ISHAAN GUPTA

San Diego, CA
8105296351
i3gupta@ucsd.edu
github.com/IshaanSD
linkedin.com/in/IgAI

Passionate about unraveling the complexities of the genome, its variation and evolution, particularly implications in the immune system, using Bioinformatics Algorithms and ML

EDUCATION

Ph.D, Computer Science (Research area: Bioinformatics), UC San Diego, La Jolla (2023 – present)
Charles Lee Powell Fellow, Performing *Genomics + Immunology + AI* research under Dr. Pavel Pevzner
BS, Computer Science : specialization in Bioinformatics, UC San Diego, La Jolla *(2019 – 2022)*
GPA: 3.92, Magna Cum Laude with CS Honors, IEEE Eta Kappa Nu Honors Society

Experience

Research Assistant (Dr. Pavel Pevzner) *04/2021 – 08/2023*
UCSD CSE Department, San Diego, CA
· Applying my **Undergraduate Honors Thesis** (2022) to find and analyze novel antibody genes in primates and human haplotypes
· Optimized UniAligner (**C++** aligner for highly-repetitive regions) - reduced runtime **by 80%** for 1M bp; added interactive dotplots feature
· Developed **Nanopore-based protein sequencing pipeline** by applying Signal Processing and ML (Clustering, PCA) approaches

Software Engineer I *06/2022 – 12/2022*
Illumina (Systems Integration), San Diego, CA
· Improved traceability of **LIMS software** central to Illumina's high-throughput **Genotyping and Methylation** pipelines (**Agile SDLC)**
· Developed and deployed highly scalable **microservices** in **Java 7/8 (Spring+JDBC)** integrated using **API**s and **AMQP**
· Found and fixed critical bugs in legacy code, troubleshot software integration issues, automated tests, curated database for support

Bioinformatics and Machine Learning Intern *07/2021 – 09/2021*
Abterra Biosciences, San Diego, CA
· Achieved **>95% performance** in **Proteomics** task for Antibody sequencing by experimenting multiple **Deep learning** approaches
· Optimized data structures and used **Data Manipulation, parallel processing** and ML libraries in **Java 8** (DL4J)

Machine Learning Project Consultant *11/2020 – 03/2021*
Model Medicines, La Jolla, CA
· Developed **BigQuery**-powered **data pipeline** supporting **SQL queries** and **neural networks** for **screening** COVID-19 drug candidates.
· Made 3 **web scrapers** to read scientific journals and evaluate **novelty of potential drugs** using search keywords and their synonyms

Undergraduate Research Assistant *03/2020 – 03/2021*
Mali Lab, Bioengineering UCSD, San Diego, CA
· Performed research on **Immune-orthogonal epitopes** for safe and efficient CRISPR/Cas9 **gene therapies**
· Built **Variants Analysis** pipeline for pooled library screens and web scraper using **Bash** and **Python** scripts and **Command line tools**

Projects

- **Autocorrelation for ecDNA hubs (Graduate-level course project)**
  Analyzed spatial organization of extrachromosomal DNA, and confirmed formation of hubs that help co-promoting oncogenic expression by visualizing *Autocorrelation* function for *image processing* of nuclei-stained cell images
- **GANs for Cancer Image Augmentation**
  Implemented *Generative Adversarial Network* using PyTorch to mimic mammograms with tumor and generate augmented data
- **COVID Mutations Analysis (**https://youtu.be/C27B4mYRpXg**)**
  Used NCBI database, BLAST CLI, and JSON parser to find and analyze phylogeny of mutations that increase COVID-19 infectivity
  Led project on *phylogeny of Sars-CoV-2 variants* to understand the evolution of different strains at *geographic* level
- **Notes2Map (Python API + JS Web app)**
  Developed web app to process lecture notes and transcripts and generate Mind Map of keywords for better visualization
  Used *Flask* to make *REST APIs* handling *NLP* and *Graph Algorithms* in Python and communicate with *ReactJS UI* app.

Mentoring/Leadership

- **Vice President, Undergraduate Bioinformatics Club** *08/2021 – 07/2022*
  Managing various club events for helping Bioinformatics students socialize, prepare for industry/academic positions
- **Data Science Student Society Workshop Chair and Bioinformatics Bootcamp Chair** *08/2020 – 07/2021*
  Developed 10 dry lab workshops on Bioinformatics skills, ML, Shell, Python, R, SQL; hosted labs on an **AWS EC2** instance

SKILLS

- Java (Stream, lambda, DL4J, JUnit), OOP, Spring, JDBC, AMQP
- Python (Pandas, PyTorch, TensorFlow, SciPy), C++, Shell, R, C
- SQL (MySQL, Postgres), Stored Proc, MongoDB, Cassandra (CQL)
- HTML, CSS, JS (React, Node.js, Express), Web components, puppeteer
- Agile (Scrum/Kanban) practices, Git, CI/CD pipeline, Docker
- Hadoop, Apache Spark (MLlib), Data Pipeline: Airflow, Kafka, RabbitMQ
- AWS: Redshift, S3, EC2, SQS, Elastic Beanstalk
- Machine learning: CNN, GANs, NLP, Recommender Systems
- Bioinformatics: scRNA analysis, plink GWAS, ChIP Seq, Bioconductor

COURSEWORK

- Algorithms for Computational Biology (Graduate-level)
- Deep Learning, Recommender Sys, Computer Vision
- Software Engineering, Parallel Computing (CUDA)
- Adv. Bioinformatics Lab, Biomolecular Big Data

CERTIFICATIONS

- Programming Foundations: Design Patterns
- Spring Data; Effective Intg. Testing with Spring Boot
- Spark and Hadoop; NoSQL Databases
- SQL for Data Science; CNNs;