Michael C. Hughes

+1-617-371-76 one one • mike@michaelchughes.com
www.michaelchughes.com • michaelchughes

Research Expertise

Machine Learning *2012-*

- Unsupervised learning, Semi-supervised learning, Supervised learning
- Probabilistic models: Bayesian nonparametrics, topic models, hidden Markov models, deep generative models
- Posterior estimation methods: variational methods, Markov chain Monte Carlo

Clinical Informatics *2017-*

- Tasks: Phenotype discovery, personalized outcome prediction, automated assisted diagnosis
- Data Types: Time series of vitals and lab results from EHR, diagnosis and procedure codes, sociodemographics, imaging (esp. echocardiograms)
- Applications: intensive care early warning systems, suggesting treatments for depression, heart disease diagnosis

Human Activity Analysis *2012-*

- Sensor time-series, motion capture, video, images

Education

Brown University
Ph.D., Computer Science. *2016*

Brown University
M.S., Computer Science. *2012*

Olin College of Engineering
B.S. Electrical & Computer Engineering *2010*

Research Experience

Assistant Professor of Computer Science
Tufts University, Medford, MA *2018 - present*

- Research on statistical machine learning methods and applications to health informatics.
- Advised Ph.D., M.S., and B.S. students in machine learning research projects.
- Taught advanced undergraduate courses: COMP 135 Intro to ML and COMP 136 Statistical Pattern Recog.
- Developed new course for graduate students with research interests: COMP 150 Bayesian Deep Learning).
- Appointed as the Ann W. Lambertus and Peter Lambertus Assistant Professor in 2019

Postdoctoral fellow: Machine learning to improve clinical decisions in healthcare
Adviser: Prof. Finale Doshi-Velez (Harvard) *2016 - 2018*

- Developed semi-supervised models for characterizing and treating depression (with Dr. Perlis and Dr. McCoy).
- Applied time-series models to predict ventilator interventions in the ICU for public dataset of ¿36,000 patients.
- Created methods for training deep models so they are more interpretable to clinicians or other users.

Postdoc project: Estimating carbon biomass from LiDAR waveforms

Adviser: Prof. Erik Sudderth & Prof. Jim Kellner (Brown U., Ecology & Evolutionary Biology) 2016

- Predicted forest biomass from LiDAR waveforms to better understand land use and climate change.
- Modeled waveforms and biomass predictions jointly via nonparametric regression using our BNPy toolbox.
- Intended for use in NASA's upcoming Global Ecosystem Dynamics Investigation (GEDI).

Ph.D. thesis: Reliable and scalable variational inference for Bayesian nonparametrics

Adviser: Prof. Erik Sudderth *2016*

- Thesis Title: Reliable and scalable variational inference for nonparametric mixtures, topics, and sequences
- Developed optimization algorithms for Bayesian nonparametric models that scale to millions of examples.
- Optimized lower bound on marginal likelihood, thus penalizing too simple and too complex explanations.
- Escaped local optima via data-driven proposals that add useful new clusters and remove redundant ones.
- Applied to topic models of 2 million NY Times articles and sequential models of the whole human genome.
- Implemented algorithms in open-source package: Bayesian Nonparametrics for Python (BNPy).

Master's project: Sequential Models for Video and Motion Capture

Adviser: Prof. Erik Sudderth *2012*

- Developed methods to discover common actions from many videos of humans performing household exercises.
- Improved existing inference algorithms with data-driven Metropolis-Hastings proposals.

Highlighted Publications

1. "Detecting Heart Disease from Multi-View Ultrasound Images via Supervised Attention Multiple Instance Learning." Zhe Huang[d], Benjamin S. Wessler, and Michael C. Hughes. Machine Learning for Healthcare Conference (MLHC), 2023.
2. "Non-Parametric and Regularized Dynamical Wasserstein Barycenters for Sequential Observations." Kevin C. Cheng[d], Shuchin Aeron, Michael C. Hughes, Eric L. Miller. IEEE Transactions on Signal Processing, 2023.
3. "Approximate inference by broadening the support of the likelihood." Michael T. Wojnowicz, Martin D. Buck, Michael C. Hughes. Symposium on Advances in Approximate Bayesian Inference (AABI), 2023.
4. "NovelCraft: A Dataset for Novelty Detection and Discovery in Open Worlds." Patrick Feeney[d], Sarah Schneider[d], Panagiotis Lymperopoulos, Liping Liu, Matthias Scheutz, and Michael C. Hughes. Transactions on Machine Learning Research (TMLR), 2023.
5. "Fix-A-Step: Semi-supervised Learning from Uncurated Unlabeled Data." Zhe Huang[d], Mary-Joy Sidhom[u], Benjamin S. Wessler, and Michael C. Hughes. Artificial Intelligence and Statistics (AISTATS) [oral, top 2% of 1500+ submissions], 2023.
6. "Automated Detection of Aortic Stenosis using Machine Learning." Benjamin S. Wessler, Zhe Huang[d], Gary Long, ... and Michael C. Hughes. Journal of the American Society of Echocardiography, 2023.
7. "Predicting Spatiotemporal Counts of Opioid-related Fatal Overdoses via Zero-Inflated Gaussian Processes." Kyle Heuton[d], Shikhar Shrestha, Thomas J. Stopka, Jennifer Pustz, Li-Ping Liu, and Michael C. Hughes. NeurIPS Workshop on Gaussian Processes, Spatiotemporal Modeling, and Decision-making Systems, 2022.
8. "Prediction-Constrained Markov Models for Medical Time Series with Missing Data and Few Labels."

Preetish Rath[d], Gabriel Hope[d], Kyle Heuton[d], Erik B. Sudderth, and Michael C. Hughes. Learning from Time Series For Health (TS4H) Workshop at NeurIPS, 2022.

