Machine Learning for Health (ML4H) 2020: Advancing Healthcare for All

Suproteem K. Sarkar SUPROTEEMSARKAR@G.HARVARD.EDU
Harvard University

Subhrajit Roy SUBHRAJITROY@GOOGLE.COM
Google Brain

Emily Alsentzer EMILYA@MIT.EDU
Massachusetts Institute of Technology, Harvard Medical School

Matthew B. A. McDermott MMD@MIT.EDU
Massachusetts Institute of Technology

Fabian Falck FABIAN.FALCK@STATS.OX.AC.UK
University of Oxford

Ioana Bica IOANA.BICA@ENG.OX.AC.UK
University of Oxford

Griffin Adams GRIFFIN.ADAMS@COLUMBIA.EDU
Columbia University

Stephen Pfohl SPFOHL@STANFORD.EDU
Stanford University

Brett Beaulieu-Jones BRETT_BEAULIEU-JONES@HMS.HARVARD.EDU
Harvard Medical School

Tristan Naumann TRISTAN@MICROSOFT.COM
Microsoft Research

Stephanie L. Hyland STEPHANIE.HYLAND@MICROSOFT.COM
Microsoft Research

1. Introduction

The sixth Machine Learning for Health (ML4H) workshop[1] was held virtually on December 11, 2020, in conjunction with the Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS 2020). This proceedings contains the 24 papers accepted to the ML4H Proceedings.

The theme of this year's workshop was "Advancing Healthcare for All." The application of machine learning to healthcare is often characterised by advances in models applied to improving patient outcomes. By developing these methods on high-quality datasets, researchers hope to better diagnose, forecast, and otherwise characterize individuals' health. At the same time, when we build tools for assisting highly-specialised caregivers, we may limit many of the benefits of machine learning to those who can access such care. The fragility of healthcare access both globally and locally makes the theme of healthcare for all critical in ensuring research yields a positive impact.

1. https://ml4health.github.io/2020/

In accordance with this theme, the workshop encouraged submissions of new work across several areas: accessible diagnostic and prognostic systems, health equity, fairness and bias, generalization across populations or systems, improving patient participation in health, augmenting and supporting the capabilities of healthcare workers, rare or underserved diseases, democratizing ML4H research, and non-traditional delivery of healthcare.

ML4H 2020 offered 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. Accepted publications of both types were given a platform for presentation, whether through an oral or poster presentation. The goal was to provide a venue to publish high-quality work, while still enabling the lively discussions that have made ML4H successful in the past.

In this front matter, we start in Section 2 by describing the workshop, including a brief discussion of the event, a summary of its inaugural mentorship programs, and a detailed summary of the paper selection process and submission statistics. In Section 3 we analyze the accepted works, and offer commentary on trends in research observed in this field, building on analyses of the 2018 and 2019 workshops (Beaulieu-Jones et al., 2019; Dalca et al., 2020). In Section 4, we comment on the composition of the ML4H community and discuss the workshop's theme of advancing healthcare for all. Finally, we close with acknowledgments, including a list of organizers and reviewers for ML4H 2020.

2. Workshop

ML4H 2020 was held virtually on December 11, 2020, as a part of NeurIPS 2020. In accordance with the virtual format, Keynotes and spotlight talks were pre-recorded using SlidesLive, panels were held live using Zoom, and poster sessions were hosted in Gather.

2.1. Program

The ML4H 2020 workshop featured five invited talks across academia and industry, six oral presentations from authors accepted at the venue, a research talk from sponsor Roche, as well as panel discussions between the invited speakers. The speakers included Mark Dredze, Noémie Elhadad, Judy Gichoya, Andrew Ng, and Ziad Obermeyer. The program included poster sessions for accepted papers and extended abstracts.

2.2. Paper Selection

Submission Statistics. Despite research interruptions from the COVID-19 pandemic and the switch to a virtual venue, the workshop continued to feature strong interest from the ML4H community, with 202 total submissions. The total number of submissions was down from 309 in 2019 and 239 in 2018. The program committee consisted of 34 meta-reviewers and 284 reviewers who completed 1,047 total reviews. At least four reviews were conducted for each proceedings track submission and at least three for each extended abstract track submission. As a workshop first, each submission also received a meta-review this year. The organizers felt the addition of meta-reviewers reduced the variance in review quality and led to more consistent acceptance decisions. Meta-reviewers reported they were satisfied with the review process.

