Thomas Hartvigsen
Cambridge, Massachusetts
https://thartvigsen.github.io
twhartvigsen@wpi.edu

Research Interests

Machine Learning, Data Mining, Time Series, Explainability, NLP, Machine Learning for Healthcare

Education

Worcester Polytechnic Institute, *Worcester, MA*
PhD, Data Science 2021[1]
MS, Data Science 2018
Advised by Elke Rundensteiner and Xiangnan Kong

SUNY Geneseo, *Geneseo, NY*
BA, Applied Mathematics, minor in Biomathematics 2016

Research Experience

Worcester Polytechnic Institute 2016-present
Research Assistant with Elke Rundensteiner and Xiangnan Kong
Machine learning and data mining for time series and text

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, NeurIPS, AAAI, ACL, CIKM, ECML, IEEE BigData, HEALTHINF, and IEEE BHI.

Refereed

1. *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 (26% acceptance rate).

2. *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).

3. *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).

[1] expected in December

4. *Learning Saliency Maps to Explain Deep Time Series Classifiers*.
   Prathyush Parvatharaju, Ramesh Doddaiah, **Thomas Hartvigsen**, Elke Rundensteiner.
   ACM International Conference on Information and Knowledge Management (**CIKM**), 2021 (271/1251 = 21% acceptance rate).

5. *Variational Open-Set Recognition*.
   Luke Buquicchio, Walter Gerych, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, **Thomas Hartvigsen**, Elke Rundensteiner, Emmanuel Agu. IEEE International Conference on Big Data (**BigData**), 2021 (19.8% acceptance rate).

6. *Human-like Explanation for Text Classification with Limited Attention Supervision*.
   Dongyu Zhang, Cansu Sen, Jidapa Thadajarassiri, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner. IEEE International Conference on Big Data (**BigData**), 2021 (19.8% acceptance rate).

7. *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).

8. *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).

9. *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).

10. *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**.

12. *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).

13. *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).

14. *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).

15. *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.

16. *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.

17. *Early Prediction of MRSA Infections using Electronic Health Records*.
    Thomas Hartvigsen, Cansu Sen, Sarah Brownell, Erin Teeple, Xiangnan Kong, Elke Rundensteiner.

International Conference on Health Informatics (**HEALTHINF**), 2018. 🏆 **Best Student Paper runner up**.

18. *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.

IN-SUBMISSION

19. *Continuous-Time Attention Network for Irregularly-Sampled Time Series Classification*.
Thomas Hartvigsen, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.
20. *Early Classification of Irregular Time Series*.
Thomas Hartvigsen, Walter Gerych, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner.
21. *Knowledge Amalgamation for Multi-Label Classification via Label Dependency Transfer*.
Jidapa Thadajarassiri, **Thomas Hartvigsen**, Walter Gerych, Xiangnan Kong, Elke Rundensteiner.
22. *Recovering The Propensity Score from Biased Positive Unlabeled Data*.
Walter Gerych, **Thomas Hartvigsen**, Emmanuel Agu, Elke Rundensteiner.
23. *SAIL: Recurrent Classifier Chains with Incomplete Labels*.
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**.

SELECTED TALKS

Harvard University, invited — Cambridge, MA
Adaptive-Halting Policy Networks for Early Classification — 2020
Host: Prof. Finale Doshi-Velez

Florida State University, invited — Panama, FL
Adaptive-Halting Policy Networks for Early Classification — 2020
Host: Prof. Karen Works

The MITRE Corporation, invited — Bedford, MA
Adaptive-Halting Policy Networks for Early Classification — 2020

Computational Sustainability Doctoral Consortium — Virtual Event
Adaptive-Halting Policy Networks for Early Classification — 2020

Worcester Polytechnic Institute, 3MT Competition — Worcester, MA
Early Classification of Clinical Time Series — 2020

University of Minnesota, Institute for Mathematics and its Applications — Minneapolis, MN
Adaptive-Halting Policy Networks for Early Classification — 2019

Northeastern University, New England Machine Learning Day — Boston, MA
Adaptive-Halting Policy Networks for Early Classification, poster — 2019

Worcester Polytechnic Institute, Arts and Sciences Week, invited — Worcester, MA
Recurrent Models for Clinical Time Series — 2019

SELECTED HONORS & AWARDS

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

Graduate Student Travel Award ($1000), WPI 2019
🏆 **People's Choice Poster Award**, Graduate Research Innovation and Exchange, WPI 2017
GAANN PhD Fellowship (Tuition Award + Annual Stipend), U.S. Dept. of Education 2016-2021

TEACHING/MENTORING

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

Students Advised

- Prathyush Parvatharaju (MS Thesis), WPI 2019-2021
  - **Masters Thesis**: *Learning Saliency Maps to Explain Deep Time Series Classifiers*
- Ramesh Doddaiah (PhD student), WPI 2020-2021
- Aleksa Perucic (MS Thesis), WPI 2019-2020
  - **Masters Thesis**: *SIFT - A Deep Network for Irregular Multivariate Time Series*
- 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

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