9. "Easy Variational Inference for Categorical Models via an Independent Binary Approximation." Michael T. Wojnowicz, Shuchin Aeron, Eric L. Miller, and Michael C. Hughes. International Conference on Machine Learning (ICML), 2022.
10. "Optimizing Early Warning Classifiers to Control False Alarms via a Minimum Precision Constraint." Preetish Rath[d] and Michael C. Hughes. Artificial Intelligence and Statistics (AISTATS), 2022.
11. "The Tufts fNIRS Mental Workload Dataset & Benchmark for Brain-Computer Interfaces that Generalize." Zhe Huang[d], Liang Wang[d], Giles Blaney, Christopher Slaughter[u], Devon McKeon, Ziyu Zhou, Robert Jacob, and Michael C. Hughes. Neural Information Processing Systems (NeurIPS) Track on Datasets and Benchmarks, 2021.
12. "Dynamical Wasserstein Barycenters for Time-series Modeling." Kevin C Cheng[d], Shuchin Aeron, Michael C. Hughes, and Eric Miller. Neural Information Processing Systems (NeurIPS), 2021.
13. "A New Semi-supervised Learning Benchmark for Classifying View and Diagnosing Aortic Stenosis from Echocardiograms." Zhe Huang[d], Gary Long, Benjamin Wessler, and Michael C. Hughes. Machine Learning for Healthcare (MLHC), 2021.
14. "Approximate Bayesian Computation for an Explicit-DurationHidden Markov Model of COVID-19 Hospital Trajectories." Gian Marco Visani[u], Alexandra Hope Lee[b], Cuong Nguyen[m], David M. Kent, John B. Wong, Joshua T. Cohen, and Michael C. Hughes. Machine Learning for Healthcare (MLHC), 2021.
15. "Assessment of a Prediction Model for Antidepressant Treatment Stability Using Supervised Topic Models." Michael C. Hughes, Melanie F. Pradier, Andrew Slavin Ross, Thomas H. McCoy Jr, Roy H. Perlis, Finale Doshi-Velez. JAMA Network Open, 2020.
16. "MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III." Shirly Wang[m], Matthew B. A. McDermott, Geeticka Chauhan, Marzyeh Ghassemi, Michael C. Hughes, and Tristan Naumann. The ACM Conference on Health, Inference, and Learning (CHIL), 2020.
17. "Feature Robustness in Non-stationary Health Records: Caveats to Deployable Model Performance in Common Clinical Machine Learning Tasks." Bret Nestor[d], Matthew B. A. McDermott, Willie Boag, Gabriela Berner, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, and Marzyeh Ghassemi. Machine Learning for Healthcare, 2019.

Superscripts indicate mentored student's status: u = undergraduate, m = masters, d = Ph.D. student, b = post-bacc, c = medical student. Complete publication list at end of this document.

Highlighted Preprints

1. "SINCERE: Supervised Information Noise-Contrastive Estimation REvisited." Patrick Feeney[d] and Michael C. Hughes. arXiv, 2023.
2. "Modeling Graph Node Correlations with Neighbor Mixture Models." Linfeng Liu[d], Michael C. Hughes, and Li-Ping Liu. arXiv, 2021.
3. "Learning Consistent Deep Generative Models from Sparse Data via Prediction Constraints." Gabriel Hope[d], Madina Abdrakhmanova, Xiaoyin Chen, Michael C. Hughes, Erik B. Sudderth. arXiv, 2020.
4. "Prediction-Constrained Training for Semi-Supervised Mixture and Topic Models." Michael C. Hughes, Leah Weiner[d], Gabriel Hope[d], Thomas H. McCoy, Roy H. Perlis, Erik B. Sudderth, and

Finale Doshi-Velez. arXiv, 2017.

5. "Fast Learning of Clusters and Topics via Sparse Posteriors." Michael C. Hughes and Erik B. Sudderth. arXiv, 2016.

Honors and Awards for Research

Best Poster Award Time Series Workshop @ ICML 2021 *2021*

- Awarded for our work on Prediction-Constrained Hidden markov models

Best Paper Award, SoCal NLP Symposium 2018 *2018*

- Awarded for 2 page summary of our AISTATS 2018 paper.

Nominee for AMIA Clinical Informatics Research Award *2017*

- 1 of 7 papers nominated at AMIA's 2017 Joint Summits on Translational Science, out of ¿50 papers.

NSF Graduate Research Fellowship Award *2011*

- Three year award to fund Ph.D. studies. Covers tuition and provides research stipend.

NDSEG Graduate Research Fellowship Award *2011*

- Three year funding award. Declined to accept NSF fellowship.

Honors and Awards for Peer Review

Top 10 Percent Reviewer Award, AISTATS 2023 *2023*

- Recognized as one of top 10 percent of all 2500 expert reviewers at a top international machine learning conference.

Highlighted Reviewer Award, ICLR 2022 *2022*

- Recognized as a top reviewer at ICLR, a top-tier international machine learning conference.

Top 10 Percent Reviewer Award, AISTATS 2022 *2022*

- Recognized as one of top 10 percent of all 2500 expert reviewers at a top-tier international machine learning conference.

Top 10 Percent Reviewer Award, NeurIPS 2020 *2020*

- Recognized as one of top 10 percent of more than 3500 expert reviewers at the top international machine learning conference.

Top 400 Reviewer Award, NeurIPS 2019 *2019*

- Recognized as one of top 400 of more than 3500 expert reviewers at the top international machine learning conference.

Top 200 Reviewer Award, NeurIPS 2018 *2018*

- Recognized as one of top 200 of more than 3500 expert reviewers at the top international machine learning conference.