Out of the 95 papers submitted to the proceedings track, 24 were accepted into the proceedings (25.2% acceptance rate), and 15 were transferred to the extended abstract track. Out of the 107 papers considered for the extended abstract track, 45 were accepted (49.2% acceptance rate, after includ-

ing the papers transferred from the proceedings track). Four proceedings submissions and two extended abstract submissions were accepted for oral presentations.

The extended abstracts were given the opportunity to be included in an ML4H arXiv index.

2.3. Mentorship Programs

Submission Mentorship Program. In an effort to improve submission quality and foster both current and future collaboration, ML4H piloted a Submission Mentorship Program this year spanning the three weeks before the submission deadline. The workshop paired 48 mentees intending to submit with 25 mentors based on mutual research interests. Mentors were asked to consider factors including:

- Is the main idea of the submission clearly explained and well-motivated?
- Is the submission well-positioned with respect to the related works? Are the related works investigated in-depth and are the contributions of the paper properly highlighted?
- Is the problem that the submission is addressing well-formulated and does the submission provide enough technical details to describe the proposed solution/model?
- Are the performance metrics well defined and the rationale for using them properly justified?

Overall, participants reported that the effort was a success. More than 89% of the mentors and mentees expressed interest in participating in the program again in the future years. Moreover, 87% of the participants said that the program had a moderate or major impact on the final submission. However, mentors and mentees both identified time constraints as the biggest roadblock preventing a successful mentor-mentee experience—in the future, the program could begin earlier to allow for multiple rounds of feedback from the mentors.

Reviewer Mentorship Program. The purpose of the ML4H Reviewer Mentorship Program was to train junior reviewers, foster new connections and relationships in the ML4H community, and ultimately improve the quality of the review process.

The core of the mentorship program was a feedback session where senior reviewer mentors provided feedback to a reviewer mentee on their reviews. Given the short reviewing timeline, mentees submitted drafts of their reviews to their mentor over email one week before the review deadline and feedback sessions occurred the following week, and mentors were free to structure the feedback session how they preferred. It was expected that mentors read their mentees' assigned papers and reviews and formulated their feedback prior to the feedback session with their mentee. However, the role of the mentor was to provide feedback to ensure that reviews were high-quality, constructive, and fair, rather than to serve as an additional reviewer. Overall, 35 mentors and 55 mentees participated.

In a follow-up survey with 13 mentor and 21 mentee responses, 85% of participants reported that they would participate again given the opportunity, and more than 79% reported that they felt that the feedback session was successful or very successful at improving review quality. Moreover, 65% of participants reported that they established a meaningful connection with their mentor/mentee. In addition, 24% of participants reported some difficulty in establishing contact with their assigned mentor or mentee, which poses an opportunity for improvement

in future iterations of the program. Future organizers should consider how to best promote accountability among those participating in the mentorship program. Qualitative feedback for future workshops included simplifying the mentorship recruitment process, recruiting additional mentors, and reducing the number of papers mentors discussed with each mentee.

3. Analysis of Works

3.1. Structured Data Analysis

Alongside their submissions, authors were asked to include structured data about their methods and focuses. Figure 1 shows the frequencies of author-reported topics for papers accepted to ML4H, with electronic health records and computer vision as the most commonly reported topics.

Authors were also asked to label the core contribution areas of their work. Among accepted submissions, the most common reported contribution was "Core method development for a health-focused problem," with 50 papers. Additionally, 15 papers focused on "Direct application to clinical practice," 9 papers on "Direct application to clinical operations," and 5 papers on "Direct application to basic biomedical science," while 6 papers listed other core contribution areas.

