Soumya Ghosh

☎ 303-956-5732
✉ soumyatghosh@gmail.com
ϟ http://www.soumyaghosh.com/

Education

Ph.D., Computer Science — 2009 - 2015
Brown University — Providence, RI
Advisor: Erik Sudderth
Thesis: Bayesian nonparametric discovery of layers and parts from scenes and objects.

M.Sc., Computer Science — 2007 - 2009
University of Colorado — Boulder, CO

M.Sc., Computer Science — 2004 - 2006
University of Houston — Houston, TX

B.E., Computer Engineering — 2000 - 2004
University of Mumbai — Mumbai, India

Research Experience

Sep 2018 - Current — Principal Investigator, **MIT-IBM Watson AI Lab**, Cambridge, MA
Feb 2016 - Current — Research Scientist, **IBM Research**, Cambridge, MA

Bayesian Deep Learning: Developed methods for reliable model selection and improving calibration of Bayesian neural networks, disentangling learned representations in deep generative models, and federated learning of neural networks.

Sensitivity and Robustness of ML methods: Developed fast and accurate approximations to methods for quantifying sensitivity of machine learning models to perturbations in training data and modeling assumptions.

Probabilistic Methods for Longitudinal Data: Explored models and methods for drawing inferences from irregularly sampled temporal data arising in healthcare settings. Applied developed models for studying the progression of neurodegenerative diseases, including Huntington's and Parkinson's diseases.

Service: Advised Postdoctoral researchers and interns on research projects and directions. Actively partook in hiring and recruitment activities for the growth of the lab.

Oct 2014 - Jan 2016 — Post Doctoral Research Scientist, **Disney Research**, Boston, MA

Online Learning of Neural Networks: Developed learning algorithms for scalable, online learning of Bayesian neural nets.

Gesture Recognition: Explored methods for recognizing gestures from videos and acceleration signals extracted from wearable sensors.

User Behavior Analysis: Statistical models for analyzing guest patterns in Disney parks.

Research Experience

Fall 09 – Summer 14 — Research assistant at Dept. of Computer Science, **Brown University**, Providence, RI

Statistical Modeling: Designed novel probabilistic models for unsupervised discovery of topics from text collections, regions from images/videos and parts from 3D objects.

Bayesian Nonparametrics: Developed novel models that increase in complexity with increasing amounts of data for analyzing spatio-temporally correlated data.

Inference: Built effective, robust and reliable stochastic search and MCMC based inference algorithms for latent variable models.

Technology Transfer: Developed and transferred research software to the Office of Naval Research.

Jun-Aug 2013 — Research intern at **Microsoft Research (MSR)** New England, Cambridge, MA

High dimensional classification: Explored dictionary learning algorithms for learning efficient representations for classification of high dimensional continuous data (images).

Large Scale Density Modeling: Developed parallel Expectation Maximization algorithms for learning large scale mixtures of factor analyzers from several million high dimensional data points.

Jun-Aug 2012 — Research intern at **Disney Research**, Pittsburgh, PA

Unsupervised Learning: Developed statistical models and Metropolis Hastings (MCMC) samplers for discovering the number and extent of regions exhibiting coherent appearance and motion in video sequences.

Jan 07- May 09 — Research assistant at Dept. of Computer Science, **University of Colorado**, Boulder, CO

Document modeling: Utilized topic models for measuring the relevancy of user contributed documents to subject-themed digital libraries.

Autonomous robot navigation (DARPA funded LAGR program): Designed computer vision algorithms and features for obstacle avoidance and traversable path identification to aid robots navigate in unstructured environments.

Jun-Aug 2008 — Research intern at **Bosch Research**, Pittsburgh, PA

Learning from class proportions: Developed algorithms that utilize prior knowledge of class proportions to better learn finite mixture models. These algorithms require only a fraction of training data and are more noise tolerant.

Jun-Jul 2007 — Summer researcher at **Lunar & Planetary Institute**, Houston, TX

Geomorphic mapping: Created tools for categorizing planetary surfaces into geomorphic landforms (craters, ridges, etc.) using naive Bayes, SVM and bagged decision tree classifiers.

