George H. Chen

4800 Forbes Ave HBH 2216, Pittsburgh PA 15213 ✉georgechen@cmu.edu

Research Interests

Trustworthy machine learning, survival analysis, time series analysis, nonparametric methods, health applications

Academic Appointments

Carnegie Mellon University Pittsburgh, PA

Associate Professor without tenure, Heinz College of Information Systems & Public Policy — July 2024 – present
Assistant Professor, Heinz College of Information Systems & Public Policy — Jan. 2017 – June 2024
Affiliated Faculty (courtesy appointment), Machine Learning Department — Aug. 2017 – present

Massachusetts Institute of Technology Cambridge, MA

Digital Learning Fellow & Postdoctoral Associate, Electrical Engineering and Computer Science Department — Sept. 2015 – Dec. 2016

Education

Massachusetts Institute of Technology Cambridge, MA — Sept. 2010 – June 2015

Ph.D. in Electrical Engineering and Computer Science, June 2015 — GPA: 5.0/5.0
- Thesis: *Latent Source Models for Nonparametric Inference* (advisors: Polina Golland, Devavrat Shah)
  Received the George M. Sprowls award for best thesis in Computer Science

Electrical Engineer (post-master's degree focused on coursework), June 2014 — GPA: 5.0/5.0
S.M. in Electrical Engineering and Computer Science, June 2012 — GPA: 5.0/5.0

University of California, Berkeley Berkeley, CA — Aug. 2006 – May 2010

B.S. with dual majors: — GPA: 3.98/4.0
Electrical Engineering and Computer Sciences (*Highest Honors*)
Engineering Mathematics and Statistics (*Highest Honors*)

Selected Industry Experience

Co-Founder and Advisor, CoolCrop (startup) Vadodara, India — June 2016 – present

Providing refrigeration and marketing analytics to farmers. We provide refrigerators to farmers in rural India along with market forecasts to help them decide when and where to sell their harvest. We now have 48 installations across 7 states in India, servicing around 9000 farmers. I help develop machine learning approaches to forecast produce prices at different markets. I still actively advise students/interns and serve on the board of directors.

Selected Awards and Honors

NSF CAREER award 2021 for "Real-Time Nonparametric Machine Learning for Healthcare with Guarantees" (award number 2047981)

Best paper award at the AAAI Workshop on AI for Behavior Change 2021

Best paper (theoretical track) at the INFORMS Data Mining and Decision Analytics Workshop 2019

MIT Goodwin Medal 2015
Top campus-wide teaching award given to graduate students; selectivity: typically 1 student across all of MIT chosen per year (occasionally 2 students)

MIT George M. Sprowls Award 2015
  Best Ph.D. thesis in Computer Science at MIT; selectivity: a few graduating Ph.D. students chosen each year from the EECS department

MIT IDEAS Global Challenge 2014: $10,000 Grand Prize Winning Team
  Annual MIT innovation and entrepreneurship competition focused on public service; selectivity: 4 teams out of 37

MIT EECS Frederick C. Hennie III Teaching Award 2013
  Departmental teaching award; selectivity: ~1-3 students in the EECS department chosen per year

National Defense Science and Engineering Graduate Fellowship 2012-2015 (award year: 2012)
  Three-year funding for graduate school; selectivity: ~200 US graduate students per year

National Science Foundation Graduate Research Fellowship 2011-2012 (award year: 2010)
  Three-year funding for graduate school (I forfeited most of this due to other fellowships); selectivity: ~2,000 US graduate students per year

Siebel Scholarship 2011-2012
  Funding for final year of my master's degree program; selectivity: within MIT EECS, 5 graduate students chosen per year

MIT Irwin Mark Jacobs and Joan Klein Jacobs Presidential Fellowship 2010-2011
  Funding for first year of graduate school; selectivity: ~15 out of the ~150 admitted EECS students per year

Eta Kappa Nu Outstanding Electrical and Computer Engineering Student Award 2010
  For academic achievements, moral character, and service: 1 student chosen per year across universities with accredited electrical and computer engineering programs (nominations are done by Eta Kappa Nu chapters but the recipient need not be a member of Eta Kappa Nu)