3.2. Data and Methods

Figure 2 presents a breakdown of the most frequent author-reported data types used among accepted papers. Medical imaging and structured Electronic Health Record (EHR) data were the most commonly used data types, and epidemiological data was also popular this year. Out of this year's accepted papers, 47% reported using at least one dataset that is publicly available. The most common datasets used were MIMIC-III (Johnson et al., 2016) and the UK Biobank (Sudlow et al., 2015). Furthermore, 66% of accepted papers committed to releasing code for their projects.

This year, authors were also asked to submit information about the statistical tests used for their analyses. Out of the accepted papers, 58% reported computing variances/confidence intervals. Full statistics are reported in Table 1. The organizers aim to continue to record this information over the years.

Among the papers accepted to ML4H this year, 70% of submissions reported performing interpretability analyses. There was a 51% acceptance rate for papers performing these analyses, versus a 30% acceptance rate for those that did not. While other factors may have contributed to these differences in acceptance rates, the organizers do hope that interpretability analyses continue to be performed when applicable.

In addition, 17% of accepted papers reported performing multi-site validations. Acceptance rates were similar across both these groups of papers—44% for papers performing these analyses, and 42% for those that did not.

Figure 3 details the frequencies of methods used in accepted papers. Neural networks continued to be a major component of ML4H papers, with increased prominence of self-attention architectures this year.

3.3. Clinical Coverage

Continuing with the precedent set in previous years, ML4H collected information about clinical conditions covered in submissions. Authors were asked to self report their submissions' clinical coverage. A comparison of conditions covered in accepted papers against previous years' disease categories is presented in Table 2. Compared to prior years (2019 and 2020), more papers focused on specific diseases and conditions, and there

Figure 1: Frequencies of author-reported topics for accepted papers.

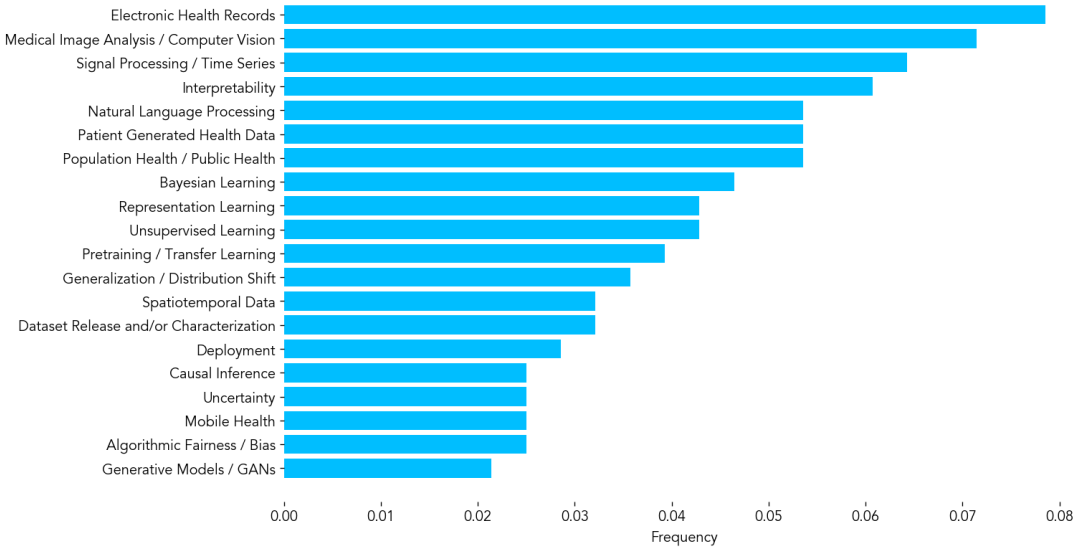


Figure 2: Frequencies of author-reported data types used in accepted papers.

