Machine Learning for Health (ML4H) 2019: What Makes Machine Learning in Medicine Different?

Adrian V. Dalca ADALCA@MIT.EDU
EECS, Massachusetts Institute of Technology
Radiology, Harvard Medical School

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

Emily Alsentzer EMILYA@MIT.EDU
CSAIL, Massachusetts Institute of Technology
Department of Biomedical Informatics, Harvard Medical School

Samuel G. Finlayson SAMUEL_FINLAYSON@HMS.HARVARD.EDU
CSAIL, Massachusetts Institute of Technology
Department of Biomedical Informatics, Harvard Medical School

Michael Oberst MOBERST@MIT.EDU
CSAIL, Massachusetts Institute of Technology

Fabian Falck FFALCK@CS.CMU.EDU
School of Computer Science, Carnegie Mellon University

Corey Chivers COREY.CHIVERS@PENNMEDICINE.UPENN.EDU
Predictive Health Care Group, University of Pennsylvania Health System

Andrew L. Beam ANDREW_BEAM@HMS.HARVARD.EDU
Department of Epidemiology, Harvard T.H. Chan School of Public Health

Tristan Naumann TRISTAN@MICROSOFT.COM
Microsoft Research

Brett Beaulieu-Jones BRETT_BEAULIEU-JONES@HMS.HARVARD.EDU
Department of Biomedical Informatics, Harvard Medical School

1. Introduction

The fifth Machine Learning for Health (ML4H 2019) workshop[1] was held on December 13, 2019, in conjunction with the thirty-third Conference on Neural Information Processing Systems (NeurIPS 2019), at the Vancouver Convention Center, Vancouver, Canada. This proceedings contains the 17 papers accepted to the inaugural ML4H Proceedings.

The goal of this year's workshop was twofold. First, it aimed to foster collaborations that meaningfully impact medicine by bringing together clinicians, health data experts, and machine learning researchers. Second, it provided a proceedings venue for technically excellent, mature work. The latter goal was motivated by increased interest in technical

1. The ML4H workshop website is https://ml4health.github.io/2019/.

innovations in machine learning that are tailored to the unique challenges of the healthcare domain. To this end, ML4H 2019 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 proceedings introduction, we start in Section 2 by describing the workshop itself, including a brief discussion of the event as well as a detailed summary of the paper selection process and submission statistics. In Section 3 we briefly analyze the accepted works, and offer commentary on the trends of research observed in this field, building on the analysis of the 2018 workshop (Beaulieu-Jones et al., 2019). Finally, we close with acknowledgments, including a list of all organizers and reviewers for ML4H 2019.

2. Workshop

2.1. Program

The ML4H 2019 workshop featured six invited talks from academia and industry, eight spotlight presentations from authors accepted at the venue, and a panel discussion with invited speakers. The speakers included Daphne Koller, Founder and CEO of insitro (Koller, 2019); Cian Hughes and Nenad Tomasev from Google Health (Hughes and Tomasev, 2019); Emily Fox, Director of Health AI at Apple and Associate Professor at the University of Washington (Fox, 2019); Luke Oakden-Rayner, Director of Medical Imaging Research at Royal Adelaide Hospital (Oakden-Rayner, 2019); Anna Goldenberg, Associate Professor at University of Toronto and Senior Scientist at SickKids Research Institute (Goldenberg, 2019); and Dale Webster and Lily Peng from Google Health and Google Brain AI (Webster and Peng, 2019). Speakers discussed applications of machine learning across diverse data modalities and topic areas, highlighting the growing breadth of ML4H both in research and application.

ML4H 2019 also featured a panel (Belgrave et al., 2019), in which panelists commented on the overall state of the field and answered audience questions, as well as spotlight talks from selected papers (Hughes and Tomasev, 2019; Asif, 2019; Rozenberg, 2019; Xu, 2019; Kapur, 2019; Jaeger, 2019; Wei, 2019; Raghu, 2019; Nagpal and Li, 2019).

2.2. Paper Selection

Suggested Topics. Authors were invited to submit papers relating to all aspects of ML4H, with a specific emphasis on the following themes: (1) Data and Labels are Noisy and/or Missing; (2) Causality and Confounding; (3) Do No Harm: Trust, Generalizability, Interpretability, and Reproducibility; (4) Deployment Challenges within Healthcare; (5) Dataset Shift; (6) Fairness and Bias; (7) Multi-modality, High-dimensionality; and (8) Privacy.

