Suraj Rajendran

420 E 70th St, New York, NY 10021

https://github.com/surajraj99 sur4002@med.cornell.edu 224-587-0371

Education

Weill Cornell Medical College — **New York City, NY**
Tri-Institutional PhD Program in Computational Biology — *Expected Graduation: June 2026*

Georgia Institute of Technology — **Atlanta, GA**
BS in Biomedical Engineering GPA: 3.96/4.0 — *May 2021*

- Minor in Computing and Intelligence GPA: 4.0/4.0
- President's Undergraduate Research Award (PURA) Fellowship
- Faculty Honors

Research Experience

Weill Cornell Medical College — **New York City, NY**
Graduate Researcher — *August 2021 - Present*

Foundation Model Framework for Predictive Embryology in IVF Procedures

- Engineered foundation models with masked autoencoder on 5 million embryo images, establishing a base for later tasks
- Advanced blastocyst quality prediction by employing the developed model, enhancing the accuracy of embryo assessments
- Utilized the model for segmentation and implantation success prediction, performing higher than benchmarks

Structural Variant Detection in Short Read Cancer Data Using Deep Learning

- Developed a deep learning based structural variant caller to detect subclonal and somatic structural variants in NGS data
- Detected novel and biologically 5 relevant variants in SK-BR-3 cell line that were previously not reported
- Outperformed other variant callers at predicting subclonal variants at varying genome coverages and contexts

Determining Steroid Efficacy in Sepsis through Target Trial Emulations

- Created a ML method to find stratify subphenotypes based on trajectory, distinguishing responses to corticosteroids
- Analyzed a large ICU dataset from three sources to enhance findings on corticosteroid efficacy across subphenotypes
- Employed causal inference tools and trial emultion methods to achieve psuedo-randomization and proper outcome metrics
- Utilized a Cox model to study 3 outcomes, deriving hazard ratios for tailored treatment strategies

Investigating Impact of Data Heterogeneity in Federated Learning

- Delineated a protocol to test the impact of data heterogeneity in local, pooled, and federated model settings
- Collected Acute Kidney Injury (AKI) and Sepsis data from 7 different hospitals within the New York medical system
- Adapted the SecureBoost algorithm to investigate the effect of data discrepancies within federated XGBoost
- Ranked 354 medications, lab test measurements, and vital signs, uniquely important at each New York hospital

Predicting Ploidy Status of Embryos Using Deep Video Classification

- Developed a pipeline to train models on in vitro fertilization time-lapse videos to determine ploidy status for 2000 embryos
- Collaborated with embryologists to determine ways to standardize videos of embryos to ensure non-biased predictions
- Created a pre-trained CNN-LSTM architecture based on ImageNet to process and predict ploidy status
- Explored state of the art video classification models such as 3DConvNet and I3D to determine efficacy of application

Wake Forest School of Medicine — **Winston-Salem, NC**
Bioinformatics Researcher, PI: Dr. Umit Topaloglu — *May 2019 - September 2021*

Privacy Preserving Methods Through Holographic Transformations

- Performed frequency domain transformations on multiple standard ML datasets to mask information more efficiently
- Created pipeline for frequency domain masking of datasets, allowing for information control by active parties
- Validated transformed dataset security through simulating a Generative Adversarial Network (GAN) attack

Identification of immunotherapy related adverse events (irAEs)

- Used machine learning to develop prediction models that will aid providers in identifying patients at high risk
- Implemented novel word embedding techniques like BioBERT to clinical notes in order to extract greater meaning
- Developed an LSTM model with over 80% accuracy at multiclass prediction of different irAEs

Predicting COVID-19 Diagnosis Using Televisit and Progress Notes

- Investigated the application of 3 deep neural network models on patient notes to extract COVID-19 symptoms
- Preprocessed and cleaned patient notes via Python's NLTK NLP, regex, and autocorrect python libraries
- Trained DNN models on notebook instance provided by Google Cloud's High-Performance Computing services
- Classified over 40,000 presumptive patients as COVID positive or negative using best performing DNN models