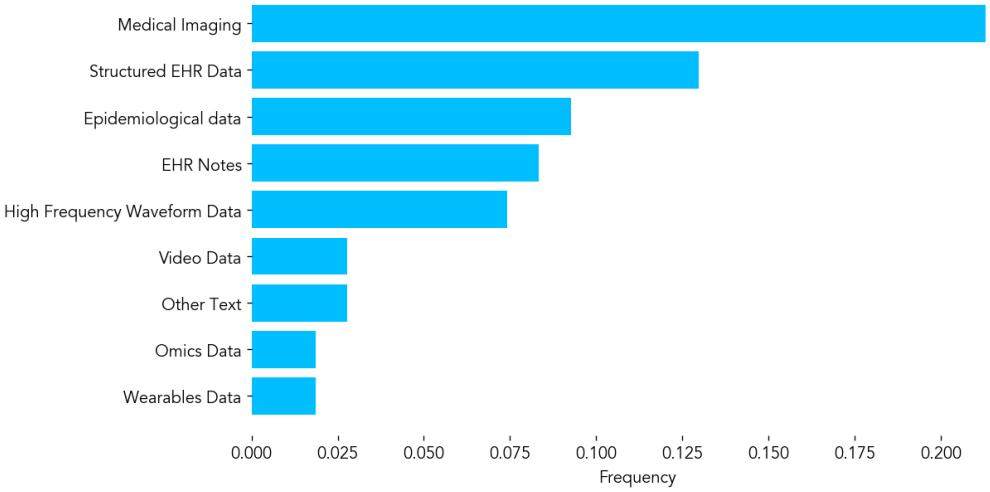
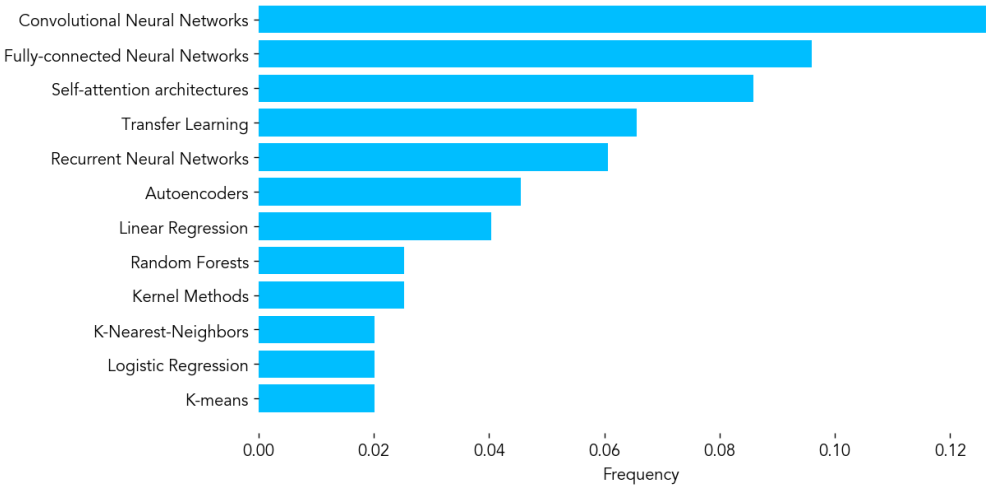


Table 1: Statistical tests conducted in accepted papers.

	2020
Variances/CI around point estimates were reported, but no hypothesis testing or statistical power analyses were performed in any model comparisons	34.1%
Variances/CI around point estimates were reported and statistical hypothesis testing or power analyses were performed in any model comparisons	23.5%
Only point estimates were computed and compared	21.2%
Not applicable	21.2%

Figure 3: Frequencies of author-reported methods used in accepted papers.



were more papers working within cardiovascular, pulmonary, and neurodegenerative diseases.

3.4. Paper Topics

We performed topic modeling over the free-text content of accepted papers using Latent Dirichlet allocation. We considered both the full proceedings papers and the extended abstracts for this study and the marginal topic distribution is shown in Figure 4. In addition to the topics from ML4H 2019 (Dalca et al., 2020), we also noted a new topic for "Covid-19" this year. According to the model's classifications, there has been an increase in papers related to "Causal Inference", "Uncertainty in ML", "Adversarial Robustness", "Neurological Disorders", and "TB and antibiotic resistance". On the other hand, there has been a reduction in papers related to "Genomics", "Genetics", "SDM" and "Privacy/Security" to the extent that no substantial topics were found related to these areas.

