Jean Feng

▬▬▬ Position

July 2020–
Present
Assistant Professor In-Residence, Department of Epidemiology and Biostatistics, *University of California, San Francisco*

Nov 2022–
Present
Assistant Professor, UCSF-UC Berkeley Joint Program in Computational Precision Health

▬▬▬ Education

Sept 2015–
June 2020
PhD, Biostatistics, *University of Washington*, Seattle, WA
Advisors: Noah Simon, Frederick Albert Matsen IV

2012–2013 **MS, Computer Science**, *Stanford University*, Stanford, CA

2009–2013 **BS, Computer Science**, *Stanford University*, Stanford, CA

▬▬▬ Publications

Siavash Zamirpour, Alan E Hubbard, Jean Feng, Atul J Butte, Romain Pirracchio, and Andrew Bishara. Development of a machine learning model of postoperative acute kidney injury using Non-Invasive Time-Sensitive intraoperative predictors. *Bioengineering*, 2023.

Jaeyun Jane Wang, Jean Feng, Camilla Gomes, Lucia Calthorpe, Amir Ashraf Ganjouei, Fernanda Romero-Hernandez, Andrea Benedetti Cacciaguerra, Taizo Hibi, Mohamed Abdelgadir Adam, Adnan Alseidi, Mohammad Abu Hilal, Nikdokht Rashidian, and International Post-Hepatectomy Liver Failure Study Group. Development and validation of prediction models and risk calculators for Post-Hepatectomy liver failure and postoperative complications using a diverse international cohort of major hepatectomies. *Ann. Surg.*, 2023.

Jean Feng, Adarsh Subbaswamy, Alexej Gossmann, Harvineet Singh, Berkman Sahiner, Mi-Ok Kim, Gene Pennello, Nicholas Petrick, Romain Pirracchio, and Fan Xia. Towards a Post-Market monitoring framework for machine learning-based medical devices: A case study. *Workshop on Regulatable Machine Learning at the 37th Conference on Neural Information Processing Systems*, 2023.

Alexandra B Steverson, Paul J Marano, Caren Chen, Yifei Ma, Rachel J Stern, Jean Feng, Efstathios D Gennatas, James D Marks, Matthew S Durstenfeld, Jonathan D Davis, Priscilla Y Hsue, and Lucas S Zier. Predictors of All-Cause Thirty-Day readmissions in patients with heart failure at an urban safety net hospital: The importance of social determinants of health. *American Journal of Medicine Open*, 2023.

Ali Sabbagh, Samuel L Washington, 3rd, Derya Tilki, Julian C Hong, Jean Feng, Gilmer Valdes, Ming-Hui Chen, Jing Wu, Hartwig Huland, Markus Graefen, Thomas Wiegel, Dirk Bohmer, Janet E Cowan, Matthew Cooperberg, Felix Y Feng, Mack Roach, 3rd, Bruce J Trock, Alan W Partin, Anthony V D'Amico, Peter R Carroll, and Osama Mohamad. Development and external validation of a machine learning model for prediction of lymph node metastasis in patients with prostate cancer. *Eur Urol Oncol*, 2023.

Ali Sabbagh, Derya Tilki, Jean Feng, Hartwig Huland, Markus Graefen, Thomas Wiegel, Dirk Böhmer, Julian C Hong, Gilmer Valdes, Janet E Cowan, Matthew Cooperberg, Felix Y Feng, Tarek Mohammad, Mohamed Shelan, Anthony V D'Amico, Peter R Carroll, and Osama Mohamad. Multi-institutional development and external validation of a machine learning model for the prediction of distant metastasis in patients treated by salvage radiotherapy for biochemical failure after radical prostatectomy. *European Urology Focus*, 2023.

