Biomedical Informatics Grand Rounds

Wednesday, November 1, 2023
3:00 pm – 4:00 pm

The State of Protein Structure Prediction and Friends

Mohammed AlQuraishi, PhD

Assistant Professor
Department of Systems Biology at Columbia University, NY

Remote Access
Join Zoom Meeting https://stonybrook.zoom.us/j/95617197636?pwd=KytzZ2pVRG9SZGpKZUtpNXJISjNjZz09
Meeting ID: 956 1719 7636 Passcode: 924293

In-Person talk: Medical and Research Translation (MART)Building, Room location 7M-0602

Abstract: AlphaFold2 revolutionized structural biology by accurately predicting protein structures from sequence. Its implementation however (i) lacks the code and data required to train models for new tasks, such as predicting alternate protein conformations or antibody structures, (ii) is unoptimized for commercially available computing hardware, making large- scale prediction campaigns impractical, and (iii) remains poorly understood with respect to how training data and regimen influence accuracy. Here we report OpenFold, an optimized and trainable version of AlphaFold2. We train OpenFold from scratch and demonstrate that it fully reproduces AlphaFold2's accuracy. By analyzing OpenFold training, we find new relationships between data size/diversity and prediction accuracy and gain insights into how OpenFold learns to fold proteins during its training process.

Bio: Mohammed AlQuraishi is an Assistant Professor in the Department of Systems Biology and a member of Columbia's Program for Mathematical Genomics, where he works at the intersection of machine learning, biophysics, and systems biology. The AlQuraishi Lab focuses on two biological perspectives: the molecular and systems levels. On the molecular side, the lab develops machine learning models for predicting protein structure and function, protein-ligand interactions, and learned representations of proteins and proteomes. On the systems side, the lab applies these models in a proteome-wide fashion to investigate the organization, combinatorial logic, and computational paradigms of signal transduction networks, how these networks vary in human populations, and how they are dysregulated in human diseases, particularly cancer. His research has been published in Nature Method, Nature Biotechnology, Nature Genetics and PNAS.

Dr. AlQuraishi earned an MS in statistics and a PhD in genetics from Stanford University. He subsequently joined Harvard Medical School as a Departmental Fellow and a Fellow in Systems Pharmacology, where he developed the first end-to-end differentiable model for learning protein structure from data. He joined the Columbia Faculty in 2020.

Educational Objectives:

1. Learn about the AlphaFold2 machine learning model for protein structure prediction.
2. Learn about how well such models generalize to new regions of protein structure space unseen during training.
3. Learn about state-of-the-art methods in related areas of molecular modeling, including modeling of conformational ensembles and protein-ligand docking.

Disclosure Statement: The faculty and planners have no relevant financial relationship with ineligible companies, whose primary business is producing, marketing, selling, re-selling, or distributing health care products used by or on patients.

Continuing Medical Education Credits: The School of Medicine, State University of New York at Stony Brook, is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians. The School of Medicine, State University of New York at Stony Brook designates this live activity for a maximum of **1** ***AMA PRA Category 1 Credits™.*** Physicians should only claim credit commensurate with the extent of their participation in the activity.