Federated Learning Using Cloud Computing

- Developed environment to test the efficiency and performance of three transfer learning methods
- Created a protocol for sharing models via GitHub from one institution to another without sharing private data
- Delineated the results of the developed federated learning mechanisms in a manuscript submitted to JCO CCI

Detecting Smoking Status Using Natural Language Processing

- Created an NLP Pipeline that cleans EHR data using Python libraries such as gensim, spaCy, and Keras
- Developed 6 machine learning models that classified patients based on their smoking status using progress notes
- Determined optimal parameters for learning rate, batch size, and loss functions using different search techniques

Biophotonics Lab **Atlanta, GA**
Undergraduate Researcher, PI: Dr. Shu Jia *August 2019 - May 2021*
Enhancing Temporal Resolution of Optical Microscopy

- Created a novel and practical method to increase temporal resolution two-fold in biomedical imaging
- Developed a mathematical model to determine the intensity of light at different temporal subframes in an sCMOS
- Drafted and parameterized the attributes of a custom fiber array that transforms images from 2D to 1D
- Created a simulation that utilized the developed mathematical model to output a super-resolution image
- Designed and implemented theoretical setup to validate efficiency and functionality of method

Fast and Accurate sCMOS Noise Correction

- Transcribed code for Automatic Correction of sCMOS-related Noise (ACsN) from MATLAB to Python
- Combined camera physics and layered sparse filtering to reduce most relevant noise sources in a sCMOS sensor
- Improved the camera performance, enabling fast, and quantitative optical microscopy with video-rate denoising

Case Comprehensive Cancer Center **Cleveland, OH**
Research Scholar *June 2020 - August 2020*
Investigating Current Methodologies and Protocols in Cancer Research

- Completed various exploratory projects, statistical analyses, and user data cleaning utilizing libraries from R
- Participated in weekly seminars that ranged in topics from methods of drug delivery to proper conduct of research
- Networked with principal investigators and discussed the nature of their research as well their future directions
- Collaborated with other scholars in presenting recent biomedical papers through Journal Clubs

Yunker Lab for Evolutionary Mechanics **Atlanta, GA**
Undergraduate Researcher *January 2019 - June 2019*
Investigating the Rise of Multicellularity Among Yeast Cells in Varying Flow Regimes

- Performed daily selection speed tests on snowflake yeast for 8 weeks in order to track multicellular growth
- Implemented a protocol for experiments in a glycerol solution to test evolutionary mechanics in different flows
- Developed a quantitative model to compare the theoretical flow of cells in a turbulent media to experimental values

Professional Experience

Regeneron **Rensselaer, NY**
Data Scientist Intern *May 2024 - Aug 2024*
Streamlining Business Processes using AI and LLMs

- Reduced text processing time by 95% using NLP and LLMs for analyzing unstructured data for categorizing various texts.
- Led contradiction detection project for regulatory documents, enhancing integrity by 80% with comparison model
- Implemented analytics and data visualization tools using GPT-4 and PandasAI, to transform data into actionable insights

Colgate-Palmolive **Topeka, KS**
Research and Development Bioinformatics Intern *June 2021 - Aug 2021*
Diagnosis Mapping and Searching within Veterinary Notes

- Developed two pipelines for standardizing new diagnoses entered by clinicians by using NLP and SNOMED relations
- Created an algorithm to find diagnoses within medical notes supplementing veterinarian diagnosis
- Added thresholds and features to the pipelines to allow for user modulation based on use application
- Validated results of the pipelines with clinicians and the ground truth to verify high performance and accuracy
- Presented pipelines and algorithms to professionals and created documentation to facilitate future use of the built programs
- Constructed a multimodal Long Short-Term Memory (LSTM) model for early prediction of renal disease

Mayo Clinic **Jacksonville, FL**
Convergence Science Intern *December 2020 - June 2021*
Lateral Access Torso Model for Surgical Simulation

- Developed an anatomical simulator for lateral access lumbar spine surgery to provide practical training of surgeons
- Implemented a novel Python-based neurostimulation algorithm that provides random auditory and visual feedback
- Integrated radiopaque vertebrae with synthetic tissue to provide realistic feedback within anatomical simulator
- Conducted prior art research to determine patentability and freedom to operate within the surgery simulator field
- Marketed and presented simulator to over 100 medical professionals, investors, and surgeons

