

## RESEARCH INTERESTS

---

Machine Learning, Data Mining, Time Series, Explainability, Natural Language Processing, Healthcare

## EDUCATION

---

**Worcester Polytechnic Institute**, Worcester, MA

PhD, Data Science

expected 2021

MS, Data Science

*Advised by Elke Rundensteiner and Xiangnan Kong*

**SUNY Geneseo**, Geneseo, NY

BA, Applied Mathematics, minor in Biomathematics

2016

## EXPERIENCE

---

**Microsoft**

2021

*PhD Intern with Dipankar Ray*

Detecting hate speech generated by large language models

**UMass Medical School**

2019

*Research Intern with Jomol Matthew*

Machine Learning to help doctors write better clinical trials faster

**University of Arizona**

2015

*NSF REU Intern with Shirley Papuga*

Modeling the effects of drought in the Sonoran desert via camera traps

## 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.*

## PUBLICATIONS

---

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

IN SUBMISSION

30. *Continuous-Time Attention Network for Irregularly-Sampled Time Series Classification.*

**Thomas Hartvigsen**, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.

29. *Early Classification of Irregular Time Series.*

**Thomas Hartvigsen**, Walter Gerych, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.

28. *Knowledge Amalgamation for Multi-Label Classification via Label Dependency Transfer.*

Jidapa Thadajarassiri, **Thomas Hartvigsen**, Walter Gerych, Xiangnan Kong, Elke Rundensteiner.

27. *Positive Unlabeled Learning with a Sequential Selection Bias.*

Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Elke Rundensteiner, Emmanuel Agu.

26. *Recovering The Propensity Score from Biased Positive Unlabeled Data.*

Walter Gerych, **Thomas Hartvigsen**, Emmanuel Agu, Elke Rundensteiner.

25. *Exact Multi-Label Classification for Incompletely-Labeled Data.*

Walter Gerych, **Thomas Hartvigsen**, Emmanuel Agu, Elke Rundensteiner.

24. *SkipSNN: Efficiently Classifying Sparse and Noisy Spike Trains.*  
Hang Yin, Xiangnan Kong, Liping Liu, Xin Dai, **Thomas Hartvigsen**.
23. *Crowd-MIA: A Crowdsourced Dataset for Multi-grained Weakly Supervised Learning.*  
Ruofan Hu, Dongyu Zhang, Dandan Tao, **Thomas Hartvigsen**, Hao Feng, Elke Rundensteiner.
22. *Human-like Explanation for Text Classification with Limited Attention Supervision.*  
Dongyu Zhang, Cansu Sen, Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
21. *Variational Open-Set Recognition.*  
Luke Buquicchio, Walter Gerych, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, **Thomas Hartvigsen**, Elke Rundensteiner, Emmanuel Agu.
20. *Multi-State Brain Network Discovery.*  
Hang Yin, Xinyue Liu, Xiangnan Kong, **Thomas Hartvigsen**, Yanhua Li.

#### REFEREED

19. *Recurrent Bayesian Classifier Chains for Exact Multi-label Classification.*  
Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner.  
Advances in Neural Information Processing Systems (**NeurIPS**), 2021.
18. *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.  
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2021 (238/1541 = 15.4% acceptance rate).
17. *Semi-Supervised Knowledge Amalgamation for Sequence Classification.*  
Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.  
AAAI Conference on Artificial Intelligence (**AAAI**), 2021 (1692/7911 = 20% acceptance rate).
16. *Learning Saliency Maps to Explain Deep Time Series Classifiers.*  
Prathyush Parvatharaju, Ramesh Doddaiiah, **Thomas Hartvigsen**, Elke Rundensteiner.  
ACM International Conference on Information and Knowledge Management (**CIKM**), 2021 (271/1251 = 21% acceptance rate).
15. *Recurrent Halting Chain for Early Multi-label Classification.*  
**Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.  
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2020 (216/1279 = 16.9% acceptance rate).
14. *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.  
Annual Meeting of the Association for Computational Linguistics (**ACL**), 2020 (571/2244 = 17.6% acceptance rate).
13. *Learning to Selectively Update State Neurons in Recurrent Networks.*  
**Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.  
ACM International Conference on Information and Knowledge Management (**CIKM**), 2020 (18% acceptance rate).
12. *Learning Similarity-Preserving Word Meta-Embedding.*  
Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.  
IEEE International Conference on Big Data (**BigData**), 2020 (15.5% acceptance rate).
11. *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.  
International Conference on Health Informatics (**HEALTHINF**), 2020. 🏆 **Best Poster**.

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

<sup>†</sup>undergraduate co-author (9 total).

<sup>\*</sup>undergraduate paper under my supervision (3 total).

## SELECTED TALKS

<b>Harvard University</b> , invited <i>Adaptive-Halting Policy Networks for Early Classification</i> Host: Prof. Finale Doshi-Velez	Cambridge, MA 2020
<b>Florida State University</b> , invited <i>Adaptive-Halting Policy Networks for Early Classification</i> Host: Prof. Karen Works	Panama, FL 2020
<b>The MITRE Corporation</b> , invited <i>Adaptive-Halting Policy Networks for Early Classification</i>	Bedford, MA 2020
<b>Computational Sustainability Doctoral Consortium</b> <i>Adaptive-Halting Policy Networks for Early Classification</i>	Virtual Event 2020
<b>Worcester Polytechnic Institute, 3MT Competition</b> <i>Early Classification of Clinical Time Series</i>	Worcester, MA 2020

University of Minnesota, Institute for Mathematics and its Applications  
*Adaptive-Halting Policy Networks for Early Classification*

Minneapolis, MN  
2019

Northeastern University, New England Machine Learning Day  
*Adaptive-Halting Policy Networks for Early Classification*, poster

Boston, MA  
2019

Worcester Polytechnic Institute, Arts and Sciences Week, invited  
*Recurrent Models for Clinical Time Series*

Worcester, MA  
2019

## SELECTED HONORS & AWARDS

---

CIKM Student Travel Award, ACM	2020
KDD Student Travel Award, NSF and ACM	2020
🏆 Best Poster, International Conference on Health Informatics	2020
Graduate Student Travel Award (\$1000), WPI	2020
IMA Travel Award (\$500), University of Minnesota	2019
KDD Student Travel Award (\$500), NSF and ACM	2019
Graduate Student Travel Award (\$1000), WPI	2019
🏆 Outstanding Graduate Research Award, WPI	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

## TEACHING/MENTORING

---

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

### Students Advised

• Prathyush Parvatharaju (MS Thesis), WPI	2019-Now
– <b>Masters Thesis:</b> <i>Learning Saliency Maps to Explain Deep Time Series Classifiers</i>	
• Ramesh Doddaiiah (PhD student), WPI	2020-Now
• Aleksa Perucic (MS Thesis), WPI	2020
– <b>Masters Thesis:</b> <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), Valporaiso 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
<b>Developed and led workshop on Deep Learning with PyTorch for Undergrads, WPI.</b>	2019

## SERVICE

---

<b>Program Committee:</b> AAAI ('21, '22), CVPR ('21), ICCV ('21), ACL ('21), EMNLP ('21)	
<b>External Reviewer:</b> KDD ('18, '19, '20)	
<b>Conference Volunteer:</b> KDD ('19, '20, '21), NeurIPS ('20)	
<b>Organized Deep Learning Reading Group, WPI</b>	2019-2020
<b>Graduate Student Advisory Council to the Dean of Arts &amp; Sciences, WPI</b>	2018-2020
<b>Graduate Student Government Senate, WPI</b>	2018
<b>Data Science Graduate Student Council, WPI</b>	2016-2019