

Sanskriti Singh

ssingh7@mit.edu | (530) 364-8327 | LinkedIn | Personal Website | GitHub

Education

Massachusetts Institute of Technology (MIT)

Cambridge, MA

B.S. in Computer Science and Engineering

Expected May 2027

Coursework: Math for CS, Deep Learning, Machine Learning, Statistics, Advanced Algorithms, Linear Algebra, Probability

Lab Assistant: Intro to Machine Learning

Honors & Awards: Top 300 Scholar (Regeneron STS '24), SCVSEFA 1st Place ('19-'23), BioGENEius (HM, 3rd, HM; '21-'23), NCWIT National Winner ('23)

Skills

Frameworks: Python (NumPy, Pandas, PyTorch, TensorFlow, Keras, Scikit-learn, Flask, Django, FastAPI, Hugging Face models, OpenCV)

Tools: AWS, Docker, Git, FreeSurfer, Matplotlib, Seaborn, Plotly, Jupyter, Google Colab, VSC, Pycharm, JSON, Linux, LaTeX

Core Areas: Deep Learning, Computer Vision, NLP, LLMs, Multimodal Learning, Representation Learning, Generative Models, Time-Series Analysis, Medical Imaging, Contrastive Learning, Semi/Self-Supervised Learning, Statistical Inference, Model Explainability

Professional Experience

Kognitos

San Jose, CA

Software Engineering Intern

Jun 2025 – Aug 2025

- **Developed** automated **LLM pipelines** using **AWS models** and heuristic search to dynamically trace and repair exception trees; implemented **regex-based structuring** and caching to enhance error message precision by **99%** and reduce API latency by **40%**.
- **Implemented and deployed** a **Django + Darklaunchy API** microservice with **PostgreSQL** and **Docker** integration, streamlining tag management across **500+ distributed production assets** and reducing latency within internal pipelines by **35%**.

Technical Experience

Automated News-to-Research Pipeline (Full-Stack Website, Python)

Jun 2025 – Present

- **Engineered** a full-stack **Flask** application that scrapes **100+ news sources** every 72 hours using asynchronous **API calls** and clusters articles with **autoencoders** to detect possible emerging research domains; **PCA**, and **HDBSCAN** to develop interactive 3d plots
- **Applied NLP** with **Sentence Transformers** for semantic embedding and automated research question generation (>95% relevance), retrieving relevant **arXiv** and other academia papers and synthesizing 600–800 word responses via **open-source language models**.

MIT CSAIL — Undergraduate Research Assistant

Aug 2024 – Present

- **Implemented LSTM** and **Transformer** architectures for credit card fraud detection across **1M+ financial transactions**, optimizing **attention heads** and **learning rates** to improve model **F1 score by 12%**; research performed under **Dr. Amar Gupta** at CSAIL.
- **Designed multimodal latent-space embeddings** under **Dr. Manolis Kellis** to cluster **6,000+ patient records** via **unsupervised** methods including PCA/autoencoder/etc., integrating **genomic data** to identify cardiovascular disease progression patterns.

Water Consumption via Unsupervised ML and Distance Algorithms

Sep 2023 – Sep 2024

- **Built** a scalable time-series pipeline using **K-Means clustering** and **Dynamic Time Warping (DTW)** on segmented high-frequency smart water meter data, extracting temporal features to classify **9+ appliance usage patterns** with **>90% accuracy**.
- **Delivered actionable insights** through **disaggregation analytics**, **anomaly detection**, and **consumption forecasting**, enabling **5–10% reductions** in personal household water use (via billing metrics) and improving long-term resource efficiency.

Research

Yale Interventional Oncology Lab — Research Intern

Jun 2023 – Oct 2023

- **Designed** a patch-based 3D **Teacher-Student GAN** for **Eovist MRI synthesis** using a **U-Net DCNN backbone** with residual bottlenecks, strided convolutions, and skip connections, achieving **30% segmentation improvement (DICE +23%)** over baseline.
- **Automated preprocessing** for **6,000+ MRI slices** (augmentation, normalization, and model training) using **PyTorch** and DICOM-NIfTI.

CheX-Nomaly: Adaptive Localization Model for Chest Radiographs (arXiv)

Sep 2022 – Jan 2024

- **Developed** a **Siamese contrastive localization model** trained on **18K VinDr-CXR** scans (max of 58 bbox/scan) to decouple disease labels from bounding boxes, achieving **mean IoU = 0.7487** and improving baseline of **mAP by +0.007**.
- **Generalized** across unseen thoracic diseases, reaching **90% classification accuracy** and reducing mislocalization error by **80%**.

Automated Coronary Calcium Scoring (IEEE, MIT URTC)

Sep 2021 – Oct 2022

- **Created** a **semi-supervised U-Net** segmentation model for **non-gated CT scans**, improving CAC risk classification from **41% → 64%**.
- **Preprocessed** 200 nongated and 900 gated CT volumes to compute **Agatston scores** and stratify cardiovascular risk levels (5 groups).

PneumoXtention: CNN with Attention for Pneumonia Detection (IEEE ISPA)

Nov 2019 – Sep 2021

- **Built** dual **13-layer CNNs with heatmap attention**, achieving **F1 = 0.82** and **AUROC = 0.95** across **26K CXR images** - ChestX-ray14.
- **Validated** pneumonia localization accuracy (**92% vs. 72% radiologists**), demonstrating strong model interpretability vs real-time.

Publications

- **S. Singh**, “PneumoXtention: A CNN Compensating for Human Fallibility in Pneumonia Detection,” IEEE ISPA, 2021.
- **S. Singh**, “Automated Coronary Calcium Scoring via Semi-supervised U-Net Models,” IEEE URTC, 2022.
- **S. Singh**, “CheX-Nomaly: Segmenting Lung Abnormalities via Contrastive Learning,” arXiv, 2023.
- **S. Singh**, “Minimizing False Positives and Negatives in Binary Classification,” arXiv, 2022.
- **S. Singh**, “A Novel Mask R-CNN for Heterogeneous Brain Tumor Segmentation,” arXiv, 2022.