Traits AI **San Diego, CA**
Software Engineering Intern *August 2018 - September 2019*
Voice Assistants for Google Assistant and Amazon Alexa

- Coded script and UI which users would be interacting with on the Google Assistant Agents or Amazon Alexa
- Created 3 Voice Activated Conversational AI using Google SDK, Node.js, and the Jovo Database
- Monitored user interaction with Google Agents and Alexa Skills using Dialogflow Analytics and DynamoDB
- Managed and regulated the AWS Lambda servers on which Actions and Skills were hosted
- Designed seamless UI/UX for Traits AI website, Google Actions, and Alexa Skills

Publications

Published and Accepted Manuscripts

- **Rajendran S**, Brendel M, Barnes J, Zhan Q, Malmsten J, Zisimopoulos P, Sigaras A, Ofori-Atta K, Meseguer M, Miller KA, Hoffman D, Rosenwaks Z, Zaninovic N, Elemento O, Hajirasouliha I. *Automatic Ploidy Prediction and Quality Assessment of Human Blastocyst Using Time-Lapse Imaging.* Nature Communications. (2024)
- Mandracchia B, Zheng C, **Rajendran S**, Liu W, Forghani P, Xu C, Jia S. *High-speed optical imaging with sCMOS pixel reassignment.* Nature Communications. (2024)
- **Rajendran S**, Pan W, Sabuncu MR, Zhou J, Wang F. *Learning Across Diverse Biomedical Data Modalities and Cohorts: Challenges and Opportunities for Innovation.* Cell Patterns. (2024)
- Pan W, Xu Z, **Rajendran S**, Wang F. *An adaptive federated learning framework for clinical risk prediction with electronic health records from multiple hospitals.* Cell Patterns. (2024)
- **Rajendran S**, Brendel M, Barnes J, Zhan Q, Malmsten J, Rosenwaks Z, Meseguer M, Zaninovic N, Elemento O, Hajirasouliha I. *Predicting Embryo Ploidy Status Using Time-lapse Images.* Human Reproduction 38. (2023)
- Cao Y, **Rajendran S**, Sundararajan P, Law R, Bacon S, Sumner S, Masuda N. *Web-Based Social Networks of Individuals With Adverse Childhood Experiences: Quantitative Study.* Journal of Medical Internet Research. (2023)
- Su C, Hou Y, Zhou M, **Rajendran S**, Maasch J, Abedi Z, Zhang H, Bai Z, Cuturrufo A, Guo W, Chaudhry F, Ghahramani G, Tang J, Cheng F, Li Y, Zhang R, DeKosky S, Bian J, Wang F et al. *Biomedical Discovery through the integrative Biomedical Knowledge Hub (iBKH)* iScience. (2023)
- **Rajendran S**, Xu Z, Pan W, Ghosh A, Wang F. *Data Heterogeneity in Federated Learning with Electronic Health Records: Case Studies of Risk Prediction for Acute Kidney Injury and Sepsis Diseases in Critical Care.* PLOS Digital Health. (2023)
- Barnes J, Brendel M, Gao V, **Rajendran S**, Jim K, Li Q, Malmsten J, Sierra J, Zisimopoulos P, Sigaras A, Khosravi P, Meseguer M, Zhan Q, Rosenwaks Z, Elemento O, Zaninovic N, Hajirasouliha I. *Development of non-invasive artificial intelligence models for the prediction of human blastocyst ploidy.* Lancet Digital Health. (2023)
- Pullen M, Valero-Moreno F, **Rajendran S**, Shah V, Bruneau B, Martinez J, Ramos-Fresnedo A, Quinones-Hinojosa A, Fox C. *Creation of a Proof-of-Concept 3D-Printed Spinal Lateral Access Simulator* Cureus 14 (2022)

