Mihir Bafna

mbafna@gatech.edu | 858-603-6580 | LinkedIn | GitHub

Education

Georgia Institute of Technology — Graduation May 2023

B.S. Computer Science (Senior Standing) - GPA: 4.0

Relevant Coursework - Deep Learning, Machine Learning, Linear Algebra, Applied Combinatorics, Probability Theory, Graph Theory, Data Structures, Design & Analysis of Algorithms, Computer Architecture, Discrete Mathematics, Objects and Design, Multivariable Calculus

Publications/Patents

- Rajkumar U., Javadzadeh S., **Bafna** M., Wu D., Yu R., Bafna V., Shang J. (2022). DeepViFi: Detecting Oncoviral Infections in Cancer Genomes using Transformers. Accepted by **ACM-BCB**, 2022
- Nguyen, Mora-Blanco, Turner, Christiansen, **Bafna**. 2021. Computer-implemented methods for quantitation of features of interest in whole slide imaging. PCT/US2021/022308, filed March 15, 2021. **Provisional patent**.

Experience

Jian Ma Lab (ML & CompBio) | **Carnegie Mellon University** — Pittsburgh, PA

Student Researcher (Just Started) — *Dec 2021 – Present*

- Using Graph Representational Neural Networks (GNNs) and Hypergraph GNNs to understanding the 3-D structure of the human genome (compartments, TADs, etc.) from sc-Hi-C sequencing data
- Using Hypergraph GNNs for understanding genetic interactions, specifically trigenic, and enrichment in protein complex formation

Zhang Lab (ML & Bioinformatics) | **Georgia Institute of Technology** — Atlanta, GA

Student Researcher — *Nov 2021 – Present*

- Analysis of cell clusterings from the learned latent factors of the matrix factorization for identifying marker features (genes, regions, etc)
- Starting project on understanding extracellular interactions (cell-cell inference, subnetwork comparison) with Graph Neural Networks and Spatial Transcriptomics data

Boundless Bio Inc — San Diego, CA

SWE/Bioinformatics Research Intern — *Aug 2020 – Dec 2021*

- **Provisional patent** for creating metaDetect: a computer vision algorithm for identifying metaphase spreads in stained whole slide images of cancer cells using image analysis filtering/techniques.
- Created automated pipeline for whole slide imaging → metaDetect (CV) → ecDNA quantification (CNN)
- **Writing publication** for Whole Slide Imaging and the metaDetect algorithm (above)
- Implemented better workflow and cloud in ecDNA Analytics
- Used machine learning (precision recall / roc curves) for optimization of image analysis parameters
- Frameworks|Languages: scikit-learn, OpenCV, Jupyter Notebooks, Google Cloud | (Python)

CompBio Lab | **UC San Diego** — San Diego, CA

Student Researcher — *July 2021 – Sep 2021*

- Developed Machine Learning (Random Forest) model to classify Human Papillomavirus genomes into their four broad classes (Alpha, Beta, Gamma, Other).
- Created a faster and more accurate classifier with machine learning gradient boosting techniques (LightGBM & XGBoost)
- **Paper accepted** for DeepViFi: Detecting Novel Viral Integration in Cancer Genomes using a transformer architecture and LightGBM for further family classification
- Frameworks|Languages: LightGBM, XGBoost, scikit-learn, Pandas, NumPy | (Python, JS)

Adibi Lab (Medical Imaging) | **Georgia Institute of Technology** — Atlanta, GA

Student Researcher — *Jan 2021 – Present*

- Using deep learning to detect lung diseases from X-ray images (Segmentation/Classification)
- Created saliency maps (GRAD-CAM method) to understand and visualize our CNN model's predictions.
- Frameworks|Languages: TensorFlow, PyTorch, Flask, React, Azure | (Python, JS)

Projects

CovidAutoEncoder | ***Python, Pytorch, Seq2Vec, PCA, UMAP*** GitHub Repository

- Created and tested various autoencoder architectures (Convolutional, Variational, Linear) to learn latent representations of SARS-Cov-2 genomes and most importantly: the latent differences between the variant genomes
- Once the lower dimensional latent representations were learned by training the autoencoders on the Alpha and Delta (more prevalent) strains, we encoded Omicron genomes into these representations and were able to clearly distinguish the variants with a simple PCA.

ecDNA Analytics | ***Flask, Python, Javascript*** GitHub Repository

- Web platform that integrates the metaDetect and ecSeg tools with respective visualizations for the purpose of extrachromosomal DNA analysis and annotations.
- View at ecDNA Analytics

MatrixOperationsLibrary | ***Python*** GitHub Repository

- Created module that contains functions for theoretical interpretations in Linear Algebra
- Essentially a bootleg NumPy that helped me deeply understand the purpose of efficient matrix computations and their use in computer science

LANOS | ***Javascript, React, Express, Node.js, MongoDB*** GitHub Repository

- Startup platform for connecting HS students with career/volunteer opportunities.
- Developed full stack matching platform as part of the Startup Lab and Idea 2 Prototype Create-X programs at Tech.

VetAssist | ***React Native, MongoDB, Alexa Skills*** GitHub Repository

- Mental health support platform that allows veterans to connect with others nearby and integrates sentiment analysis of their daily virtual journals to assess mental health
- Virtual journaling was created with the Alexa Skills library so that the veterans could speak to a virtual "friend" to journal their day
- Winner of HealthTech Hacks GT

AtlantaEnvJustice | ***Flask, Python, Javascript*** GitHub Repository

- Interactive map website to highlight the disproportionate modes of quality and environmental injustice in Atlanta.
- Uses the mapbox api and an aesthetic UI to bring to light certain injustices (racial, financial, etc.) regarding environmental inequality in counties around the Atlanta metropolitan area.

Awards/Honors

Georgia Tech Faculty Honors, (4.0 GPA) | Awarded Each Semester — Dec 2020, May 2021, Dec 2021, May 2022
HealthTech Hacks @ GT | First Place Winner — March 2021
Grand Challenges @ GT | First Place Winner for best project idea — May 2021
Westview HS Honors | Top Math Student in HS Graduating Class (600) — June 2020
Westview HS Honors | Top 3% GPA in HS Graduating Class (600) — June 2020

Skills

Programming Languages: Python, Java, Javascript, HTML5/CSS, C, Matlab, R

Tools | **Frameworks:** TensorFlow, PyTorch, OpenCV, Jupyter, numpy, Scikit, Flask, Git, MERN Stack, Google Cloud, Raspberry Pi, NVIDIA Jetson, Arduino, pysam, Biopython