



PASAN DISSANAYAKE

✉ pasand@umd.edu  linkedin.com/in/pasandissanayake  pasandissanayake.github.io

RESEARCH INTERESTS

My research interests lie in the area of trustworthy machine learning. I am particularly interested in using tools from information theory, statistics and optimization to develop novel strategies for responsible AI that is aligned with our pillars of trust. I also seek empirical validation of these strategies through carefully designed experiments.

EDUCATION

University of Maryland, College Park, MD Started Aug. 2022

PhD in Electrical and Computer Engineering (CGPA: 4.0/4.0)

Adviser: Prof. Sanghamitra Dutta

Research areas: Trustworthy machine learning, explainability, privacy, fairness

University of Moratuwa, Sri Lanka Graduated Aug. 2021

BSc (Hons.) in Electronic and Telecommunication Engineering (GPA: 4.04/4.2)

Research adviser: Dr. Prathapasinghe Dharmawansa

PUBLICATIONS

- **P. Dissanayake** and S. Dutta, “Model Reconstruction Using Counterfactual Explanations: A Perspective From Polytope Theory”, *NeurIPS 2024*
- **P. Dissanayake**, F. Hamman, B. Halder, I. Sucholutsky, Q. Zhang, S. Dutta, “Formalizing Limits of Knowledge Distillation Using Partial Information Decomposition”, Compression Workshop at NeurIPS 2024
- B. Halder, F. Hamman, **P. Dissanayake**, Q. Zhang, I. Sucholutsky, S. Dutta, “Quantifying Spuriousness of Biased Datasets Using Partial Information Decomposition,” DMLR workshop at ICML 2024
- F. Hamman, **P. Dissanayake**, S. Mishra, F. Lecue, S. Dutta, “Quantifying Prediction Consistency Under Model Multiplicity in Tabular LLMs” (under review)
- M. Nomeir, **P. Dissanayake**, S. Meel, S. Dutta, S. Ulukus, “Private Counterfactual Retrieval” (under review)
- E. Noorani, **P. Dissanayake**, F. Hamman, S. Dutta, “Counterfactual Explanations for Model Ensembles Using Entropic Risk Measures” (under review)
- **P. Dissanayake**, P. Dharmawansa and Y. Chen, “Distribution of the Scaled Condition Number of Single-Spiked Complex Wishart Matrices,” in *IEEE Transactions on Information Theory*, vol. 68, no. 10, pp. 6716-6737, Oct. 2022
- P. Dharmawansa, **P. Dissanayake** and Y. Chen, “The Eigenvectors of Single-Spiked Complex Wishart Matrices: Finite and Asymptotic Analyses,” in *IEEE Transactions on Information Theory*, vol. 68, no. 12, pp. 8092-8120, Dec. 2022

PRESENTATIONS

Model extraction using counterfactual explanations (Poster)

Artificial Intelligence Interdisciplinary Institute Leadership Maryland Summit (Oct. 2024)

Northrop Grumman University Research Symposium, McLean, VA (Oct. 2023)

North American School of Information Theory, Philadelphia, PA (June 2023)

AWARDS AND HONORS

- **Outstanding Teaching Assistant (2024) - ECE, University of Maryland**

- **Dean's Fellowship (2022) - University of Maryland**
- **Dean's List (in all 8 semesters) - University of Moratuwa**
For maintaining a semester GPA of 3.8 or higher during the considered semester
- **Google Hash Code (2019) Country rank: 3**
Team programming competition based on real Google engineering problems
- **MoraXtreme (2017) Second runners-up**
Team programming competition organized by IEEE Student Branch of University of Moratuwa

RESEARCH PROJECTS

Quantifying task-related knowledge in knowledge distillation

Partial information decomposition is used to quantify the distillable and distilled knowledge of a teacher corresponding to a given student and a downstream task within a knowledge distillation setting. Experiments are focused on image datasets such as CIFAR10, CIFAR100 and CUB-200-2011. PyTorch implementations of WideResNet models are used in the experiments.

Counterfactual explanations and model extraction attacks

Explored how counterfactual explanations can be utilized for model extraction attacks, even when the API provides explanations only to the queries originating from one side of the decision boundary. Experiments were implemented using Tensorflow, and carried-out on tabular datasets. Published at NeurIPS2024. Currently working on schemes to retrieve counterfactual explanations privately.