- **Rajendran S**, Ong T, Zameza P, Wolfe S, Topaloglu U, Duncan P, Anwar M, Samuel R, Budigi B, Lack C, Sarwal A. *Including social determinants of health in prognostic models for intracerebral hemorrhage.* Critical Care Medicine 50 (1) (2022)
- Topaloglu M, Morrell E, **Rajendran S**, Topaloglu U. *In the Pursuit of Privacy: The Promises and Predicaments of Federated Learning in Healthcare.* Frontiers in Artificial Intelligence. (2021)
- **Rajendran S**, Obeid J, Binol H, D'Agostino R, Foley K, Zhang W, Austin P, Brakefield J, Gurcan M, Topaloglu U. *A Cloud Based Federated Learning Implementation Across Medical Centers.* JCO Clinical Cancer Informatics (2021)
- Margalski D, Lycan T, **Rajendran S**, Topaloglu U. *Machine learning for prospective identification of immunotherapy related adverse events (irAEs).* Journal of Clinical Oncology (2020) 38
- **Rajendran S**, Topaloglu U. *Extracting Smoking Status from Electronic Health Records Using NLP and Deep Learning.* AMIA Jt Summits Translational Sci Proc. (2020) 507-516

Presentations

Poster Presentations

- **Rajendran S**, Holt J. *Effect of Flow Regime on Snowflake Yeast Geometry.* South Eastern Regional Yeast Conference (SERYM) (2019)

Oral Presentations

- **Rajendran S**, Brendel M, Barnes J, Zhan Q, Malmsten J, Rosenwaks Z, Meseguer M, Zaninovic N, Elemento O, Hajirasouliha I. *Automatic Ploidy Prediction and Quality Assessment of Human Blastocysts Using Time-Lapse Imaging.* Nature Communications. (2024)
- **Rajendran S**, Sundararajan P, *Methods and Results for Models Developed to Score Student Responses Using BERT.* National Assessment of Educational Progress Colloquium. (2022)
- **Rajendran S**, Sundararajan P, *Predicting criminal recidivism using specialized feature engineering and XGBoost.* National Institute of Justice Symposium. (2021)
- **Rajendran S**. *Extracting Smoking Status from Electronic Health Records Using NLP and Deep Learning.* AMIA Jt Summits Translational Sci Proc. (2020)

Service & Outreach

Lumiere Education — **New York City, NY**
Research Mentor — *March 2023 - Present*
- Mentored a student in machine learning for healthcare, enhancing their understanding of self-supervision techniques
- Supervised a project on diagnosis prediction using medical imaging data, boosting diagnostic accuracy by 2-3
- Co-authored a conference-accepted research paper, highlighting the student's project to an international audience
- Fostered the student's growth, contributing to their recognition at a prestigious conference in India

NIH BEAMS Challenge — **New York City, NY**
Team Leader — *October 2023 - December 2023*
- Spearheaded a lesson plan around "Telephone" focusing on making concepts around DNA accessible to middle school students
- Created an interactive and story-driven curriculum that simplified the principles of genetic science and CRISPR technology
- Collaborated with a team of educators to develop educational materials that will be shared with schools in 2024

Dept. of Health & Humans Services Blood Donation Campaign — **New York City, NY**
Team Leader — *October 2022 - February 2023*
- Developed a winning proposal for the "Giving=Living" campaign, aimed at promoting blood donations to address shortages
- Conducted formative research and identified 4 barriers to blood donation, leading to a reduction in negative experiences
- Designed a community-driven approach that increases short-term blood supply and ensures long-term sustainability
- Collaborated with stakeholders to ensure donations among diverse populations, specifically Black and Latino communities

Addressing Inequities in Academic Recognition for Disadvantaged Groups — **Atlanta, GA**
Team Leader — *December 2021 - April 2022*

- Proposed a policy to ensure that students with disabilities get academic recognition for completed courses at Georgia Tech
- Interviewed 5 Georgia Tech individuals, including staff at Office of Disability Services and Diversity and Inclusion Fellows
- Presented to Georgia Tech College of Engineering Diversity & Inclusion Council, along with the Dean and Associate Deans
- Analyzed consequences of instituting policy and determined that 50%+ of students with disabilities will be positively affected

