

RESEARCH INTERESTS

Machine Learning, Data Mining, Time Series, Natural Language Processing, Healthcare, Fairness in AI systems

ACADEMIC APPOINTMENTS

Massachusetts Institute of Technology, Cambridge, MA Starting 01/2022
Postdoctoral Associate, CSAIL, PI: Marzyeh Ghassemi

EDUCATION

Worcester Polytechnic Institute, Worcester, MA 08/2016-12/2021
PhD, Data Science
MS, Data Science 12/2018
Advised by Elke Rundensteiner and Xiangnan Kong

SUNY Geneseo, Geneseo, NY
BA, Applied Mathematics, minor in Biomathematics 08/2012-05/2016

RESEARCH EXPERIENCE

Worcester Polytechnic Institute, Research Fellow, PI: Prof. Elke Rundensteiner 08/2016-12/2021
Microsoft, PhD Intern with Dr. Dipankar Ray and Dr. Hamid Palangi 05/2021-08/2021
UMass Medical School, Research Intern, PI: Dr. Jomol Matthew 08/2018-09/2019
University of Arizona, Research Intern, PI: Prof. Shirley Papuga 05/2015-08/2015

GRANTS

NSF-III: Timely Classification for Actionable Predictions (Under Review)
PI: Elke Rundensteiner, Co-PI: Xiangnan Kong.
This grant proposal is written based on my research and I am responsible for 90% of the writing.

SELECTED HONORS & AWARDS

🏆 **Best Poster**, International Conference on Health Informatics 2020
🏆 **Outstanding Graduate Research Award**, WPI 2019
🏆 **Best Poster**, Graduate Research Innovation and Exchange, WPI 2019
🏆 **People's Choice Poster Award**, Graduate Research Innovation and Exchange, WPI 2017
GAANN Fellowship (Annual Tuition + Stipend Award), U.S. Dept. of Education 2016-2021

PUBLICATIONS

I have published in KDD, NeurIPS, AAAI, ACL, CIKM, ECML, IEEE BigData, HEALTHINF, and IEEE BHI.

REFEREED

20. *Recovering the Propensity Score from Biased Positive Unlabeled Data.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner.
AAAI, 2022.
19. *Positive Unlabeled Learning with a Sequential Selection Bias.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Elke Rundensteiner, Emmanuel Agu.
SDM, 2022.

18. *Recurrent Bayesian Classifier Chains for Exact Multi-label Classification.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner.
NeurIPS, 2021.
17. *Learning Saliency Maps to Explain Deep Time Series Classifiers.*
Prathyush Parvatharaju, Ramesh Doddaiiah, Thomas Hartvigsen, Elke Rundensteiner.
CIKM, 2021.
16. *Energy-Efficient Models for High-Dimensional Spike Train Classification using Sparse Spiking Neural Networks.*
Hang Yin, John Boaz Lee, Xiangnan Kong, Thomas Hartvigsen, Sihong Xie.
KDD, 2021.
15. *Semi-Supervised Knowledge Amalgamation for Sequence Classification.*
Jidapa Thadajarassiri, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
AAAI, 2021.
14. *Variational Open-Set Recognition.*
Luke Buquicchio, Walter Gerych, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Thomas Hartvigsen, Elke Rundensteiner, Emmanuel Agu.
IEEE BigData, 2021.
13. *Human-like Explanation for Text Classification with Limited Attention Supervision.*
Dongyu Zhang, Cansu Sen, Jidapa Thadajarassiri, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BigData, 2021.
12. *Recurrent Halting Chain for Early Multi-label Classification.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
KDD, 2020.
11. *Human Attention Maps for Text Classification: Do Humans and Neural Networks Focus on the Same Words?*
Cansu Sen, Thomas Hartvigsen, Biao Yin, Xiangnan Kong, Elke Rundensteiner.
ACL, 2020.
10. *Learning to Selectively Update State Neurons in Recurrent Networks.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
CIKM, 2020.
9. *Learning Similarity-Preserving Word Meta-Embedding.*
Jidapa Thadajarassiri, Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BigData, 2020.
8. *Clinical Performance Evaluation of a Machine Learning System for Predicting Hospital-Acquired Clostridium Difficile Infection.*
Erin Teeple, Thomas Hartvigsen, Cansu Sen, Kajal Claypool, Elke Rundensteiner.
HEALTHINF, 2020. 🏆 **Best Poster**.
7. *Adaptive-Halting Policy Network for Early Classification.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
KDD, 2019.
6. *Patient-Level Classification of Clinical Note Sequences Guided by Attributed Hierarchical Attention.*
Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BigData, 2019.
5. *Learning Temporal Relevance in Longitudinal Medical Notes.*
Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BigData, 2019.
4. *Comparing General and Locally-Learned Word Embeddings for Clinical Text Mining.*
Jidapa Thadajarassiri, Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BHI, 2019.

3. *Detecting MRSA Infections by Fusing Structured and Unstructured Electronic Health Record Data*.
Thomas Hartvigsen, Cansu Sen, Elke Rundensteiner.
BIOSTEC, 2018.
2. *Early Prediction of MRSA Infections using Electronic Health Records*.
Thomas Hartvigsen, Cansu Sen, Sarah Brownell, Erin Teeple, Xiangnan Kong, Elke Rundensteiner.
HEALTHINF, 2018. 🏆 **Best Student Paper runner up**.
1. *CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining*.
Cansu Sen, Thomas Hartvigsen, Kajal Claypool, Elke Rundensteiner.
ECML, 2017.

