Machine Learning for Health (ML4H) 2022

Antonio Parziale[1] ANPARZIALE@UNISA.IT
Monica Agrawal[2] MAGRAWAL@MIT.EDU
Shengpu Tang[2] TANGSP@UMICH.EDU
Kristen Severson[3] KSEVERSON@MICROSOFT.COM
Luis Oala[4] LUIS.OALA@HHI.FRAUNHOFER.DE
Adarsh Subbaswamy[4] ASUBBASWAMY@JHU.EDU
Sayantan Kumar[5] SAYANTAN.KUMAR@WUSTL.EDU
Elora Schoerverth[6] ELORA.SCHOERVERTH@HHI.FRAUNHOFER.DE
Stefan Hegselmann[7] STEFAN.HEGSELMANN@UNI-MUENSTER.DE
Helen Zhou[7] HLZHOU@ANDREW.CMU.EDU
Ghada Zamzmi[8] ALZAMZMIGA@NIH.GOV
Purity Mugambi[8] PMUGAMBI@UMASS.EDU
Elena Sizikova[8] ELENA.SIZIKOVA@FDA.HHS.GOV
Girmaw Abebe Tadesse[8] GIRMAW.ABEBE.TADESSE@IBM.COM
Yuyin Zhou[8] YZHOU284@UCSC.EDU
Taylor Killian[9] TWKILLIAN@CS.TORONTO.EDU
Haoran Zhang[9] HAORANZ@MIT.EDU
Fahad Kamran[10] FHDKMRN@UMICH.EDU
Andrea Hobby[11] ADH81@GEORGETOWN.EDU
Mars Huang[8,11] MSCHUANG@STANFORD.EDU
Ahmed Alaa[12] AMALAA@BERKELEY.EDU
Harvineet Singh[12] HS3673@NYU.EDU
Irene Y. Chen[13] IYCHEN@CSAIL.MIT.EDU
Shalmali Joshi[13] SHALMALI@SEAS.HARVARD.EDU

Organizing committee for ML4H 2022
[1] Publications Chair, [2] Workflow Chair, [3] Workflow Subchair, [4] Program Chair, [5] Speaker Coordinator Subchair, [6] Program Virtualization Subchair, [7] Social Chair, [8] Social Subchair, [9] Virtualization Chair, [10] Communication Chair, [11] Communication Subchair, [12] Finance Chair, [13] General Chair

1. Introduction

The second Machine Learning for Health (ML4H) symposium was held both virtually and in-person on November 28, 2022, in New Orleans, Louisiana, USA. Similar to the previous year's symposium (Roy et al., 2021), ML4H was organized as a stand-alone event co-located with the Neural Information Processing Systems (NeurIPS) conference. This year, ML4H accepted a total of 80 submissions including 52 extended abstracts and 28 full-length proceedings papers. The symposium invited submissions comprising machine learning research on relevant problems in health and biomedicine. ML4H 2022 featured two submission tracks: a proceedings track, which encompassed full-length submissions of technically mature and rigorous work, and an extended abstract track, that would accept less mature, but innovative

research for discussion. Accepted publications of both types were given a platform for presentation, through poster presentations. Each poster was presented both virtually and in-person. The goal was to provide a venue to publish high-quality work, while still enabling the lively discussions that have made the ML4H workshops worthwhile in the past. In this front matter, we provide an overview of the ML4H 2022 symposium, including various mentorship programs, the paper selection process, and the submission statistics (Section 2). In Section 3 we analyze the accepted works, and offer commentary on trends in research observed in this field, building on analyses of the previous ML4H workshops and symposiums (Sarkar et al., 2020; Roy et al., 2021). In Section 4, we comment on the composition of the ML4H community. Finally, we close with acknowledgments, including a list of meta-reviewers and reviewers for ML4H 2022.