Phoebe N Miller, Fernanda Romero-Hernandez, Patricia Conroy, Lucia Calthorpe, Mignote Yilma, Sarah Mohamedaly, Yvonne M Kelly, Jean Feng, Kenzo Hirose, Kimberly Kirkwood, Ajay V Maker, Carlos Corvera, Eric Nakakura, Adnan Alseidi, and Mohamed A Adam. Hand-Assisted versus pure Minimally-Invasive distal pancreatectomy: Is there a downside to lending a hand? *World J. Surg.*, 2023.

Jackie J Lin, Patricia C Conroy, Fernanda Romero-Hernandez, Mignote Yilma, Jean Feng, Kenzo Hirose, Eric Nakakura, Ajay V Maker, Carlos Corvera, Kimberly Kirkwood, Adnan Alseidi, and Mohamed A Adam. Hypertension requiring medication use: a silent predictor of poor outcomes after pancreaticoduodenectomy. *J. Gastrointest. Surg.*, 2023.

Leila Haghighat, Satvik Ramakrishna, James W Salazar, Jean Feng, Joey Chiang, Ellen Moffatt, and Zian H Tseng. Homelessness and incidence and causes of sudden death: Data from the POST SCD study. *JAMA Intern. Med.*, 2023.

Jean Feng and Noah Simon. Ensembled sparse-input hierarchical networks for high-dimensional datasets. *Statistical Analysis and Data Mining*, 2022.

Jean Feng, Rachael V Phillips, Ivana Malenica, Andrew Bishara, Alan E Hubbard, Leo A Celi, and Romain Pirracchio. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. *npj Digital Medicine*, 2022.

Jean Feng, Gene Pennello, Nicholas Petrick, Berkman Sahiner, Romain Pirracchio, and Alexej Gossmann. Sequential algorithmic modification with test data reuse. *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, 2022.

Jean Feng, Alexej Gossmann, Berkman Sahiner, and Romain Pirracchio. Bayesian logistic regression for online recalibration and revision of risk prediction models with performance guarantees. *Journal of the American Medical Informatics Association*, 2022.

Daniel Lazzareschi, Ravindra L Mehta, Laura M Dember, Juliane Bernholz, Alparslan Turan, Amit Sharma, Sachin Kheterpal, Chirag R Parikh, Omar Ali, Ivonne H Schulman, Abigail Ryan, Jean Feng, Noah Simon, Romain Pirracchio, Patrick Rossignol, and Matthieu Legrand. Overcoming barriers in the design and implementation of clinical trials for acute kidney injury: a report from the 2020 kidney disease clinical trialists meeting. *Nephrol. Dial. Transplant*, 2022.

Andre Esteva, Jean Feng, Douwe van der Wal, Shih-Cheng Huang, Jeffry P Simko, Sandy DeVries, Emmalyn Chen, Edward M Schaeffer, Todd M Morgan, Yilun Sun, Amirata Ghorbani, Nikhil Naik, Dhruv Nathawani, Richard Socher, Jeff M Michalski, Mack Roach, Thomas M Pisansky, Jedidiah M Monson, Farah Naz, James Wallace, Michelle J Ferguson, Jean-Paul Bahary, James Zou, Matthew Lungren, Serena Yeung, Ashley E Ross, Howard M Sandler, Phuoc T Tran, Daniel E Spratt, Stephanie Pugh, Felix Y Feng, and Osama Mohamad. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *npj Digital Medicine*, 2022.

Jean Feng, Arjun Sondhi, Jessica Perry, and Noah Simon. Selective prediction-set models with coverage guarantees. *Biometrics*, 2021.

Jean Feng, Scott Emerson, and Noah Simon. Approval policies for modifications to Machine Learning-Based software as a medical device: A study of bio-creep. *Biometrics*, 2021.

Jean Feng, William S DeWitt, Aaron McKenna, Noah Simon, Amy Willis, and Frederick A Matsen. Estimation of cell lineage trees by maximum-likelihood phylogenetics. *Annals of Applied Statistics*, 2021.

Jean Feng. Learning to safely approve updates to machine learning algorithms. *Proceedings of the Conference on Health, Inference, and Learning*, 2021.

