Sanjana Srivastava

+1-4348331321 | sanjana11147@gmail.com | LinkedIn | GitHub | G Scholar | Website

Education

University of Virginia **May 2025**
Master of Science in Data Science *GPA: 4.0*

Indian Institute of Technology Roorkee **May 2022**
Master & Bachelor of Technology in Geological Technology & Mathematics *CGPA: 8.2*

Interests & Skills

Interests: Data Analytics, Computer Vision, Generative AI, LLMs, Multi-Modal Modeling, Healthcare Research
Languages: Python, R, SQL, HTML, CSS, C/C++, Java, JavaScript, React, Django, Latex
Packages: MySQL, PyTorch, TensorFlow, Scikit-learn, Git, AWS, Hadoop, Spark, RAG, Tableau, Docker, Linux

Experience

Data Scientist **Oct 2024 – May 2025**
Deloitte (Capstone Project)

- Built an LLM evaluation pipeline using NLP and LLM-as-a-judge (GPT, Claude) for Deloitte's multi-domain clients
- Benchmarked DeepEval, MLFlow, Bedrock, TruLens, and RAGAS across 10+ datasets in a holistic leaderboard
- Identified and evaluated 8 key metrics including summarization, accuracy, hallucination, and toxicity detection

Machine Learning Researcher **June 2022 – May 2024**
Healthcare AI Lab, University of Virginia

- Advisors - Dr. Sana Syed, Dr. Donald E. Brown
- Developed ML models for disease diagnosis using tissue images, clinical, & transcriptomic data for clinical clients
- Collaborated with medical professionals to apply ML techniques for gut function analysis and prescriptive modeling
- Published and presented at top research conferences iScience, ESPGHAN, DDW, NASPGHAN, and ASTMH

Software Engineering Intern **June 2021 – Aug 2021**
BNY Mellon

- Built a production-grade XML-to-CSV parser in Java, enabling data transformation for financial systems
- Implemented A/B testing for UI workflows and proposed performance optimization across internal dashboards

Software Engineering Intern **June 2019 – Aug 2019**
ZestMoney

- Constructed a Java payment gateway, reducing transaction latency & increasing conversion rate by 5%
- Redesigned Android UI using MVVM architecture and Retrofit, accelerating user signups by 30% and enhancing UX

Projects

Evaluating efficacy of synthetic images generated using diffusion models | *Code*

- Developed a U-Net-based diffusion model to generate 256×256 histology image patches conditioned on 6 nuclei types
- Improved U-Net segmentation performance with synthetic training data, yielding a +5.2% gain in Dice score

Text-driven latent diffusion model for localized fashion image editing | *Code*

- Developed a text-conditioned diffusion model enabling region-specific edits using natural language prompts
- Achieved IoU of 0.76 on garment region localization via a CLIP-guided segmentation pipeline

Humorous image captioning system | *Code*

- Created a CLIP-based encoder-decoder model fine-tuned to generate captions on a custom dataset of 200+ memes
- Integrated a RAG-based image retrieval mechanism with prompt-tuned LLM for context-aware caption generation

Deep learning understanding of diseases using medical imaging and video data | *Code*

- Trained supervised & unsupervised CNN models on tissue images, achieving $> 80\%$ accuracy on disease detection
- Applied Gaussian Mixture Models and PCA to uncover 4 recurring visual clusters in biopsy image datasets

Correlating disease gene signature with imaging data | *Code*

- Built a CNN-based feature extraction pipeline with gene code clustering to correlate visual and genomic biomarkers
- Identified top 10 image features associated with 3 key gene code clusters, aiding interpretability of tissue morphology

Buy and Sell application | *Code*

- Deployed a web app for a community of 10,000 campus users, facilitating listing, search, filtering, and subscriptions
- Processed over 1,000 listings, reduced item discovery time by $> 60\%$ through efficient categorization and search

Petrographic characterization of a chondrite sample | *Advisor - Dr. Nachiketa Rai*

- Successfully investigated mineralogy present in the chondrite section, leveraging Decision Trees and Linear Regression
- Performed data mining and data analysis on back-scattered electron images for predictive modeling

Publications & Talks

Efficacy Of Synthetic Histopathological Images In Enhancing Nuclei Segmentation Tasks
S. Srivastava, A. Shrivastava, S. Rhoads, P.T. Fletcher, S. Syed, D.E. Brown
Under Review

Comparative Study Of Large Language Model Evaluation Frameworks With A Focus on NLP vs LLM-As-A-Judge Metrics
S. Srivastava, A. Alabdulwahab, C. Japic, C. Le, ..., A. Zhang
Link | Systems and Information Engineering Design Symposium (SIEDS) 2025

What Is Normal? Characterization Of Control Pediatric Duodenal Biopsies Using Clinical Data, Machine Learning Image Analysis, And Transcriptomics
F. Rhoads, J. Sessions, **S. Srivastava**, F. Zulqarnain, V. Jain, ..., S. Syed
Poster | European Society for Paediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) 2024

Machine-Learning-Based Integrative–'Omics Analyses Reveal Immunologic And Metabolic Dysregulation In Environmental Enteric Dysfunction
F. Zulqarnain, X. Zhao, K. Setchell, Y. Sharna, P. Fernandes, **S. Srivastava**, ..., S. Syed
Link | iScience 2024

Identifying Metabolic Signatures Of Environmental Enteric Dysfunction In Pakistani Mothers And Children Using Tissue-Specific Metabolic Modeling
S. Syed, F. Zulqarnain, **S. Srivastava**, W. Khan, I. Nisar
Poster | Bill & Melinda Gates Foundation Grand Challenges Annual Meeting 2023

Quantitative Morphometry And Machine Learning Model To Explore Duodenal And Rectal Mucosal Tissue Of Children With Environmental Enteric Dysfunction
M. Khan, Z. Jamil, L. Ehsan, F. Zulqarnain, **S. Srivastava**, ..., S. Syed
Link | American Journal of Tropical Medicine and Hygiene (ASTMH) 2023

The Intersection Of Video Capsule Endoscopy And Artificial Intelligence: Addressing Unique Challenges Using Machine Learning
S. Guleria, B. Schwartz, Y. Sharma, P. Fernandes, J. Jablonski, S. Adewole, **S. Srivastava**, ..., S. Syed
Link | Preprint

Awards & Achievements

- IIT Roorkee Graduation Award: Nayyar Award for **Excellence in Communication** given to 1 student with exceptional communication skills from the graduating batch
- Indian Ministry of Human Resource and Development (MHRD) Assistantship for meeting the graduate CGPA criterion, demonstrating **Academic Excellence** exceeding the required limit
- Cleared Joint Entrance Examination (IIT - JEE) with **99.6 percentile**

Extracurricular Activities

Teaching Assistant for a graduate-level Statistical Learning course at the University of Virginia
Project Leader at Information Management Group - Official Coding Society of IIT Roorkee
Editor-in-Chief at Geek Gazette - Official Technical Magazine of IIT Roorkee
Actor/Director/Producer at Dramatics Section, IIT Roorkee
Council Member of Hostel Management at Kasturba Bhawan, IIT Roorkee
Sports Team Member of Table Tennis National Sports Organization, IIT Roorkee