Jun-Aug 2005 — Summer intern at **National Radio Astronomy Observatory**, WV

Telescope analytics: Identified and captured leading indicators of radio telescope failures. Developed an extensible system for periodic logging of the relevant factors in a MySQL database.

Research Grants

- *Parkinson's Disease Progression Model.*
  Michael J. Fox Foundation For Parkinson's Research
  Principal Investigator. August 2018 - July 2019.

Publications

Invited Papers

1. **Soumya Ghosh** and Finale Doshi-Velez. Discussions On Horseshoe Regularisation For Machine Learning In Complex And Deep Models. *International Statistical Review*, April 2020. https://doi.org/10.1111/insr.12377.

2. Tomasz F. Stepinski, Ricardo Vilalta and **Soumya Ghosh**. Machine Learning Tools for Automatic Mapping of Martian Landforms. *IEEE Intelligent Systems* 22, 6, 100–106, Nov 2007.

Refereed Publications

1. Kristen A Severson, Lana M Chahine, Luba A Smolensky, Kenney Ng, **Soumya Ghosh***, Jianying Hu*. Discovery of Parkinson's disease states using machine learning and longitudinal data. *The Lancet Digital Health (to appear)*, 2021. * co-last authors.

2. Meet P. Vadera, **Soumya Ghosh**, Kenney Ng, Benjamin M. Marlin. Post-hoc loss-calibration for Bayesian neural networks. *Uncertainty in Artificial Intelligence (UAI)*, 2021.

3. Siddharth Biswal, **Soumya Ghosh**, Jon Duke, Bradley Malin, Walter Stewart, Cao Xiao, Jimeng Sun. EVA: Generating Longitudinal Electronic Health Records Using Conditional Variational Autoencoders. *Machine Learning for Healthcare (MLHC)*, 2021.

4. **Soumya Ghosh***, William T. Stephenson*, Tin D. Nguyen, Sameer K. Deshpande, Tamara Broderick. Approximate Cross-Validation for Structured Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. * equal contributions.

5. Sebastian Claici, Mikhail Yurochkin, **Soumya Ghosh**, Justin Solomon. Model Fusion with Kullback–Leibler Divergence. *37th International Conference on Machine Learning (ICML)*, 2020.

6. Kristen A Severson, Lana M Chahine, Luba A Smolensky, Kenney Ng, Jianying Hu, **Soumya Ghosh**. Personalized Input-Output Hidden Markov Models for Disease Progression Modeling. *Machine Learning for Healthcare (MLHC)*, 2020.

7. Bum Chul Kwon, Vibha Anand, Kristen A. Severson, **Soumya Ghosh**, Zhaonan Sun, Brigitte I. Frohnert, Markus Lundgren, and Kenney Ng. DPVis: Visual Analytics with Hidden Markov. *IEEE Transactions on Visualization and Computer Graphics*, 2020.

8. **Soumya Ghosh**, Jiayu Yao, Finale Doshi-Velez. Model selection in Bayesian neural networks via Horseshoe priors. *Journal of Machine Learning Research (JMLR)*, 2019.

9. Mikhail Yurochkin, Mayank Agarwal, **Soumya Ghosh**, Kristjan Greenewald, Nghia Hoang. Statistical model aggregation via parameter matching. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

10. Mikhail Yurochkin, Mayank Agarwal, **Soumya Ghosh**, Kristjan Greenewald, Nghia Hoang, Yasaman Khazaeni. Bayesian Nonparametric Federated Learning of Neural Networks. *36th International Conference on Machine Learning (ICML)*, 2019.

11. Kristen Severson, **Soumya Ghosh**, Kenney Ng. Unsupervised Learning with Contrastive Latent Variable Models. *33rd AAAI Conference On Artificial Intelligence (AAAI)*, 2019.

12. Zhaonan Sun, **Soumya Ghosh**, Ying Li, Yu Cheng, Amrita Mohan, Cristina Sampaio, Jianying Hu. A probabilistic disease progression modeling approach and its application to integrated Huntington's disease observational data. *JAMIA Open, 2(1):123-30,* 2019. JAMIA High Impact Open Access collection 2019. (https://academic.oup.com/jamiaopen/pages/impact).