Brian D Williamson and Jean Feng. Efficient nonparametric statistical inference on population feature importance using shapley values. *International Conference on Machine Learning*, 2020.

Jean Feng and Noah Simon. An analysis of the cost of hyper-parameter selection via split-sample validation, with applications to penalized regression. *Statistica Sinica*, 2020.

Jean Feng, David A Shaw, Vladimir N Minin, Noah Simon, and Frederick A Matsen, IV. Survival analysis of DNA mutation motifs with penalized proportional hazards. *Ann. Appl. Stat.*, 2019.

Kristian Davidsen, Branden J Olson, William S DeWitt, 3rd, Jean Feng, Elias Harkins, Philip Bradley, and Frederick A Matsen, 4th. Deep generative models for T cell receptor protein sequences. *Elife*, 2019.

Jean Feng, Brian Williamson, Noah Simon, and Marco Carone. Nonparametric variable importance using an augmented neural network with multi-task learning. *International Conference on Machine Learning*, 2018.

Jean Feng and Noah Simon. Gradient-based regularization parameter selection for problems with nonsmooth penalty functions. *J. Comput. Graph. Stat.*, 2018.

Pre-prints

Jean Feng, Alexej Gossmann, Romain Pirracchio, Nicholas Petrick, Gene Pennello, and Berkman Sahiner. Is this model reliable for everyone? testing for strong calibration. 2307.15247.

Jean Feng, Alexej Gossmann, Gene Pennello, Nicholas Petrick, Berkman Sahiner, and Romain Pirracchio. Monitoring machine learning (ML)-based risk prediction algorithms in the presence of confounding medical interventions. 2211.09781.

Jean Feng and Noah Simon. Sparse-Input neural networks for high-dimensional nonparametric regression and classification. *arXiv*, 2019.

Abstracts

A Sabbagh, D Tilki, J Feng, J C Hong, M H Chen, J Wu, H Huland, M Graefen, T Wiegel, D Böhmer, S Washington, III, J Cowan, M R Cooperberg, F Y Feng, P Carroll, B Trock, A W Partin, A V DAmico, and O Mohamad. Machine learning for the prediction of distant metastases following postprostatectomy salvage radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 2022.

A Sabbagh, S Washington Iii, D Tilki, J C Hong, J Feng, M H Chen, J Wu, H Huland, M Graefen, T Wiegel, D Böhmer, J Cowan, M R Cooperberg, F Y Feng, B Trock, A W Partin, A V DAmico, P Carroll, and O Mohamad. Machine learning for the prediction of lymph node metastasis in patients with prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 2022.

I Friesner, J Feng, S Kalnicki, M K Garg, N Ohri, and J C Hong. Machine Learning-Based prediction of hospitalization using daily step counts for patients undergoing chemoradiation. *Int. J. Radiat. Oncol. Biol. Phys.*, 2022.

Denise Cecil, Jean Feng, Alex Paynter, Jessica Perry, Noah Simon, Nicholas Drovetto, Lauren Corulli, Erin Rodmaker, Susan Strenk, David Fredricks, and Mary Disis. 1308 bacteria specific IL-10 secreting t-cells derived from the gut are cross-reactive with tumor antigens and accelerate tumor growth in mouse models. *J Immunother Cancer*, 2022.

Presentations

Invited Oral Presentations and Seminars

2023   *Statistical Tools for Auditing Machine Learning Algorithms Across Subgroups and Time*, IMS Young Mathematical Scientists Forum

2023   *Statistical Tools for Auditing Machine Learning Algorithms Across Subgroups and Time*, Bay Area Biotech-Pharma Statistics Workshop

2023   *Sequential algorithmic modification with test data reuse*, Medical Imaging and Data Resource Center Seminar

2023   *Is this risk prediction model reliable for everyone? A test for strong model calibration*, Joint Statistical Meetings

2023   *Monitoring machine learning (ML)-based risk prediction algorithms in the presence of confounding medical interventions*, Joint Statistical Meetings