Submission Statistics. Despite a shorter submission timeline, the workshop saw continued growth in the number of submissions, receiving 111 full papers and 198 extended abstracts, up from 239 total submissions in 2018. The program committee consisted of 294

Table 1: Clinician involvement in submissions.

	2018	**2019**
None	65.1%	40.95%
Consultant/Acknowledged	3.6%	19.05%
Primary author	7.8%	8.57%
Secondary co-author	23.4%	31.43%

reviewers and completed a total of 1208 reviews. At least five reviews were completed for each proceedings track submission, and at least three for each extended abstract submission. Nineteen full papers were accepted for this proceedings (17% acceptance rate), of which seventeen chose to appear in the final proceedings, and another 29 were modified and accepted as extended abstracts with poster presentation. Of the submitted extended abstracts, 68 (34%; including the 29 modified full papers) were accepted for poster presentation. The extended abstracts were given the option to be included in an `arXiv` index at https://arxiv.org/abs/2002.01584 (McDermott et al., 2020).

3. Analysis of Accepted Works

3.1. Structured Data Analysis

As part of this year's submission process, we asked authors to self-report on various aspects of their paper (e.g., what level of involvement clinician collaborators had, or what their disease focus was). Reviewers were also asked to gauge the validity of these responses during review. Using these data, we characterize several trends in the 2019 accepted works compared to the ML4H 2018 analyses (Beaulieu-Jones et al., 2019).

Two important trends were an increased proportion of papers with clinical contributors and the increased usage of separate external validation datasets. Table 1 shows that in comparison to ML4H 2018, 24.15% fewer overall papers had no clinician involvement while a majority of accepted works this year had clinicians as either consultants or secondary co-authors (authors other than first, equal contributions or last). Table 2 shows an increase in the number of accepted works using multiple datasets and external datasets for validation. Half (50.0%) of the accepted works in 2019 used multiple datasets compared to just 17.4% in 2018.

This year's workshop also saw an increase in the number of works that were not disease-specific from 25.9% to 42.7% (Table 3). Additionally, both Cardiovascular and Diabetes saw large decreases from 2018 to 2019. Table 4 shows the types of data analyzed in accepted works. In 2018, "Other" made up only 3.6% of works but increased by more than 10% as the diversity of datatypes increased.

3.2. Topic Analysis

We also performed topic modeling over the free-text content of accepted papers. In Figure 1, we compare topics to those of the previous year (Beaulieu-Jones et al., 2019; Antropova et al., 2018). We observe a significant increase in papers under both the "General ML4H"

Table 2: Utilization of additional datasets for validation and potential for generalization.

	2018	**2019**
Single Dataset	83.6%	50.0%
Multiple Datasets	12.0%	37.1%
Separate Dataset for Evaluation	5.4%	12.9%

Table 3: Most common clinical conditions studied.

	2018	**2019**
Oncology	12.7%	15.5%
Neurodegenerative	8.5%	6.8%
Pulmonary	3.6%	4.9%
Cardiovascular	11.4%	2.9%
Diabetes	6.0%	0.9%
No Specific Disease	25.9%	42.7%