Current Funding Support

C2D2AI: Pilot Investigation for Neuroimaging Biomarkers

Alzheimer's Drug Discovery Foundation (ADDF) *9/1/23 - 1/31/25*

- Full title: Covert Cerebrovascular Disease Detected by Artificial Intelligence (C2D2AI): Pilot Investigation for Pragmatic Neuroimaging Biomarkers for Future Stroke and Dementia Risk
- Co-Investigators: David Kent (PI, Tufts Medical) and Wansu Chen (Kaiser Permanente Southern Cal.)
- Total Amount: $599,788

Machine Learning Models for Human Performance Prediction

U.S. Army NSRDEC, Natick, MA (via Tufts CABCS) *9/1/20 - present*

- Full title: Statistical and Machine Learning Models for Data Reduction and Human Performance Prediction
- Co-Investigators: Eric Miller (co-PI, Tufts ECE) and Shuchin Aeron (co-PI, Tufts ECE)
- Total Amount: roughly $496,098 per year
- 1 year contract renewed each year for 4 years in a row
- Part of larger cross-institution MASTR-E project funded by US Army

Pending Grant Applications

Focused Imaging as a New Diagnostic strategy for Aortic Stenosis

NIH R01, 5 years of funding *Proposed 04/24-03/29*

- Team: Ben Wessler (co-PI, Tufts Medical) and David Kent (co-I, Tufts Medical)
- MCH leads ML method development and deployment across Aim 1 and 2
- Submitted 06/23. Notification expected in Nov.

Decision-Aware Learning of Adaptive Probabilistic Models

NSF CAREER, 5 years of funding *Proposed 02/24 - 01/29*

- Title: Decision-Aware Learning of Adaptive Probabilistic Models from Limited Supervision
- Submitted 07/23. Notification expected in Jan.

Covert Cerbrovascular Disease detected by AI

NIH R01, 5 years of funding *Proposed 09/23 - 08/28*

- Subtitle: Pragmatic neuroimaging biomarkers for future stroke and dementia risk
- Team: David Kent (co-PI, Tufts Medical) and Wansu Chen (co-I, Kaiser Permanente)
- MCH leads ML method development and deployment across Aim 1 and 2
- Submitted 02/23. Awaiting final decision as of late Oct. 2023
- Initial scores indicate top 12% of all submissions

Past Funding Support

Autonomous Cognitive Technologies for Novelty in Open Worlds

DARPA SAIL-ON Program *11/15/19 - 5/15/23*

- SAIL-ON Program: Science of Artificial Intelligence and Learning for Open-world Novelty
- Team at Tufts: Matthias Scheutz (PI, CS), Liping Liu (CS), Jivko Sinapov (CS)
- Team at Arizone State: Chitta Baral (CSE), Subbarao Kambhampati (CSE)

Amortized Inference for Large-Scale Graphical Models
NSF CISE: Robust Intelligence: Small *9/1/19 - 8/31/22*
- Co-Investigators: Liping Liu (PI, Tufts CS) and Thomas Stopka (Tufts Public Health)
- Total Amount: $399,923

A Benchmark De-identified Echocardiogram Database
Pilot Grant from Tufts CTSI *5/1/21 - 4/30/22*
- Full title: A Benchmark De-identified Echocardiogram Database for Studying Automated Diagnoses
- Co-Investigators: Benjamin Wessler (PI)
- Tufts CTSI = Clinical and Translational Science Institute

Estimating the societal value of COVID-19 therapeutics
Tufts Medical Center (Originating Sponsor: Pfizer) *1/1/21 - 12/31/21*
- Co-Investigators: Peter Neumann (Tufts CEVR) and Joshua Cohen (Tufts CEVR)

The value of predictive analytics during the COVID epidemic
Tufts Springboard Award (Tufts Univ. Provost's Office) *7/1/20 - 6/30/21*
- Full title: Demonstrating the value of a proposed Tufts-led predictive analytics and comparative effectiveness research network during the COVID epidemic
- Co-Investigators: David Kent (Tufts Medical) and Jessica Paulus (Tufts Medical)
- Total Amount: $50,000

Estimating Individual Treatment Effects
Tufts Collaborates Award (Internal) *7/1/19 - 6/30/20*
- Title: 'Estimating Individual Treatment Effects from Randomized Clinical Trials using Machine Learning'
- Co-Investigators: David Kent (Tufts Medical Center)
- Total Amount: $50,251

Invited Talks

Invited Talk at Takeda *10/2023*
- Title: Overcoming the Limited Availability of Labeled Data for Medical Image Classification

Invited Talk at PACE Research Group meeting, Tufts Medical Center" *07/2023*
- Title: Towards deployable automatic screening of aortic stenosis from echocardiograms

Invited Talk, WPI ECE Dept. Online Graduate Seminar *11/2022*
- Title: Diagnosing Heart Disease and Preventing Fatal Overdoses via Probabilistic Machine Learning

Invited Talk, UMass Dartmouth CS Dept. Graduate Seminar *11/2022*
- Title: Diagnosing Heart Disease and Preventing Fatal Overdoses via Probabilistic Machine Learning

Invited Talk, Apple Inc. *07/2022*
- Title: Challenges in Time Series - False Alarm Control and Gradual State Transitions

Invited Talk at "I Can't Believe it's not Better" workshop at NeurIPS 2020 *12/2020*

- Title: I Can't Believe Supervision for Latent Variable Models is not Better: The Case for Prediction constrained training
- Event: I Can't Believe It's Not Better! Workshop (ICBINB at NeurIPS 2020)
- Workshop summary: Bridging the gap between theory and empiricism in probabilistic machine learning
- Talk summary: Makes case for our recent work on prediction constrained training, from AISTATS 2018, AISTATS 2020, and in preparation work

Invited Talk at Northwell Health ML group *07/2020*

- Title: Optimizing Machine Learning Models for Interpretable, Actionable Predictions on Electronic Health Records
- Event: Regular virtual meeting of a ML research working group at Northwell Health (large healthcare provider in NYC)
- Summarizes recent MLHC 2019, CHIL 2020, AISTATS 2020, and JAMA Netw. Open 2020 papers

Invited Talk at U. Arizona *02/2020*

- Title: Overcoming model misspecification in structured clustering and reinforcement learning with prediction constrained training
- Event: Regular meeting of a research working group at U. Arizona funded by NSF TRIPODS award
- Summarizes our recent AISTATS 2020 paper

Invited Short Talk at Duke Clinical Research Institute Think Tank meeting *01/2020*