2. Symposium

The ML4H 2022 symposium was held both virtually and in-person as a hybrid event on November 28, 2022. In accordance with the virtual and in-person format, keynotes and panel discussions were held live at the venue and streamed for virtual attendees via SlidesLive. Oral presentations for best-papers proceeded similarly. Virtual attendees had the opportunity to ask questions through a live chat. We also held four virtual and nine in-person roundtables. Virtual roundtables proceeded on Zoom. Virtual poster sessions were hosted in Gather.Town while in-person poster sessions were held at the venue. Mentorship programs including reviewer mentorships, author mentorship and career mentorships were held asynchronously in accordance with paper submission and reviewing timelines.

2.1. Program

Eight speakers gave presentations at ML4H 2022 of which three were keynotes (Mihaela van der Schaar, Emmanuel Candès and Ben Recht) and five were invited talks (Himabindu Lakkaraju, Zachary Lipton, Nicola Pezzotti, Berkman Sahiner and Edwin Fong). The speakers also contributed to two panel sessions that highlighted concerns in the machine learning for health research and industry communities. The first panel, *Utility and Shortcomings of XAI in Healthcare*, provided a space for Symposium participants to weigh the pros and cons of interpretability and explanation methods for machine learning models in the context of healthcare applications. The second panel, *Paths to Practice for ML Technologies*, provided participants with an overview of different perspectives on bringing machine learning technology into production in healthcare applications. In addition, participants were able to discuss accepted works at two poster sessions, both in-person and virtually, as well as broader themes in ML4H at 13 research roundtables that were hosted by committees of senior and junior roundtable chairs (see Section 2.4 for more details). Furthermore, winning submissions of the AHLI Breast Cancer Datathon presented their work to the audience. The day concluded with an evening social at the Bryant Park Nola where ML4H 2022 participants could unwind and finish the day in a relaxed atmosphere.

2.2. Paper Selection

Submission Statistics This year 139 manuscripts were submitted to the ML4H symposium, with a slight increase with respect to the 123 of the 2021 edition. That confirms that the communities of computer scientists and clinicians consider ML4H a suitable venue where discussing about the

applications of machine learning technologies to the health field.

The program committee consisted of 21 meta-reviewers and 163 reviewers who completed 475 total reviews. At least three, and up to 5, reviews were conducted for each proceedings track and extended abstract track submission.

Out of the 87 papers submitted to the proceedings track, 28 were accepted into the proceedings (32.2% acceptance rate). The reviewers could recommend to transfer the paper from the proceedings track to the extended abstract track. 23 submitted full papers were transferred to the extended abstract track. Out of the 52 papers submitted to the extended abstract track, 29 were accepted (55.8% acceptance rate). As a result, there were 52 extended abstracts in total. The extended abstracts were given the opportunity to be included in an ML4H arXiv index at https://ml4health.github.io/online_proceedings.html

2.3. Mentorship Programs

ML4H ran three mentorship programs: (1) the submission mentorship program, (2) the reviewer mentorship program, and (3) the career mentorship program. The overarching goal was to share the knowledge and experience of senior researchers to make the field of machine learning for healthcare more accessible and improve important work for the community.

Submission Mentorship To foster collaboration and provide authors with support for creating high-quality ML4H submissions, we hosted a submission mentorship program with a rolling application deadline. In total, 61 mentees and 42 mentors registered to participate in the program between the first week of July and the last week of August. From July, matches between mentees and mentors were sent out weekly on a rolling basis. The criteria for matching mentors with mentees are common research interests and the seniority level. During the mentoring period, mentees had bi-weekly meetings with the mentees to provide feedback/ suggestions for improvement on aspects such as (1) the direction of the design; (2) models and experiments; (3) results analysis; and (4) presentation/organization of the final paper.

Soon after the ML4H submission deadline, we sent out a survey to all participants to gain a better understanding of the participants' experience and concerns and gather feedback to be integrated into the submission program for next year. Overall, respondents reported that the effort was a success, with the majority (71%) reporting good overall experience (respondent score ≥ 3 in a scale of 1-5). More than 76% expressed interest in participating in the program again in the future years. Moreover, 42% of the mentees said that the program had a moderate to major impact on the final submission. Respondents also reported several positive feedback such as the excellent process of matching between mentors and mentees, the good communication from organizers, and the mentor/mentee enthusiasm.