UC Berkeley Engineering Science Departmental Citation 2010
  Top graduating senior in UC Berkeley's Engineering Science program; selectivity: 1 student per year

UC Berkeley Outstanding Graduate Student Instructor Award 2009
  Campus-wide teaching award (received as an undergraduate teaching assistant); selectivity: ~10% of teaching assistants

Intel International Science and Engineering Fair 2006: 1st place team
  For high school research in computer science; selectivity: 7 teams out of 236
  As part of the award, an asteroid was named after me: `https://ssd.jpl.nasa.gov/` (search for "21437")

Books/Monographs

[B2] "An Introduction to Deep Survival Analysis Models for Predicting Time-to-Event Outcomes"
  George H. Chen
  Foundations and Trends in Machine Learning, December 2024

[B1] "Explaining the Success of Nearest Neighbor Methods in Prediction"
  George H. Chen, Devavrat Shah
  Foundations and Trends in Machine Learning, May 2018

Journal Papers (* denotes equal contribution)

[J7] "Fairness in Survival Analysis with Distributionally Robust Optimization"
  Shu Hu*, George H. Chen*
  Journal of Machine Learning Research (JMLR), August 2024
  Journal paper version of conference paper [C19] published at ML4H 2022

[J6] "Neural Topic Models with Survival Supervision: Jointly Predicting Time-to-Event Outcomes and Learning How Clinical Features Relate"
George H. Chen*, Linhong Li*, Ren Zuo, Amanda Coston, Jeremy C. Weiss
Artificial Intelligence in Medicine, June 2024
Journal paper version of paper [C17] published at AIME 2020

[J5] "Survival Kernels: Scalable and Interpretable Deep Kernel Survival Analysis with an Accuracy Guarantee"
George H. Chen
Journal of Machine Learning Research (JMLR), February 2024
Best paper finalist (applied track) at the INFORMS Data Mining and Decision Analytics Workshop, October 2022

[J4] "Influence via Ethos: On the Persuasive Power of Reputation in Deliberation Online"
Emaad Ahmed Manzoor, George H. Chen, Dokyun Lee, Michael D. Smith
Management Science, May 2023
Best paper at the AAAI Workshop on AI for Behavior Change 2021

[J3] "ECOD: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions"
Zheng Li*, Yue Zhao*, Xiyang Hu, Nicola Botta, Cezar Ionescu, George H. Chen
IEEE Transactions on Knowledge and Data Engineering, March 2022

[J2] "Consumer Behavior in the Online Classroom: Using Video Analytics and Machine Learning to Understand the Consumption of Video Courseware"
Mi Zhou, George H. Chen, Pedro Ferreira, Michael D. Smith
Journal of Marketing Research, December 2021

[J1] "Targeting Villages for Rural Development Using Satellite Image Analysis"
Kush R. Varshney, George H. Chen, Brian Abelson, Kendall Nowocin, Vivek Sakhrani, Ling Xu, Brian L. Spatocco
Big Data, March 2015

Conference Papers (* denotes equal contribution, α-β denotes alphabetical author ordering)

[C25] "Generalized Prompt Tuning: Adapting Frozen Univariate Time Series Foundation Models for Multivariate Healthcare Time Series"
Mingzhu Liu, Angela H. Chen, George H. Chen
Machine Learning for Health (ML4H), December 2024

[C24] "Improving Fairness in Deepfake Detection"
Yan Ju*, Shu Hu*, Shan Jia, George H. Chen, Siwei Lyu
IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 2024

[C23] "Temporal Supervised Contrastive Learning for Modeling Patient Risk Progression"
Shahriar Noroozizadeh, Jeremy C. Weiss, George H. Chen
Machine Learning for Health (ML4H), December 2023

[C22] "Neurological Prognostication of Post-Cardiac-Arrest Coma Patients Using EEG Data: A Dynamic Survival Analysis Framework with Competing Risks"
Xiaobin Shen, Jonathan Elmer, George H. Chen
Machine Learning for Healthcare (MLHC), August 2023