- Talk: Preferred Quality Metrics for Clinical Prediction Models
- Event: Leveraging Artificial Intelligence and Machine Learning Methods and Approaches to Transform Clinical Trial Design, Planning, and Execution
- Host Organization: Duke Clinical Research Institute
- An invitation-only event in Washington D.C. gathering 40 ML experts, clinical experts, and policy makers

Invited Talk at Meeting of Critical Care Directors in Madrid, Spain *01/2020*

- Talk Title: Optimizing Machine Learning Models for Interpretable, Actionable Predictions
- Event: Reunión Sobre Nuevas Tecnologías en el Tratamiento de Datos Clínicos Electrónicos
- Translation: Meeting on New Technologies for Processing Electronic Health Records
- Hosts: RGI Informatics (Dr. Richard Goldstein, CEO) and Fuenlabrada University Hospital (Joaquin Álvarez, head of ICU).
- An invitation-only hosted event in Madrid for 30 directors of intensive care units around Madrid, Spain

Invited Mentor at 2019 PLA General Hospital - MIT Critical Data Datathon *11/2019*

- 4th annual PLAGH-MIT Datathon
- Event held in Beijing, China with 25 teams of local clinicians and computational scientists
- Team goal: Answer clinical question on historical public dataset (MIMIC) over 1 weekend
- Event goal: Develop local teams' skills via intense practice with expert oversight
- My role: Advise teams toward principled and clinically-useful analysis

Invited Talk at BNP 2019 *06/2019*

- Title: Scalable and Reliable Variational Inference for Dirichlet Process Clustering with Sparse Assignments
- Venue: 12th International Conference on Bayesian Nonparametrics
- Summarizes the effective learning methods behind our BNPy toolbox

Invited Tutorial at MLHC 2018 *08/2018*

- Machine Learning for Clinicians: Advances for Multi-Modal Health Data at MLHC '18
- Designed to help clinicians understand enough modern machine learning to collaborate successfully with ML researchers.

Invited Panelist *12/2016*

- Software panel at Advances in Approximate Bayesian Inference workshop at NIPS '16.

Professional Service

Senior Area Chair

- 2024 - CHIL
- 2023 - CHIL

Area Chair

- 2023 - MLHC and ML4H
- 2022 - CHIL and MLHC
- 2021 - CHIL and MLHC

Senior Program Committee / Meta-Reviewer

- 2021 - AAAI
- 2020 - AAAI

Program Committee / Reviewer

- 2023 - NeurIPS, NeurIPS Datasets&Benchmarks, AISTATS, ICLR
- 2022 - NeurIPS, NeurIPS Datasets&Benchmarks, AISTATS, ICLR
- 2020 - NeurIPS, AISTATS, ICLR, MLHC
- 2019 - NeurIPS (reviewer award), AISTATS, ICLR
- 2018 - NeurIPS (reviewer award), AAAI, AISTATS, ICLR, AMIA CRI
- 2017 - NeurIPS, ICML, AAAI
- 2016 - NeurIPS
- 2015 - NeurIPS, ICML
- 2014 - NeurIPS, ICML
- 2013 - NeurIPS (reviewer award)

Workshop Organizer: ML4H at NeurIPS 2018 *2018*

- Machine Learning for Health workshop at NeurIPS '18 (NeurIPS ML4H 2018).
- Full-day workshop with invited keynotes, accepted papers/posters, and lively panel discussions.
- Provided a forum for interdisciplinary interaction between clinicians, statisticians, and computer scientists.
- Helped with website, PR, and continuity in peer-review process from previous years.

Workshop Organizer: BNP at NeurIPS 2018 *2018*

- All of Bayesian Nonparametrics workshop at NeurIPS '18 (NeurIPS BNP 2018).
- Full-day workshop with invited keynotes, accepted papers/posters, and lively panel discussions.
- Helped with peer-review process for accepted posters, https://sites.google.com/view/nipsbnp2018/schedule.

Workshop Organizer: ML4H at NeurIPS 2017 *2017*

- Machine Learning for Health workshop at NIPS '17 (NIPS ML4H 2017)
- Full-day workshop with invited keynotes and panels involving clinicians, statisticians, and computer scientists.
- Organized peer-review process for 118 submitted papers.

Workshop Organizer: BNP at NeurIPS 2016 *2016*

- Practical Bayesian Nonparametrics workshop at NIPS '16.
- Full-day workshop with invited speakers, contributed talks, two panel discussions, and lively poster session.
- Led decisions on ¿25 submitted papers based on peer review.

Mentorship

Postdoctoral Mentees

Tufts University

- Michael T. Wojnowicz, from 2020 - 2023
- – Tufts DISC Data Scientist, primary adviser M. C. Hughes
- – Papers: ICML '22
- – Next position: Research associate at Harvard Univ., Dept. of Biostatistics, advised by Jeff Miller

Doctoral Mentees

Tufts University

- Zhe Huang, CS, from 2019 - present (spring 2024 expected)
- – Topic: Overcoming Limited Labeled Data in the Wild
- – Awards: Outstanding Academic Scholarship Award from Tufts School of Engineering (one per year)
- – Papers: AISTATS '23 (oral), MLHC '23, MLHC '21, NeurIPS D&B '21
- Preetish Rath, CS, from 2019 - present (fall 2024 expected)
- – Topic: Addressing False Alarms and Missingness in Clinical Prediction Models
- – Papers: AISTATS '22, IMLH workshop '23
- Patrick Feeney, CS, from 2020 - spring 2025 (expected)
- – Topic: Self-supervised and Supervised Contrastive Learning in Open Worlds
- – Papers: TMLR '23, arXiv '23
- Kyle Heuton, CS, from 2020 - fall 2025 (expected)
- – Topic: Training spatiotemporal models to learn where to intervene
- – Papers: IMLH workshop '23
- Ethan Harvey, CS, from 2023 - fall 2026 (expected)
- – Topic: Time-to-event prediction of dementia from neuroimages

