Richard Paul Yim M.S.

+1 (424) 390-6473
yim.richardp@gmail.com
http://www.richpaulyim.com

Education

The University of North Carolina at Chapel Hill — Chapel Hill, North Carolina
Ph.D. in Biostatistics — 2025-2030

- NIEHS T32 Predoctoral Trainee in Environmental Health Biostatistics

University of California, Davis (UC Davis) — Davis, California
M.S. in Applied Mathematics — 2021 - 2023

- **Supplemental Graduate-level Coursework in Biostatistics/Statistics:** *Probability Theory I/II, Survival Analysis, Generalized Linear Models, Longitudinal Data Analysis, Design and Analysis of Clinical Trials, Computational Statistics, Causal Inference for Observational Studies*
- Core Sequences: *Real Analysis, Differential Equations, Numerical Analysis*
- Thesis Paper: Neural Profiles of Two- and Three-State Cellular Automata on Two-Dimensional Lattice Domains
  [Link: `https://escholarship.org/uc/item/24v881xz`]

University of California, Los Angeles (UCLA) — Los Angeles, California
B.S. in Mathematics and Statistics (Double Major) — 2017 - 2021

- **Select Coursework:** *Honors Linear Algebra, Honors Algebra I/II, Real Analysis I/II, Graduate-level Numerical Linear Algebra I/II, Design and Analysis of Experiments, Mathematical Statistics, Statistical Learning, Reinforcement Learning*

Experience

University of California, San Francisco (UCSF) — *Clinical Data Scientist*
August 2023 - August 2025 | *San Francisco, CA*

- Researched the performance of both open-source (Hugging Face) and closed-source (OpenAI) LLMs for information extraction from EHR derived clinical notes and reports as well QLoRA fine-tuning using HPC clusters and GPU compute
- Developed longitudinal models for early diagnosis of systemic mastocytosis in collaboration with Mayo Clinic with `Torch` and `caret`
- Working within Bakar Computational Health Sciences Institute (BCHSI) under supervision of Dr. Vivek Rudrapatna on inflammatory bowel diesease research

Principal Financial Group — *Data Analyst Intern*
May 2022 - August 2022 | *Sacramento, CA*

- Utilized cloud compute (AWS SageMaker and Athena) to apply nonparametric methods from spectral clustering to random forests, segmenting financial advisors and clients based on behavioral and demographic data
- Developed a process for unsupervised clustering and segmentation on over 160K (∼50% of population) SEC registered financial advisors for targeted marketing; results and analyses formally integrated in Sacramento sales office
- Identified characteristic investment decisions of financial advisors, and communicated critical and actionable findings to nontechnical internal stakeholders

Skills

Programming/Scripting Languages: `R/Python/SQL/Bash` (Professional), `Matlab` (Academic), `JS/HTML/CSS` (Enthusiast).

Tools and Technologies: `Emacs/Vim`, Docker, Git, GNU/Linux, AWS (S3, EC2, Sagemaker), Microsoft Azure VM, HPC scripting, LaTeX.

Frameworks: PyTorch, TensorFlow/Keras, ThreeJS, Tidyverse, SciPy, Scikit-Learn, caret, Hugging Face Transformers, dplyr/pandas, RStan.

Modeling Skills: GNNs, LLMs, RL, GLM, GEE, proportional hazards, mixed effects, Bayesian hierarchical modeling, statistical/machine learning methods (e.g., neural networks, tree-based methods, supervised/unsupervised clustering, dimensionality reduction methods)

Manuscripts

*Indicates first co-authorship.

1. **R. Yim**, A. Silverman, S. Wang, V. Rudrapatna. Optimal strategies for adapting open-source large language models for clinical information extraction: a benchmarking study in the context of ulcerative colitis research. (**Published**, *BMJ Digital Health & AI*, 2025.) [Link: `https://bmjdigitalhealth.bmj.com/content/1/1/e000014`]
2. A. Goyal.*, Z. Wu.*, **R. Yim**, B. Chen, Z. Xu, H. Lyu. An interpretable latent linear model for nonlinear coupled oscillators on graphs. (**In Revision**, *AIP – Chaos*, 2024.) [Link: `https://arxiv.org/abs/2311.14910`]
3. **R. Yim**.*, H. Bassi.*, R. Kodukula, C. Zhu, H. Lyu. Learning to predict synchronization of pulse coupled oscillators on heterogeneous graphs. (**Published**, *Scientific Reports – Nature*, 2022.) [Link: `https://www.nature.com/articles/s41598-022-18953-8`]
4. **R. Yim**. Statistical Learning for Best Practices in Tattoo Removal. (**Published**, *SIAM Undergraduate Research Online*, 2021.) [Link: `https://www.siam.org/media/odbb50ep/s142132r.pdf`]

Projects

Homeboy Industries — *Statistics Research Intern*
December 2020 - July 2021 | *Los Angeles, CA*

- **Awarded $2,000** for cohort study in joint collaboration with faculty from UCLA Computational Applied Mathematics and USC Keck School of Medicine
- Performed end-to-end analysis on six years of longitudinal data on over 500 patients that have received laser assisted tattoo removal treatment with and without complications (over 2,000 tattoos) regression analysis and nonparametric methods
- Created periodic deliverables to data manager and presented final findings to laser removal clinicians, strongly recommending conservative laser treatment practices

UCLA Computational Applied Mathematics REU — *Undergraduate Researcher*
July 2020 - September 2020 | *Los Angeles, CA*

- **Awarded $7,000** to study Firefly Cellular Automata (FCA) and other pulse-coupled oscillator models using machine learning to predict long-term dynamics
- Collaborated with other undergraduate researchers, developing a novel application of Long-Term RNNs, and designed sophisticated simulation data
- Utilized Amazon Web Services (AWS) S3 buckets and EC2 instances for sophisticated training of neural networks over hundreds of thousands of state-space dynamics data points

Teaching

University of California, Davis (UC Davis) — *Teaching Assistant*
September 2022 - June 2023 | *Davis, CA*

- Fall 2022: Mat 17A (Calculus for Biology/Medicine), Mat 21A (Calculus)
- Winter 2023: Mat 17B (Calculus for Biology/Medicine), Mat 21C (Calculus)
- Spring 2023: Mat 127B (Real Analysis II), Mat 17B (Calculus for Biology/Medicine)

American River College — *Instructor Assistant*
November 2022 - June 2023 | *Sacramento, CA*

- Tutor for remedial support courses holding several weekly office hours sessions

Olga Radko Endowed Math Circle at UCLA — *Lead Instructor*
September 2018 - August 2021 | *Los Angeles, CA*

- Lead instructor for intermediate 2A students (6-7th graders) holding courses in-person (pre-COVID-19 pandemic) and over Zoom
- Organized and curated material for weekly lessons in topics such as logic, geometry and graph theory