[C21] "A General Framework for Visualizing Embedding Spaces of Neural Survival Analysis Models Based on Angular Information"
George H. Chen
Conference on Health, Inference, and Learning (CHIL), June 2023

[C20] "BOND: Benchmarking Unsupervised Outlier Node Detection on Static Attributed Graphs"
Kay Liu*, Yingtong Dou*, Yue Zhao*, Xueying Ding, Xiyang Hu, Ruitong Zhang, Kaize Ding, Canyu Chen, Hao Peng, Kai Shu, Lichao Sun, Jundong Li, George H. Chen, Zhihao Jia, Philip S. Yu
Neural Information Processing Systems (NeurIPS) (Datasets and Benchmarks track), November-December 2022

[C19] "Distributionally Robust Survival Analysis: A Novel Fairness Loss Without Demographics"
Shu Hu*, George H. Chen*
Machine Learning for Health (ML4H), November 2022

[C18] "TOD: Tensor-Based Outlier Detection, a General GPU-Accelerated Framework"
Yue Zhao, George H. Chen, Zhihao Jia
Proceedings of the VLDB Endowment, Vol 16, No. 3, November 2022

[C17] "Neural Topic Models with Survival Supervision: Jointly Predicting Time-to-Event Outcomes and Learning How Clinical Features Relate"
Linhong Li, Ren Zuo, Amanda Coston, Jeremy C. Weiss, George H. Chen
International Conference on Artificial Intelligence in Medicine (AIME), August 2020

[C16] "Predicting Mortality Risk in Viral and Unspecified Pneumonia to Assist Clinicians with COVID-19 ECMO Planning"
Helen Zhou*, Cheng Cheng*, Zachary C. Lipton, George H. Chen, Jeremy C. Weiss
Artifical Intelligence in Medicine, August 2020
(Also presented at the *International Conference on Machine Learning (ICML) Workshop on Global Health*, July 2020)

[C15] "Deep Kernel Survival Analysis and Subject-Specific Survival Time Prediction Intervals"
George H. Chen
Machine Learning for Healthcare, July 2020

[C14] "Missing Not at Random in Matrix Completion: The Effectiveness of Estimating Missingness Probabilities Under a Low Nuclear Norm Assumption"
Wei Ma*, George H. Chen*
Neural Information Processing Systems (NeurIPS), December 2019
Best paper (theoretical track) at the INFORMS Data Mining and Decision Analytics Workshop 2019

[C13] "Truck Traffic Monitoring with Satellite Images"
Lynn H. Kaack, George H. Chen, M. Granger Morgan
Computing and Sustainable Societies (COMPASS), July 2019
(Also presented at the *International Conference on Machine Learning (ICML) Workshop on Climate Change*, June 2019)

[C12] "Nearest Neighbor and Kernel Survival Analysis: Nonasymptotic Error Bounds and Strong Consistency Rates"
George H. Chen
International Conference on Machine Learning (ICML), June 2019

[C11] "An Interpretable Produce Price Forecasting System for Small Farmers in India using Collaborative Filtering and Adaptive Nearest Neighbors"
Wei Ma, Kendall Nowocin, Niraj Marathe, George H. Chen
Information and Communication Technologies and Development (ICTD), January 2019

[C10] "Toward Reducing Crop Spoilage and Increasing Small Farmer Profits in India: a Simultaneous Hardware and Software Solution"
George H. Chen, Kendall Nowocin, Niraj Marathe
Information and Communication Technologies and Development (ICTD), November 2017

[C9] "A Latent Source Model for Patch-Based Image Segmentation"
George H. Chen, Devavrat Shah, Polina Golland
Medical Image Computing and Computer Assisted Interventions (MICCAI), October 2015

[C8] "A Latent Source Model for Online Collaborative Filtering"
(*α-β*) Guy Bresler, George H. Chen, Devavrat Shah
Neural Information Processing Systems (NeurIPS), December 2014
Selected for spotlight (one of 62/1678 submissions)