Co-advised Doctoral Research Mentees

Tufts University

- Kevin Cheng, ECE, from 2019 - fall 2022
- – Primary adviser: Eric Miller (Tufts ECE)
- – Topic: Optimal Transport methods for Time Series modeling
- – Papers: NeurIPS '21, IEEE Trans Signal Proc. '23, ICASSP '20
- – Next position: Takeda, Principal AI/ML Research Scientist

Masters Program Research Mentees

Tufts University

- Yu Liu, MS DS with thesis, from 2020 - 2021
- – Topic: An Evaluation Pipeline for Heterogeneous Treatment Effect Prediction
- – Next Position: Merck
- Xi Chen, MS CS with project, from 2020 - 2021
- – Topic: Bayesian Nonparametric Mixture Models for Missing Data
- – Next Position: CS PhD student at Rutgers

Post-bacc Research Mentees

Tufts University

- Ally Lee, BS CS from Tufts, in 2020
- – Topic: Bayesian Analysis of Autoregressive Models for Multi-Site Hospital Admission Forecasting
- – Next Position: Software Engineer at Hubspot, Boston, MA
- Lily H. Zhang, BS from Harvard, from 2019 - 2020
- – Topic: 'Any Parameter Encoders for Topic Models: Variational Encoders that amortize across models as well as data'
- – Next Position: PhD candidate in CS at NYU, advised by Rajesh Ranganath

Undergraduate Research Mentees

Tufts University

- Mary-Joy Sidhom, BS CS with honors thesis, in 2022
- – Topic: Deep Learning for Doppler Echocardiography from Limited Labeled Data
- – Papers: AISTATS '23 (oral)
- – Next Position: Software Engineer II at ASML
- Hezekiah Branch, BS CS at Tufts, in 2020
- – Tufts LSAMP fellow
- – Topic: Supervised Learning for Clinical Multivariate Time Series

Harvard University SEAS

Research Mentor *2016-2017*

- Mentored undergraduate senior thesis projects on Bayesian nonparametric inference.
- Frederick Widjaja. 2017 honors thesis: Streaming Variational Inference for the Indian Buffet Process.
- Madhu Vijay. 2017 honors thesis: Characterizing Posterior Uncertainty for the Indian Buffet Process.

Brown University

Research Mentor *2014-2016*

- Mentored students on projects related to Bayesian nonparametric clustering and the BNPy Python package.
- William Stephenson. 2015 undergraduate honors thesis: Variational Inference for Hierarchical Dirichlet Process based Nonparametric Models.
- Sonia Phene. 2015 undergraduate honors thesis: Multiprocessor Parallelization of Variational Inference for Bayesian Nonparametric Topic Models.
- Mengrui Ni. 2015 masters project: Variational Inference for Beta-Bernoulli Dirichlet Process Mixture Models.
- Mert Terzihan. 2015 masters project.

Teaching

Tufts CS Dept.
Course: CS 135 Intro to Machine Learning *Fall 2023*
- Taught core principles of machine learning to 101 students
- Course format: 3 open-ended projects, 5 homeworks (conceptual and code questions), 2 exams

Tufts CS Dept.
Course: CS 136 Statistical Pattern Recognition *Spring 2023*
- Taught advanced statistical learning course to 36 students
- Course format: 5 math-intensive homeworks, 5 coding-intensive homeworks, 5 quizzes

Tufts CS Dept.
Course: CS 152 Bayesian Deep Learning *Fall 2022*
- Taught special topics course to 26 students
- Course format: weekly homeworks for first month, then 2-month open-ended team project

Tufts CS Dept.
Course: COMP 136 Statistical Pattern Recognition *Spring 2021*
- Taught advanced statistical learning course to 29 students
- Course format: 5 math-intensive homeworks, 5 coding-intensive homeworks, 5 short quizzes, 2 exams

Tufts CS Dept.
Course: COMP 135 Intro to Machine Learning *Fall 2020*
- Taught core principles of machine learning to 95 students
- Course format: 3 open-ended projects, 5 homeworks (conceptual and code questions), and 5 quizzes

Tufts CS Dept.
Course: COMP 136 Statistical Pattern Recognition *Spring 2020*
- Taught advanced statistical learning course to 35 students
- Course format: 5 math-intensive homeworks, 5 coding-intensive homeworks, 5 short quizzes, 2 exams

Tufts CS Dept.
Course: COMP 150 Bayesian Deep Learning *Fall 2019*
- Taught advanced topics seminar to 23 students
- Course format: weekly homeworks for first month, then 2-month open-ended team project

Tufts CS Dept.
Course: COMP 135 Introduction to Machine Learning *Spring 2019*
- Taught core principles of machine learning to about 50 students
- Course format: 3 open-ended projects, weekly homeworks, and 2 exams

Tufts CS Dept.

Course: COMP 150 Bayesian Deep Learning *Fall 2018*

- Taught advanced topics seminar to about 18 students
- Course format: weekly homeworks for first month, then 2-month open-ended team project
- One project resulted in publication at IEEE conference (ICDL-EpiRob 2019)

Lead Graduate TA for CS 142: Intro to Machine Learning *Fall 2013*

- Led weekly 1 hour recitation session to review key concepts for 50+ students.
- Designed homework assignments and exam questions.

Outreach Experience

Tufts DIAMONDS Program **Medford, MA**

Research Mentor *2021, 2022, 2023*

- Mentored two students / summer in data science research projects.

TEALS and Boston Latin Academy **Roxbury, MA**

Volunteer AP Computer Science Instructor *2014-2016*

- Taught 1-2 classes / week for 2 years via TEALS "CS in every high school" initiative sponsored by Microsoft.
- Developed hands-on lessons to excite students from diverse backgrounds about computational thinking.
- Mentored full-time teacher Ingrid Roche as she transitioned from media arts to AP computer science (Java).

Harvard Humanitarian Initiative **Cambridge, MA**

Signal Program Fellow *2014*

- Developed prototype detector for common housing structures in sub-Saharan Africa from satellite images.
- Intended for humanitarian oversight of conflict areas where burning structures is common attack pattern.
- Featured in TEDx talk: `https://youtu.be/u7l9rBwOnwU`

Olin College Engineering Discovery **Needham, MA**

Co-Founder and Curriculum Director *2007-2010*

- Managed 15 undergrads in developing hands-on lessons for 4th-8th graders.
- Hosted workshops for 30 children to design, build, and launch bottle rockets.
- Pioneered green energy workshop which earned over $750 in outside funding.

