Cancer AI Research: Computational Approaches Addressing Imperfect Data

Virtual Workshop
April 3-4, 2023

Workshop Description: The application of AI to cancer research holds promise to accelerate new discoveries, enable early detection, improve diagnosis, and spur development of new therapies for cancer. Machine learning and other forms of AI have made a significant impact in some areas of cancer research, but the full promise of data-driven approaches has been elusive. While there are important ongoing efforts to collect and produce large, well-annotated datasets to support the training of robust deep learning models, the heterogeneity and complexity of cancer, along with privacy and bias concerns, continues to limit the application of AI methods to many critical areas of cancer research. There is a need for foundational advances in machine learning that can operate on incomplete, noisy, unbalanced and/or biased data across the cancer research continuum.

The goals of this workshop are to (1) examine the state of the science for AI methods designed to operate on noisy, complex, or low-dimensional data, (2) explore how these methods may be applied to key areas of cancer research, and (3) discuss processes for identifying the biological questions that will motivate further advances in machine learning. This workshop will highlight the importance of leveraging advances across fields to accelerate cancer research and discovery through AI.

Day 1: April 3, 2023

(all times EDT)

11:00 - 11:30 **Welcome and opening remarks**

NCI: Origin and purpose of the workshop
Opening presentation: Caroline Uhler, Broad and MIT
"*Cancer Research and AI - A 2-Way Street*"

11:30 - 1:00 **Session 1: Expanding classical structure prediction with machine learning towards drug discovery**

Session Chair: Trey Ideker, UC San Diego
Introductory comments from the chair (5 min)

Anima Anandkumar, Cal Tech and NVIDIA
"Role of generative AI in chemistry and biology" (15 min)

Andrej Sali, UC San Francisco
"*From integrative structural biology to cell biology*" (15 min)

Jure Leskovec, Stanford University
"Machine learning for cell-type discovery and identification in spatially resolved single-cell data" (15 min)

Panel discussion (speakers + additional panelists)
Rick Stevens, Argonne National Laboratory (5 min)
Sergey Ovchinnikov, Harvard (5 min)
Discussion (30 min)

1:00 - 1:15 **BREAK**

1:15 - 2:45 **Session 2: Chemical, genetic, and mechanical perturbations for understanding mechanisms in cancer: Extrapolating beyond existing data**

Session Chair: Fabian Theis, Helmholtz Munich
Introductory comments from the chair (5 min)

Yoshua Bengio, Université de Montréal
"*Causal Discovery with GFlowNets*" (15 min)

GV "Shiva" Shivashankar, ETH Zurich
"*Mechano-Genomics of Tumor Microenvironment*" (15 min)

Smita Krishnaswamy, Yale University
"*Learning generalizable perturbation models with manifold learning and dynamics*" (15 min)

Panel discussion (speakers + additional panelists)
Paquita Vazquez, Broad Institute (5 min)
Byung-Jun Yoon, Texas A&M University and Brookhaven National Laboratory (5 min)
Discussion (30 min)

2:45 - 3:00 **BREAK**

3:00 - 4:30 **Session 3: Multimodal learning in data limited contexts: Leveraging tissue-level data for understanding cell-cell interactions in cancer**

Session Chair: Dana Pe'er, Memorial Sloan Kettering
Introductory comments from the chair (5 min)

Atul Deshpande, Johns Hopkins University
"*Mapping molecular interactions in tumor microenvironments through latent space analysis of spatial transcriptomics*" (15 min)

Elham Azizi, Columbia University
"*Probabilistic modeling of dynamics in the tumor microenvironment*" (15 min)

Livnat Jerby, Stanford University
"*Structure, location, and function: From natural to engineered cellular circuits*" (15 min)

Panel discussion (speakers + additional panelists)
Marianna Rapsomaniki, IBM Research (5 min)
Arjun Krishnan, University of Colorado (5 min)
Discussion (30 min)

Day 2: April 4, 2023
(all times EDT)

11:00 - 11:05 **Welcome to Day 2**

11:05 - 12:35 **Session 4: Making use of large-scale, structured clinical research data and image repositories**

Session Chair: Ziad Obermeyer, UC Berkeley
Introductory comments from the chair (5 min)

Chris Probert, InSitro
"*Learned Embeddings Enable Multi-Modal Clinical Data Fusion"* (15 min)

James Zou, Stanford University
"*Advancing precision oncology with real-world clinico-genomics and spatial omics*" (15 min)

Mihaela van der Schaar, University of Cambridge
"*How can we build digital twins from complex, messy, incomplete and private real-world clinical data*" (15 min)

Panel discussion (speakers + additional panelists)
Lily Peng, Verily (5 min)
Matthew Lungren, Microsoft/UCSF (5 min)
Discussion (30 min)

12:35 - 12:45 **BREAK**

12:45 - 2:15 **Session 5: Improving modeling of real-world evidence data in clinical research and clinical trial design**

Session chair: Tianxi Cai, Harvard T.H. Chan School of Public Health
Introductory comments from the chair (5 min)

Sean Khozin, MIT
"*Addressing external validity deficits in conventional cancer clinical trials with real-world evidence*" (15 min)

Limor Appelbaum Beth Israel Deaconess and Martin Rinard, MIT
"*Challenges and solutions for building cancer predictive models with EHR data*" (15 min)

Ryan Copping, Roche & Genentech
"*Getting more out of Real-World Data with Advanced Analytics*" (15 min)

Panel discussion (speakers + additional panelists)
Donna Rivera, FDA (5 min)
Khaled El Emam, University of Ottawa (5 min)
Discussion (30 min)

2:15 - 2:25 **BREAK**

2:25 - 3:15 **Session 6: Cross-cutting discussion with session chairs**

Session chair: Olivier Gevaert, Stanford University
Introductory comments from the chair (10 min)

Panel Discussion (40 min)
Caroline Uhler
Trey Ideker
Dana Pe'er
Ziad Obermeyer
Tianxi Cai

3:15 - 3:30 **Closing Remarks**

Workshop Chairs:
Caroline Uhler, MIT and Broad Institute
Olivier Gevaert, Stanford University

NCI Planning Committee:
Juli Klemm
Jennifer Couch
Sean Hanlon
Natalie Abrams
Keyvan Farahani
Emily Greenspan
Paul Han
Roxanne Jensen
Jerry Li
Catherine Schweppe