Gautam Machiraju

6044 College Ave, Apt C | Oakland, CA 94618 | 614-599-6288

gmachi@stanford.edu ◇ gmachiraju.github.io ◇ @gmachiraju

Education

Stanford University 2018 – 2024

Ph.D. Biomedical Informatics — Department of Biomedical Data Science (BDS)

Advisors: Profs. Parag Mallick (Radiology), Christopher Ré (Computer Science)
Committee: Profs. Sylvia Plevritis (BDS, Radiology), James Zou (BDS)
Funding: NIH NLM Training Grant, Stanford Data Science Scholarship, Canary-ACED Fellowship
Expertise: AI-driven data copilots for scientific discovery

University of California, Berkeley 2012 – 2016

B.A. Applied Mathematics, Minor in Bioengineering

Interests: Mathematical Biology (elected degree concentration) and numerical analysis with particular emphasis in differential equations, optimization, & algorithms

Recent Work Experience

Graduate Student Researcher @ Mallick & Ré Labs | Stanford AI Lab (SAIL) 2018 – Present

- **Published** novel techniques to foster explainability capabilities for virtually any foundation model (LLMs, VLMs, etc.) in the form of efficient (linear-time), few-shot, and modality-generalizable feature attribution; applications identify class-specific regions in high-dimensional data (e.g. images, graphs, sequences) with only class labels: e.g. segmentation of tumors in gigapixel pathology, binding pockets in protein structures, & evidential text in large text documents **[ICML 2022, arXiv]**
- **Published** explainability evaluation frameworks for architecture selection **[ECCV 2022]**
- Applying methods to aid understanding of scientific phenomena, e.g. identify (1) novel prognostic biomarkers of cancer progression in multiplexed histopathology [with GE Global Research] & (2) rare genetic variants for Mendelian disease
- *Keywords:* AI4science, explainability & interpretability, foundation models, deep learning, weak supervision & few-shot learning

Graduate Research Consultant @ Thieme, Geldsetzer, & Baiocchi Labs | Stanford 2021 – present

- Computer vision for monitoring global health & human development. Moonlighting projects include:

(1) **Published** app-based mpox detection with mobile-friendly vision models trained on social media corpus of skin lesion images **[Nat Med 2022]**
(2) Deep regression models to predict maternal & child health outcomes of remote villages in low- & middle-income countries using satellite imagery, government surveys, & other remote sensing data
(3) Few-shot detection of human trafficking compounds in the Brazilian Amazon rainforest using satellite imagery; foundation model (IBM/NASA) fine-tuning & feature attribution (see above thesis work)
(4) Clinical-grade foundation models built to pass USMLE medical board exams; multi-modal models supporting long text prompts/captions and medical imagery

- *Keywords:* computer vision, foundation models, deep learning

Student Researcher @ IBM Research Summer – Fall 2023

- Stanford-based collaboration to enable visually grounded foundation models for downstream use in clinical settings (AR, ambient intelligence); open vocabulary object detection to enable memory surrogates and prompted retrievers
- **Submitted**: fact-checking vision-language foundation models for improved image tagging & retrieval
- *Keywords:* foundation models, computer vision, natural language processing

Graduate Teaching Assistant @ Department of BDS, Stanford University 2019 – 2020

- TA for BIOMEDIN 214: Representations & Algorithms for Computational Molecular Biology (with Russ Altman)

- Fall 2019: Developed coding assignments & exams, led weekly office hours, delivered lectures on CS fundamentals, Python, & deep learning; received **teaching award** from student evals
- Fall 2020: Transitioned curriculum to remote-friendly format during COVID-19 pandemic

Bioinformatics Research Assistant @ Mallick Lab, Stanford Radiology 2016 – 2018

- **Published** mathematical model (ODEs) for biomarker shedding kinetics using tumor microenvironment data; enabled simulations to assess serum biomarker viability **[Nat Sci Rep 2020]**
- **Submitted**: NLP of PubMed articles to construct biomarker database; few-shot fine-tuning & entailment of LLMs to identify text relations
- Anomaly detection for multivariate time-series (multi-omics + wearables) to identify drivers of viral disease onset [with DARPA]
- *Keywords:* numerical methods, foundation models, computer vision, natural language processing, time-series analysis

