

RESEARCH INTERESTS

Deep Learning, Time Series, Data Mining, Reinforcement Learning, Explainability, Machine learning for Health

EDUCATION

Worcester Polytechnic Institute , Worcester, MA	2016-present
PhD, Data Science	Expected Dec. 2021
MS, Data Science	2018
<i>Advised by Elke Rundensteiner and Xiangnan Kong</i>	
SUNY Geneseo , Geneseo, NY	
BA, Applied Mathematics, minor in Biomathematics	2016

EXPERIENCE

Worcester Polytechnic Institute	2016-present
<i>Research assistant with Elke Rundensteiner</i>	
Machine learning and data mining for time series and text	
Microsoft, Azure Machine Learning	2021
<i>PhD Intern with Dipankar Ray</i>	
Detecting hate speech generated by large language models	
UMass Medical School	2019
<i>Research Intern with Jomol Matthew</i>	
Machine Learning to help doctors write better clinical trials faster	
University of Arizona	2015
<i>NSF REU Intern with Shirley Papuga</i>	
Modeling the effects of drought on creosote bushes in the Sonoran desert via camera-trap images.	

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 (KDD'19 and KDD'20) and I am responsible for 90% of the writing.

SELECTED HONORS & AWARDS

KDD Student Travel Award , NSF and ACM	2020
🏆 Best Poster , International Conference on Health Informatics	2020
IMA Travel Award (\$500) , University of Minnesota	2019
Graduate Student Travel Award (\$1000) , WPI	2019
Outstanding Graduate Research Award , WPI Data Science	2019
🏆 Best Poster (\$500) , Graduate Research Innovation and Exchange, WPI	2019
People's Choice Poster Award , Graduate Research Innovation and Exchange, WPI	2017
GAANN Fellowship (Tuition Award + Annual Stipend) , U.S. Dept. of Education	2016

PUBLICATIONS

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

UNDER REVIEW

30. **Thomas Hartvigsen**, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner. *Continuous-Time Attention Network for Irregularly-Sampled Time Series Classification*.

29. **Thomas Hartvigsen**, Walter Gerych, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner. *Early Classification of Irregular Time Series*.
28. Jidapa Thadajarassiri, **Thomas Hartvigsen**, Walter Gerych, Xiangnan Kong, Elke Rundensteiner. *Knowledge Amalgamation for Multi-Label Classification via Label Dependency Transfer*.
27. Hang Yin, Xiangnan Kong, Liping Liu, Xin Dai, **Thomas Hartvigsen**. *SkipSNN: Efficiently Classifying Sparse and Noisy Spike Trains*.
26. Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Elke Rundensteiner, Emmanuel Agu. *Positive Unlabeled Learning with a Sequential Selection Bias*.
25. Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Elke Rundensteiner. *Recurrent Bayesian Classifier Chains for Exact Multi-label Classification*.
24. Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Elke Rundensteiner. *Recovering The Propensity Score from Biased Positive Unlabeled Data*.
23. Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Elke Rundensteiner, Emmanuel Agu. *Exact Multi-Label Classification for Incompletely-Labeled Data*.
22. Ruofan Hu, Dongyu Zhang, Dandan Tao, **Thomas Hartvigsen**, Hao Feng, Elke Rundensteiner. *Crowd-MIA: A Crowdsourced Dataset for Multi-grained Weakly Supervised Learning*.
21. Dongyu Zhang, Cansu Sen, Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Human-like Explanation for Text Classification with Limited Attention Supervision*.
20. Luke Buquicchio, Walter Gerych, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, **Thomas Hartvigsen**, Elke Rundensteiner. *Variational Open-Set Recognition*.
19. Hang Yin, Xinyue Liu, Xiangnan Kong, **Thomas Hartvigsen**, Yanhua Li. *Multi-State Brain Network Discovery*.