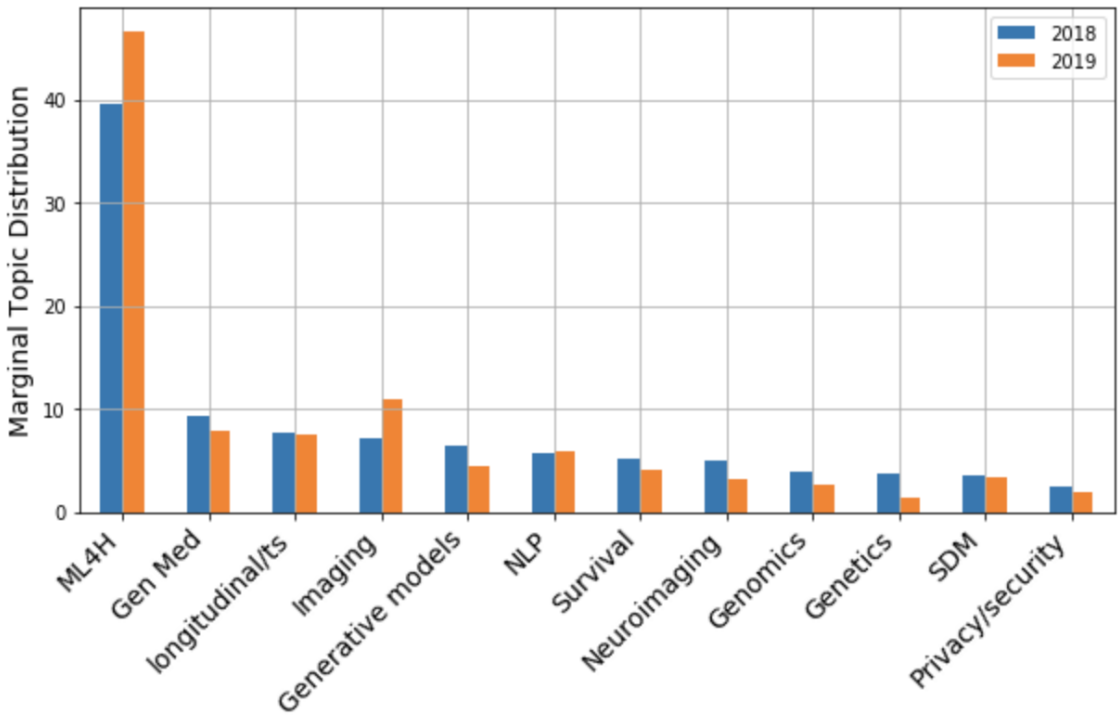

Figure 1: LDA topic distribution across accepted works between the ML4H workshops in 2018 and 2019.

Table 4: Types of data used in accepted work.

	2018	**2019**
Images	23.5%	25.5%
Structured (e.g. EHR, Epidemiological)	32.5%	24.2%
Text (e.g. Biomedical & Clinical Notes)	20.5%	22.9%
Other	3.6%	13.7%
Time Series (e.g. Vitals, Wearables)	10.2%	5.9%
-omics	6.0%	5.2%
Video	3.6%	2.6%

topic as well as the "Imaging" topic, and significantly fewer submissions under "Genetics" and "Generative models." The increases in imaging and decreases in genetics and generative models likely represent changes in the underlying research landscape. Increased imaging, in particular, could be a result of the increasing data availability of imaging data, in particular chest X-ray radiographs. An increased number of papers in the "unspecified" category is likely a result of a minor data shift between the two years leaving papers uncategorized, or the introduction of new topic areas that were not sufficiently popular in prior years to be discerned by the topic model.

4. Acknowledgments

We are extremely thankful for the support of the workshop organizers, the entire program committee, and our generous sponsor, Flagship Pioneering. In addition to supporting the workshop event itself, thanks to Flagship Pioneering, we were able to award 57 travel grants with a prioritization for trainees presenting at ML4H.

4.1. List of Organizers

Brett Beaulieu-Jones (General Chair)
Adrian V. Dalca (Senior Program Chair)
Emily Alsentzer (Program Chair)
Matthew B.A. McDermott (Program Chair)
Samuel G. Finlayson (Program Chair)
Stephanie Hyland (D&I Chair)
Jasvinder Kandola (Communications Chair)
Corey Chivers (Finance)
Amelia Averitt (Travel Grants)
Fabian Falck (Program)
Michael Oberst (Program)
Irene Chen (PR)
Tristan Naumann (Senior Advisor)
Andrew Beam (Senior Advisor)
Madalina Fiterau (Senior Advisor)

4.2. 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:

Nabeel Abdur Rehman
Anas Abidin
Mazdak Abulnaga
Monica Agrawal
Nitin Agrawal
Nauman Ahad
Adnan Akbar
Antonios Alexos
Bryce Allen
Raphael Aruleba
Azin Asgarian
Reza Ashrafi
Guillaume Ausset
Marta Avalos-Fernandez
Amelia Averitt
Turgay Ayer
Mohammad Taha Bahadori
Aparna Balagopalan
Ioana Baldini
Seojin Bang
Aubrey Barnard
Matt Barnes
Ali Bashashati
Arun Baskaran
Maxime Bassenne
Meseret Dastaw Bayeleygne
Andrew Beam
Oliver Bear Don't Walk
Danielle Belgrave
Max Berrendorf
Jean-Emmanuel Bibault
Ioana Bica
Davis Blalock
Willie Boag
Anthony Bonifonte
Michal Borsky
Nathaniel Braman
Christopher Bridge
Michael Brudno
Alison Callahan
Pashmina Cameron
Francesco Paolo Casale
Numan Celik
Elias Chaibub Neto
Ankush Chakrabarty
Chun-Hao Chang
Chun-Hung Chao
Marie-Laure Charpignon
Zhengping Che
George Chen
Irene Chen
Jieshi Chen
Lujie Chen
Richard Chen
Yang-En Chen
Corey Chivers
Edward Choi
Olivia Choudhury
Andrew Dai
Nandita Damaraju
Sajad Darabi
Saloni Dash
Edward De Brouwer
Jacob Deasy
Charles Delahunt
Sewale Demeke
Iman Deznabi
Diana Diaz Herrera
Xiaorong Ding
Mariia Dobko
Elham Dolatabadi
Jocelyn Dunstan
Michael Dusenberry
Chen Fang
Mehdi Fatemi
Alex Fedorov
Scott Fleming
Jessica Forde
Jason Fries
Chufan Gao
Laura-Jayne Gardiner
Leon Gatys
Nathan Gaw
Eskender Gebremicheal
Michael Gensheimer
Craig Glastonbury
Lovedeep Gondara
Jen Gong
Omer Gottesman
Yang Guo
Shaun Gupta
Katharina Hoebel
Thomas Hooven
Xinyuan Hu
Yimer Hussien
Ben Huynh
Stephanie Hyland
Hamza Ichchou
Jyoti Islam
Issa Issa
Fattaneh Jabbari
Paul Jäger
Hamza Javed
Vincent Jeanselme
Di Jin
Qiao Jin
Nelson Johansen
Alistair Johnson
Shalmali Joshi
Kenneth Jung
Mohammad Kachuee
Aakash Kaku
Jayashree Kalpathy-Cramer
Gael Kamdem De Teyou
Jaz Kandola
Satyananda Kashyap
Sehj Kashyap
Rachneet Kaur
Sarveswaran Kaushik
Julyan Keller-Baruch
Jonas Kemp
Aria Khademi
Muhammad Khan
Taylor Killian
Seungchan Kim
Günter Klambauer
Benjamin Kompa
Rahul Krishnan
Pavitra Krishnaswamy
Vinod Kurmi
Christoph Kurz
Emmanuel Lagarde
Adam Lavertu
David Ledbetter
Hyunkwang Lee

Gal Levy-Fix
Kathleen Lewis
Wenyuan Li
Xinyu Li
Paul Pu Liang
Ruizhi Liao
Baihan Lin
Dianbo Liu
Sheng Liu
Yuan Liu
Justin Lovelace
Ming Yang Lu
Ming Yu Lu
Pierre Machart
Gautam Machiraju
Rafid Mahmood
Maggie Makar
Gustavo Malkomes
Jacky Mallett
Ben Marafino
Brett Marinelli
Razvan Marinescu
Daniel Mas Montserrat
Vishwali Mhasawade
Sami Mian
Andrew Miller
So Yeon Min
Yontan Mintz
Pablo Moreno-Munoz
Mehul Motani
Sumit Mukherjee
Balamurali Murugesan
Zaid Nabulsi
Chirag Nagpal
Vivek Natarajan
Tristan Naumann
Meike Nauta
Pragyansmita Nayak
Daniel Neil
Antoine Neuraz
Minh Nguyen
Amber Nigam
Sarvesh Nikumbh
Beau Norgeot
Luke O'Connor
Luke Oakden-Rayner
Eric Oermann
Olugbenga Olamijuwon
Mustapha Oloko-oba
Subba Reddy Oota
Agni Orfanoudaki
David Ouyang
Danielle Pace
Shruti Palaskar
Sai Prabhakar Pandi Selvaraj
Kruti Pandya
Sonali Parbhoo
Mahboobeh Parsapoor
Dibyajyoti Pati
Arijit Patra
Badri Patro
Saulo Pedro
Olga Perepelkina
Ioakeim Perros
Jacob Pfau
Stephen Pfohl
Sarah Poole
Sandhya Prabhakaran
Purva Pruthi
Aniruddh Raghu
Deepta Rajan
Narges Razavian
Gurunath Reddy
Gregory Rehm
Emma Rocheteau
Usman Roshan
Hagai Rossman
Frank Rudzicz
Hamed Sadeghi
Nicola Sahar
Kuheli Sai
Noor Sajid
John Santerre
Guillermo Sapiro
Allen Schmaltz
Jessica Schrouff
Marc-Andre Schulz
Khawla Seddiki
Daniel Severo
Kristen Severson
Khader Shameer
Md Shamsuzzaman
Divya Shanmugam
Pulkit Sharma
Shanya Sharma
Yiqiu Shen
Toru Shirakawa
Bhanu Singh
Harvineet Singh
Rina Singh
Rohit Singla
Frank Soboczenski
Nataliya Sokolovska
Hossein Soleimani
Rafael Sousa
Ramya Srinivasan
Adarsh Subbaswamy
Ram Subramanian
Vaishnavi Subramanian
Lina Sulieman
Mimi Sun
Tony Sun
Zhaonan Sun
Pradyumna Byappanahalli Suresha
Madhumita Sushil
Antonio Sze-To
Girmaw Abebe Tadesse
Sarah Tan
Ryutaro Tanno
Colin Targonski
Devin Taylor
Thomas Taylor
Neil Tenenholtz
Anshul Thakur
Maxime Thibault
Elizabeth Tran
Anusua Trivedi
Anup Tuladhar
Wouter van Amsterdam
Tom Van Steenkiste
Chad Vanderbilt
Rohit Vashisht
Rielly Victor