Bioengineering SDE Intern @ Strateos (formerly Transcriptic) Summer 2016

Software engineering & scientific computing to refine robotic platform for wet-lab automation; projects in numerical optimization of robotic gantry movement and queueing & search of platform's job requests

Current Projects

Thesis work:

- Gautam Machiraju, Alexander Derry, Arjun Desai, Neel Guha, Amir-Hossein Karimi, James Zou, Russ Altman, Christopher Ré, Parag Mallick. **Prospector Heads: Equippable Concept-based Feature Attribution**, in preparation for ICML 2024.
- Gautam Machiraju, Erin Craig, Amir-Hossein Karimi, James Zou, Fiona Ginty, Robert Tibshirani, Sylvia Plevritis, Christopher Ré, Parag Mallick. **Spatial Differential Expression & In Silico Verification Methods for Biomarker Discovery in Multiplexed Imagery**, In preparation for *Nature Methods*.

  Applying Prospectors (see above) to spatial biology in the search for clinically relevant biomarkers of cancer progression.

Non-thesis work:

- Tahir Miriyev, Gautam Machiraju, et al. **Weakly Supervised Semantic Segmentation Enabled by Clinical Captioning**, In preparation for MICCAI 2024.
- Alexander Henry Thieme, Tahir Miriyev, Gautam Machiraju, et al. **Constructing Medical Vision-Language Models with Clinical Evaluation**, In preparation for *Nature Medicine*
- Haojie Wang, Gautam Machiraju, Ali Lenjani, Turan Orenbas, Luke Carani, Pascal Geldsetzer. **Predicting Maternal and Child Health Indicators using Publicly Available Remote Sensing, Survey, and Geo-tagged Data**, In preparation for *The Lancet*.

Submitted Works & Peer-reviewed Publications

Note: additional pre-prints can be found on personal website.

High-dimensional Scientific Data & Explainability (thesis work):

1. Hunter Boyce, Gautam Machiraju, Parag Mallick. **Spatial Statistics for Spatial Biology** (In Review).
2. Gautam Machiraju, Arjun Desai, James Zou, Christopher Ré, Parag Mallick. **Prospectors: Leveraging Short Contexts to Mine Salient Objects in High-dimensional Imagery**, *International Conference on Machine Learning* (ICML) workshop on *Interpretable Machine Learning for Healthcare* (IMLH) 2023.

3. Gautam Machiraju, Sylvia Plevritis, Parag Mallick. **A Dataset Generation Framework for Evaluating Megapixel Image Classifiers & their Explanations**, *European Conference on Computer Vision* (ECCV) 2022.

Developing & Adapting Foundation Models for Applications (non-thesis work):

4. Gautam Machiraju, Razi Mahmood, Tanveer Syeda-Mahmood. **Guess 'n' Check: Fact-checking Vision-Language Models for Image Tagging** (In Review).

5. Varun Tandon, Gautam Machiraju, Parag Mallick. **Grammar Matters: Exploring Grammatical Variation's Role in Improving Fine-Tuned LLMs for Biomedical Relation Extraction** (In Review).

Public Health & Clinical Informatics (non-thesis work):

6. Alexander Henry Thieme, Yuanning Zheng, Gautam Machiraju, et al. **Development and Evaluation of an Image-based Deep Learning Algorithm to Classify Skin Lesions from Mpox Virus Infection**, *Nature Medicine* (2023).

7. Minh Nguyen, Conor Corbin, Tiffany Eulalio, Nicolai Ostberg, Gautam Machiraju, Ben Marafino, Michael Baiocchi, Christian Rose, Jonathan Chen. **Developing Machine Learning Models to Personalize Care Levels among Emergency Room Patients for Hospital Admission**, *JAMIA* (2021).

Mathematical Modeling for Cancer Biomarkers (non-thesis work):

8. Sharon Hori, William Greenwald, Hewitt Chang, Tony Sun, Gautam Machiraju, Parag Mallick, Sanjiv Sam Gambhir, Ramasamy Paulmurugan. **Intracellular versus Extracellular Proteins as Candidate Circulating Biomarkers for Cancer Early Diagnostic Screening** (In Review).

