

RESEARCH INTERESTS

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

RESEARCH EXPERIENCE

MIT Postdoctoral Associate at CSAIL with Marzyeh Ghassemi	01/2022 - present
Worcester Polytechnic Institute Research Fellow with Elke Rundensteiner and Xiangnan Kong	08/2016 - 12/2021
Microsoft PhD Intern with Dipankar Ray and Hamid Palangi	05/2021 - 08/2021
UMass Medical School Research Intern with Jomol Matthew	08/2018 - 09/2019
University of Arizona NSF REU Intern with Shirley Papuga	05/2015 - 08/2015

EDUCATION

Worcester Polytechnic Institute , Worcester, MA PhD, Data Science MS, Data Science <i>Advised by Elke Rundensteiner and Xiangnan Kong</i>	08/2016 - 12/2021
SUNY Geneseo , Geneseo, NY BA, Applied Mathematics, minor in Biomathematics	08/2012 - 05/2016

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

🏆 AAAI Oral Spotlight	2022
🏆 Best Poster , International Conference on Health Informatics	2020
🏆 Outstanding Graduate Research Award , WPI	2019
🏆 Best Poster , Graduate Research Innovation and Exchange, WPI	2019
IMA Travel Award , University of Minnesota	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

Summary: KDD (3), AAAI (2), ACL (2), NeurIPS, CIKM (2), SDM, ECML, BigData (5), HEALTHINF (2), BHI.

REFEREED

21. **TOXIGEN: A Large-Scale Machine-Generated Dataset for Adversarial and Implicit Hate Speech Detection.**
Thomas Hartvigsen, Saadia Gabriel, Hamid Palangi, Maarten Sap, Dipankar Ray, Ece Kamar.
ACL, 2022. To appear.

20. *Recovering the Propensity Score from Biased Positive Unlabeled Data.*
Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner.
AAAI, 2022. **Oral Spotlight.**
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. To appear.
18. *Recurrent Bayesian Classifier Chains for Exact Multi-label Classification.*
Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner.
NeurIPS, 2021.
17. *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.
16. *Semi-Supervised Knowledge Amalgamation for Sequence Classification.*
Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
AAAI, 2021.
15. *Learning Saliency Maps to Explain Deep Time Series Classifiers.*
Prathyush Parvatharaju, Ramesh Doddaiiah, **Thomas Hartvigsen**, Elke Rundensteiner.
CIKM, 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, Kaja 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. *Continuous-Time Attention Network for Irregularly-Sampled Time Series Classification.*
Thomas Hartvigsen, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.
7. *Stop&Hop: Early Classification of Irregular Time Series.*
Thomas Hartvigsen, Walter Gerych, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.
6. *Class-Specific Explainability for Deep Time Series Classifiers.*
Ramesh Doddaiiah, Prathyush Parvatharaju, Elke Rundensteiner, **Thomas Hartvigsen**.
5. *Knowledge Amalgamation for Multi-Label Classification via Label Dependency Transfer.*
Jidapa Thadajarassiri, **Thomas Hartvigsen**, Walter Gerych, Xiangnan Kong, Elke Rundensteiner.
4. *The Road to Explainability is Paved with Bias: Measuring the Fairness of Explanations.*
Aparna Balagopalan, Haoran Zhang, Kimia Hamidieh, **Thomas Hartvigsen**, Frank Rudzicz, Marzyeh Ghassemi.
3. *SAIL: Recurrent Classifier Chains with Incomplete Labels.*
Walter Gerych, **Thomas Hartvigsen**, Emmanuel Agu, Elke Rundensteiner.
2. *SkipSNN: Efficiently Classifying Noisy Spike Trains.*
Hang Yin, Xiangnan Kong, Liping Liu, **Thomas Hartvigsen**, Xin Dai.
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

TALKS

Massachusetts Institute of Technology, MIT Horizons	April 2022
Massachusetts Institute of Technology, Prof. Marzyeh Ghassemi's group	November 2021
Harvard University, Prof. Finale Doshi-Velez's group	September 2020
Florida State University, Data Science Seminar	June 2020
MITRE, Data Science Seminar	March 2020
IBM Research, SystemML group	January 2020
University of Minnesota, IMA Workshop	September 2019
Northeastern University, NEML Day	September 2019
Worcester Polytechnic Institute, Arts & Sciences Showcase	May 2019

TEACHING/MENTORING

I have supervised one PhD qualifier, two masters theses, and eight undergraduate students.

Students Advised:

• Prathyush Parvatharaju, MS, WPI (went to CodaMetrix)	2019-2021
– Masters Thesis: <i>Learning Saliency Maps to Explain Deep Time Series Classifiers</i>	
• Ramesh Doddaiiah, PhD, WPI	2020-now
– PhD Qualifier: <i>Class-Specific Explainability for Deep Time Series Classifiers</i>	
• Aleksa Perucic, MS, WPI (went to Amazon)	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, Valporaiso 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

Conference Program Committee:

- ACL Rolling Reviewer (2021-present)
- AAAI ('21, '22)
- CVPR ('21)
- ICCV ('21)
- ACL ('21, '22)
- EMNLP ('21)

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

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

Deep Learning Reading Group, Organizer, 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

REFERENCES

Available upon request