[C7] "A Latent Source Model for Nonparametric Time Series Classification"
(*α-β*) George H. Chen, Stanislav Nikolov, Devavrat Shah
Neural Information Processing Systems (NeurIPS), December 2013

[C6] "Sparse Projections of Medical Images onto Manifolds"
George H. Chen, Christian Wachinger, Polina Golland
Information Processing in Medical Imaging (IPMI), June-July 2013

[C5] "Indoor Localization and Visualization Using a Human-Operated Backpack System"
Timothy Liu, Matthew Carlberg, George Chen, Jacky Chen, John Kua, Avideh Zakhor
International Conference on Indoor Positioning and Indoor Navigation (IPIN), September 2010

[C4] "Indoor Localization Algorithms for a Human-Operated Backpack System"
George Chen, John Kua, Stephen Shum, Nikhil Naikal, Matthew Carlberg, Avideh Zakhor
International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), May 2010

[C3] "Classifying Urban Landscape in Aerial LIDAR Using 3D Shape Analysis"
Matthew Carlberg, Peiran Gao, George Chen, Avideh Zakhor
International Conference on Image Processing (ICIP), November 2009

[C2] "2D Tree Detection in Large Urban Landscapes Using Aerial LIDAR Data"
George Chen, Avideh Zakhor
International Conference on Image Processing (ICIP), November 2009

[C1] "Image Augmented Laser Scan Matching for Indoor Dead Reckoning"
Nikhil Naikal, John Kua, George Chen, Avideh Zakhor
International Conference on Intelligent Robots and Systems (IROS), October 2009

Workshop Papers (excludes papers whose final versions are published in journal or conference venues)

[W2] "Survival-Supervised Topic Modeling with Anchor Words: Characterizing Pancreatitis Outcomes"
George H. Chen, Jeremy C. Weiss
Neural Information Processing Systems (NeurIPS) Workshop on Machine Learning for Health, December 2017
(Also presented at *Society for Medical Decision Making North American Meeting*, October 2017)

[W1] "Deformation-Invariant Sparse Coding for Modeling Spatial Variability of Functional Patterns in the Brain"
George H. Chen, Evelina G. Fedorenko, Nancy G. Kanwisher, Polina Golland
Neural Information Processing Systems (NeurIPS) Workshop on Machine Learning and Interpretation in Neuroimaging,
December 2011

Teaching Experience

Instructor, *Unstructured Data Analytics* (CMU 94-775/95-865)   Every Fall & Spring term since Fall 2017 (95-865)
                                                                                                            Spring terms 2018–2022, 2025 (94-775)

I teach two versions of this half-semester course that's on analyzing text and images using machine learning: 94-775 is for public policy students, and 95-865 is for information systems students. I developed 94-775/95-865 largely from scratch. It is heavily practical with homework assignments completed in Python Jupyter notebooks and averages roughly one code demo per lecture. My lectures cover basic natural language processing, exploratory data analysis including manifold learning and topic modeling, and predictive data analysis including deep neural networks for analyzing images and time series. Starting Fall 2023, I also cover text generation using n-gram models and generative pre-trained transformers (GPTs). Main differences between the two course versions: 94-775 has a final group project that must address a public policy problem. To make time for this project, 94-775 students are not required to learn the deep learning part of the course in depth (they are not graded on any assignment that requires deep learning). In contrast, 95-865 students are required to learn to code in PyTorch (they have a homework assignment and an exam that both use PyTorch); 95-865 students do not do a final project.

Instructor (Computer Science), Middle East Entrepreneurs of Tomorrow (MEET)   Summers 2015, 2016, 2023

MEET brings promising high school Palestinian and Israeli students together for a three-year program that teaches them computer science, entrepreneurship, and a deeper understanding of the conflict in the region. MEET aims to educate and empower these young leaders to create positive political and social change in the Middle East. I co-taught a three-week introduction to web development in Jerusalem for the students in their second year of MEET. For their final project, the students worked in binational teams to develop socially responsible web apps that promote cultural diversity, which they presented to Google Tel Aviv. Our material coverage included HTML, CSS, Bootstrap, Python, SQLAlchemy, Flask, and JavaScript. I created new lectures and labs on building web apps with Flask.