The increase in papers related to causal inference, probability estimation, adversarial robustness and uncertainty likely represents a growing interest to explore models beyond point estimates in machine learning for healthcare.

4. The ML4H Community

The ML4H workshop encourages the growth of the machine learning for health community across people of all backgrounds. The workshop introduced submission and reviewer mentorship programs this year, which encouraged participants from underrepresented backgrounds to apply. Additionally, 83% of submissions and 77% of accepted papers came from first authors who had not previously had work accepted to ML4H. The organizers continue to encourage participants from all backgrounds to submit to ML4H.

4.1. Representation

In an effort to better understand the ML4H community, the workshop collected demographic information for the first time this year. 54% of submitting authors identified as Man, while 35% identified as Woman and 10% preferred not to specify. 34% of submitting authors identified as Asian or Asian American, 28% identified as White, 17% preferred not to specify, and less than 5% identified as American Indian or Alaska Native, Black or African American, or Native Hawaiian or Other Pacific Islander. Moreover, 5% of submitting authors identified as being of Hispanic, Latino, or Spanish origin, and 10% preferred not to answer this question.

Finally, the workshop began to collect and present information about the institutional affiliations of submitting authors. Figure 5 shows the most commonly represented institutions across accepted papers.

By collecting and reporting this information, we hope to better track future efforts to improve the diversity of the ML4H community.

4.2. Clinician Involvement

ML4H continued to feature high rates of clinician involvement, with 60% of submissions involving clinicians in some form. Clinician involvement is compared to the 2018 and 2019 workshops in Table 3. There were slightly fewer submissions this year with clinicians as primary authors, and slightly more submissions with clinicians as secondary authors, involved as consultants, or acknowledged. Additionally, clinician involvement was mildly associated with higher acceptance rates, at 44% versus 41%.

Table 2: Clinical coverage of accepted papers.

	2018	2019	2020
Oncology	12.7%	15.5%	10.6%
Neurodegenerative	8.5%	6.8%	11.8%
Pulmonary	3.6%	4.9%	12.9%
Cardiovascular	11.4%	2.9%	9.4%
Diabetes	6.0%	0.9%	4.7%
No Specific Disease	25.9%	42.7%	23.5%

Figure 4: LDA topic distribution of accepted works in ML4H 2020.

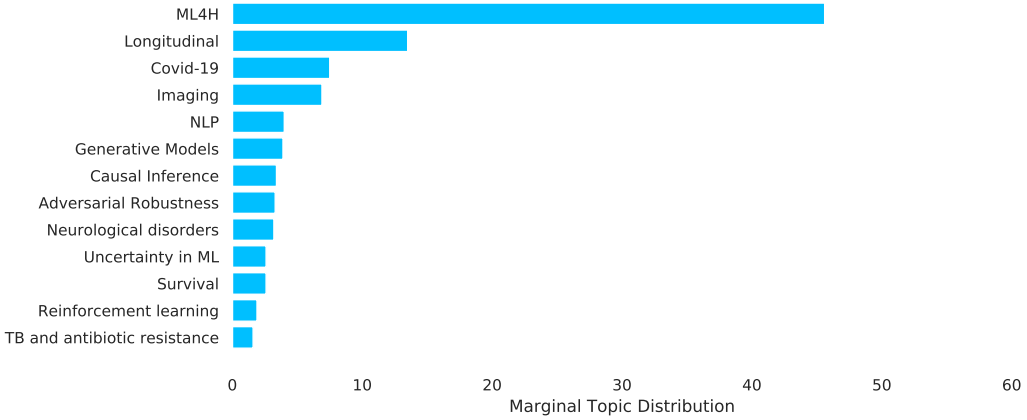


Figure 5: Most commonly represented institutions across accepted papers.