13. Jiayu Yao, Weiwei Pan, **Soumya Ghosh**, Finale Doshi-Velez. Quality of Uncertainty Quantification for Bayesian Neural Network Inference. *ICML workshop on uncertainty and robustness in deep learning*, 2019.

14. Melanie F Pradier, Weiwei Pan, Jiayu Yao, **Soumya Ghosh**, Finale Doshi-Velez. Latent Projection BNNs: Avoiding weight-space pathologies by learning latent representations of neural network weights. *NeurIPS workshop on Bayesian deep learning,* 2018.

15. **Soumya Ghosh**, Jiayu Yao, Finale Doshi-Velez. Structured Variational Learning of Bayesian Neural networks with Horseshoe Priors. *35th International Conference on Machine Learning (ICML)*, 2018.

16. Ajjen Joshi, **Soumya Ghosh**, Sarah Gunnery, Linda Tickle-Degnen, Stan Sclaroff, Margrit Betke. Context-Sensitive Prediction of Facial Expressivity Using Multimodal Hierarchical Bayesian Neural Networks. *13th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2018.

17. Bin Liu, Ying Li, Zhaonan Sun, **Soumya Ghosh**, Kenney Ng. Early Prediction of Diabetes Complications from Electronic Health Records: A Multi-task Survival Analysis Approach. *32nd AAAI Conference On Artificial Intelligence (AAAI)*, 2018.

18. **Soumya Ghosh** and Finale Doshi-Velez. Model Selection in Bayesian Neural Networks via Horseshoe Priors. *NIPS Workshop on Bayesian Deep Learning,* 2017.

19. Ajjen Joshi, **Soumya Ghosh**, Margrit Betke, Stan Scarloff, Hanspeter Pfister. Personalizing Gesture Recognition Using Hierarchical Bayesian Neural Networks. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

20. **Soumya Ghosh**, Zhaonan Sun, Ying Li, Yu Cheng, Amrita Mohan, Cristina Sampaio, Jianying Hu. An Exploration of Latent Structure in Observational Huntington's Disease Studies. *AMIA Joint Summits on Translational Science, CRI*, 2017.

21. Zhaonan Sun, Ying Li, **Soumya Ghosh**, Yu Cheng, Amrita Mohan, Cristina Sampaio, Jianying Hu. A Data-Driven Method for Generating Robust Symptom Onset Indicators in Huntington's Disease Registry Data. *American Medical Informatics Association Annual Symposium (AMIA)*, 2017.

22. Vibha Anand, Amos Cahan, **Soumya Ghosh**. Clinical Trials.Gov: A Topical Analyses. *AMIA Joint Summits on Translational Science, CRI*, 2017.

23. Vibha Anand, **Soumya Ghosh**, Amit Anand. Is there a Priority Shift in Mental Health Clinical Trials? *World Congress on Medical and Health Informatics (Medinfo)*, 2017.

24. **Soumya Ghosh**[*], Yu Cheng[*], Zhaonan Sun. Deep State Space Models for Computational Phenotyping. *IEEE International Conference on Health Informatics (ICHI)*, 2016. [*] equal contributions.

25. **Soumya Ghosh**, Francesco Delle Fave, Jonathan Yedidia. Assumed Density Filtering Based Methods for Scalable Learning of Bayesian Neural Networks. *30^{th} AAAI Conference On Artificial Intelligence (AAAI)*, 2016.

26. **Soumya Ghosh** and Erik Sudderth. Approximate Bayesian Computation for Distance-Dependent Learning. *NIPS Workshop on Bayesian Nonparametrics: The Next Generation (NIPSW)*, 2015.

27. **Soumya Ghosh**, Michalis Raptis, Leonid Sigal, Erik Sudderth. Nonparametric Clustering with Distance Dependent Hierarchies. *30^{th} Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.

28. **Soumya Ghosh**, Erik B. Sudderth, Matthew Loper, Michael J. Black. From Deformations to Parts: Motion-based Segmentation of 3D Objects. *Advances in Neural Information Processing Systems 25 (NIPS)*, 2012.

29. **Soumya Ghosh**, Erik B. Sudderth. Nonparametric learning for layered segmentation of natural images. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012.

