

Sahil Arora
linkedin.com/in/sahilarora3
arolihas.github.io

Georgia Institute of Technology
MS Machine Learning
BS Computer Science (Intelligent Modeling and Simulation)
CPE Lyon – Certificate in Molecular Biology

August 2015 – May 2021
GPA 3.84/4.00
Highest Honors
May 2017 – June 2017

EXPERIENCE

VantAI Machine Learning Engineer

July 2021 – Current

- Developed geometric deep learning models that characterize the molecular surface of proteins and identify likely non-partner specific binding sites by training on structural features with curated datasets
- Used contrastive learning on dMaSIF embeddings of point cloud representations to achieve 86% accuracy on observed PPI
- Created command line utility and Dockerized endpoint for efficient prediction of potential binding sites on human proteome
- Applied model to downstream analysis for surface segmentation, partner specific prediction, and scoring simulated poses
- Trained pairwise alignment-free compatibility model that predicted interaction likelihood of interfaces between surface patches on the interface, achieving 99% sensitivity and 65% specificity on key internal targets against E3s
- Validated compatibility model on STRINGdb and BioGRID protein interaction databases, achieving 0.95 and 0.15 enrichment factor respectively (fraction of E3:target interactions with predicted probability > 0.65 that are present in the database)
- Built dashboard for visualizing surface patches, learned embeddings, and chemical features in predicted protein-protein interactions of key targets to accelerate prioritization screening with external clients in drug discovery process

Georgia Institute of Technology Head Teaching Assistant

August 2020 – May 2021

- Managed team of 15 TAs and created assignments and quizzes for 500+ person online masters Deep Learning course

IBM – Chief Information Office Software Engineering Internship

June 2020 – May 2021

- Wrote and optimized Selenium tests in Java for a Cucumber test automation framework in a Salesforce platform

Merck Pharmaceuticals – Manufacturing Division Machine Learning Internship

June 2020 – May 2021

- Developed CNN for classification of defective vaccine vials to capture potentially up to \$0.5M per year in lost revenue
- Created Flask web application with SQLite3 backend to upgrade ticket management system for contracting service

Virginia Systems and Technology Software Engineering Internship

January 2019 – May 2019

- Implemented features for an Angular/Typescript/Java web app simulating flight path triangulation of targets

Novo Nordisk – Clinical Medical and Regulatory Operations Data Analytics Internship

May 2018 – August 2018

- Built modal interface for modifying budget line items with Excel VBA to optimize Business Management team processes

RESEARCH

Biological Systems and Engineering Laboratory

August 2019 – August 2020

- Applied feature selection with scikit-learn to identify minimal metabolite concentration sets and classify cell phenotypes

BioMEMS and Biomechanics Laboratory

January 2018 – August 2019

- Implemented fast online peak detection algorithm for AFM probe-surface contact with force measurement data
- Attached gold nanorods to silica particles via thiols that move under directed laser light to study photothermal propulsion

PROJECTS

Motif Embedded Disease Pathway Prediction in Protein-Protein Interaction Networks (Tensorflow+Keras, stellargraph)

- Embedded motif frequency distribution in GCN to improve node classification task from 8% to 56% F1-micro score

Knee MRI Diagnosis Modeling (PyTorch, fastai)

- Applied multiheaded attention and semi-supervised learning to improve AlexNet classification of tears to 95% AUC

SafeSpace Mobile App (Kotlin, Google Cloud, CouchDB, wireshark)

- Built app that reports how busy a place is with WiFi tracking and crowdsourced locations for COVID-19 hackathon

LEADERSHIP

Data Science at GT External Affairs Member

August 2019 – December 2020

- Created marketing plan and contacted sponsors as part of a team for Hacklytics data science hackathon

International Society of Pharmaceutical Engineering Founder and Vice President

August 2017 – August 2019

- Led organization of over fifty members and held networking events to connect students to faculty and industry

SKILLS

Languages: Python, Java, R, Dart (Flutter), C++, SQL

Machine Learning: PyTorch, TensorFlow, scikit-learn, NumPy, SciPy, pandas, jupyter, matplotlib, plotly

Web: JavaScript (Angular, Node), TypeScript, Flask, d3.js

Software Engineering: Google Cloud, Docker, Argo, Gradle, Maven, Jenkins, Jira, Agile

Operating Systems: Unix (Ubuntu, MacOS, CentOS), Windows