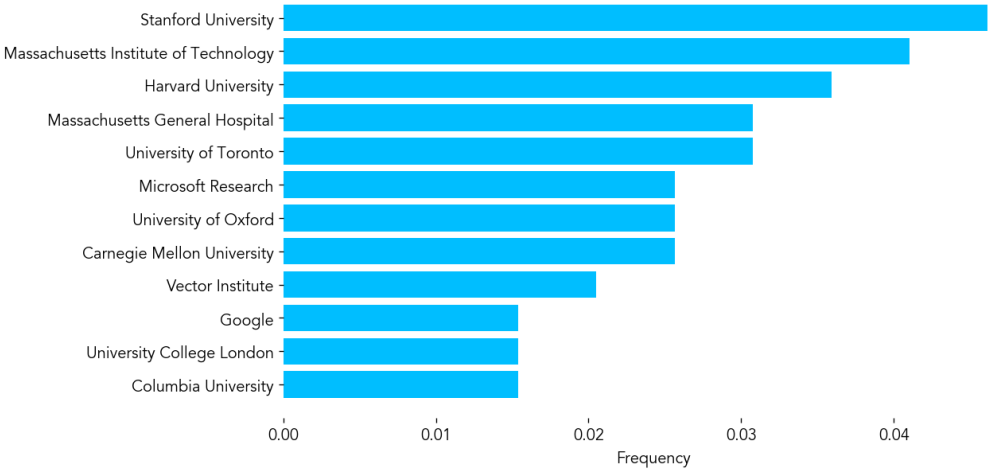


Table 3: Clinician involvement in paper submissions.

	2018	2019	2020
None	65.1%	41.0%	40.2%
Consultant/Acknowledged	3.6%	19.1%	21.2%
Primary author	7.8%	8.6%	5.3%
Secondary author	23.4%	31.4%	33.3%

5. Conclusion

The sixth Machine Learning for Health workshop (ML4H 2020) aimed to highlight the role of machine learning in expanding access to and quality of healthcare for all people. We further reflected on how to improve access to our own community, through mentorship programs for both submissions and reviewers. The workshop featured a varied program of invited and spotlight talks, panels and poster sessions. Despite the COVID-19 pandemic and the switch to a virtual format, interest in the intersection of machine learning and healthcare remains strong, and the workshop aimed to highlight and support this field and community.

6. Acknowledgements

6.1. List of Organizers

Stephanie Hyland (General Chair)
Emily Alsentzer (Senior Program Chair)
Matthew B. A. McDermott (Senior Program Chair)
Irene Y. Chen (Outreach/Comms Chair)
Subhrajit Roy (Publication Chair)
Charles Onu (Virtual Chair)
Allen Schmaltz (Finance Chair)
Ehi Nosakhare (Speaker Chair)
Jasvinder Kandola (Awards Chair)
Fabian Falck (Submission System Coordinator)
Suproteem K. Sarkar (Statistics/Surveys Coordinator)
Stephen Pfohl (Reviewer Mentorship Coordinator)
Ioana Bica (Submission Mentorship Coordinator)
Griffin Adams (Student Mentorship Coordinator)
Dani Kiyasseh (Email Coordinator)
Benjamin Akera (Virtualization)
Oliver J. Bear Don't Walk IV (Virtualization)
Aahlad Manas Puli (Finance Coordinator)

6.2. Advisory Committee

Brett Beaulieu-Jones (Harvard Medical School)
Andrew L. Beam (Harvard University)
Danielle Belgrave (Microsoft Research)
Adrian V. Dalca (Massachusetts General Hospital, Harvard Medical School, Massachusetts Institute of Technology)
Anna Goldenberg (University of Toronto, Vector institute, The Hospital for Sick Children)
Tristan Naumann (Microsoft Research)

6.3. Meta reviewers

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

Alexander Kotov, Allen Schmaltz, Andrew M. Dai, Brett Beaulieu-jones, Di Jin, Dianbo Liu, Edward Choi, Fabian Falck, George H. Chen, Girmaw abebe Tadesse, Griffin Adams, Ioana Baldini, Jason Fries, Jessica Schrouff, Jinsung Yoon, Joe Futoma, Kuheli Sai, Masoud Rouhizadeh, Matt Barnes, Mehdi Fatemi, Mohammad taha Bahadori,

Mukund Sudarshan, Neil Tenenholtz, Prithwish Chakraborty, Ruizhi Liao, Sarah Poole, Sarah Tan, Stephen Pfohl, Subba reddy Oota, Sudhir Rachapalle, Tzu-ming Hsu, William Boag, Yonatan Mintz, Yuan Liu.

