Divya Shanmugam

divyas@cornell.edu • dmshanmugam.github.io

Research

I develop methods to overcome the limits of imperfect data and models, motivated by challenges in healthcare.

Academic positions and degrees

Cornell Tech, New York, NY
Postdoctoral Researcher, Department of Computer Science — 2024 – Present

Massachusetts Institute of Technology, Cambridge, MA
Ph.D., Electrical Engineering and Computer Science (05/2024)
Thesis: *Advancing Equity and Reliability in Machine Learning.* — 2018 – 2024

Massachusetts Institute of Technology, Cambridge, MA
M.Eng., Electrical Engineering and Computer Science (05/2018)
Thesis: *Representation Learning for Improved Distance and Risk Metrics.* — 2017 – 2018

Massachusetts Institute of Technology, Cambridge, MA
B.S., Electrical Engineering and Computer Science (05/2017) — 2013 – 2017

Honors

Rising Star in EECS (MIT)	2025
Best Paper Award, Conference on Health, Inference, and Learning (CHIL)	2025
Honorable Mention Best Findings Paper, Machine Learning for Health (ML4H)	2023
Jane Street Fellowship (Honorable Mention)	2022
Best Talk, GW6 Research Summit	2021
National Science Foundation Graduate Research Fellowship	2017

Publications

** denotes equal contribution.*

Improving Spontaneous Labor Prediction with Electronic Health Record Data — 2025
M. Krishnamoorthy*, D. Shanmugam*, D. Tjandra, A. Peahl, C. Pancaro, E. Ziedan, A. Kowalski, J. Wiens.
In progress.

The Trillion Dollar Algorithm: Lessons from Machine Learning for Medicare Advantage Risk Adjustment — 2025
D. Shanmugam*, M. Johnson*, D. Meyers, J. Wiens, E. Pierson.
Under review at NEJM AI.

Machine learning reveals hidden diagnoses among underserved patients — 2025
D. Shanmugam, B. Hardy, A. Wang, S. Divikaran, E. Pierson**, M. Barnett**.
Under review.

Identifying Mechanisms of Disparities within Cascades of Cardiovascular Care after an Emergency Department Visit 2025
A. Wang, D. Shanmugam, S. Divikaran, E. Pierson, M. Barnett.
Under review.

Survival Analysis with Limited Overlap and Censoring Distribution Shift 2025
M. Krishnamoorthy, D. Shanmugam, D. Tjandra, A. E. Kowalski, J. Wiens.
Under review.

Evaluating multiple models using labeled and unlabeled data 2025
D. Shanmugam*, S. Sadhuka*, M. Raghavan, J. Guttag, B. Berger**, E. Pierson**.
Neural Information Processing Systems (NeurIPS) 2025.

Learning Disease Progression Models That Capture Health Disparities 2025
E. Chiang, D. Shanmugam, A. Beecy, G. Sayer, N. Uriel, D. Estrin, N. Garg, E. Pierson.
Conference on Health, Inference, and Learning (CHIL) 2025.

Test-time augmentation improves efficiency in conformal prediction 2025
D. Shanmugam, H. Lu, S. Swaminarayan, J. Guttag.
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2025.

Generative Artificial Intelligence in Medicine 2025
D. Shanmugam, M. Agrawal, R. Movva, I. Y. Chen, M. Ghassemi, M. Jacobs, E. Pierson.
Annual Review of Biomedical Data Science, 2025.

Using large language models to promote health equity 2025
E. Pierson*, D. Shanmugam*, R. Movva*, J. Kleinberg*, et al.
New England Journal of Medicine AI, 2025.

Quantifying disparities in intimate partner violence: a machine learning method to correct for underreporting 2024
D. Shanmugam, K. Hou, E. Pierson.
npj Women's Health, 2024.

Longitudinal Changes in Desire and Attraction Among Women Who Have Discontinued Hormonal Contraceptives 2024
J. Gassen, S. Mengelkoch, D. Shanmugam, J. Pierson, A. van Lamsweerde, E. Benhar, S. E. Hill.
Hormones and Behavior, 2024.

Coarse race data conceals disparities in clinical risk score performance 2023
R. Movva*, D. Shanmugam*, K. Hou, P. Pathak, J. Guttag, N. Garg, E. Pierson.
Machine Learning for Healthcare Conference (MLHC) 2023.

The Relationship Between Photoperiod and Ovulation Rate: A Multi-Site Study Using NaturalCycles Data 2023
D. Shanmugam, M. Espinosa, J. Gassen, A. van Lamsweerde, J. Pearson, E. Behar, S. E. Hill.
Scientific Reports, 2023.

Kaleidoscope: Semantically-grounded, Context-specific Model Evaluation 2023
H. Suresh, D. Shanmugam, A. Bryan, T. Chen, A. D'Amour, J. Guttag, A. Satyanarayan.
ACM CHI Conference on Human Factors in Computing Systems (CHI) 2023.