Instructor, *Computational Probability and Inference* (MIT 6.008.1x, offered on edX)   Fall 2016

New intro-college/advanced-high-school-level course covering basic probability, probabilistic graphical models, and learning probability distributions. These three main topics were covered with a heavy emphasis on coding. The course prerequisites were comfort in Python programming and calculus. I developed nearly all the course notes, 75% of the videos, numerous new exercises, all the autograders, and a new two-part final project on learning structure in agriculture data and forecasting market prices. This course is based on the first half of an existing residential MIT course 6.008 that I helped develop (see below for the course *Introduction to Inference*) but differs in that its presentation has been made to be accessible to a much broader audience.

Teaching Assistant, *Introduction to Inference* (MIT 6.S080, now renumbered as 6.3800)   Fall 2012, Spring 2014

New undergraduate core Electrical Engineering and Computer Science course introducing statistical inference and probabilistic graphical models. I taught for the class during the first two semesters that it had ever been offered. I developed substantial portions of the courseware, including Khan Academy style videos for students, a series of PYTHON robot tracking coding projects, new recitation notes, new problem sets, and more. For my teaching and courseware development, I won a departmental teaching award as well as the top campus-wide teaching award given to graduate students.

Instructor, *Statistics for Research Projects* (MIT 6.S085)   January 2014

Two-week introduction to statistics with an emphasis on recognizing when and how to apply different statistical tools to research problems. I updated the lectures to fold in ideas from predictive analytics and mathematical statistics.

Teaching Assistant, *Algorithms for Inference* (MIT 6.438, now renumbered as 6.7810)   Fall 2011

Introductory graduate-level course on probabilistic graphical models. I made Khan Academy style videos and helped typeset the first complete set of lecture notes for the class. I also delivered three lectures for the Fall 2013 class.

Teaching Assistant, *Structure and Interpretation of Systems and Signals* (UCB EE20)   Fall 2007, Spring 2008, Fall 2008

Undergraduate core Electrical Engineering and Computer Sciences course introducing the math behind processing images, audio, and video, and behind systems for communication (e.g., radio) and control (e.g., robotic manipulators). My third time teaching for this course, I was the head teaching assistant, managing seven other teaching assistants. My teaching effectiveness was recognized by a campus teaching award.

Student and Post-Doc Supervision

Current PhD students (at CMU's Heinz College, PhD student advising is largely similar to in mathematics, economics, and statistics; students are independent and not required to operate in a "lab model" as is typical in CS/engineering):

- Shahriar Noroozizadeh (co-advised by Jeremy C. Weiss)
- Xiaobin Shen

Current master's student collaborator: Mingzhu Liu

Current external collaborator: Shaopeng Zhang

Past postdoc: Shu Hu, Assistant Professor at Purdue University in Indianapolis

Past PhD students (♣ indicates a PhD student who worked with me on a secondary master's in machine learning, i.e., I was their master's research advisor but not their PhD research advisor):

- Helen S. Zeng (primarily advised by Michael D. Smith), Assistant Professor at UC Davis Graduate School of Management
- Yue Zhao (PhD 2023; co-advised by Leman Akoglu and Zhihao Jia), Assistant Professor at USC in Computer Science
- Emaad Manzoor (PhD 2021; co-advised by Dokyun Lee), Assistant Professor at Cornell University Johnson Graduate School of Management
- Mi Zhou (PhD 2020; co-advised by Michael D. Smith and Pedro Ferreira), Assistant Professor at UBC Sauder School of Business
- ♣ Wei Ma (master's in ML 2018/PhD 2019), Assistant Professor at Hong Kong Polytechnic University in the Department of Civil and Environmental Engineering
- ♣ Lynn H. Kaack (master's in ML 2018/PhD 2019), Assistant Professor at the Hertie School

Past master's students, and first employment/position after graduation:

- Thomas Tam (MSPPM 2023), Sunstella Foundation/Jewish Healthcare Foundation
- Brenda Palma (MISM 2022), Dell
- Xiaotong (Maggie) Lu (MISM 2020), McKinsey
- Runtong (Fred) Yang (MISM 2019), Capitol One
- Ren Zuo (MISM 2018), Cornerstone Research

Past undergraduate students, and first employment/position after graduation:

- Linhong (Lexie) Li (B.S. 2020), McKinsey
- Junyan Pu (B.S. 2020), CMU master's degree program in CS

Selected Service

I taught a survival analysis tutorial "A Tour of Survival Analysis, from Classical to Modern" at the following venues:
- *SIGMETRICS* 2021, virtual
- *Conference on Health, Inference, and Learning (CHIL)* 2020, virtual (co-taught with Jeremy C. Weiss)

Conference organizing committees:
- *Conference on Health, Inference, and Learning (CHIL)* 2024 research roundtables chair (joint with Xiaoxiao Li)
- *Conference on Health, Inference, and Learning (CHIL)* 2022 proceedings chair (joint with Gerardo Flores and Tom Pollard)
- *Conference on Health, Inference, and Learning (CHIL)* 2021 proceedings chair (joint with Emily Alsentzer and Matthew McDermott) and track chair (joint with Rumi Chunara and Alistair Johnson for the track "Impact and Society")

Workshop and symposia organizing committees:
- *AAAI Fall Symposium 2023: Survival Prediction Algorithms, Challenges & Applications*, Arlington, VA (joint with Russ Greiner, Chirag Nagpal, Weijing Tang, Kevin Xu)
- *KDD Machine Learning for Consumers and Markets Workshop* 2021, virtual (joint with Wen Wang, Han Zhao, and Dokyun Lee)
- *APPAM 2019 Pre-Conference Workshop on ML and Public Policy* 2019, Denver, CO (co-led a session with Jens Ludwig "What New Challenges Arise When Applying ML Tools to Policy Problems?")
- *NeurIPS Nearest Neighbors Workshop* 2017, Long Beach, CA (joint with Devavrat Shah and Christina Lee Yu)

Reviewer (this list is not exhaustive):
- *Machine Learning in Healthcare (MLHC)*: 2017–2021, 2023–2024, area chair 2025
- *Machine Learning for Health (ML4H)*: 2017–2019, metareviewer/area chair 2020–2023
- *Conference on Health, Inference, and Learning (CHIL)*: metareviewer in 2021
- *Neural Information Processing Systems (NeurIPS)*: 2016–2022, area chair 2023–2025
- *International Conference on Machine Learning (ICML)*: 2018–2024, area chair 2025
- *International Conference on Learning Representations (ICLR)*: 2021–2025
- *The Annals of Statistics*: 2019
- *The Annals of Applied Statistics*: 2024
- *Journal of the American Statistical Association*: 2024
- *Statistics and Public Policy*: 2025
- *Management Science*: 2024
- *INFORMS Journal on Computing*: 2022
- *Management Information Systems Quarterly*: 2020

Head of Curriculum, Middle East Entrepreneurs of Tomorrow (MEET)   Oct. 2015 – Oct. 2016

Worked with the other Head of Curriculum (Stephanie Wang) and the team stationed in Jerusalem to establish the entrepreneurship and computer science curriculum for the 2016-2017 MEET school year that brings together Palestinian and Israeli high school students to learn and work together

Member, Task Force on the Future of MIT Education   Apr. 2013 – July 2014

Discussed and made recommendations for the future of MIT education both on and off campus, accounting for the growing prevalence of digital learning tools; I also led student outreach events and helped assemble student and faculty surveys on MIT education

Vice President of Resources, MIT Sidney-Pacific Graduate Community   Apr. 2012 – Apr. 2013

Helped run the largest MIT graduate dorm Sidney-Pacific (700 residents) as one of the five executive officers, working with 50 other graduate student officers, the housemasters, and the building manager to put on copious amounts of events, start a community garden, launch a new website, overhaul building resources, handle severe conflicts, and more; I managed 10 officers and the $140,000 budget