Sandhya Tripathi

Department of Anesthesiology, Washington University School of Medicine, MO 63110
Email: sandhyat@wustl.edu Webpage: sandhyat.github.io

EDUCATION

Indian Institute of Technology Bombay, Mumbai, India

PhD in Operations Research (Machine learning focused) at Industrial Engineering and Operations Research (IEOR) Jul 2014 – Jan 2020

- Research Area: Design and evaluation of loss functions in classification problems.
- Supervisor: Prof. N. Hemachandra
- Teaching Assistant: Data Analytics in Operations Research, Online Machine Learning, Computer Programming & Algorithms (Lab component), Decision Analysis and Game Theory, Engineering Statistics, Optimization Techniques.
- **IEOR Alumnus Endowment: Excellence in Doctoral Dissertation Award (2020-22)**

M.Sc. in Operation Research at IEOR Jul 2012 – Jun 2014

- Research Area: Finance, Regression Analysis, Queuing Theory.
- Supervisor: Prof. N. Hemachandra
- Coursework: Engineering Statistics, Stochastic Processes, Optimization Techniques, Game Theory, Machine Learning, Foundations of Intelligent and Learning Agents, Statistical Techniques in Data Mining.
- Cumulative GPA: 8.96/10 (4.0/4.0 GPA)

Delhi University, Delhi, India

B.Sc. (Honors) in Statistics at Lady Shri Ram College for Women Jul 2009 – Jun 2012

- Cumulative GPA: 84.37% (3.0/4.0 GPA)

EXPERIENCE

Washington University School of Medicine , St. Louis, USA

Post Doctoral Research Scholar at Department of Anesthesiology Feb 2020 – Till Date

Clinical Model Development and Deployment

- Developed solutions for the problem of merging patient medical records from different hospitals or health recording systems using state-of-the-art deep learning techniques.
- Queried large open-source medical dataset MIMIC in *PostgreSQL* for solution's performance demonstration.
- Validated these techniques by implementing self-designed *PyTorch* models on high-performance computing platforms in portable *Docker* containers to demonstrate improvement over existing methods.
- Developed tailor-made code in *R* using statistical techniques to process intra-operative medication data for manual entry or drug unit mismatch errors.
- Designed supervised prediction models for post-surgical complications (such as in-hospital mortality, cardiac arrest etc) in *Python* to inform clinicians of individualized patient risk. Utilized LSTM and attention-based deep networks to process intra-operatively recorded medications and vitals for model deployment using *torch serve*.

Evaluation of clinically deployed models

- Designed risk prediction explanation methods using machine learning interpretability techniques while collaborating with domain expert colleagues on *GitHub*.
- Evaluated the deployed machine learning models that were being used in a telemedicine centre as part of a large randomized control trial for racial, sex-based or age-based discrimination.
- Conducted interviews with the clinicians to understand the impact of AI-generated risk predictions along with explanations on their decisions.

Impact of social vulnerability on surgery outcomes

- Quantified zipcode based social economic status for patients from St Louis area using *GIS* tools.
- Conducted statistical analysis to study the effect of social vulnerability on the post surgical complications.

Indian Institute of Technology Bombay, Mumbai, India

Research Scholar at IEOR Jul 2014 – Jan 2020

Learning in the presence of label/annotation noise

- Designed algorithms to learn in the presence of label noise (due to crowdsourcing or sensor breakdown) in a cost-sensitive (arise in safety-critical systems) binary classification setup.
- When either of the two label classes is corrupted/flipped with the same rate, proposed a modified version of conventional squared loss along with theoretical guarantees and empirically demonstrated its significantly better performance on a variety of real-world datasets.
- When the two label classes are corrupted/flipped with different rates, proposed an adaptation of Generative Adversarial Networks (GANs) that is robust to corrupted real-world MNIST and Fashion-MNIST image datasets.

Scalable classifiers based on exponential loss function

- Developed a theoretical framework using exponential loss for scalable binary classifiers.
- Demonstrated performance improvements in terms of accuracy and time across a large number of varying feature and sample-sized datasets.