30. **Soumya Ghosh**, Andrei B. Ungureanu, Erik B. Sudderth, David Blei. Spatial distance dependent Chinese restaurant processes for image segmentation. *Advances in Neural Information Processing Systems 24 (NIPS)*, 2011.

31. **Soumya Ghosh**, Jane Mulligan. A segmentation guided label propagation scheme for autonomous navigation. *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

32. **Soumya Ghosh**, Tomasz F. Stepinski, Ricardo Vilalta. Automatic Annotation of Planetary Surfaces With Geomorphic Labels. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 175–185, 2010.

33. **Soumya Ghosh**, Joseph Pfeiffer III, Jane Mulligan. A general framework for reconciling multiple weak segmentations of an image. *Workshop on Applications of Computer Vision (WACV)*, 2009.

34. Steven Bethard, **Soumya Ghosh**, James Martin, Tamara Sumner. Topic Model Methods for Automatically Identifying Out-of-Scope Resources. *Joint Conference on Digital Libraries (JCDL)*, 2009.

35. **Soumya Ghosh**, Soundar Srinivasan, Burt Andrews. Using Weak Supervision in Learning Gaussian Mixture Models. *International Joint Conference on Neural Networks (IJCNN)*, 2009.

36. Tomasz F. Stepinski, **Soumya Ghosh**, Ricardo Vilalta. Machine learning for automatic mapping of planetary surfaces. *Proceedings of the 19th National Conference on Innovative Applications of Artificial Intelligence (IAAI)*, 2007

37. Tomasz F. Stepinski, **Soumya Ghosh**, Ricardo Vilalta. Automatic recognition of landforms on Mars using terrain segmentation and classification. *International Conference on Discovery Science (DS)*, 2006.

Patents

- Meet P. Vadera, Uri Kartoun, Soumya Ghosh, Kenney Ng. Post-hoc loss-calibration for Bayesian neural networks. *Filed and pending*, 2021.
- Kristjan Greenewald, Mayank Agarwal, Mikhail Yurochkin , **Soumya Ghosh**, Nghia Hoang, Yasaman Khazaeni. A method for combining pre-trained neural networks into a memory and computation efficient global model. *Filed and pending*, 2019.
- Zhaonan Sun, **Soumya Ghosh**, Ying Li, Yu Cheng, Jianying Hu. Generating robust symptom onset indicators. *Filed and pending*, 2017.
- Yu Cheng, **Soumya Ghosh**, Jianying Hu, Ying Li, Zhaonan Sun. Identifying and indexing discriminative features for disease progression in observational data. *Filed and pending*, 2017.

Teaching Experience

- Teaching Assistant: CSCI 1950-f, Introduction to Machine Learning, Spring 2011.
- Postdoctoral Researcher Advising:
  - Kristen A. Severson, Center for Computational Health, IBM Research, 2018 – 2020.
  - Bin Liu, Center for Computational Health, IBM Research, 2017 – 2019.
- Student (Intern) Advising:
  - Giridhar Gopalan, (Ph.D. candidate at Dept. of Statistics, Harvard University), Summer 2015.
  - Ajjen Joshi, (Ph.D. candidate at Dept. of Computer Science, Boston University), Summer 2015.
  - Michael Colomb, (Undergraduate, UIUC), Summer 2017.
  - Ethan Evans, (Ph.D. candidate at Dept. of Chemistry, MIT), Winter 2018.
  - Siddharth Biswal, (Ph.D. candidate of Dept. of Computer Science, Georgia Tech), Summer 2018.
  - Joe Davison, (Graduate student, SEAS, Harvard University), Summer 2019.
  - Meet Vadera, Ph.D. candidate of Dept. of Computer Science, University of Massachusetts, Amherst), Summer 2020.

Awards and Honors

CVPR Outstanding Reviewer	2021.
IBM Outstanding Technical Achievement Award for scalable and reusable models of disease progression	2021.
CVPR Outstanding Reviewer	2019.
IBM Outstanding Technical Achievement Award	2019.
First Place, ICHI Data Challenge	2016.
UAI travel grant	2014.
NIPS travel grant	2011.
AAAI travel grant	2007.
Brown University graduate fellowship	2009-10.
JN TATA endowment scholarship	2004-05.
National Scholarship Scheme, Govt. of India	1998.