However, mentors and mentees identified a few challenges that impact the mentor-mentee experience. Examples of these roadblocks include time constraints due to the short mentoring period and different time zones and further guidelines on how to navigate authorship considerations. In future years of these programs, we will take all these issues into consideration by increasing the mentoring period, assigning participants in the same or similar time zones, and providing clear guidelines regarding authorship. We also plan to gather feedback in the middle of the program to allow detecting early issues and providing timely assistance.

Reviewer Mentorship The goal of the reviewer mentorship program is to mentor junior reviewers for ML4H. Mentors are senior researchers who are experienced reviewers. The minimum requirement for a reviewer mentor are: 3rd year PhD student, and, previously reviewed at least 6 papers and/or published at least 5 papers. Both mentors and mentees indicate interest through a signup form, in which they provide their; training level, review experience, preferred research subjects, and for mentors, the preferred number of mentees. The ML4H organizing team matches mentees with mentors by their shared interests in research, and their training status. For instance, reviewer mentees who are postdocs are matched with professors, senior industry researchers and/or senior clinicians with whom they share at least one research subject. The matching algorithm also honours the mentors' requested number of mentees. Once matched, the mentor and mentees are expected to self-coordinate to share and read the mentees' assigned papers, read and discuss the mentees' reviews, and, discuss the author responses and rebuttals to the submitted reviews. At the end of the program, participants are invited to provide feedback that will be used to improve the offering in the next iteration of ML4H.

This year's program matched 41 mentees with 27 mentors. The mentees ranged in training level with the lowest being Masters students, and highest being Postdoctoral researcher. The mentors' training status ranged from 3rd year PhD student to Professor/Senior Industry researcher/Medical doctor. An examination of early feedback on this year's program reveals that the majority of the respondents (85%) were happy with their matches (respondent score ≥ 3 on a scale of 1-5), of which, a large proportion (67%) were satisfied with their match (rating of 4 or 5). With respect to the value derived from the mentorship, a large proportion (67%) of the respondents indicated that they had a meaningful connection with their mentor/mentee and that the feedback they gave or received was helpful. Overall, the majority of the respondents ($\sim$87%) indicated that they would participate in the program in the future.

To ensure that the program better serves the participants in ML4H 2023, the team will incorporate two key items from the feedback provided by this year's participants. First, to ensure that mentors better plan their time commitment to the program, the signup form will provide a question on how many papers they can provide feedback for (in addition to asking how many mentees they want to mentor). Secondly, at the risk of overcommunicating, the program organizers will send reminders to the participants at the start of each phase of the program. Feedback indicated that this would especially be appreciated by mentors who are more senior and often very busy.

Career Mentorship The goal of the career mentorship program is to pair mentees with mentors who can provide advice on career-related topics (e.g., developing a long-term research plan, doing healthcare research in industry, work-life balance). The program consisted of a one-hour group mentoring workshop and a an hour-long 1-1 mentoring session to provide advice on the topics of interest. 105 mentees and 23 mentors signed up to participate in the program.

2.4. Roundtables

Following the positive feedback received on the roundtables in ML4H 2021, which all were organized virtually, we proceed to have similar sessions in ML4H 2022. To this end, we crowd-sourced interesting topics in ML for healthcare, and top ranked topics are selected. The structure of a roundtable session follows that of ML4H 2021, and consists of

Senior and Junior chairs. The Senior chairs are selected and invited among experts in each topic domain, and they are tasked with leading the session. Junior chairs are mainly students, and they are responsible to moderate the session. We include both virtual (4) and in-person (9) roundtables that target virtual and in-person attendees of the symposium, respectively. Detailed lists of topics and their chairs are provided below for both in-person and virtual roundtables.