The Science Marvels **New York City, NY**
Speaker *December 2021*
- Collaborated with Clinton High School and Science Marvels to present a seminar on machine learning projects
- Introduced high school students to various STEM topics to build networks and develop machine learning skills
- Provided insight to students on how to get involved with machine learning research and competitions

Alpha Chi Omega - Professional Chemistry Honor Society **Atlanta, GA**
Science Outreach Chair *January 2020 - December 2020*
- Collaborated with Georgia Tech and other STEM organizations in the Atlanta area to host events that celebrate chemistry
- Arranged food drive with local community center to provide underprivileged residents with aid during COVID-19
- Organized tutoring sessions through which fraternity members could support local K-8 students in a myriad of subjects

Save the Water **Surfside, FL**
Project Leader of the Research Branch *April 2015 - May 2018*
- Guided the RD Group through many projects including the implementation of the DILOS Program for school students
- Coordinated weekly research to create 50+ scientific articles on water pollution in different parts of the world
- Interviewed and trained newly hired associates on proper process of research leading to a 25% decrease in turnover rate
- Maintained the organization's website by optimizing SEO score for different articles and webpages

Projects

Investigating Actionable Molecules for Biosynthesis *February 2023 - March 2023*
- Performed extensive market research to identify a feasible molecule the client can produce through synthetic methods
- Conducted market analysis of pterostilbene, the identified molecule, and calculated its growth in the supplement market
- Generated a detailed competitive landscape of the pterostilbene space and actionable strategies to market the supplement
- Delivered actional business strategies to the client and was selected as top 5 teams out of an initial pool of 40 applicants

Developing Marketing and Licensing Strategies for Integral Molecular *October 2022 - November 2022*
- Worked with a team of graduate students to create plans for marketing a bispecific antibody therapy for multiple myeloma
- Conducted market research and analyzed competitors in the multiple myeloma space, accounting for clinical development
- Recommended actionable strategies to Wharton professors, industry professionals, and Integral Molecular's CEO

Analyzing ACEs Using Natural Language Based Prediction Networks *Decemeber 2021 - Present*
- Trained an LSTM network on textual input from social media to predict Adverse Childhood Experiences (ACEs)
- Performed cluster analysis to determine similarities between Reddit posts focused on traumatic childhood events
- Generated a graph network with 50+ nodes depicting Twitter users and their likelihood of having experienced an ACE
- Conducted social network analysis (SNA) to determine communities in which external intervention was necessary

Aether Analytics - Anonymous Job Search Service *April 2022 - September 2022*
- Worked in a team of 4 engineering graduates to pursue a startup venture for improving the job search experience
- Identified a problem of interest based on unmet needs >100 customer interviews including students and companies
- Developed a web platform for hosting users using DJANGO and MongoDB which we iteratively improved via beta testing
- Presented final product to investors and academics at Georgia Tech CREATE-X Demo Day

Department of Education Automated Scoring Challenge *November 2021 - January 2022*
- Constructed 10+ machine learning models to score constructed response items for the NAEP's reading assessments
- Utilized BERT language models through Pytorch in combination with natural language cleaning processes for classification
- Fulfilled the performance requirements to use automated models in a real-world setting within 5% margin of error
- Accounted for racial and gender disparities within student data using data augmentation to minimize model bias

Spike.io - Diabetes Prevention Application *October 2021 - December 2021*
- Ideated a solution to aid people with pre-diabetes in mending their lifestyle to prevent onset of diabetes
- Designed a sample user interface for the Spike.io application to present potential use cases

- Developed a business plan for attaining data and marketing Spike.io to a wider audience of health conscious individuals

Predicting Criminal Recidivism Using Feature Engineering and XGBoost *June 2021 – October 2021*

- Utilized state of the art machine learning techniques to assist in predicting recidivism to aid in evaluating prison efficiency
- Preprocessed and standardized large datasets to ready them for a XGBoost model with fine-tuned parameters
- Added features to the dataset to ensure that models were not biased against certain demographics

Deriving Actionable Strategies Using Machine Learning *January 2021 – February 2021*