Industry Experience

Google **Mountain View, CA**

Software Engineering Intern *Summer 2013*

- Improved walking/biking/running classifier using smartphone accelerometer data.
- Led collection of dataset from dozens of individuals for classifier evaluation via custom Android app.

All Conference Publications (in reverse chronological order)

1. "Detecting Heart Disease from Multi-View Ultrasound Images via Supervised Attention Multiple Instance Learning." Zhe Huang[d], Benjamin S. Wessler, and Michael C. Hughes. Machine Learning for Healthcare Conference (MLHC), 2023.

2. "Approximate inference by broadening the support of the likelihood." Michael T. Wojnowicz, Martin D. Buck, Michael C. Hughes. Symposium on Advances in Approximate Bayesian Inference (AABI), 2023.
3. "Fix-A-Step: Semi-supervised Learning from Uncurated Unlabeled Data." Zhe Huang[d], Mary-Joy Sidhom[u], Benjamin S. Wessler, and Michael C. Hughes. Artificial Intelligence and Statistics (AISTATS) [oral, top 2% of 1500+ submissions], 2023.
4. "Easy Variational Inference for Categorical Models via an Independent Binary Approximation." Michael T. Wojnowicz, Shuchin Aeron, Eric L. Miller, and Michael C. Hughes. International Conference on Machine Learning (ICML), 2022.
5. "Optimizing Early Warning Classifiers to Control False Alarms via a Minimum Precision Constraint." Preetish Rath[d] and Michael C. Hughes. Artificial Intelligence and Statistics (AISTATS), 2022.
6. "The Tufts fNIRS Mental Workload Dataset & Benchmark for Brain-Computer Interfaces that Generalize." Zhe Huang[d], Liang Wang[d], Giles Blaney, Christopher Slaughter[u], Devon McKeon, Ziyu Zhou, Robert Jacob, and Michael C. Hughes. Neural Information Processing Systems (NeurIPS) Track on Datasets and Benchmarks, 2021.
7. "Dynamical Wasserstein Barycenters for Time-series Modeling." Kevin C Cheng[d], Shuchin Aeron, Michael C. Hughes, and Eric Miller. Neural Information Processing Systems (NeurIPS), 2021.
8. "Taming fNIRS-based BCI Input for Beter Calibration and Broader Use." Liang Wang[d], Zhe Huang [d], Ziyu Zhou, Devon McKeon, Giles Blaney, Michael C. Hughes, and Robert J. K. Jacob. ACM Symposium on User Interface Software and Technology (UIST), 2021.
9. "A New Semi-supervised Learning Benchmark for Classifying View and Diagnosing Aortic Stenosis from Echocardiograms." Zhe Huang[d], Gary Long, Benjamin Wessler, and Michael C. Hughes. Machine Learning for Healthcare (MLHC), 2021.
10. "Approximate Bayesian Computation for an Explicit-DurationHidden Markov Model of COVID-19 Hospital Trajectories." Gian Marco Visani[u], Alexandra Hope Lee[b], Cuong Nguyen[m], David M. Kent, John B. Wong, Joshua T. Cohen, and Michael C. Hughes. Machine Learning for Healthcare (MLHC), 2021.
11. "Stochastic Iterative Graph Matching." Linfeng Liu[d], Michael C. Hughes, Soha Hassoun, and Li-Ping Liu. International Conference of Machine Learning (ICML), 2021.
12. "MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III." Shirly Wang[m], Matthew B. A. McDermott, Geeticka Chauhan, Marzyeh Ghassemi, Michael C. Hughes, and Tristan Naumann. The ACM Conference on Health, Inference, and Learning (CHIL), 2020.
13. "POPCORN: Partially Observed Prediction-Constrained Reinforcement Learning." Joseph Futoma, Michael C. Hughes, and Finale Doshi-Velez. AISTATS, 2020.
14. "Optimal Transport Based Change Point Detection and Time Series Clustering." Kevin Cheng[d], Shuchin Aeron, Michael C. Hughes, Erika Hussey, and Eric Miller. IEEE ICASSP 2020, 2020.
15. "Regional Tree Regularization for Interpretability in Deep Neural Networks." Mike Wu[d], Sonali Parbhoo, Michael C. Hughes, Ryan Kindle, Leo Celi, Maurizio Zazzi, Volker Roth, and Finale Doshi-Velez. AAAI, 2020.
16. "Feature Robustness in Non-stationary Health Records: Caveats to Deployable Model Performance in Common Clinical Machine Learning Tasks." Bret Nestor[d], Matthew B. A. McDermott, Willie Boag, Gabriela Berner, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, and Marzyeh Ghassemi. Machine Learning for Healthcare, 2019.