Data Augmentation for Electrocardiograms 2022
A. Raghu, D. Shanmugam, E. Pomerantsev, J. Guttag, C. Stultz.
ACM Conference on Health, Inference, and Learning (CHIL) 2022.

Learning to Limit Data via Scaling Laws: A Computational Interpretation for the Legal Principle of Data Minimization 2022
D. Shanmugam, S. Shabanian, F. Diaz, M. Finck, A. Biega.
ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2022.

Better Aggregation in Test-Time Augmentation 2021
D. Shanmugam, D. Blalock, G. Balakrishnan, J. Guttag.
IEEE/CVF International Conference on Computer Vision (ICCV) 2021. (Oral, top 3%).

Multiple Instance Learning for ECG Risk Stratification 2019
D. Shanmugam, D. Blalock, J. Guttag.
Machine Learning for Healthcare Conference (MLHC) 2019. (Oral).

Selected talks

Reliable machine learning with health data

Applied Data Science, Cornell Tech, New York, NY 2025
AI & Public Policy Seminar, Cornell Tech, New York, NY 2025
Wadhwani School of Data Science & AI Seminar, IIT Madras, Virtual 2025

Methods to Evaluate the Quality of Clinical Care

INFORMS, Atlanta, GA 2025
International Conference on Statistics and Data Science (ICSDS), Vancouver, CA 2025
Center for Population Health (CPH), UC Berkeley, Berkeley, CA 2025
UVA School of Data Science, Charlottesville, VA 2025
Healthy ML Group, MIT, Cambridge, MA 2025
Diversity in Health Data: Achieving Benefit for All, Virtual 2025

The Medicare Advantage Algorithm

AI & Public Policy Seminar, Cornell University, Ithaca, NY 2025

Advancing Equity & Reliability in Machine Learning

Statistical Methods for Health Equity, Data Science for Health Equity, Virtual 2024

All the Data We Cannot See

AI and Medicine Workshop, Mt. Sinai, New York, NY 2024

Computational Healthcare for Equity and iNclusion Lab, UC Berkeley, Berkeley, CA 2023

Quantifying Inequality in Underreported Conditions

MIT Algorithmic Fairness Reading Group, Cambridge, MA 2023
MIT AI Ethics Seminar, Cambridge, MA 2022
University of Chicago Crime and Education Lab, Virtual 2021
Cornell Information Science Seminar, Virtual 2021
Microsoft Research New England, Virtual 2021

At the Intersection of Conceptual Art and Deep Learning: The End of Signature

List Center, Cambridge, MA 2022

Learning to Limit Data Collection using Scaling Laws

Microsoft Research Montreal, Virtual 2020

Machine Learning, Data Collection, and Women's Health

Texas Christian University, Virtual 2020

Mentorship

Helen Lu (2021–2024); William Hou (2022–2023); Anna Bryan (2021–2022); Tiffany Chen (2021–2022); Angela Zhang (2021–2022); Neha Hulkund (2020–2021); Roshni Sahoo (2018–2020); Skylar Gordon (2018–2019); Xinyi Guo (2018–2019).

Reviewing

Conference on Fairness, Accountability, and Transparency (FAccT) 2022 – 2025
Conference on Health, Inference, and Learning (CHIL) 2020, 2022
Machine Learning for Healthcare (MLHC) 2020 – 2023
Machine Learning for Health Symposium (ML4H) 2019 – 2022
International Conference on Machine Learning (ICML) 2021 – 2022
Computer Vision and Pattern Recognition (CVPR) 2021
Women in Machine Learning (WiML), NeurIPS Workshop 2018, 2020
Neural Information Processing Systems (NeurIPS) 2020

Panels

Reviews and Rebuttals Mentorship Panel (ML4H) — Panelist 2025
Round Table on Fairness (ML4H) — Discussant 2025
AI and Geriatrics Workshop (Johns Hopkins) — Invited participant 2025
AI and Society Panel (Olin College) — Panelist 2023
Career Mentorship Panel (MIT) — Panelist 2021
Graduate Student Panel (McCormick Hall) — Panelist 2020
Graduate Student Panel (MIT Women in EECS) — Panelist 2019
Lightning Talks (MIT Women in EECS) — Speaker 2017

Leadership

Organizing Committee, Workshop Chair, MLHC 2025
Organizing Committee, Publicity Chair, FAccT 2025
Organizing Committee, Workflow Chair, ML4H 2023
Applied Machine Learning Seminar Organizer, MIT 2023
Undergraduate Mentorship Improvement Initiative, MIT 2020 – 2021
GW6 Event Coordinator, MIT 2018 – 2019
MIT AI Mentorship Program Coordinator, MIT 2018 – 2020