Interpretability in classification via Shapley value of a cooperative classification game

- Modelled a novel classification game for explaianable & interpretable feature selection.

PUBLICATIONS

1) **Tripathi, S.**, Fritz, B. A., Abdelhack, M., Avidan, M. S., Chen, Y., and King, C. R. (2024). Multi-view representation learning for tabular data integration using inter-feature relationships. Journal of Biomedical Informatics 2024 Feb 10;151:104602. doi: 10.1016/j.jbi.2024.104602.

2) **Tripathi, S.** and King, C.R. (2024). Contrastive learning: Big Data Foundations and Applications. In Proceedings of the 7th Joint International Conference on Data Science and Management of Data (11th ACM IKDD CODS and 29th COMAD) (pp. 493-497).

3) Abdelhack, M., Zhang, J., **Tripathi, S.**, Fritz, B. A., Avidan, M. S., Chen, Y., and King, C. R. (2023). A Modulation Layer to Increase Neural Network Robustness Against Data Quality Issues, Transactions on Machine Learning Research.

4) **Tripathi, S.**, Fritz, B. A., Abdelhack, M., Avidan, M. S., Chen, Y., and King, C. R. (2022). Algorithmic bias in machine learning-based delirium prediction. (Extended Abstract) Machine Learning for Health (ML4H) symposium 2022, November 28th, 2022, New Orleans, United States.

5) Hemachandra, N., Patil, K., and **Tripathi, S.** (2020). Equilibrium points and equilibrium sets of some GI/M/1 queues. Queueing Systems, 96(3), 245–284.

6) Petety, A., **Tripathi, S.**, and Hemachandra, N. (2020). Attribute noise robust binary classification (Student Abstract). In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, pp. 13897–13898).

7) **Tripathi, S.**, Hemachandra, N., and Trivedi, P. (2020). Interpretable feature subset selection: A Shapley value based approach. In IEEE International Conference on Big Data (Big Data) (pp. 5463–5472). IEEE.

8) **Tripathi, S.**, and Hemachandra, N. (2019). Cost sensitive learning in the presence of symmetric label noise. In Advances in Knowledge Discovery and Data Mining (pp. 15–28). Springer International Publishing.

9) **Tripathi, S.**, and Hemachandra, N. (2018). Scalable linear classifiers based on exponential loss function. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (pp. 190–200).

COMMUNICATED & WORKING PAPERS

1) Abdelhack, M.*, **Tripathi, S.***, Chen, Y., Avidan, M. S., and King, C. R. (2023). Social Vulnerability and Surgery Outcomes: A Cross-sectional Analysis.Preprint. Under Review.

2) **Tripathi, S.**, Fritz, B. A., Avidan, M. S., Chen, Y., and King, C. R. (2021a). Clinical model report cards: A transparent approach for model uncertainty and fairness quantification during deployment. Working paper.

3) **Tripathi, S.**, Fritz, B. A., Abdelhack, M., Avidan, M. S., Chen, Y., and King, C. R. (2020). (Un) fairness in post-operative complication prediction models. arXiv preprint arXiv:2011.02036.

4) **Tripathi, S.**, and Hemachandra, N. (2020). GANs for learning from very high class conditional noisy labels. arXiv preprint arXiv:2010.09577.

PRESENTATIONS & POSTERS

1) **IEEE BigData 2020 (Virtual)** Presented our conference paper titled “Interpretable feature subset selection”. Dec 2020.

2) **PAKDD 2019, Macau** Presented our conference paper titled “Cost-sensitive learning in the presence of symmetric label noise”. Apr 2019.

3) **IRCN, University of Tokyo** Attended and presented a poster titled “GANs based label noise robust classification” in the Neuro-inspired Computation Course. Mar 2019.

4) **CoDS-COMAD 2018, Goa, India** Presented the work titled “Scalable linear classifiers based on exponential loss function” in the Joint International Conference on Data Science and Management of Data. Jan 2018.

5) **WINE 2017, IISC Bengaluru, India** Presented poster and lightning talk about the work titled “Shapley Value based Binary Classifiers” at International Conference on Web and Internet Economics. Dec 2017.