2.4.1. In person Roundtables

1. Are our ML models really making an impact in the hospital? What do caregivers and clinicians want and what is still missing?
    - Senior Chairs: Roxana Dangeshou and Siyu Shi
    - Junior Chairs: Jennifer Chien and Sujay Nagaraj
2. Evaluation of healthcare data prior to applying ML, e.g., representation analysis, annotation quality, OOD, clusters of IDs
    - Senior Chairs: Nicola Pezzotti and Pin-Yu Chen
    - Junior Chairs: Neha Hulkund and Shreyas Bhave
3. How to ensure generalizability of ML in healthcare?
    - Senior Chairs: Emmanuel Candès, Stephen R. Pfohl and Edwin Fong
    - Junior Chairs: Michael Oberst and Amruta Pai
4. How do we inject domain knowledge into DL models, in particular when not much data is available?
    - Senior Chairs: Aakanksha Naik, Ben Lengerich, Ying Xu and Eran Halperin
    - Junior Chairs: Caleb Ellington and Wisdom Ikezogwo
5. How to effectively integrate multiple data sources (e.g., EHR, images, genomics) for ML applications in healthcare?
    - Senior Chairs: Jonathan Bidwell, Dominik Dahlem, and Mark Sendak
    - Junior Chairs: Jason Dou
6. How can we utilize foundation models (very large pre-trained models) for healthcare?
    - Senior Chairs: Payel Das, Zachary C. Lipton, and Byung-Hak Kim
    - Junior Chairs: Monica Agrawal and Changye Li
7. Using ML for population health
    - Senior Chairs: Nathaniel Hendrix and Dimitrios Spathis
    - Junior Chair: Piniel Argaw
8. How to incentivize creation or publication of new data collections and facilitate international collaboration?
    - Senior Chairs: Jun Seita and Bastiaan Quast
    - Junior Chairs: Mehak Gupta and Xinhui Li
9. Post-approval monitoring and validation of AI systems in health care
    - Senior Chairs: Berkman Sahiner, Anna Decker and Harvineet Singh
    - Junior Chairs: Marta Lemanczyk and Yuhui Zhang

2.4.2. Virtual Roundtables

1. How machine learning can help prevent and respond to infectious disease outbreaks, were are we now?
    - Senior Chair: Megan Coffee
    - Junior Chair: Christian Garbin
2. Are our ML models really making an impact in the hospital? What do caregivers and clinicians want and what is still missing?
    - Senior Chair: Collin Stultz
    - Junior Chair: Elizabeth Healey
3. How to effectively integrate multiple data sources for machine learning applications in healthcare?
    - Senior Chair: Mert R. Sabuncu
    - Junior Chair: Heejong Kim
4. How to evaluate of healthcare data and labels prior to applying machine learning?
    - Senior Chair: Ziad Obermeyer
    - Junior Chair: Claire Boone

3. Analysis of Works

3.1. Structured Data Analysis

3.1.1. Author-reported information

During submission, authors were asked to answer structured questions about subject areas of their manuscripts and data modalities. Authors were allowed to select multiple topics for their papers. Figure 1 shows the frequency with which topics were selected by authors. The top five categories were Supervised Learning, Explainability & Interpretability, Public & Social Health, Representation Learning, and Generative Models. For the third year in a row, Explainability & Interpretability has been confirmed as one of the most investigated topics of the symposium. The interest in Generative models has increased significantly since the 2020 edition of the symposium. This confirms a general trend in the machine learning research area.

The authors also reported the data modalities related to the submission. Authors were allowed to make a multiple choice among the categories: Text, Images, Time series, Waveforms, Omics, Mobile Health, Multimodal, Other. The top three selections were Time series, Images and Text with a percentage of 30.9%, 25.9% and 24.7%, respectively. Finally, 44% of the authors indicated that they are able to make data and code available publicly.

3.1.2. Reviewers and Meta-reviewers expertise

Reviewers and Meta-reviewers were asked to indicate their areas of expertise when they were invited to join the panel of experts that would have evaluated the manuscripts submitted to the symposium.

Figure 2 shows the distribution of experts among the different areas. Supervised learning and Transfer learning were the two most selected areas of expertise among reviewers, while Supervised Learning and Explainability & Interpretability were the most selected among meta-reviewers.

