

Bhavay Aggarwal

236 16th Street NW, Atlanta, GA 30363

📧 bhavaygg 🌐 bhavay-aggarwal

Email : baggarwal9@gatech.edu

Mobile : +1-(404)-747-8290

EDUCATION

- **Georgia Institute of Technology** Atlanta, GA
Ph.D. in Bioinformatics (Expected May 2027) | M.S. in Bioinformatics, 2023 | GPA: 3.91 August 2022 – Present
Courses: ML for Computational Biology, ML for Chemistry, Data Science for Epidemiology, Computational Genomics
- **Indraprastha Institute of Information Technology** Delhi, India
Bachelor of Engineering in Computer Science and Biosciences; GPA: 7.91/10.0 August 2018 – June 2022
Courses: Advanced Machine Learning, Deep Learning, Information Retrieval, Software Development using Open Source

PUBLICATIONS

- Hao, Y., Kim, Y., **Aggarwal, B.** and Sinha, S., 2026. **MANTIS: Analytics toolkit for spatial metabolomics with matching spatial transcriptomics data.** bioRxiv, pp.2026-01.
- Kumar, A., Moctezuma, F.R., **Aggarwal, B.**, Zhang, N., Coskun, A.F. and Sinha, S., 2026. **CELLWHISPER: INFERENCE OF DIRECT CELL-CELL COMMUNICATION FROM SPATIAL TRANSCRIPTOMICS.** bioRxiv, pp.2026-01.
- **Aggarwal, B.** and Sinha, S., 2025. **CellSP enables module discovery and visualization for subcellular spatial transcriptomics data.** Communications Biology, 8(1), p.1528.
- Schneider, R.S., Nieves, E.B., **Aggarwal, B.**, Bowles-Welch, A.C., Stevens, H.Y., Kippner, L.E., Boden, S.D., Mautner, K., Drissi, H., Roy, K. and Lam, W.A., 2025. **On-chip 3D potency assay for prediction of clinical outcomes for cell therapy candidates for osteoarthritis.** Nature Communications, 16(1), p.4915.
- Kumar, A., Schrader, A.W., **Aggarwal, B.**, Boroojeny, A.E., Asadian, M., Lee, J., Song, Y.J., Zhao, S.D., Han, H.S. and Sinha, S., 2024. **Intracellular spatial transcriptomic analysis toolkit (InSTAnT).** Nature communications, 15(1), p.7794.
- Priyadarshani, P., Van Grouw, A., Liversage, A.R., Rui, K., Nikitina, A., Tehrani, K.F., **Aggarwal, B.**, Stice, S.L., Sinha, S., Kemp, M.L. and Fernández, F.M., 2024. **Investigation of MSC potency metrics via integration of imaging modalities with lipidomic characterization.** Cell Reports, 43(8).
- **Aggarwal, B.** and Ray, A., 2021. **AgAnt: A computational tool to assess Agonist/Antagonist mode of interaction.** bioRxiv, pp.2021-11.
- Sharma, T., Upadhyay, U., Kalra, J., Arora, S., Ahmad, S., **Aggarwal, B.** and Bagler, G., 2020, April. **Hierarchical clustering of world cuisines.** In 2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW) (pp. 98-104). IEEE.

EXPERIENCE

- **Georgia Institute of Technology** Atlanta, GA
Graduate Research Assistant - Sinha Lab January 2023 - Present
 - Hands-on experience across bulk RNA-seq, single-cell, and spatial omics datasets, including transcriptomic and metabolomic modalities.
 - Creating new methods to understand cell characterization and communication from spatial transcriptomic data.
 - Developed **CellSP**, a tool for robust and interpretable analysis of subcellular spatial transcriptomic data. 📄
 - Developing statistical and deep learning models (Vision Transformers, GNNs) to identify patterns associated with transcript localization in subcellular spatial transcriptomics data.
 - Contributed to **InSTAnT** by refactoring the codebase for parallelization and efficiency, adding new functionality, and analyzing spatial transcriptomics data from different technologies. 📄

- Designed machine learning models to identify patterns of stem cell activation in cell manufacturing processes using lipidomics and morphological descriptors.
- **Technologies Used:** *Python, R, PyTorch, Scikit-Learn.*

- **GlaxoSmithKline**

Collegetown, PA
May 2023 - August, 2023


Computational Biology Intern

- Contributed to the MegaMap project by designing statistical approaches to map genetic perturbations to disease signatures from large-scale RNA-seq experiments.
- Inferred robust gene-disease relationships and identified a minimal set of genes capable of characterizing multiple disease states.
- Performed rigorous statistical testing to validate significant gene-disease associations against real-world clinical data.
- **Technologies Used:** *Python, R*

- **Indian Institute of Science (IISc)**

Remote, India
May 2021 - December 2021

Research Intern - ATCG Lab

- Implemented novel distance metrics within approximate nearest neighbor search libraries (e.g., NMSLIB) for genomic data.
- Engineered a distributed genomic distance calculation framework using Ray, improving indexing time by 90%.
- **Technologies Used:** *Python, C/C++, Ray, Dask.* 

- **Wikilimo**

Remote, India
February 2021 - April 2021


Machine Learning Intern

- Developed data processing pipelines for cleaning, normalization, and interpolation of environmental time-series data.
- Formulated and optimized machine learning models for wildfire prediction and time-series forecasting, leveraging explainable AI techniques (SHAP) for model interpretation.
- **Technologies Used:** *Pytorch Lightning, WandB, Optuna, SHAP, Xarray.*

- **Indraprastha Institute of Information Technology Delhi**

Delhi, India
May 2020 - July 2020

Research Intern - Complex Systems Lab

- Designed ML methods to model inter-ingredient relationships, leveraging Transformers (BERT, RoBERTa) to capture recipe structure and ingredient context.
- **Technologies Used:** *Hugging Face, Pytorch, Keras.* 

- **All India Institute of Medical Sciences**

Delhi, India
February 2020 - October 2020



Research Intern

- Created machine learning models to segment brain and neck tumors from X-ray images using super voxels.
- **Technologies Used:** *Pytorch, OpenCV.*

SKILLS

Languages: Python, R, Java, C/C++, Javascript, MySQL, HTML, CSS
Libraries & Frameworks: PyTorch, Keras, Scikit-Learn, Hugging Face, Ray, Dask, React, Django
Developer Tools: Git, Slurm, AWS, GCP

PROJECTS

- **AgAnt:** Developed deep learning models for Receptor-Ligand binding activity prediction, using GNNs and Transformers to learn enhanced protein representations via transfer learning. 
- **Outbreak analysis of Escherichia coli O157:** Contributed to a bioinformatics web interface combining genome assembly, functional annotation, and comparative genomics to enable rapid outbreak analysis of *E. coli* strains.
- **Disease Symptom Search Trend Analysis:** Performed spatiotemporal analysis of Google search trends for various disease symptoms to forecast regional COVID-19 cases using machine learning. 
- **pyLFDA:** Developed an open-source Python library for performing pairwise lipid force distribution analysis in molecular dynamics simulations. 