Divya Shanmugam

divyas@mit.edu, dmshanmugam.github.io

Education

Massachusetts Institute of Technology Expected 05/2024
Ph.D, Electrical Engineering and Computer Science

Massachusetts Institute of Technology 05/2018
Master of Engineering, Electrical Engineering and Computer Science
Thesis title: *Representation Learning for Improved Distance and Risk Metrics*

Massachusetts Institute of Technology 05/2017
B.S., Electrical Engineering and Computer Science

Experience

Research Intern, Machine Learning & Statistics Group, MSR New England 06/2020 - 09/2020
Modeling the Prevalence of Intimate Partner Violence via EHR Data: Estimating the relative prevalence of IPV across subgroups using a Bayesian approach. We produce a method that can recover the relative prevalence *exactly* under the assumption of covariate shift. (*under review*)

Research Intern, Fairness, Accountability, and Transparency Group, MSR Montreal 06/2020 - 09/2020
Learning to Limit Data Collection: Operationalizing the GDPR principle of data minimization—the responsibility to collect data ethically—in the context of machine learning. We present a method to guide data collection based on a desired level of model performance. (*under review*)

Ph.D. Candidate, Clinical and Applied Machine Learning Group 09/2018 - present
When & Why Test-Time Augmentation Works: Data augmentation is commonly used to increase performance for image classification networks. Why? We present the conditions conducive to test-time augmentation and provide an improved method. Accepted to ICCV 2021 (oral presentation, top 3% of submissions).

Multiple Instance Learning for ECG Risk Stratification: Existing risk metrics for cardiovascular death rely on hand-crafted features. We use multiple instance learning to identify features predictive of cardiovascular risk directly from an ECG signal. Accepted to MLHC 2019 (spotlight presentation).

Research Intern, Borealis AI 06/2018 - 09/2018
Learning on Noisy Data: We developed a model to correct label noise in training data using an energy-based autoencoder.

Research Assistant, Clinical and Applied Machine Learning Group 07/2017 - 06/2018

Metric Learning for Time Series: Applied machine learning towards improved distance and risk metrics for time series. Presented at Women in Machine Learning Workshop 2017, Machine Learning for Health Workshop 2017.

Dialect-based Disparities in NLP: Researched language diversity across economic class towards the development of dialect-agnostic representations for text. Presented at Women in Machine Learning Workshop 2017.

Research Intern, D.E. Shaw Research 06/2016 - 09/2016

Accelerating Graphical Rendering: Restructured molecular dynamics graphics algorithm to enable 5x parallelism using quaternion representations.

Research Assistant, Computation and Biology Group 06/2016 - 09/2016

Fast Metagenomic Sequencing: Worked on methods to expedite metagenomic analysis by exploiting redundancy.

Publications

* denotes equal contribution.

1. **D. Shanmugam**, E. Pierson "Quantifying Inequality in Underreported Medical Conditions". *(under review)*
2. **D. Shanmugam**, S. Shabanian, F. Diaz, M. Finck, A. Biega "Learning to Limit Data via Scaling Laws: Data Minimization Compliance in Practice". *(under review)*
3. **D. Shanmugam**, D. Blalock, G. Balakrishnan, J. Guttag, "Better Aggregation in Test-Time Augmentation". *ICCV 2021* (oral, top 3%)
4. **D. Shanmugam**, D. Blalock, J. Guttag, "Multiple Instance Learning for ECG Risk Stratification". *MLHC-19* (oral)
5. J. Sahota*, **D. Shanmugam***, J. Ramanan, S. Eghbali, M. Brubaker, "Addressing Feature-Dependent Label Noise: A Generative Framework" *(preprint)*
6. **D. Shanmugam**, D. Blalock, J. Guttag, "Jiffy: A Convolutional Approach to Multivariate Time Series Classification". *(Master's thesis)*

Workshops and poster sessions

1. *Learning to Limit Data Collection for Data Minimization Compliance*
   Women in Machine Learning Workshop, NeuRIPS 2020
2. *Unsupervised Domain Adaptation in the Absence of Source Data*
   Uncertainty & Robustness in Deep Learning Workshop, ICML 2020
3. *Image Segmentation of Liver Stage Malaria Infection with Spatial Uncertainty Sampling*
   Workshop on Computational Biology, ICML 2019

4. *Multiple Instance Learning for Cardiac Risk Stratification.*
   Women in Machine Learning Workshop, NeurIPS 2018 (oral presentation)
5. *Multiple Instance Learning for ECG Risk Stratification.*
   Machine Learning for Health Workshop, NeurIPS 2018
6. *ECG Risk Stratification Using Multiple Instance Learning.*
   MIT DSAIL 2018
7. *Jiffy: A Convolutional Approach to Learning Time Series Similarity.*
   MIT MasterWorks 2018
8. *A Convolutional Approach to Learning Time Series Similarity.*
   Women in Machine Learning Workshop, NeurIPS 2017
9. *Identifying and Accounting for Disparities in Language Due to Economic Class.*
   Women in Machine Learning Workshop, NeurIPS 2017
10. *Compressive Metagenomics*
    MIT Microbiome Center Symposium 2016

Invited talks

1. *Quantifying Inequality in Underreported Conditions.* University of Chicago Crime and Education Lab, Virtual, December 2021 (upcoming)
2. *Quantifying Inequality in Underreported Conditions.* Cornell Information Sciences Seminar, Virtual, November 2021
3. *Estimating the Relative Prevalence of Underreported Medical Conditions.* Microsoft Research New England, Virtual, July 2021
4. *Learning to Limit Data Collection using Scaling Laws.* Microsoft Research Montreal, Virtual, August 2020
5. *Machine Learning, Data Collection, and Women's Health.* Texas Christian University, Virtual, June 2020
6. *Multiple Instance Learning for ECG Risk Stratification.* University of Michigan, Ann Arbor, Michigan, August 2019

Professional Service

Teaching

Teaching assistant: Introduction to Machine Learning, MIT — SPR 2018
Teaching assistant: Introduction to Machine Learning, MIT — FAL 2017

Mentorship

Anna Bryan, UROP	2021-now
Tiffany Chen, UROP	2021-now
Helen Lu, UROP	2021-now
Angela Zhang, UROP	2021-now
Neha Hulkund, UROP	2020-2021
Roshni Sahoo, SuperUROP	2018-2020
Skylar Gordon, AI Mentee	2018-2019
Xinyi Guo, AI Mentee	2018-2019

Reviewing

International Conference on Machine Learning	2021
Computer Vision and Pattern Recognition Conference	2021
Neural and Information Processing Systems	2020
Conference on Health, Inference, and Learning	2020
Machine Learning for Healthcare Conference	2020, 2021
Machine Learning for Health NeurIPS workshop	2019, 2020
Women in Machine Learning NeurIPS workshop	2018, 2020

Panels

Career Mentorship Panel (MIT Undergraduate Research Technology Conference)	2021
Graduate Student Panel (McCormick Hall)	2020
Graduate Student Panel (MIT Women in EECS)	2019
Lightning Talks (MIT Women in EECS)	2017

Service

Undergraduate Mentorship Improvement Initiative	2019-2020
GW6 Event Coordinator	2018-2019
MIT AI Mentorship Program Coordinator	2018-2020

Awards

NSF Graduate Research Fellowship 2017

Languages

Programming: Python (tensorflow, keras, pytorch), Go, C
Spoken: English (Proficient), Spanish (Intermediate), Tamil (Beginner)