2023 *Monitoring machine learning (ML)-based risk prediction algorithms in the presence of confounding medical interventions*, International Chinese Statistical Association Applied Statistics Symposium

2023 *Quality assurance and improvement for Machine Learning-based clinical decision support systems*, UC Berkeley Biostatistics Seminar

2023 *Monitoring machine learning (ML)-based prediction algorithms in the presence of confounding medical interventions*, AMIA 2023 Informatics Summit

2022 *Efficient nonparametric statistical inference for population variable importance*, IMS International Conference on Statistics and Data Science (ICSDS)

2022 *Opportunities at the intersection of Machine Learning and Epidemiology*, Society for Epidemiologic Research (SER) Digital

2022 *Quality assurance and improvement for Machine Learning-based clinical decision support systems*, Colorado School of Public Health

2022 *Statistical Methods for Monitoring and Updating AI/ML-Based Software as a Medical Device*, ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop

2022 *Sequential algorithmic modification with test data reuse*, Uncertainty in Artificial Intelligence

2022 *Approval policies for modifications to Machine Learning-Based Software as a Medical Device: A study of bio-creep*, International Biometrics Society Journal Club

2021 *Safe approval policies for continual learning systems in healthcare*, Brown University

2021 *Bayesian logistic regression for online recalibration and revision of risk prediction models with guarantees*, ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop

2021 *Variable Selection and Architecture Search for Neural Networks*, ASA Statistical Learning and Data Science Webinar

2021 *Bayesian logistic regression for online recalibration and revision of risk prediction models with guarantees*, Western North American Region (WNAR) Annual Meeting

2021 *Learning to safely approve updates to machine learning algorithms*, ACM Conference on Health, Inference, and Learning

2021 *Safe approval policies for continual learning systems in healthcare*, University of Waterloo

2020 *Efficient nonparametric statistical inference on population feature importance using Shapley values*, International Conference on Machine Learning

2020 *Training Procedures and Regulatory Policies for Safe Machine Learning Models in Healthcare*, University of California, San Francisco

2020 *Training Procedures and Regulatory Policies for Safe Machine Learning Models in Healthcare*, The University of Texas, MD Anderson Cancer Center

2020 *Training Procedures and Regulatory Policies for Safe Machine Learning Models in Healthcare*, University of California, Irvine

2020 *Approval policies for modifications to Machine Learning-Based Software as a Medical Device: A study of Bio-creep*, International Conference on Health Policy Statistics

2019	*Sparse-Input Neural Networks for High-dimensional Nonparametric Regression and Classification*, Western North American Region (WNAR) Annual Meeting
2018	*Nonparametric variable importance using an augmented neural network with multi-task learning*, International Conference on Machine Learning
2018	*Sparse-Input Neural Networks for High-dimensional Nonparametric Regression and Classification*, University of Washington Biostatistics Colloquium
2018	*Sparse-Input Neural Networks for High-dimensional Nonparametric Regression and Classification*, Joint Statistical Meetings
2017	*Sparse-Input Neural Networks for High-dimensional Nonparametric Regression*, ICML Workshop on Principled Approaches to Deep Learning
2011	*Haptic Belt with Pedestrian Detection*, Neural Information Processing Systems

Contributed Oral Presentations

| 2023 | *Monitoring machine learning (ML)-based risk prediction algorithms: Addressing the challenge of confounding medical interventions*, 2023 Innovations in Regulatory Science Summit |
| 2019 | *Uncertainty-Aware Black-Box Predictors with Coverage Guarantees*, Joint Statistical Meetings |

Funding History

| 2020-2023 | UCSF-Stanford CERSI Program; Role: PI
"Safe algorithmic change protocols for modifications to AI/ML-based Software as a Medical Device."
$170,000 in direct costs |
| 2023-2026 | Patient-Centered Outcomes Research Institute; Role: PI
"Diagnostic tools for quality improvement of machine learning-based clinical decision support systems."
$750,000 in direct costs |