3.2. Topic Modeling

We performed topic modeling over the text content of accepted papers using Latent Dirichlet allocation (Blei et al., 2003; Syed, 2019). We considered both the full proceedings papers and the extended abstracts for this study and the marginal topic distribution is shown in Figure 3.

In addition to the topics from the previous two editions of the symposium (Sarkar et al., 2020; Roy et al., 2021), "Self-supervised

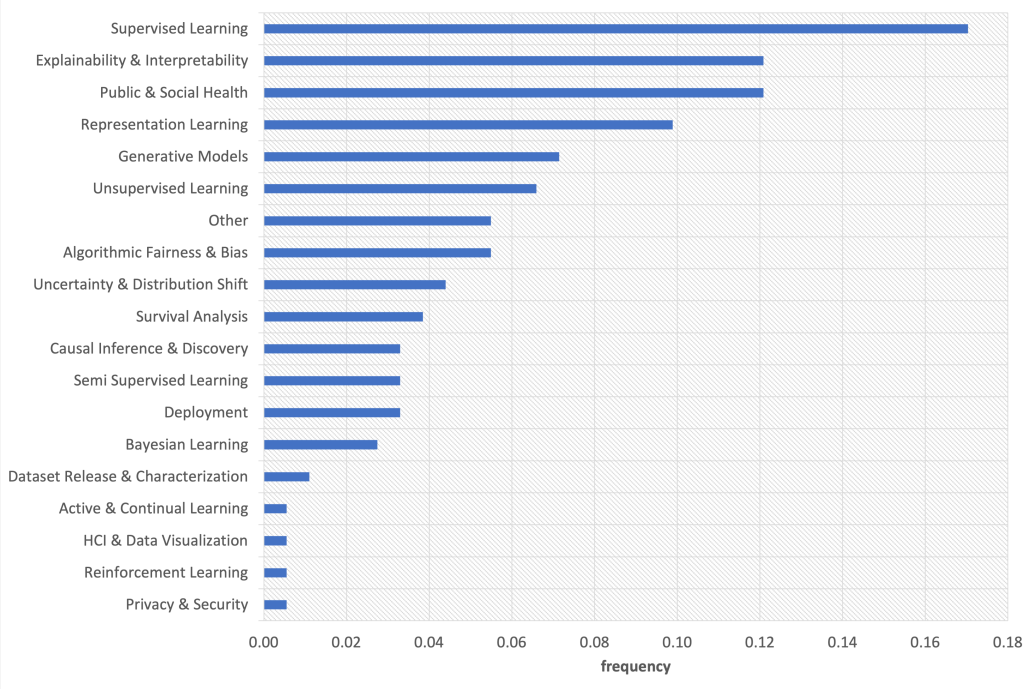

Figure 1: Author-reported subject area of accepted papers

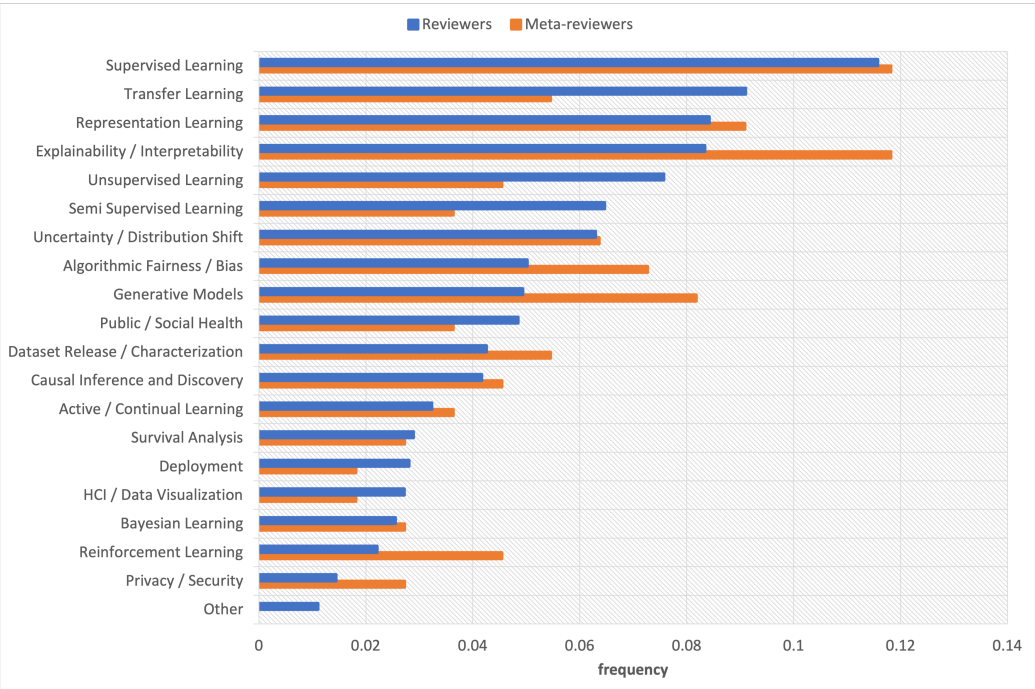

Figure 2: Areas of expertise reported by reviewers and meta-reviewers

learning" came out from topic modeling. In the previous editions was observed an increase in papers related to the topics of Causal Inference and Uncertainty and this trend is confirmed for the current edition. Moreover, we observed an increase in papers related to Medical Imaging and Survival analysis with respect to the previous years.

4. The ML4H Community

The first ML4H workshop was organized in 2016 to "*engender discussion between machine learning and clinical researchers about how statistical learning can enhance both the science and the practice of medicine*" (Shalit et al., 2016). Since then, the ML4H community is grown year by year strengthening its transdisciplinary and international nature.

This year, as of November 20th, 326 attendees were registered for the event. During the registration, people were asked to answer a set of questions that we used to outline the heterogeneity of the community and to achieve greater equity, diversity, and inclusion in setting up the event program.

4.1. Geographical distribution