17. "Supervised Machine Learning Algorithms Using Patient Related Factors to Predict in-Hospital Mortality Following Acute Myeloid Leukemia Therapy." Nauman Saleem Siddiqui[c], Andreas Klein, Amandeep Godara, Cindy Varga, Rachel J. Buchsbaum, and Michael C. Hughes. Proceedings of 61st Annual Meeting of the American Hematology Society, 2019.
18. "Sensorimotor Cross-Behavior Knowledge Transfer for Grounded Category Recognition." Gyan Tatiya[d], Ramtin Hosseini[d], Michael C. Hughes, and Jivko Sinapov. Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2019.
19. "Semi-Supervised Prediction-Constrained Topic Models." Michael C. Hughes, Gabriel Hope[d], Leah Weiner[d], Thomas H. McCoy Jr, Roy H. Perlis, Erik B. Sudderth, and Finale Doshi-Velez. Artificial Intelligence and Statistics (AISTATS), 2018.
20. "Beyond Sparsity: Tree Regularization of Deep Models for Interpretability." Mike Wu[u], Michael C. Hughes, Sonali Parbhoo, Maurizio Zazzi, Volker Roth, and Finale Doshi-Velez. Association for Advancement of Artificial Intelligence (AAAI), 2018.
21. "From Patches to Images: A Nonparametric Generative Model." Geng Ji[d], Michael C. Hughes, and Erik B. Sudderth. International Conference on Machine Learning (ICML), 2017.
22. "Right for the Right Reasons: Training Differentiable Models by Constraining their Explanations." Andrew Slavin Ross[m], Michael C. Hughes, and Finale Doshi-Velez. International Joint Conference on Artificial Intelligence (ICJAI), 2017.
23. "Predicting Intervention Onset in the ICU with Switching State Space Models." Marzyeh Ghassemi, Mike Wu[u], Michael C. Hughes, Peter Szolovits, and Finale Doshi-Velez. AMIA Summit on Clinical Research Informatics, 2017.
24. "Scalable Adaptation of State Complexity for Nonparametric Hidden Markov Models." Michael C. Hughes, William Stephenson[u], and Erik B. Sudderth. Neural Information Processing Systems (NIPS), 2015.
25. "Reliable and Scalable Variational Inference for the Hierarchical Dirichlet Process." Michael C. Hughes, Dae Il Kim, and Erik B. Sudderth. Artificial Intelligence & Statistics (AISTATS), 2015.
26. "Memoized Online Variational Inference for Dirichlet Process Mixture Models." Michael C. Hughes and Erik B. Sudderth. Neural Information Processing Systems (NIPS), 2013.
27. "Effective Split-Merge Monte Carlo Methods for Nonparametric Models of Sequential Data." Michael C. Hughes, Emily Fox, and Erik B. Sudderth. Neural Information Processing Systems (NIPS), 2012.
28. "The Nonparametric Metadata Dependent Relational Model." Dae Il Kim, Michael C. Hughes, and Erik B. Sudderth. International Conference on Machine Learning (ICML), 2012.

All Journal Publications (in reverse chronological order)

1. "Non-Parametric and Regularized Dynamical Wasserstein Barycenters for Sequential Observations." Kevin C. Cheng[d], Shuchin Aeron, Michael C. Hughes, Eric L. Miller. IEEE Transactions on Signal Processing, 2023.
2. "NovelCraft: A Dataset for Novelty Detection and Discovery in Open Worlds." Patrick Feeney[d], Sarah Schneider[d], Panagiotis Lymperopoulos, Liping Liu, Matthias Scheutz, and Michael C. Hughes. Transactions on Machine Learning Research (TMLR), 2023.
3. "Automated Detection of Aortic Stenosis using Machine Learning." Benjamin S. Wessler, Zhe Huang[d], Gary Long, ... and Michael C. Hughes. Journal of the American Society of Echocardiography, 2023.

4. "The role of machine learning in clinical research: transforming the future of evidence generation." E. Hope Weissler, Tristan Naumann, Tomas Andersson, Rajesh Ranganath, Olivier Elemento, Yuan Luo, Daniel F. Freitag, James Benoit, Michael C. Hughes, Faisal Khan, Paul Slater, Khader Shameer, Matthew Roe, Emmette Hutchison, Scott H. Kollins, Uli Broedl, Zhaoling Meng, Jennifer L. Wong, Lesley Curtis, Erich Huang and Marzyeh Ghassemi. Trials, 2021.
5. "Optimizing for Interpretability in Deep Neural Networks with Simulable Decision Trees." Mike Wu[d], Sonali Parbhoo, Michael C. Hughes, Volker Roth, and Finale Doshi-Velez. Journal of Artificial Intelligence Research (JAIR), 2021.
6. "Enzyme Promiscuity Prediction Using Hierarchy-Informed Multi-Label Classification." Gian Marco Visani[u], Michael C. Hughes, and Soha Hassoun. Bioinformatics, 2021.
7. "On Matched Filtering for Statistical Change Point Detection." Kevin Cheng[d], Eric L Miller, Michael C Hughes, and Shuchin Aeron. IEEE Open Journal of Signal Processing, 2020.
8. "A Framework for Sensorimotor Cross-Perception and Cross-Behavior Knowledge Transfer for Object Categorization." Gyan Tatiya[d], Ramtin Hosseini[d], Michael C. Hughes, and Jivko Sinapov. Frontiers in Robotics and AI, 2020.
9. "Assessment of a Prediction Model for Antidepressant Treatment Stability Using Supervised Topic Models." Michael C. Hughes, Melanie F. Pradier, Andrew Slavin Ross, Thomas H. McCoy Jr, Roy H. Perlis, Finale Doshi-Velez. JAMA Network Open, 2020.
10. "Predicting change in diagnosis from major depression to bipolar disorder after antidepressant initiation." Melanie F. Pradier, Michael C. Hughes, Thomas H. McCoy Jr, Sergio A. Barroilhet, Finale Doshi-Velez, and Roy H. Perlis. Neuropsychopharmacology, 2020.
11. "Predicting Treatment Discontinuation after Antidepressant Initiation." Melanie F. Pradier, Thomas H. McCoy, Michael C. Hughes, Roy H. Perlis, and Finale Doshi-Velez. Translational Psychiatry, 2020.
12. "Refinery: An Open Source Topic Modeling Web Platform." Daeil Kim, Benjamin F. Swanson, Michael C. Hughes, and Erik B. Sudderth. JMLR Machine Learning Open Source Software (MLOSS), 2017.
13. "Joint Modeling of Multiple Time Series via the Beta Process with Application to Motion Capture Segmentation." Emily Fox, Michael C. Hughes, Erik B. Sudderth, and Michael I. Jordan. Annals of Applied Statistics, Vol. 8(3), 2014.