6) **ECQT 2016, Toulouse** Presented the work titled “Equilibrium sets of some $GI/M/1$ queues” in European Conference on Queuing Theory. Jul 2016.

7) **MLSS 2015, University of Kyoto** Attended and presented a poster titled "Some Equilibrium Prediction Problems in $M/M/1$ Queues" at International Machine Learning Summer School. Aug 2015.

TECHNICAL SKILLS

Programming Skills: Python, Pytorch, AMPL, Docker (Proficient).

Statistical and Simulation Tools: R, SPSS, Anylogic, Scilab, Mathematica (Intermediate).

Databases & Version control: MySQL, PostgreSQL (Beginner), Github.

TEACHING & MENTORING

1) Organized and taught a tutorial, 'Contrastive learning: Big Data Foundations and Applications' to the researchers from both academia and industry at ACM CODS-COMAD 2024, Bengaluru, India. Jan 2024.
2) Conducted a tutorial, 'Contrastive learning: Big Data Foundations and Applications' at IEEE BigData 2023, Sorrento, Italy. Dec 2023.
3) Jointly mentored a master's student from the WUSTL-CS program (Shuting Cui) to solve the problem of merging medical records that are structured in a time-dependent way and are very sparse. Jan 2022-Feb 2023
4) Organized and taught an online tutorial on 'Label noise: Problems and solutions' at IEEE DSAA 2020. Oct 2020.
5) Jointly mentored an undergraduate summer student from NISER Bhubaneswar (Aditya Petety) on the problem of learning in the presence of feature noise (sensor failures etc) leading to a student abstract presentation at AAAI 2020. May 2019-Jul 2019
6) Teaching Assistantship: Conducted tutorials, designed assignments, and evaluated exams for many courses including Engineering statistics, Optimization techniques, Topics in Machine learning, and Online learning. Jan 2015-May 2018

SHORT RESEARCH PROJECTS

BB553 Bio-informatics: Course Project under guidance of Prof. P. Bhaumik

Identification of protein-protein interface residues using ExpERM. Jul 2017 – Nov 2017

- Explored methods for predicting protein-protein interface residues using both supervised and unsupervised machine learning approaches.

BB607 Proteomics: Course Project under guidance of Prof. S. Srivastava

Human Proteome Reference Maps Jul 2017 – Nov 2017

- Presented a survey of pioneering research papers on human proteome map.

Master's Research Projects under guidance of Prof. N. Hemachandra

Estimating equilibrium point in Admission Control for M/M/1 queue May 2014 – Dec 2014

- Estimated the equilibrium point in a firm market setting without assuming known demand distribution.

Biased Estimators in Regression Analysis Jan 2014 – Apr 2014

- Compared the intercept and non-intercept model and studied the conditions under which the biased estimators (LASSO & Ridge estimator) can be obtained from the general class of estimators.

Option Pricing Models (Finance) Jul 2013 – Nov 2013

- Studied introduction about derivatives and securities and implemented Binomial Pricing Model & Black Scholes Model for Option Pricing.

LANGUAGES

Hindi, English, Japanese, Italian.

ACADEMIC SERVICES

PC member: ECML2020, IEEE BigData'20 Special Track-XAI, IJCAI-21,-22,-23, ML4H-21,-22,-23

- Top Reviewer Award at ML4H 2021. Given to only 10 out of hundreds of reviewers.

Reviewer: SADHANA IAS journal, Journal of Biomedical Informatics, Neural Networks, Engineering Applications of Artificial Intelligence

NON-ACADEMIC ACHIEVEMENTS

1) Participated in ExCELS Program organized by Office of Postdoctoral affairs at Washington University in St Louis. Oct 2022
2) Presented with President's Volunteer Service Award in recognition and appreciation for commitment to Gateway Pet Guardians through volunteer service. Jan 2022
3) Member of Indian Youth Delegation to China organized by Government of India. Jun 2016
4) Awarded NCC 'C' grade senior division certificate. Jun 2011