Quantifying prediction consistency of tabular LLMs under fine-tuning multiplicity

This was an effort to efficiently quantify model multiplicity of LLMs fine-tuned on tabular datasets. Experiments were done with BigScience T0 and Google FLAN T5 models from Huggingface. LORA and T-Few methods were used as the fine-tuning techniques. Publication under review.

Single-spiked complex Wishart matrices

Studied the properties of two important statistics of the complex Wishart ensemble, with a focus on the centralized (zero-mean) case with a single-spiked covariance structure. Simulations were done using Mathematica. Published at IEEE Trans. Info. Theory.

RELEVANT COURSEWORK

- CMSC764 - Advanced Numerical Optimization (UMD, Spring 2024)
- CMSC742 - Algorithms in Machine Learning: Guarantees and Analyses (UMD, Fall 2023)
- ENEE662 - Convex Optimization (UMD, Fall 2023)
- ENEE621 - Estimation and Detection Theory (UMD, Spring 2023)
- ENEE627 - Information Theory (UMD, Spring 2023)
- ENEE620 - Random Processes in Communication and Control (UMD, Fall 2022)
- EN4573 - Pattern Recognition and Machine Intelligence (UoM, Semester 8)

EMPLOYMENT

Research Assistant Department of Electrical and Computer Engineering, University of Maryland, College Park, MD	Jan 2024 to Present
Teaching Assistant Department of Electrical and Computer Engineering, University of Maryland, College Park, MD	Aug. 2022 to Dec. 2023
Lecturer (on Contract) Department of Electronic and Telecommunication Engineering, University of Moratuwa, Sri Lanka	Dec. 2021 to Aug. 2022

TEACHING EXPERIENCE

University of Maryland, College Park, MD

- ENEE420 - Communication Systems (Fall 2022)
- ENEE324 - Engineering Probability (Spring 2023, Fall 2023)
- ENEE436 - Foundations of Machine Learning (Fall 2023)

University of Moratuwa, Sri Lanka

- EN4720 - Security in Cyber-Physical Systems
- EN2550 - Fundamentals of Image Processing and Machine Vision
- EN3053 - Digital Communications - I

COMPUTER LANGUAGES AND TOOLS

Programming: Python, C++, Java **Technical Computing:** MATLAB, Mathematica
Data/ML Frameworks: Tensorflow, PyTorch, Keras, Scikit-Learn, Pandas, Weights&Biases

CERTIFICATIONS

- Machine Learning - Stanford University on Coursera
- Data Structures - University of California, San Diego on Coursera
- Algorithmic Toolbox - Offered by the University of California, San Diego on Coursera
- Certificate in Software Development in Java - National Institute of Business Management, SL
- Networking Essentials (NE-CS3032-IN-16-EN) - Offered by Cisco Networking Academy

SERVICES

- Reviewer - International Conference on Learning Representations (ICLR) 2025
- Reviewer - Workshop on Explainable AI in Finance (XAIFIN) 2024
- Reviewer - IEEE International Symposium on Information Theory (ISIT) 2024
- Reviewer - ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2024
- Reviewer - IEEE Wireless Communications Letters (*view on Publons*)
- Project Lead - UMD Trustworthy AI in Law and Society (TRAILS) AI Summer Camp 2024
- Member of the panel of judges - UMD Louis Stokes Alliance for Minority Participation (LSAMP) Fall Research Symposium 2023
- Mentor - Montgomery Blair High School Magnet Student Program 2023

EXTRA-CURRICULAR

- Secretary - Drama Society, University of Moratuwa for the year 2020/2021
- Co-chairperson - Organizing committee of “Abhina 2019”, the annual talent show of the Dept. of Electronic and Telecommunication Engineering
- Volunteer - Soyuru Sathkaraya 2018: underprivileged school development project organized by Engineering Faculty Students’ Union, University of Moratuwa
- Volunteer - E-Care 2018: Underprivileged school development project organized by Electronics Club, Dept. of Electronic and Telecommunication, University of Moratuwa
- Volunteer - Grama Prabodhaya 2017: Underprivileged school development project and junior high school seminar series organized by Rotaract Club, University of Moratuwa