Each registrant could optionally indicate their country of residence and 73% of them provided this information. Among the participants that indicated their country, 61% reside in the continent of America, 28% reside in the continent of Europe, 11% live in the continent of Asia, and the remaining part lives in Africa and Oceania. The most represented nation is The United States of America while the runner-up is the United Kingdom.

4.2. In-person vs Online participation

This year's symposium was the first edition organized as a hybrid event, combining the tradition of the first four in-person editions and the experience gained during the last two online editions.

Almost 58% of registrants opted for in-person attendance, and $\sim 52\%$ of them indicated that their country of residence is outside the continent of America.

4.3. Primary community

We asked the registrants to select their primary community among the following options: *Machine Learning / Computer Science*, *Health / Medicine*, and *Other*. When the option *Other* was selected, we asked to provide more details. 80% of attendees indicated *Machine Learning / Computer Science* as their primary community, 17% selected *Health / Medicine*, and the remaining part specified a community that could be considered interdisciplinary.

4.4. Experience and background

The ML4H community is made up of both young and experienced people from academia and industry. 92% of the registrants provided their age group. Slightly more than 62% of them are aged between 21 and 30, 30% between 31 and 40, approximately 5% between 41 and 50, and 2% between 51 and 60. Approximately 63% of participants are students. The participants registered as "non-student" have different backgrounds: more than 40% are from academia, more than 39% are from industry, and the background of the remaining part is both from academia and industry.