Awards

2020	University of Washington Thomas R. Fleming Excellence in Biostatistics Award
2020	International Conference on Health Policy Statistics, Student Travel Award For manuscript: *Approval policies for modifications to Machine Learning-Based Software as a Medical Device*
2018	Joint Statistical Meetings Section on Statistical Learning and Data Science, Student Paper Award For manuscript: *Sparse-input neural networks for high-dimensional nonparametric regression and classification*
2018	University of Washington Biostatistics Donovan J. Thompson Award for Best Combined Performance on Ph.D. Theory and Applied Qualifying Examinations
2015–2017	Big Data for Genomics and Neuroscience Training Grant

Software

EASIER-Net	Python and R packages for fitting neural networks for high-dimensional data
	`https://github.com/jjfeng/easier_net`
	`https://github.com/jjfeng/easier_net_R`

GapML	Python package for analyzing cell-lineage tracing data from GESTALT `https://github.com/matsengrp/gestaltamania`
SPINN	Python package for estimating sparse-input neural networks `http://github.com/jjfeng/spinn`
samm	Python package for estimating somatic hypermutation rates of nucleotide motifs `http://github.com/matsengrp/samm`

Teaching

2021-Present	Instructor, Biostat 216: Machine Learning in R for the Biomedical Sciences, UCSF
2022	Lecturer, Epidemiology, Biostatistics and Population Science (EBPS), UCSF School of Medicine Bridges Curriculum
2020-Present	Instructor, Machine Learning Boot Camp: Analyzing Biomedical and Health Data, Columbia University
2020	Instructor, Supervised statistical learning, 6th Seattle Symposium in Biostatistics

Student Advising

2023-Present	Harvineet Singh, Postdoctoral researcher
2023-Present	Jinhee Lee, UCSF-UC Berkeley Computational Precision Health Designated Emphasis program, UCSF-UC Berkeley Bioengineering PhD program
Spring 2022	Margaret Tsui, Biomedical Informatics PhD program, Qualifying Exam Committee
Summer 2022	Amanda Everitt, Biomedical Informatics PhD program, Qualifying Exam Committee

Faculty Advising

2023-Present	Andrew Bishara, Assistant Professor, K23 grant

Service

Conferences and workshops
o Track Chair, Conference on Health, Inference, and Learning (CHIL) 2023
Referee Service
o Conference on Neural Information Processing Systems
o NeurIPS Workshop on Distribution Shifts
o Frontiers in Digital Health
o Lancet Digital Health
o Nature Medicine
o International Conference on Machine Learning
o International Conference on Learning Representations
o Journal of Computational and Graphical Statistics
o Statistics in Biopharmaceutical Research
o Annals of Applied Statistics

- Annals of Statistics
- Statistics in Medicine
- Neural Networks
- ASA Section on Statistical Learning and Data Science Student Paper Committee
- Communications Medicine

Conferences
- Joint Statistical Meetings, Session Chair
- Western North American Region (WNAR) Annual Meeting, Session Chair
- Eastern North American Region (ENAR) Annual Meeting, Roundtable leader

University Service
- UCSF Artificial Intelligence Oversight Committee, March 2023–Present
- UCSF Artificial Intelligence/Machine Learning Demonstration Projects Steering Committee, Winter-Spring 2022
- UCSF Initiative for Digital Transformation in Computational Biology & Health Grant Review Committee, September 2021
- Department of Epidemiology and Biostatistics, Digital Health Initiative Steering Committee, 2020–Present
- Department of Epidemiology and Biostatistics, Admissions Committee for the Master's Degree in Health Data Science, 2022–Present

Work Experience

2019	**Research Intern**, *Insitro*, South San Francisco, CA Developed statistical models of genomic data.
2012–2015	**Software engineer**, *Coursera*, Mountain View, CA Built the professional certificate program and payment system. Technical lead on projects with 3-5 people. Mentored interns and junior engineers.

Other

Summer 2023	Guest speaker for Kode with Klossy Summer program
2022-present	Guest lecturer for UCSF AI4ALL Summer program