

Tom Hartvigsen

twhartvigsen@wpi.edu

[thartvigsen.github.io](https://github.com/thartvigsen)

INTERESTS: Data Mining, Recurrent Neural Networks, Time Series, Reinforcement Learning, Interpretability.

EDUCATION

Worcester Polytechnic Institute, Worcester, MA

Ph.D., Data Science

Expected Dec. 2021

Advisors: Prof. Elke Rundensteiner, Prof. Xiangnan Kong

Dissertation Title: *Observation and Prediction Timing in Time Series Classification.*

Committee Members: Prof. Randy Paffenroth (WPI), Prof. Jenna Wiens (U. Mich.)

M.S., Data Science

2018

Advisors: Prof. Elke Rundensteiner, Prof. Xiangnan Kong

Thesis: *Adaptive-Halting Policy Network for Early Classification.*

SUNY Geneseo, Geneseo, NY

B.A., Applied Mathematics

May 2016

BioMathematics minor

Advisors: Prof. Chris Leary, Prof. Kirk Anne

EXPERIENCE

PhD Research Fellow, Worcester Polytechnic Institute

Aug 2016 - present

Solving time series classification problems for time-sensitive domains. I also supervised two MS theses and three summers of NSF-funded undergrad research.

Advisors: Prof. Elke Rundensteiner, Prof. Xiangnan Kong

Research Intern (Machine Learning - NLP), UMass Medical School

Sep 2018 - Aug 2019

Developed an automatic text summarization tool for clinical trial eligibility criteria to be used in recommendation for new clinical trials.

Supervisor: Dr. Jomol Matthew

NSF REU Intern, University of Arizona, Department of Environmental Science

Summer 2015

Built an instance-segmentation model for remotely-captured images of creosote bushes to model the effects of drought over time in the sonoran desert.

Supervisor: Prof. Shirley Papuga

Teaching Assistant, SUNY Geneseo

2015 - 2016

Teaching assistant for Modeling Biological Systems twice and BioStatistics once. I also developed and led one 2-hour lecture/in-class exercise in R.

Supervisors: Prof. Chris Leary, Prof. Gregg Hartvigsen

Research Assistant, SUNY Geneseo

2013 - 2016

Modeled infection spread on graphs, built a graph dataset from IMDB, mined song lyrics for text features useful for discriminating genres and artists.

Supervisors: Prof. Chris Leary, Dr. Kirk Anne

PUBLICATIONS

MANUSCRIPTS

1. *Continuous-Time Attention Policy Network for Irregularly-Sampled Time Series.*
Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
In submission to KDD 2021.

2. *Learning Saliency Maps to Explain Deep Time Series Classifiers.*
Prathyush Parvatharaju, Ramesh Doddiah, Elke Rundensteiner, Thomas Hartvigsen.
In submission to KDD 2021.
3. *Recurrent Bayesian Classifier Chains for Multi-label Classification.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Elke Rundensteiner.
In submission to KDD 2021.
4. *Semi-Supervised Attention for Explainable Text Classification.*
Jidapa Thadajarassiri, Dongyu Zhang, Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
In submission to KDD 2021.
5. *Deep Positive Unlabeled Learning with a Sequential Bias.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Elke Rundensteiner, Emmanuel Agu.
In submission to KDD 2021.
6. *Exact Multi-label Classification with Incompletely-Labeled Data.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Elke Rundensteiner, Emmanuel Agu.
In submission to KDD 2021.
7. *Variational Open-Set Recognition.*
Luke Buquicchio, Walter Gerych, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Thomas Hartvigsen, Elke Rundensteiner.

PEER-REVIEWED

16. *Semi-Supervised Knowledge Amalgamation for Sequence Classification.*
Jidapa Thadajarassiri, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
AAAI 2021.
15. *Recurrent Halting Chain for Early Multi-label Classification.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
ACM SIGKDD 2020, research track.
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.
ACL 2020, long paper.
13. *Learning to Selectively Update State Neurons in Recurrent Networks.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
CIKM 2020, long paper.
12. *Learning Similarity-Preserving Word Meta-Embedding.*
Jidapa Thadajarassiri, Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BigData 2020, long paper.
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.
HEALTHINF 2020, long paper, best poster award.
10. *Adaptive-Halting Policy Network for Early Classification.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
ACM SIGKDD 2019, research track.
9. *Patient-Level Classification of Clinical Note Sequences Guided by Attributed Hierarchical Attention.*
Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BigData 2019.

8. *Learning Temporal Relevance in Longitudinal Medical Notes.*
Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BigData 2019.
7. *Comparing General and Locally-Learned Word Embeddings for Clinical Text Mining.*
Jidapa Thadajarassiri, Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE BHI 2019.
6. *Early Diagnosis Prediction with Recurrent Neural Networks.*
Daniel Johnston[†], Liubou Klindziuk[†], Lolita Nazarov[†], Thomas Hartvigsen, Elke Rundensteiner.
IEEE 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.
BIOSTEC 2018.
4. *Handling Missing Values in Multivariate Time Series Classification.*
Julia Friend[†], Alec Hauck[†], Sruthi Kurada[†], Cansu Sen, Thomas Hartvigsen, Elke Rundensteiner.
IEEE URTC 2018.
3. *Early Prediction of MRSA Infections using Electronic Health Records.*
Thomas Hartvigsen, Cansu Sen, Sarah Brownell[†], Erin Teeple, Xiangnan Kong, Elke Rundensteiner.
HEALTHINF 2018.
2. *MRSA Infection Prediction System.*
Sarah Brownell[†], Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner.
IEEE URTC 2017.
1. *CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining.*
Cansu Sen, Thomas Hartvigsen, Kajal Claypool, Elke Rundensteiner.
ECML 2017.