Skills

Core Skills	Statistical modeling and inference, machine learning, Bayesian analysis, graphical models, linear algebra, optimization, Markov Chain Monte Carlo (MCMC) and computer vision.
Programming	Proficient in Python (NumPy, SciPy). Past experience in Java, R, SQL, C/C++ & shell scripting.
Tools	scikit-learn, scikit-image, NLTK, Autograd, Tensorflow, PyTorch, and PyMC.

Talks

- Approximate Cross-Validation for Structured Models.
  - MIT-IBM Watson AI Lab meeting, July 2020.
  - World Meeting of the International Society for Bayesian Analysis, June 2021.
- Horseshoe Priors for Bayesian Neural Networks. EECS, Tufts University, October 2018.
- Deep Generative Models. Fraenkel lab, MIT, February 2018.
- Deep Learning — A cautionary tale. North East Computational Health Summit, IBM Research, Yorktown Heights, 2017.
- Introduction to classification. Guest lecture at MMSCI, Harvard Medical School, 2017.
- Learning and inference in distance dependent models. IVC seminar series, Boston University, Nov 2015.
- Bayesian nonparametric discovery of layers and parts from scenes and objects.
  - IBM Research, Yorktown Heights, NY, Apr 2014.
  - HP, Palo Alto, CA, Apr 2014.
  - Bosch Research, Palo Alto, CA, May 2014.

  - Exxon Mobil, Corporate Strategic Research, Clinton, NJ, May 2014.
  - BBN Technologies, Cambridge, MA, May 2014.
  - Disney Research, Boston, MA, May 2014.
  - Schlumberger-Doll Research center, Cambridge, MA, May 2014.
- Bayesian nonparametrics for image, video and user analysis. Disney Research, Boston, Feb 2015.
- Statistical models for spatially correlated data. Microsoft Research, New England, Jun 2013.
- Nonparametric learning for layered segmentation of natural images. Disney Research, Pittsburgh, Jun 2012.
- Motion based segmentation of articulated 3D objects. Faculty recruiting, Brown University, 2012.
- Machine learning for automatic mapping of planetary surfaces. IAAI 2007.
- Automatic Recognition of Landforms on Mars Using Terrain Segmentation & Classification. Discovery Science 2006.

Service

- Workshop Organization:
  - Practical Bayesian Methods for Big Data, IBM AI research week, MIT, Cambridge, MA. September 2019.
  - Second NorthEast Computational Health Symposium, Cambridge, MA. April 2018.
- Peer Review:
  - Grant Agencies:
    * Israel Science Foundation (**ISF**), 2021.
  - Area Chair:
    * Machine Learning for Healthcare (**MLHC**), 2021.
  - Program Committee Member / Reviewer:
    * Advances in Neural Information Processing Systems 29, 28, 27 and 26 (**NeurIPS**), 2021, 2020, 2017, 2016, 2015, 2014 and 2013.
    * International Conference on Learning Representations (**ICLR**), 2021, 2020, 2019.
    * International Conference on Machine Learning (**ICML**), 2021, 2019, 2018, 2017.
    * NeurIPS workshop on Bayesian Deep Learning, 2019.

    * NIPS workshop on Bayesian Nonparametrics: The next generation, 2015.
    * IEEE Computer Vision and Pattern Recognition (**CVPR**) 2021, 2019, 2018, 2016, 2015.
    * AAAI Conference on Artificial Intelligence (**AAAI**) 2017, 2016.
    * International Conference of Computer Vision (**ICCV**) 2015 and 2013.
    * Conference on Technologies and Applications of Artificial Intelligence (TAAI), 2014.
  - Journal Reviewer:
    * Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2020.
    * International Statistical Review, 2019.
    * Artificial Intelligence 2016.
    * IET Computer Vision, May 2013.
    * Springer Journal of Signal, Image and Video Processing, May 2012.
- PhD Admissions Committee: Brown University, Dec 2013.
- Graduate Student Orientation Committee: Brown University, September 2011.