5. Conclusions

Research in Machine Learning to improve healthcare is well underway, with the ML4H Symposium (and its previous versions) having played a key role in bringing relevant communities together. The second Machine Learning for Health symposium

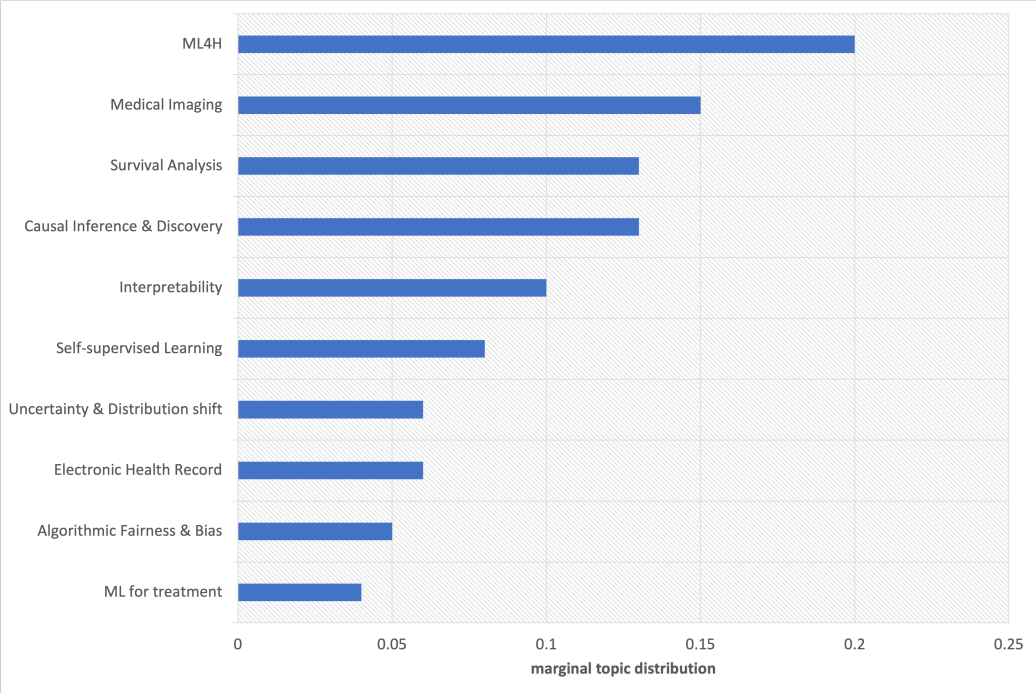

Figure 3: LDA marginal topic distribution of accepted papers

(ML4H 2022) brought together, in a hybrid event, machine learning researchers, clinicians, and healthcare data experts. In this year's program, invited speakers and attendees will discuss more nuanced challenges the field has faced, including the utility and challenges of Explainable Artificial Intelligence as well as paths to practice for Machine Learning in healthcare. The two focus sessions are accompanied by keynotes as well as interactive research roundtables, two poster sessions, and datathon presentations. ML4H continues to serve as an interdisciplinary and international melting pot of enthusiastic researchers, providing exposure to students interested in this area and an opportunity for experts from academia and industry to present cutting-edge research in the field.

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. Meta reviewers

The organizers thank all meta-reviewers for their expert guidance in reviewing submissions to ML4H:

Collin Stultz, Di Jin, Edward Choi, George Chen, Jessica Schrouff, Julia E. Vogt, Liangqiong Qu, Lujie Chen, Maggie Makar, Masoud Rouhizadeh, Matt Barnes, Negar Rostamzadeh, Prithwish Chakraborty, Ruizhi Liao, Sandhya Prabhakaran, Sarah

Tan, Shaun Canavan, Yanan Sui, Yogatheesan Varatharajah, Yonatan Mintz, Cristina Soguero-Ruiz.

6.3. Reviewers

The proceedings and discussions would not have been possible without the generous reviewers who provided meaningful feedback and evaluation of all papers and extended abstracts. We are grateful to all reviewers and meta-reviewers:

Aakash Kaku, Adam Breitholtz, Adam Yala, Adarsh Subbaswamy, Alain Ryser, Alan Kaplan, Alex Fedorov, Alexander Marx, Ali Septiandri, Alison Callahan, Amy X. Lu, Andrew Prinster, Anton Matsson, Antonio Parziale, Aparna Balagopalan, Ashwin Vaswani, Aviva Prins, Ben Lengerich, Caleb Ellington, Chaoqi Yang, Charles B. Delahunt, Chen Zhang, Christina X. Ji, Christopher P. Bridge, Daphne Schlesinger, David Zimmerer, Diana Mincu, Donna Tjandra, Doron Stupp, Ece Ozkan, Edward De Brouwer, Elena Sizikova, Elliot Schumacher, Evgenii V. Bykovetc, Fabian Hausmann, Fabian Laumer, Fahad Kamran, Frank Rudzicz, Ghada Za, Gulshan Sharma, Hallee E. Wong, Hammaad Adam, Hang Yuan, Haoran Zhang, Hyewon Jeong, Ivana Maric, Jason Poulos, Jennifer Williams, Jessica Karina De Freitas, Jiayu Yao, Jinpei Han, Jongoh Jeong, Joseph Enguehard, Junyi Gao, Juyong Kim, Ka-Ho Chow, Kathryn Wantlin, Katie Matton, Katikapalli Subramanyam Kalyan, Kazuma Kobayashi, Kenza Bouzid, Kexin Huang, Laya Rafiee, Lena Stempfle, Lily H. Zhang, Mahmut Yurt, Malvern Madondo, Manan Dey, Mario Wieser, Marta Avalos, Martin Seneviratne, Maruf Adewole, Matthew B.A. McDermott, Matthias Kirchler, Matthias Norden, Md Osman Gani, Meera Krishnamoorthy, Meghasyam Tummalacherla, Mehak Gupta, Michael Dusenberry, Michael Oberst, Miguel Xochicale, Min Lee, Mohammad Eslami, Morgan Sanchez, Nazanin Makkinejad, Newton Mwai Kinyanjui, Nick Phillips, Nicola Sahar, Nishan Srishankar, Oliver Rausch, Pablo Moreno-Muñoz, Payal Chandak, Praveen Gurunath Bharathi, Praveer Singh, Qiao Jin, Qixuan Jin, Qiyuan Hu, Rafael Teixeira Sousa, Ramya Tekumalla, Rebecca Boiarsky, Ricards Marcinkevics, Ridwan Alam, Robin Khatri, Rongguang Wang, Roy Adams, Ryan-Rhys Griffiths, Saeed Alahmari, Sandhya Tripathi, Sarah Jabbour, Sayantan Kumar, Sebastian Caldas, Serina Chang, Seung Eun Yi, Shalmali Joshi, Shreyasi Pathak, Shuyang Cao, Stefan Groha, Stefan Hegselmann, Stephen Robert Pfohl, Subhrajit Roy, Sudeshna Das, Suhani Vora, Sumana Basu, Tong Wu, Taha Ceritli, Tanaya Babtiwale, Teya Bergamaschi, Thomas Marco Sutter, Tiffany Callahan, Tom Ginsberg, Tomas Bosschieter, Tony Yue Sun, Trenton Chang, Tzu-Ting Wei, Upasana Upadhyay Bharadwaj, Urjoshi Sinha, Vahid Balazadeh Meresht, Vaishnavi Subramanian, Vincent Jeanselme, Vinod Kumar Chauhan, Vinyas Harish, Vishwa Sanjay Parekh, Viswesh Krishna, Wangzhi Dai, Weicheng Zhu, Wenqi Wei, Wouter A.C. van Amsterdam, Xiao Gu, Xiao Li, Xiaoli Yang, Xiaomeng Dong, Xiaoy Zhang, Xinyi Zheng, Yan Gao, Yan Zhuang, Yannet Interian, Yoav Itzhak Wald, Young Joon Fred Kwon, Yuan Zhao, Yuwei Sun, Yvonne Lui, Zhaozhi Qian.

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.

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.

Uri Shalit, Marzyeh Ghassemi, Jason Fries, Rajesh Ranganath, Theofanis Karaletsos, David Kale, Peter Schulam, and Madalina Fiterau. Machine learning for health workshop. In *NIPS 2016 Workshops*, pages 20–22. 2016. URL http://media.nips.cc/Conferences/2016/NIPS-2016-Workshop-Book.pdf.

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