[†]undergraduate advisee.

HONORS AND AWARDS

CIKM Student Travel Grant, ACM	2020
KDD Student Travel Grant, NSF and ACM	2020
Best Poster, HEALTHINF	2020
Graduate Student Travel Grant, WPI	2020
IMA Travel Grant, University of Minnesota	2019
KDD Student Travel Grant, NSF and ACM	2019
Graduate Student Travel Grant, WPI	2019
Best Poster, Graduate Research Innovation and Exchange, WPI	2019
People's Choice Poster Award, Graduate Research Innovation and Exchange, WPI	2018
Graduate Student Travel Grant, WPI	2018
People's Choice Poster Award, Graduate Research Innovation and Exchange, WPI	2017
Graduate Student Travel Grant, WPI	2017
GAANN Ph.D. Fellowship, U.S. Department of Education	2016

PRESENTATIONS AND INVITED TALKS

Computational Sustainability Doctoral Consortium	Virtual Event
<i>Adaptive-Halting Policy Networks for Early Classification</i>	October 2020
Harvard University, Data to Actionable Knowledge Group	Cambridge, MA
<i>Adaptive-Halting Policy Networks for Early Classification</i>	September 2020
Florida State University	Panama, FL
<i>Adaptive-Halting Policy Networks for Early Classification</i>	June 2020

MITRE, Data Science Group <i>Adaptive-Halting Policy Networks for Early Classification</i>	Bedford, MA March 2020
Worcester Polytechnic Institute, Data Science Department Colloquium <i>Selective Activation in Recurrent Neural Networks</i>	Worcester, MA November 2019
University of Minnesota, Institute for Mathematics and its Applications <i>Adaptive-Halting Policy Networks for Early Classification</i>	Minneapolis, MN September 2019
Worcester Polytechnic Institute, NSF REU Tutorial <i>Introduction to PyTorch and Deep Learning</i>	Worcester, MA July 2019
Northeastern University, New England Machine Learning Day <i>Adaptive-Halting Policy Networks for Early Classification</i>	Boston, MA May 2019
Worcester Polytechnic Institute, Arts and Sciences Week <i>Recurrent Models for Clinical Time Series</i>	Worcester, MA May 2019
Worcester Polytechnic Institute, Graduate Research Innovation & Exchange <i>Partial Recurrent State Updates for Irregular Multivariate Time Series</i>	Worcester, MA March 2019
Worcester Polytechnic Institute, Graduate Research Innovation & Exchange <i>Adaptively-Halting RNN for Tunable Earliness in Multivariate Time Series Classification</i>	Worcester, MA March 2018
Worcester Polytechnic Institute, Graduate Research Innovation & Exchange <i>CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining</i>	Worcester, MA March 2017
SUNY Geneseo, Modeling Biological Systems class <i>An Introduction to Percolation Modeling</i>	Geneseo, NY April 2016

TEACHING

Students Advised

• Aleksa Perucic, MS Thesis, Worcester Polytechnic Institute – Thesis Title: <i>SIFT - A Deep Network for Irregular Multivariate Time Series</i>	2019-2020
• Liubuo Klindziuk, BS, Amherst College, NSF REU	Summer 2019
• Daniel Johnston, BS, Columbia University, NSF REU	Summer 2019
• Lolita Nazarov, BS, StonyBrook University, NSF REU	Summer 2019
• Julia Friend, BS, Oberlin College, NSF REU	Summer 2018
• Alex Hauck, BS, Valparaiso University, NSF REU	Summer 2018
• Sruthi Kurada, Advanced Math & Science Academy Charter School, NSF REU	Summer 2018
• Sarah Brownell, BS, Simmons University, NSF REU	Summer 2017
• Sean Tocci, BS, UMass Dartmouth, NSF REU	Summer 2017

Teaching Assistant , SUNY Geneseo, Modeling Biological Systems (2x) and BioStats (1x).	2015-2016
Modeling Biological Systems , SUNY Geneseo	2016
<i>Guest lecturer:</i> taught Percolation Models, including an in-class exercise in R.	

SERVICE

Program Committee: AAAI ('21), CVPR ('21), ICCV ('21)	
External Reviewer: KDD ('19, '20)	
Organized/led the Deep Learning Reading Group at WPI	2019-2020
Graduate Student Council of Arts & Sciences, WPI	2018-2020
Graduate Student Government Senate, WPI	2018
Data Science Graduate Student Council, WPI	2016-2019

TECHNICAL SKILLS

Programming: Python, R, L^AT_EX, SQL.
Frameworks: PyTorch, TensorFlow, Scikit-learn, NumPy.