IN-SUBMISSION

8. *TOXIGEN: Controlling Language Models to Generate Implied and Adversarial Toxicity*.
Thomas Hartvigsen, Saadia Gabriel, Hamid Palangi, Maarten Sap, Dipankar Ray, Ece Kamar.
7. *Continuous-Time Attention Network for Irregularly-Sampled Time Series Classification*.
Thomas Hartvigsen, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.
6. *Learning to Stop Early and Classify Ongoing Irregular Time Series*.
Thomas Hartvigsen, Walter Gerych, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.
5. *Knowledge Amalgamation for Multi-Label Classification via Label Dependency Transfer*.
Jidapa Thadajarassiri, Thomas Hartvigsen, Walter Gerych, Xiangnan Kong, Elke Rundensteiner.
4. *SAIL: Recurrent Classifier Chains with Incomplete Labels*.
Walter Gerych, Thomas Hartvigsen, Emmanuel Agu, Elke Rundensteiner.
3. *SkipSNN: Efficiently Classifying Sparse and Noisy Spike Trains*.
Hang Yin, Xiangnan Kong, Liping Liu, Xin Dai, Thomas Hartvigsen.
2. *Crowd-MIA: A Crowdsourced Dataset for Multi-grained Weakly Supervised Learning*.
Ruofan Hu, Dongyu Zhang, Dandan Tao, Thomas Hartvigsen, Hao Feng, Elke Rundensteiner.
1. *Multi-State Brain Network Discovery*.
Hang Yin, Xinyue Liu, Xiangnan Kong, Thomas Hartvigsen, Yanhua Li.

SUPERVISED UNDERGRADUATE PAPERS

3. *Early Diagnosis Prediction with Recurrent Neural Networks*.
Daniel Johnston[†], Liubou Klindziuk[†], Lolita Nazarov[†], Thomas Hartvigsen, Elke Rundensteiner.
IEEE URTC 2019. 🏆 **Best Paper runner up**.
2. *Handling Missing Values in Multivariate Time Series Classification*.
Julia Friend[†], Alec Hauck[†], Sruthi Kurada[†], Cansu Sen, Thomas Hartvigsen, Elke Rundensteiner.
IEEE URTC 2018.
1. *MRSA Infection Prediction System*.
Sarah Brownell[†], Thomas Hartvigsen, Elke Rundensteiner.
IEEE URTC 2017.

[†]undergraduate co-author

SELECTED TALKS

Harvard University, invited
Adaptive-Halting Policy Networks for Early Classification
Host: Prof. Finale Doshi-Velez

Cambridge, MA
2020

Florida State University, invited
Adaptive-Halting Policy Networks for Early Classification
Host: Prof. Karen Works

Panama, FL
2020

MITRE , invited <i>Adaptive-Halting Policy Networks for Early Classification</i>	Bedford, MA 2020
Computational Sustainability Doctoral Consortium <i>Adaptive-Halting Policy Networks for Early Classification</i>	Virtual Event 2020
University of Minnesota, Institute for Mathematics and its Applications <i>Adaptive-Halting Policy Networks for Early Classification</i>	Minneapolis, MN 2019
Northeastern University, New England Machine Learning Day <i>Adaptive-Halting Policy Networks for Early Classification</i> , poster	Boston, MA 2019
Worcester Polytechnic Institute, Arts and Sciences Week , invited <i>Recurrent Models for Clinical Time Series</i>	Worcester, MA 2019

TEACHING/MENTORING

I have supervised two Masters Theses and eight NSF-funded REU students.

Students Advised:

• Prathyush Parvatharaju, MS, WPI	2019-2021
– Masters Thesis: <i>Learning Saliency Maps to Explain Deep Time Series Classifiers</i>	
• Ramesh Doddaiiah, PhD, WPI	2020-2021
• Aleksa Perucic, MS, WPI	2019-2020
– Masters Thesis: <i>SIFT - A Deep Network for Irregular Multivariate Time Series</i>	
• Liubuo (Yuuna) Klindziuk, BS, Amherst College	2019
• Daniel Johnston, BS, Columbia University	2019
• Lolita Nazarov, BS, StonyBrook University	2019
• Julia Friend, BS, Oberlin College	2018
• Alex Hauck, BS, Valparaiso University	2018
• Sruthi Kurada, Advanced Math & Science Academy Charter School	2018
• Sarah Brownell, BS, Simmons University	2017
• Sean Tocci, BS, UMass Dartmouth	2017

Developed workshop on Deep Learning with PyTorch for Undergrads , WPI.	2019
---	------

SERVICE

Program Committee:

- AAAI ('21, '22)
- CVPR ('21)
- ICCV ('21)
- ACL ('21, '22)
- EMNLP ('21)
- NAACL ('22)

External Reviewer: KDD ('18, '19, '20)

Conference Volunteer: KDD ('19, '20, '21), NeurIPS ('20, '21)

Organized Deep Learning Reading Group , WPI	2019-2020
--	-----------

Graduate Student Advisory Council to the Dean of Arts & Sciences , WPI	2018-2020
---	-----------

Graduate Student Government Senate , WPI	2018
---	------

Data Science Graduate Student Council , WPI	2016-2019
--	-----------