6.4. List of 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:

Aahlad Puli, Aakanksha Rana, Aakash Kaku, Adam Yala*, Albee Ling, Alex Fedorov, Alison Callahan, Amelia Averitt*, Amy x. Lu, Anastasiia Kabeshova, Andre Nguyen, Angeline Yasodhara, Aniruddh Raghu, Anna Menacher, Aparna Balagopalan, Ari Ercole, Aris Gkoulalasdivanis*, Arman Cohan, Arun Baskaran, Arunachalam Narayanaswamy, Avik Ghose, Baihan Lin, Balamurali Murugesan, Benjamin Akera, Benjamin Huynh, Benjamin Kompa, Brett Marinelli, Brian Cozzi, Chandan Singh, Changhee Lee, Chantal Shaib, Charles Delahunt, Charles Onu, Chen Fang, Chirag Nagpal, Christian Bock, Christina Ji, Christopher Bridge, Chufan Gao, Corey Chivers, Craig Glastonbury, Dani Kiyasseh, David Belanger, David Fouhey, David Ledbetter, David Ouyang, David Zimmerer, Davina Zamanzadeh, Dawit Abtew, Deepak Sharma, Devin Taylor, Divya Shanmugam, Diyuan Lu, Dmitry V. Dylov, Edward De brouwer, Ehi Nosakhare, Elham Dolatabadi, Elias Chaibub neto, Elliot Schumacher, Emma Rocheteau*, Emmanuel Lagarde, Eric Lehman, Ethan Chi, Evan Cofer, Evgeniy Faerman, Eyal Rozenberg, Fabian Falck, Farah Shamout, Ferran Hueto puig, Fethi Jarray, Folakemi Kayode, Frank Soboczenski, Gautham Krishna Gudur, Giovanni Cina*, Gregory Rehm, Griffin Adams, Guillaume Ausset, Guillermo Sapiro, Gustavo Malkomes, Günter Klambauer, Hamid Tizhoosh, Han Yuan, Hang Yuan, Haoran Zhang, Harri Lähdesmäki, Harvineet Singh*, Hazrat Ali, Heather Berlin, Henghui Zhu, Henrique Aguiar, Hyunkwang Lee*, Ilya Valmianski, Imran Qureshi, Ioana Bica, Irene Chen, Irene Li, Ivan Klyuzhin, J. Matias Di martino, Jaan Altosaar, Jacob Deasy, Jason Wei, Jekaterina Novikova, Jelmer Wolterink, Jeremy Nixon, Jeremy Weiss, Jerry Wei, Jessica Forde, Jianbin Tang, Jieshi Chen, Jingshu Liu, Jocelyn Dunstan, John Giorgi, John Zech*, Jon White, Jonas Kemp, Jongoh Jeong, Justin Lovelace, Karl Øyvind Mikalsen, Katikapalli Subramanyam Kalyan, Kaushik Sarveswaran, Kexin Huang, Khawla Seddiki, Krishnakant Saboo, Kristen Severson, Kruti Pandya, Kyunghyun Cho, Laura-jayne Gardiner, Layne Price, Lina Sulieman*, Linda Moy, Lionel Fillatre, Louis Vaickus, Lovedeep Gondara, Luis Oala*, Lujie Chen, Maarten Poirot, Mallory Nobles, Marc Combalia, Mariona Carós, Marta Avalos, Matthew B. A. McDermott, Matthias Hüser*, Maurice Van keulen, Max Berrendorf*, Maxime Thibault, Mazdak Abulnaga, Md Shamsuzzaman, Md Abdullah al Hafiz Khan, Md Mahfuzur Rahman, Meghasyam Tummalacherla, Michael Dusenberry*, Michael Gensheimer, Michael Pinsky, Michael Sjoding, Michal Borsky, Min Lee, Minh Nguyen, Mohammad Eslami, Mohammad Khan, Monica Agrawal, Mudit Jain, Muhammad Arjumand Masood, Mukund Sudarshan, Nan Wu, Nandita Damaraju, Natalie Harris, Nataliya Sokolovska, Nathan Ng, Nazanin Makkinejad, Nelson Johansen, Nicola Sahar*, Nils Forkert, Nimshi Venkat Meripo, Numan Celik, Olga Kovaleva, Olga Perepelkina, Oliver Baclic, Oliver Bear Don't Walk iv, Orla Doyle, Pablo Morenomunoz, Pashmina Cameron, Peng Liao,