All Workshop Papers (in reverse chronological order)

1. "Learning where to intervene with a differentiable top-k operator: Towards data-driven strategies to prevent fatal opioid overdoses." Kyle Heuton[d], Shikhar Shrestha, Thomas J. Stopka, Michael C Hughes. Workshop on Interpretable Machine Learning in Healthcare (IMLH), 2023.
2. "Semi-supervised Ordinal Regression via Cumulative Link Models for Predicting In-Hospital Length-of-Stay." Alexander Arjun Lobo[m], Preetish Rath[d], Michael C Hughes. Workshop on Interpretable Machine Learning in Healthcare (IMLH), 2023.
3. "Predicting Spatiotemporal Counts of Opioid-related Fatal Overdoses via Zero-Inflated Gaussian Processes." Kyle Heuton[d], Shikhar Shrestha, Thomas J. Stopka, Jennifer Pustz, Li-Ping Liu, and Michael C. Hughes. NeurIPS Workshop on Gaussian Processes, Spatiotemporal Modeling, and Decision-making Systems, 2022.
4. "Semi-supervised Learning from Uncurated Echocardiogram Images with Fix-A-Step." Zhe Huang[d],

Mary-Joy Sidhom[u], Benjamin S. Wessler, Michael C. Hughes. Medical Imaging Meets Neurips Workshop, 2022, 2022.

5. "Prediction-Constrained Markov Models for Medical Time Series with Missing Data and Few Labels." Preetish Rath[d], Gabriel Hope[d], Kyle Heuton[d], Erik B. Sudderth, and Michael C. Hughes. Learning from Time Series For Health (TS4H) Workshop at NeurIPS, 2022.
6. "TMED 2: A Dataset for Semi-Supervised Classification of Echocardiograms." Zhe Huang[d], Gary Long, Benjamin S. Wessler, and Michael C. Hughes. DataPerf: Benchmarking Data for Data-Centric AI, a workshop at ICML 2022 , 2022.
7. "Learning Consistent Deep Generative Models from Sparsely Labeled Data." Gabriel Hope[d], Madina Abdrakhmanova, Xiaoyin Chen, Michael C. Hughes, and Erik B. Sudderth. Advances in Approximate Bayesian Inference (AABI), 2022.
8. "Easy Variational Inference for Categorical Observations via a New View of Diagonal Orthant Probit Models." Michael T. Wojnowicz, Shuchin Aeron, Eric L. Miller, and Michael C. Hughes. Tractable Probabilistic Modeling workshop at UAI, 2021.
9. "Optimizing Clinical Early Warning Systems to Meet False Alarm Constraints." Preetish Rath and Michael C. Hughes. Interpretable Machine Learning for Healthcare (IMLH) workshop at ICML 2021, 2021.
10. "Prediction-Constrained Hidden Markov Models for Semi-Supervised Classification." Gabriel Hope[d], Michael C. Hughes, Finale Doshi-Velez, and Erik B. Sudderth. Time Series Workshop at ICML, 2021.
11. "Evaluating the Use of Reconstruction Error for Novelty Localization." Patrick Feeney[d] and Michael C. Hughes. Uncertainty and Robustness in Deep Learning (UDL) workshop at ICML 2021, 2021.
12. "Forecasting COVID-19 Counts At A Single Hospital: A Hierarchical Bayesian Approach." Alexandra Hope Lee[b], Panagiotis Lymperopoulos[m], Joshua T. Cohen, John B. Wong, and Michael C. Hughes. ICLR Workshop on Machine Learning for Preventing and Combating Pandemics, 2021.
13. "Using Hierarchy-Informed Multi-Label Classification for Enzyme Promiscuity Prediction." Gian Marco Visani[u], Michael C. Hughes, and Soha Hassoun. Machine Learning in Computational Biology Workshop (MLCB), 2020.
14. "Rapid Model Comparison by Amortizing Across Models." Lily H. Zhang[b], and Michael C. Hughes. Second Symposium on Advances in Approximate Bayesian Inference (AABI 2019), 2019.
15. "Classification of Enzyme Promiscuity Using Positive, Unlabeled, and Hard Negative Examples." Gian Marco Visani, Michael C. Hughes and Soha Hassoun. Machine Learning in Computational Biology Workshop (MLCB), 2019.
16. "Prediction-Constrained POMDPs." Joseph Futoma, Michael C. Hughes, and Finale Doshi-Velez. Reinforcement Learning under Partial Observability (RLPO) workshop at NeurIPS 2018, 2018.
17. "Rethinking clinical prediction: Why machine learning must consider year of care and feature aggregation." Bret Nestor[d], Matthew B. A. McDermott, Geeticka Chauhan, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, Marzyeh Ghassemi. Machine Learning for Healthcare (ML4H) workshop at NeurIPS 2018, 2018.
18. "Prediction-Constrained Topic Models for Antidepressant Prediction." Michael C. Hughes, Gabriel Hope[d], Leah Weiner[d], Thomas H. McCoy, Roy H. Perlis, Erik B. Sudderth, and Finale Doshi-Velez. NIPS Workshop on Machine Learning for Health (NIPS ML4H), 2017.
19. "Associations between aboveground forest biomass and waveform LiDAR metrics: implications for modeling footprint-level biomass using Global Ecosystem Dynamics Investigation data." J. Kellner,

J. B. Blair, L. Duncanson, L., S. Hancock, M. A. Hofton, M. C. Hughes, S. Marselis, S., J. Armston, E. B. Sudderth, H. Tang, L. Weiner[d], and R. Dubayah. American Geophysical Union, Fall General Assembly, 2016.

20. "Supervised topic models for clinical interpretability." Michael C. Hughes, Huseyin Melih Elibol, Thomas McCoy, Roy Perlis, and Finale Doshi-Velez. NIPS Workshop on Machine Learning for Health (NIPS ML4H), 2016.
21. "BNPy: Reliable and scalable variational inference for Bayesian nonparametric models." Michael C. Hughes and Erik B. Sudderth. 3rd NIPS Workshop on Probabilistic Programming, 2014.
22. "Nonparametric Discovery of Activity Patterns from Video Collections." Michael C. Hughes and Erik B. Sudderth. CVPR Workshop on Perceptual Organization in Computer Vision (POCV), 2012.