4% selected *Health / Medicine* as their primary community, 4.8% selected *Informatics*, and 4.8% selected *Other*.

4.3. Background and Experience

The ML4H symposium continues to attract participants from a diverse set of backgrounds. 57.4% of attendees are students. Of the remainder, 57.1% have a background in academia, 35.2% in industry, and 3.8% in government. We also see a broad distribution in the levels of experience. Of the participants who provided their age group, the majority (56.3%) are between the ages of 21 and 30; 30.5% are between 31 and 40, and 9.7% are older than 40.

5. Conclusions

The research field of machine learning for healthcare is steadily growing and maturing. The third Machine Learning for Health symposium aimed to further support this process. This year, we offered a Proceedings and Findings track to submit research work. The submission rate grew by over 40% and the preliminary registrations by 17% emphasizing the relevancy of the field. Four paper awards were introduced to highlight exceptional submissions. ML4H 2023 brought together machine learning researcher, clinicians, and healthcare data scientists. Invited talks and thematic sessions pointed out timely challenges of the field. To stimulate discussions and exchange of participants ML4H offered poster session, research roundtables, and a social evening event. We introduced a new Demo Track to showcase successful translations of machine learning for health research into practical applications. A cornerstone of ML4H are its mentorship programs to support new members of the community with advice of more senior researchers. In total, 338 mentors and mentees from diverse backgrounds participated in these programs underlining its impact. ML4H continues to serve as an interdisciplinary and international exchange platform for cutting-edge research in machine learning for health.

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 thank goes to Tasmie Sarker, who coordinated the communication 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:

Allison Koenecke, Byron C Wallace, Collin Stultz, Di Jin, Edward Choi, George H. Chen, Irene Chen,

Julia E Vogt, Liangqiong Qu, Lujie Karen Chen, Masoud Rouhizadeh, Matt Barnes, Michael C Hughes, Monica Agrawal, Prithwish Chakraborty, Sarah Tan, Shaun J. Canavan, Sonali Parbhoo, Yogatheesan Varatharajah.

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, Aakash Kaku, Adam Breitholtz, Alain Ryser, Alan Kaplan, Alex Fedorov, Alexander Marx, Ali Septiandri, Alison Callahan, Amin Adibi, Amir Reza Peimani, Amy X Lu, Andrew Prinster, Anna Zapaishchykova, Ari Ercole, Arsh Verma, Arvind Pillai, Ayush Tomar, Benjamin C Warner, Besma Guesmi, Caleb Ellington, Changye Li, Chen Zhang, Christina X Ji, Clinton Wang, Colleen Elise Chan, Daniel Capellan-Martin, Daphne Schlesinger, DongHyeon Seo, Donna Tjandra, Ece Ozkan, Elise Jortberg, Elizabeth Healey, Elliot Schumacher, Fahad Kamran, Fang Wu, Fanyang Yu, Frank Rudzicz, Fridolin Haugg, Giovanni Cina, Gyubok Lee, Hallee E Wong, Han Yuan, Haoran Zhang, Harri Lähdesmäki, Harshith Padigela, Harvineet Singh, Hejie Cui, Helen Zhou, Humera Tariq, Hyewon Jeong, Hyunseung Chung, Intae Moon, Ivaxi Sheth, Jaan Altosaar, Jack Quach, Jacob Oppenheim, Jason Poulos, Jessica Karina De Freitas, Jiaying Lu, Jiechao Gao, Jinpei Han, Jong Hak Moon, Jongoh Jeong, Jungwoo Oh, Junyi Gao, Justin Cosentino, Juyong Kim, Ka-Ho Chow, Kaiwen Zha, Karim Kadry, Kathryn Wantlin, Katy Blumer, Kazuma Kobayashi, Kristen A Severson, Kyunghoon Hur, Laura Manduchi, Lili Wu, Lily H Zhang, Malvern Madondo, Marta Avalos, Maryam Daniali, Matthew James West, Matthew Peroni, Mazdak Abulnaga, Md Saiful Islam, Meera Krishnamoorthy, Meghasyam Tummalacherla, Michael Ito, Michelle M Li, Min Hun Lee, Min Jae Lee, Mohammad Azizmalayeri, Mohammad Eslami, Monica Munnangi, Morufu Olalekan Raimi, Naihao Deng, Neeraj Wagh, Nimshi Venkat Meripo, Nina Mishra, Nripsuta Ani Saxena, Oliver Rausch, Priyanka Mary Mammen, Pablo Moreno-Muñoz, Payal Chandak, Peiqi Wang, Peng Cao, Peniel Argaw, Peter Drotar, Pooja Sarin, Pragyansmita Nayak, Praveen Gurunath Bharathi, Purity Mugambi, Qiao Jin, Qixuan Jin, Rachid Riad, Rachneet Kaur, Rafael Teixeira Sousa, Ricards Marcinkevics, Robin Khatri, Roy Adams, Ryan Andrew Chi, Ryan Schuerkamp, Saeed Alahmari, Sandhya Tripathi, Sarah Brueningk, Sarah Jabbour, Sasikanth Kotti, Sebastian Caldas, Seongsu Bae, Serina Chang, Seung Eun Yi, Shreyasi Pathak, Shuge Lei, Shuo Shuo Liu, Shuvom Sadhuka, Sidhika Balachandar, Siyi Tang, Srimonti Dutta, Stefan Groha, Stephen Robert Pfohl, Sumedha Rai, Sunny Gupta, Tong Wu, Thomas Hartvigsen, Thomas Marco Sutter, Tianyu Liu, Tiffany Callahan, Tomas Bosschieter, Tonmoy Hossain, Toon Vanderschueren, Trenton Chang, Tzu-Ting Wei, Vajratiya Vajrobol, Van Quynh-Thi Truong, Victor Ikechukwu Agughasi, Vincent Jeanselme, Vinod Kumar Chauhan, Vishwali Mhasawade, Viswesh Krishna, Wangzhi Dai, Wasiwasi Mgonzo, William La Cava, Wisdom Oluchi Ikezogwo, Wouter A.C. van Amsterdam, Xiaomeng Dong, Xing Han, Yan Zhuang, Yefan Tao, Yining Hua, Yoav Itzhak Wald, Yonatan Naamad, You Cheng, Yuan Zhao, Yunxiang Zhang, Yurui Cao, Yuta Kobayashi, Yuwen Chen, Yuxin Xiao, Yvonne Lui, Zhengyuan Liu, Mozhgan Saeidi.

References

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003. URL https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Ali Madani, Ben Krause, Eric R Greene, Subu Subramanian, Benjamin P Mohr, James M Holton, Jose Luis Olmos Jr, Caiming Xiong, Zachary Z Sun, Richard Socher, et al. Large language models generate functional protein sequences across diverse families. *Nature Biotechnology*, pages 1–8, 2023.

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 sym-*

posium, 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.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180, 2023.

Shaheen Syed. Perform latent dirichlet allocation full workflow. https://github.com/shaheen-syed/LDA.git, 2019.

The White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, October 2023.