Harshita Saha

408-839-5639 • Boston, MA • hsaha@hsph.harvard.edu • harshita.saha2020@gmail.com

SKILLS

Languages: Python | Bash | R | SQL | Java | C++ | C

Frameworks and Tools: PyTorch | TensorFlow | Bioconductor | Git | Jupyter | HPC | Docker | Scrum

Selected Packages: pandas | NumPy | SciPy | scikit-learn | matplotlib | seaborn | dplyr | tidyr | ggplot2

EDUCATION

Harvard University | Cambridge, MA Sep. 2024 - May 2026
M.S. Computational Biology and Quantitative Genetics Cumulative GPA: **4.00/4.00**

University of California San Diego | San Diego, CA Sep. 2020 - June 2024
B.S. Bioinformatics w/ Data Science minor Cumulative GPA: **3.96/4.00**

Relevant Coursework: Machine Learning for Computational Biology, Deep Learning for Biomedical Data, AI in Medical Imaging, Advanced Bioinformatics, Data Analysis and Inference, Computing for Big Data, Data Management, Advanced Data Structures, Algorithm Design and Analysis, Applied Regression Analysis

EXPERIENCE

Computational Biology Research Assistant Aug. 2025 - Present
Harvard T.H. Chan School of Public Health | Boston, MA

- Analyzing mitochondrial responses to **perturbations** using transcriptomic and cellular imaging data.
- Deriving mitochondrial metrics from image **segmentation features** to define morphological states.
- Integrating **transcriptomics**, drug chemical **structure**, and **MoA** to link expression and morphology.
- Building **multimodal machine learning** and **deep learning** models for mitochondrial phenotypes.
- Developing pipelines to identify **signatures** driving mitochondrial fragmentation for target discovery.

Data Science Contractor Aug. 2025 - Dec. 2025
Novo Nordisk | Lexington, MA

- Designed and developed pipelines to extract, process, and analyze **3D engineered** muscle tissue data.
- **Automated** workflows end to end to accelerate experimental analyses **70%** and enable novel insights.
- Extended and tested scripts across **3D engineered** tissue types, supporting analyses **company wide**.
- Tested and enhanced platforms to run **segmentation** and **foundation models** on cell imaging data.
- Implemented modular **HPC** workflows with a focus on optimization, scalability, and generalizability.

Data Science Intern June 2025 - Aug. 2025
Novo Nordisk | Lexington, MA

- Created pipelines to benchmark **machine learning** models across high throughput **imaging** screens.
- Built automated scripts to identify perturbations inducing **significant cellular phenotype changes**.
- Compared **segmentation** features and **foundation model embeddings** in perturbation assessment.
- **Interpreted** foundation model embeddings via alignment with segmentation based imaging features.

Bioinformatics Research Assistant Jan. 2023 - June 2024
UC San Diego School of Medicine | San Diego, CA

- Designed and developed **bulk RNA-seq** and **scRNA-seq** analysis pipelines in Python, R, and Bash.
- Created and optimized workflows for cell **clustering**, **labeling**, **DE**, **GSEA**, and data **visualization**.

- Applied DESeq2, edgeR, Seurat, Harmony, and clusterProfiler to quantify **gene** and **pathway** effects.
- Analyzed **COVID-19 mRNA vaccine** effects on immune cells, and **miRNA** perturbations in cancer.

SELECTED PROJECTS

Multimodal Perturbation Modeling for MoA Prediction Sep. 2025 – Dec. 2025

Massachusetts Institute of Technology | Cambridge, MA

- Generated **embeddings** for transcriptomic and cellular imaging data using **representation learning**.
- Combined representations with late fusion and refined multimodal embeddings via **contrastive loss**.
- Used various embedding **classifiers** to predict MoA and evaluate performance across configurations.

Deep Learning for Peripheral Blood Cell Image Classification Apr. 2025 – May 2025

Harvard Medical School | Boston, MA

- Developed **CNN** models with 98.8% accuracy and 0.999 AUC, outperforming published baselines.
- Conducted **hyperparameter tuning** and evaluated **generalizability** to ensure stable performance.
- Used **attention** mechanisms and applied Grad-CAM for **interpretability** to validate biological basis.

Pancreatic Cancer Gene and Pathway Analysis Nov. 2024 – Dec. 2024

Harvard T.H. Chan School of Public Health | Boston, MA

- Analyzed **differential gene expression** in tumors vs. normal tissue using **TCGA** and **GTEx** data.
- Used **GSEA** and **network analysis** to identify dysregulated pathways and gene interaction patterns.
- Identified distinguishing signatures and validated key genes specific to **early stage** pancreatic cancer.

TEACHING EXPERIENCE

Biostatistics Graduate Teaching Fellow Sep. 2025 - Present

Department of Biostatistics - Harvard T.H. Chan School of Public Health | Boston, MA

- Assisting in the courses **Computing for Big Data** and **Applied Data Structures and Algorithms**.
- Supporting over **100 graduate students** across **2 semesters** via office hours and review sessions.
- Tutoring for **algorithm design** and **optimization**, **cloud computing**, and **package development**.

Data Science Undergraduate Teaching Assistant Sep. 2022 - June 2024

Halıcıoğlu Data Science Institute - UC San Diego | San Diego, CA

- Instructed **Principles**, **Practice and Application**, and **Theoretical Foundations of Data Science**.
- Provided support to over **2000 students** across **6 quarters** with an average approval rating of **95%**.
- Assisted students in Python for **machine learning**, **data science**, and **statistical data analysis**.

AWARDS

Halıcıoğlu Data Science Institute Undergraduate Excellence Award May 2024

UC San Diego | San Diego, CA