9. Gautam Machiraju, Parag Mallick, Hermann Frieboes. **Multicompartment Modeling of Protein Shedding Kinetics During Vascularized Tumor Growth**, *Nature Scientific Reports* (2020).

Awards, Scholarships, & Fellowships

- Creativity in Research Scholars (CIRS) Program, Stanford d.school (2023 cohort)
- Canary-ACED Graduate Fellowship (100% support) via International Alliance for Cancer Early Detection (2022-2023).
- Stanford Data Science Scholarship (50% support) via Stanford Data Science Institute (2021-2022, 2023-2024)
- 2022 HAI Google Cloud Credits Award ($40K) for "Algorithmically identifying histo-molecular biomarkers of cancer progression" (2022-2023)

Invited Institute, Consortium, & National Presentations

- International Conference on Machine Learning (ICML), 2023 workshop of Interpretable Machine Learning for Healthcare (IMLH). Poster: **Prospectors: Leveraging Short Contexts to Mine Salient Objects in High-dimensional Imagery**
- European Conference on Computer Vision (ECCV), 2022. Poster: **A Dataset Generation Framework for Evaluating Megapixel Image Classifiers & their Explanations**
- National Library of Medicine Informatics Trainee Conference, 2022. Focus Session presentation: **A Dataset Generation Framework for Evaluating Megapixel Image Classifiers & their Explanations**
- Canary Center at Stanford for Cancer Early Detection; April 20, 2022; Center-wide presentation: **Toward algorithmically identifying histopathological biomarkers of cancer progression**

- National Library of Medicine Informatics Trainee Conference, 2021. Focus Session presentation: **Learning multi-scale morphological features in multiplexed, megapixel microscopy images**
- Canary Center at Stanford for Cancer Early Detection; May 19, 2021; Center-wide presentation: **Learning multi-scale morphological features in multiplexed, megapixel microscopy images**
- National Library of Medicine Informatics Trainee Conference, 2020. Open Mic Session presentation: **Automated Node-Positive Classification and Feature Extraction Using Tumor Microenvironment Imaging**
- The Early Detection of Cancer Conference, 2019. Poster: **Accelerating plasma biomarker discovery with spatially-explicit mathematical modeling**
- Canary Center at Stanford for Cancer Early Detection; July 18, 2018; Center-wide presentation: **Accelerating plasma biomarker discovery with a spatially-explicit mathematical model**
- Center for Cancer Systems Biology (CCSB) Symposium, 2017. Poster: **Mathematical model of tumor cell heterogeneity and biomarker shedding origins**.
- Canary Symposium, 2017. Poster: **Mathematical model of tumor cell heterogeneity and biomarker shedding origins**
- Biomedical Computation at Stanford (BCATS) Symposium, 2017. Poster: **Mathematical model of cancer heterogeneity and biomarker shedding kinetics**

Academic Teaching, Service, & Leadership Experience

- **Research Mentor**, SMASH Rising. Co-mentoring four undergraduate students of underrepresented backgrounds in STEM. Taught programming and proteomics background material; led research to identify proteins shed into blood with in-house mouse xenografts. June 2022 – August 2022.
- **Reviewer**, European Conference on Computer Vision (ECCV), 2022
- **Reviewer**, Annual Biomedical Research Conference for Minority Students (ABRCMS), 2022
- **Mentor**, Stanford ADVANCE Mentor Collective Program (in partnership with Mentor Collective). Advising mentees from around the US who are applying to colleges or graduate schools. Part of the Stanford ADVANCE volunteer group. Received training and certification to be on-call as a mentor when requested. June 2021 – Present.
- **Co-Czar**, Biomedical Informatics Training Program. Elected student leader and liaison working on recruitment, admissions, student resources and events (e.g. departmental Roundtables with invited speakers), curriculum, DEI efforts (e.g. Antiracism Book Clubs), socials, etc. Host weekly office hours for student feedback. Sit on departmental executive committee. Estimated 500 hours of service to the department. October 2020 – October 2021.
- **Research Mentor**, SMASH Rising. Co-mentoring four undergraduate students of underrepresented backgrounds in STEM. Students are exposed to and work toward identifying candidate protein biomarkers of cancer through bioinformatics approaches. Gave daily lectures on systems biology, multi-omics, mass spectrometry, data wrangling, and data visualization. Introduce Python via Google Colab notebooks. Helped form research questions culminating in research project. June 2021 – August 2021.
- **Panelist**, Stanford ADVANCE - Rotations Workshop. July 9, 2021.
- **Committee Member**, DBDS Faculty JEDI Committee. See details below for efforts related to justice, equity, diversity, and inclusion (JEDI). September 2020 – September 2021.