REFEREED

18. Hang Yin, John Boaz Lee, Xiangnan Kong, **Thomas Hartvigsen**, Sihong Xie. *Energy-Efficient Models for High-Dimensional Spike Train Classification using Sparse Spiking Neural Networks*. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2021 (238/1541 = 15.4% acceptance rate).
17. Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Semi-Supervised Knowledge Amalgamation for Sequence Classification*. AAAI Conference on Artificial Intelligence (**AAAI**), 2021 (1692/7911 = 20% acceptance rate).
16. Prathyush Parvatharaju, Ramesh Doddaiiah, **Thomas Hartvigsen**, Elke Rundensteiner. *Learning Saliency Maps to Explain Deep Time Series Classifiers*. ACM International Conference on Information and Knowledge Management (**CIKM**), 2021 (271/1251 = 21% acceptance rate).
15. **Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner. *Recurrent Halting Chain for Early Multi-label Classification*. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2020 (216/1279 = 16.9% acceptance rate).
14. Cansu Sen, **Thomas Hartvigsen**, Biao Yin, Xiangnan Kong, Elke Rundensteiner. *Human Attention Maps for Text Classification: Do Humans and Neural Networks Focus on the Same Words?* Annual Meeting of the Association for Computational Linguistics (**ACL**), 2020 (571/2244 = 17.6% acceptance rate).
13. **Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner. *Learning to Selectively Update State Neurons in Recurrent Networks*. ACM International Conference on Information and Knowledge Management (**CIKM**), 2020 (18% acceptance rate).
12. Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Learning Similarity-Preserving Word Meta-Embedding*. IEEE International Conference on Big Data (**BigData**), 2020 (15.5% acceptance rate).

11. Erin Teeple, **Thomas Hartvigsen**, Cansu Sen, Kajal Claypool, Elke Rundensteiner. *Clinical Performance Evaluation of a Machine Learning System for Predicting Hospital-Acquired Clostridium Difficile Infection*. International Conference on Health Informatics (HEALTHINF), 2020. 🏆 **Best Poster**.
10. **Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner. *Adaptive-Halting Policy Network for Early Classification*. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2019 (170/1200 = 14.2% acceptance rate).
9. Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Patient-Level Classification of Clinical Note Sequences Guided by Attributed Hierarchical Attention*. IEEE International Conference on Big Data (BigData), 2019 (19.3% acceptance rate).
8. Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Learning Temporal Relevance in Longitudinal Medical Notes*. IEEE International Conference on Big Data (BigData), 2019 (19.3% acceptance rate).
7. Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *Comparing General and Locally-Learned Word Embeddings for Clinical Text Mining*. IEEE International Conference on Biomedical and Health Informatics (BHI), 2019.
6. Daniel Johnston, Liubou Klindziuk, Lolita Nazarov, **Thomas Hartvigsen**, Elke Rundensteiner. *Early Diagnosis Prediction with Recurrent Neural Networks*. * IEEE MIT Undergraduate Research Technology Conference (URTC), 2019. 🏆 **Best Paper runner up**.
5. **Thomas Hartvigsen**, Cansu Sen, Elke Rundensteiner. *Detecting MRSA Infections by Fusing Structured and Unstructured Electronic Health Record Data*. International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC), 2018.
4. Julia Friend, Alec Hauck, Sruthi Kurada, Cansu Sen, **Thomas Hartvigsen**, Elke Rundensteiner. *Handling Missing Values in Multivariate Time Series Classification*. * IEEE MIT Undergraduate Research Technology Conference (URTC), 2018.
3. **Thomas Hartvigsen**, Cansu Sen, Sarah Brownell, Erin Teeple, Xiangnan Kong, Elke Rundensteiner. *Early Prediction of MRSA Infections using Electronic Health Records*. International Conference on Health Informatics (HEALTHINF), 2018. 🏆 **Best Student Paper finalist**.
2. Sarah Brownell, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. *MRSA Infection Prediction System*. * IEEE MIT Undergraduate Research Technology Conference (URTC), 2018.
1. Cansu Sen, **Thomas Hartvigsen**, Kajal Claypool, Elke Rundensteiner. *CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining*. European Conference on Machine Learning (ECML), 2017.

*undergraduate paper under my supervision (3 total).

SELECTED TALKS

Harvard University, invited speaker <i>Adaptive-Halting Policy Networks for Early Classification</i> Host: Prof. Finale Doshi-Velez	Cambridge, MA 2020
Florida State University, invited speaker <i>Adaptive-Halting Policy Networks for Early Classification</i> Host: Prof. Karen Works	Panama, FL 2020
The MITRE Corporation, invited speaker <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
Worcester Polytechnic Institute, 3MT Competition <i>Early Classification of Clinical Time Series</i>	Worcester, MA 2020
University of Minnesota, Institute for Mathematics and its Applications <i>Adaptive-Halting Policy Networks for Early Classification</i>	Minneapolis, MN 2019

TEACHING/MENTORING

I have been the primary advisor for two Masters Theses and three NSF-funded REU groups.

Students Advised

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

SERVICE

Program Committee: AAAI ('21, '22), CVPR ('21), ICCV ('21), ACL ('21), EMNLP ('21)	
External Reviewer: KDD ('18, '19, '20)	
Conference Volunteer: KDD ('19, '20, '21), NeurIPS ('20)	
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