- Analyzed historical CarMax data to identify trends in customer purchases and preferences across many demographics
- Developed neural network and random forest models to predict customer decisions based on various attributes
- Determined marketing inventory strategies for CarMax to utilize to draw in distinct segments of customers
- Created an interface which allows personalized experience custom fit for each customer to maximizes their satisfaction

Plaza: Your Local Business Recommender *May 2020 – September 2020*

- Crafted a Google Assistant chatbot allows users to local businesses in a specific market and provide support to them
- Uses natural language processing and parts-of-speech identification to recognize and process different user inputs
- Integrated the Google Maps API into the assistant in order to find target businesses and pertinent information about them

SafeShop: Using AI & Sensor Fusion to Aid Businesses *May 2020 – June 2020*

- Developed an application focused on reinstilling consumer confidence and promoting business traffic post-outbreak
- Designed a prototype UI for the application that allowed user to choose from a variety of methods to discover new facilities
- Created an algorithm which combines inputs from a tracking sensor and crowdsources user data to form a “safety rating”

Development of a Device to Detect Parkinson’s Symptoms *January 2019 – May 2019*

- Created a device that can measure wrist movements using an IMU to capture the frequency and strength of rest tremors
- Constructed an algorithm which could deduce the presence of a Parkinsonian off-period based on the frequency of tremors
- Presented development process and experimental prototyping to a panel of judges and professors

Fellowships and Accolades

Fellowships

National Science Foundation Graduate Research Fellowship *August 2023 - August 2026*
Role: PhD Student
Stipend: $34,000

Honors & Awards

2023 NIH Biomedical Engineering Adapted for Middle Schoolers Challenge Winner, [$5,000 Cash Prize] *December 2023*
2023 ARPA-H Dash to Accelerate Health Outcomes [Quarterfinalist, $10,000 Cash Prize] *May 2023*
2023 UTSW Annual Healthcare Case Competition [Top 5 Finalist] *March 2023*
US Dept. of HHS Giving=Living Blood & Plasma Innovation Challenge Winner [$10,000 Cash Prize] *February 2023*
2022 University of Pennsylvania Healthcare Case Competition [Top 5 Finalist] *November 2022*
Georgia Tech Student Innovation Competition - Promoting Equity and Access [$1000 Cash Prize] *April 2022*
US Dept. of Commerce Smart Tracking Challenge Phase I Winner [$5000 Cash Prize] *April 2022*
Georgia Tech Hacklytics 2022 [2nd Place Healthcare Hack and 2nd Place Overall] *February 2022*
CDC & NASA Detecting Emerging Threats Challenge [Won 1st Place - $7500 Cash Prize] *January 2022*
DOE Automated Scoring Challenge [Won 4th Place - $1250 Cash Prize] *January 2022*
Mayo Clinic Healthcare Hackathon [Won 3rd Place - $1000 Cash Prize] *October 2021*
DOJ Recidivism Forecasting Challenge [Won 4 Awards totalling $23,000] *August 2021*
Georgia Tech $1B+ Startup Hackathon [One of 17 Finalists out of 193 Contestants] *April 2021*
Pueblo Data Mine Analytics Challenge [2nd Place out of 100 Contestants - $800 Cash Prize] *March 2021*
CarMax ML/AI Data Analytics Showcase [1st Place out of 200 Teams - $3000 Cash Prize] *February 2021*
AAMI Foundation’s Michael J. Miller Scholarship [$3000 Cash Prize] *January 2021*

Professional Memberships

Association for the Advancement of Medical Instrumentation *Since January 2021*

Skills

- **Software:** SolidWorks | google cloud | keras (tensor flow backend) | SQL | natural language processing
- **Biology:** NGS | structural variant calling | genomic analysis
- **Programming:** Python | Java | MATLAB | Node.js | R | Github
- **Biomedical:** flow analysis | clinical investigation | inferential statistics | microscopy | process piping systems | calorimetry
- **Communication:** statistics | technical writing
- **Tools:** 3D printers | laser cutters | soldering | band saw | miter saw | router | planer
- **Languages:** English – native | French – intermediate
- **Certifications:** CITI Training | Certified SOLIDWORKS Professional