Pradyumna Suresha, Pragyansmita Nayak, Prashnna Gyawali, Praveen Gurunath bharathi, Praveer Singh, Preethi Raghavan, Qiao Jin, Qiyuan Hu, Rachid Riad, Rachneet Kaur, Rafael Sousa, Rafid Mahmood, Ramya Srinivasan, Renyu Zhang, Rina Singh, Rohit Singla, Rulin Shao, Russell Kunes, Ryan-rhys Griffiths, Sabri Boughorbel, Sadid Hasan, Saeed Hassanpour, Sahra Ghalebikesabi, Sai prabhakar Pandi selvaraj, Sajad Darabi, Sami Mian, Samuel Yang, Sana Tonekaboni, Sandeep Konam, Sandhya Prabhakaran, Sangameshwar Patil, Sarvesh Nikumbh, Saulo Pedro, Saurabh Gombar, Sedigheh Eslami, Seojin Bang, Seung Eun Yi, Sewale Demeke, Shalmali Joshi, Shameer Khadar, Shanya Sharma, Sheng Liu, Shreyasi Pathak*, Shruti Palaskar, Shuangjia Zheng, Simeon Spasov*, So Yeon Min, Sonali Parbhoo, Stefan Bonn, Stefan Harrer, Stephane Gaiffas, Stephanie Hyland, Subhrajit Roy, Sudeshna Das, Suhani Vora, Sumit Mukherjee, Supriya Nagesh, Suproteem Sarkar, Taylor Killian*, Tiffany Callahan, Tom Dhaene, Tom Van steenkiste, Tong Wu*, Tony Duan, Tony Y. Sun, Toru Shirakawa, Tristan Naumann, Umar Asif, Usman Roshan, Utkarsh Sarawgi, Vaishnavi Subramanian, Venet Osmani, Victor Rielly, Victor Rodriguez, Vincent Jeanselme, Vinith Suriyakumar, Vishwali Mhasawade, Vivek Natarajan, Weicheng Zhu, Wenqi Wei, Wenyuan Li, Wisdom D'almeida, Wouter Van Amsterdam*, Xavier Amatriain, Xiaomeng Dong, Xiaoyi Zhang, Xinyu Li, Yannet Interian, Yi-han Sheu, Yiyang Yu, Yogatheesan Varatharajah, Yogesh Kumar, Yoonyoung Park, Yuannan Cai, Yuchuan Wang, Yvonne Lui, Zach Shahn, Zaid Nabulsi, Zhaonan Sun, Zhengyuan Liu, Zijun Zhang.

* denotes reviewers who had at least two reviews explicitly recognized as excellent by the meta-reviewers.

References

Brett Beaulieu-Jones, Samuel G Finlayson, Corey Chivers, Irene Chen, Matthew McDermott, Jaz Kandola, Adrian V Dalca, Andrew Beam, Madalina Fiterau, and Tristan Naumann. Trends and focus of machine learning applications for health research. *JAMA network open*, 2(10): e1914051–e1914051, 2019.

Adrian V Dalca, Matthew BA McDermott, Emily Alsentzer, Samuel G Finlayson, Michael Oberst, Fabian Falck, Corey Chivers, Andrew Beam, Tristan Naumann, and Brett Beaulieu-Jones. Machine learning for health (ml4h) 2019: What makes machine learning in medicine different? In *Machine Learning for Health Workshop*, pages 1–9, 2020.

Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *Plos med*, 12(3): e1001779, 2015.