

## RESEARCH INTERESTS

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

## RESEARCH EXPERIENCE

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

## EDUCATION

<b>Worcester Polytechnic Institute</b> , Worcester, MA PhD, Data Science MS, Data Science <i>Advised by Elke Rundensteiner and Xiangnan Kong</i>	08/2016 - 12/2021
<b>SUNY Geneseo</b> , 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

🏆 <b>Best Poster</b> , International Conference on Health Informatics	2020
🏆 <b>Outstanding Graduate Research Award</b> , WPI	2019
🏆 <b>Best Poster</b> , Graduate Research Innovation and Exchange, WPI	2019
<b>IMA Travel Award</b> , University of Minnesota	2019
🏆 <b>People's Choice Poster Award</b> , Graduate Research Innovation and Exchange, WPI	2017
<b>GAANN Fellowship (Annual Tuition + Stipend Award)</b> , 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: Controlling Language Models to Generate Implied and Adversarial Toxicity.**  
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<sup>†</sup>, 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<sup>†</sup>, Liubou Klindziuk<sup>†</sup>, Lolita Nazarov<sup>†</sup>, **Thomas Hartvigsen**, Elke Rundensteiner.  
IEEE URTC 2019. 🏆 Best Paper runner up.
2. *Handling Missing Values in Multivariate Time Series Classification.*  
Julia Friend<sup>†</sup>, Alec Hauck<sup>†</sup>, Sruthi Kurada<sup>†</sup>, Cansu Sen, **Thomas Hartvigsen**, Elke Rundensteiner.  
IEEE URTC 2018.
1. *MRSA Infection Prediction System.*  
Sarah Brownell<sup>†</sup>, **Thomas Hartvigsen**, Elke Rundensteiner.  
IEEE URTC 2017.

<sup>†</sup>undergraduate co-author

## TALKS

---

<b>Timely and Trustworthy Machine Learning</b> MIT, HealthyML group, invited	11/2021
<b>Adaptive-Halting Policy Networks for Early Classification</b> Computational Sustainability Doctoral Consortium	10/2020
Harvard University, DtAK group, invited	09/2020
Florida State University Data Science Seminar, invited	06/2020
MITRE Data Science, invited	03/2020
IBM Research, SystemML group, invited	01/2020
University of Minnesota, IMA	09/2019
Northeastern University, NEML Day poster	09/2019
ACM SIGKDD	08/2019
<b>Learning to Selectively Update State Neurons in Recurrent Networks</b> ACM CIKM	10/2020
<b>Recurrent Halting Chains for Early Multi-label Classification</b> ACM SIGKDD	08/2020
<b>Recurrent Models for Clinical Time Series</b> Worcester Polytechnic Institute, Arts & Sciences Showcase, invited	05/2019

## TEACHING/MENTORING

---

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

### Students Advised:

• Prathyush Parvatharaju, MS, WPI (went to CodaMetrix) – <b>Masters Thesis:</b> <i>Learning Saliency Maps to Explain Deep Time Series Classifiers</i>	2019-2021
• Ramesh Doddaiiah, PhD, WPI – <b>PhD Qualifier:</b> <i>Class-Specific Explainability for Deep Time Series Classifiers</i>	2020-now
• Aleksa Perucic, MS, WPI (went to Amazon) – <b>Masters Thesis:</b> <i>SIFT - A Deep Network for Irregular Multivariate Time Series</i>	2019-2020
• 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
<b>Developed workshop on Deep Learning with PyTorch for Undergrads, WPI.</b>	2019

## SERVICE

---

### Conference Program Committee:

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

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

**REFERENCES**

---

Available upon request