Nyamos Waigama
Dennis Wang
Lucy Wang
Yunlong Wang
Sangeeta Warrier
Jason Wei
Anthony Wertz
Sarah Wiegreffe
Mario Wieser
Tom Woolf
Eric Wu
Sidra Xu
Xiaojing Xu
Chhavi Yadav
Adam Yala
Dig Vijay Kumar Yarlagadda
Kamer Yuksel
Lee Zamparo
John Zech
Haoran Zhang
Luna Zhang
Xiaoyi Zhang
Degui Zhi
Henghui Zhu
Taiyu Zhu
Weicheng Zhu

References

Natalia Antropova, Andrew L Beam, Brett K Beaulieu-Jones, Irene Chen, Corey Chivers, Adrian Dalca, Sam Finlayson, Madalina Fiterau, Jason Alan Fries, Marzyeh Ghassemi, Mike Hughes, Bruno Jedynak, Jasvinder S. Kandola, Matthew McDermott, Tristan Naumann, Peter Schulam, Farah Shamout, and Alexandre Yahi. Machine learning for health (ML4H) workshop at neurips 2018. *arXiv preprint arXiv:1811.07216*, 2018.

Umar Asif. Privacy Preserving Human Fall Detection using Video Data, December 2019. URL https://slideslive.com/38923239.

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.

Danielle Belgrave, Anna Goldenberg, Cian Hughes, and Luke Oakden-Rayner. Panel, December 2019. URL https://slideslive.com/38923244.

Emily Fox. Models of Cognition: From Predicting Cognitive Impairment to the Brain Networks underlying Complex Cognitive Processes, December 2019. URL https://slideslive.com/38923233.

Anna Goldenberg. Predicting Cardiac Arrest Design for Deployment, December 2019. URL https://slideslive.com/38923245.

Cian Hughes and Nenad Tomasev. A clinically-applicable approach to the continuous prediction of acute kidney injury, December 2019. URL https://slideslive.com/38923234.

Paul Jaeger. Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection, December 2019. URL https://slideslive.com/38923239.

Arnav Kapur. Non-Invasive Silent Speech Recognition in Multiple Sclerosis with Dysphonia, December 2019. URL https://slideslive.com/38923241.

Daphne Koller. Machine learning: A new approach to drug discovery, December 2019. URL https://slideslive.com/38923235.

Matthew B.A. McDermott, Emily Alsentzer, Sam Finlayson, Michael Oberst, Fabian Falck, Tristan Naumann, Brett K. Beaulieu-Jones, and Adrian V. Dalca. ML4H Abstract Track 2019. *arXiv preprint arXiv:2002.01584*, 2020.

Chirag Nagpal and Xinyu Li. Deep Survival MAchines: Fully Parametric Survival Regression, December 2019. URL https://slideslive.com/38923236.

Luke Oakden-Rayner. Preventing disasters: Why Safety Is The Foundation Of Medical Machine Learning, December 2019. URL https://slideslive.com/38923243.

Maithra Raghu. Transfusion: Understanding Transfer Learning for Medical Imaging, December 2019. URL https://slideslive.com/38923236.

Eyal Rozenberg. Localization with Limited Annotation for Chest X-Rays, December 2019. URL https://slideslive.com/38923241.

Dale Webster and Lily Peng. Deep Learning and Healthcare, December 2019. URL https://slideslive.com/38923237.

Jerry Wei. Generative Image Translation for Data Augmentation in Colorectal Histopathology Images, December 2019. URL https://slideslive.com/38923239.

Xiaojing Xu. Pain Evaluation in Video using Extended Multitask Learning from Multidimensional Measurements, December 2019. URL https://slideslive.com/38923241.