- **Committee Member**, Student DEI Committee of the Biomedical Informatics Training Program. Student and faculty efforts to diversify admissions and culture at Stanford's Department Biomedical Data Science (DBDS). Spearheaded bi-annual departmental DEI Town Halls, antiracism book clubs, the funding of Women and Non-binary Students of Biomedical Data Science, an application review program for prospective students, and a *BMI Bootcamp* to onboard a diverse incoming MS and PhD student body through informatics and career development workshops. June 2020 – September 2021.
- **Mentor**, BIODS 360: Inclusive Mentorship in Data Science. Working with undergraduates at other institutions to introduce applied data science concepts and career paths in industry and academia. Tailoring weekly exercises to learn concepts. March 2021 – June 2021.
- **Mentor**, ADVANCE Undergraduate Institute (AUI). Mentored future Stanford applicants in interview preparation and application reviews. April 2021.
- **Graduate Teaching Assistant**, BIOMEDIN 214 (Representations & Algorithms for Computational Molecular Biology). Worked with Prof. Russ Altman to update syllabus, headed office hours and review sessions, developed exam questions and code for projects. Designed and gave two lectures: Python for bioinformatics and deep learning methods applied to biological data. Received teaching award from the Stanford Center for Professional Development (SCPD). Autumn 2019 & 2020.
- **Reviewer**, Deep Learning for Genomics, IEEE/ACM Transactions on Computational Biology and Bioinformatics 2020
- **Reviewer**, ML4Health workshop, NeurIPS 2019
- **Trainee Panelist**, AI in Radiology Research, Stanford Radiology Joint Research Retreat; Oct 29, 2019
- **Student Planning Committee Member**, Stanford Department of Biomedical Data Science (BDS) Scientific Retreat 2019; Sept 27, 2019
- **Volunteer**, The Early Detection of Cancer Conference, 2019; Sept 24-25, 2019.
- **Alumni Panelist**, Mathematics Undergraduate Student Association (MUSA), UC Berkeley; May 3, 2018
- **Organizing Committee Member**, Biomedical Computation at Stanford (BCATS) Symposium 2018
- **Research Mentor**, Canary Cancer Research Education Summer Training (CREST) Program, 2018. Mentoring students to aid in lab's NLP project (see below), ultimately with a submission to the Siemens Competition.
- **Organizing Committee Member**, Stanford Radiology Joint Research Retreat 2017
- **Research Mentor**, Canary Cancer Research Education Summer Training (CREST) Program, 2017. Culminated in building Natural Language Processing pipeline for PubMed data extraction, with ultimate goal of creating a public database of disease biomarkers.

Entrepreneurship Experience & Incubator Membership

- **Sutardja Center for Entrepreneurship and Technology, UC Berkeley**, January – May 2017
- **Free Ventures, UC Berkeley**, January – May 2015
- **Catalyst@Berkeley**, September – December 2014

Relevant Skills

Programming	Python (Pytorch, Tensorflow), shell [extensive]
	R, MATLAB [intermediate]
	SQL, CUDA, C++, Java, Spark, Scala, Javascript, HTML/CSS [basic]
Workflow	Cluster (Slurm), Cloud (GCP, Kubernetes), integrations (VSCode, W&B, Jupyter, rmate)
Design	LaTeX, vector graphics (BioRender, Adobe suite)
Mentorship	Mentored 15 URM trainees on independent research projects

Further Inquiry

For more details regarding research projects, publication & preprints, industrial work experience, academic service, coursework (in BDS, CS, EE, Stats, Maths, etc.), and entrepreneurial experience, please visit my Linkedin page (https://www.linkedin.com/in/gmachiraju) or personal website (gmachiraju.github.io).