

# Tom Hartvigsen

twhartvigsen@wpi.edu

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

INTERESTS: Deep Learning, Recurrent Neural Nets, Interpretability, Time Series, Reinforcement Learning.

## EDUCATION

---

**Worcester Polytechnic Institute**, *Worcester, MA*

Ph.D., Data Science

2021

Advisors: Elke Rundensteiner, Xiangnan Kong

**SUNY Geneseo**, *Geneseo, NY*

B.A., Applied Mathematics

2016

BioMathematics minor

Advisors: Chris Leary, Kirk Anne

## EXPERIENCE

---

**Graduate Research Fellow**, *Worcester Polytechnic Institute*

Aug 2016 - 2021

Studying and developing recurrent models for a variety of challenging sequence classification tasks.

Supervisors: Prof. Elke Rundensteiner, Prof. Xiangnan Kong

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

Sep 2018 - Aug 2019

Worked on auto-summarization of clinical trial eligibility criteria for recommendation in new 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 arid environments.

Supervisor: Prof. Shirley Papuga

**Research Assistant**, *SUNY Geneseo*

2014 - 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

---

### IN-SUBMISSION

1. *Learning to Selectively Update State Neurons in Recurrent Networks.*  
**Thomas Hartvigsen**, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
2. *Deep Biased Positive Unlabeled Learning of Sequential Data.*  
Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Elke Rundensteiner.
3. *Explainable Document Classification with Human-guided Attention.*  
Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
4. *Learning Similarity-Preserving Word Meta-Embedding.*  
Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.

### PEER-REVIEWED PUBLICATIONS

1. *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.

2. *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.  
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2020.
3. *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 award.**
4. *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.
5. *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.
6. *Learning Temporal Relevance in Longitudinal Medical Notes.*  
Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.  
IEEE International Conference on Big Data (**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 International Conference on Biomedical and Health Informatics (**BHI**), 2019.
8. *Early Diagnosis Prediction with Recurrent Neural Networks.*  
Daniel Johnston\*, Liubou Klindziuk\*, Lolita Nazarov\*, **Thomas Hartvigsen**, Elke Rundensteiner.  
IEEE MIT Undergraduate Research Technology Conference (**URTC**), 2019. **Best paper runner up.**
9. *Detecting MRSA Infections by Fusing Structured and Unstructured Electronic Health Record Data.*  
**Thomas Hartvigsen**, Cansu Sen, Elke Rundensteiner.  
Communications in Computer and Information Science (**CCIS**) 1024, 2018.
10. *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. Short-listed for Best Student Paper.
11. *Handling Missing Values in Multivariate Time Series Classification.*  
Julia Friend\*, Alec Hauck\*, Sruthi Kurada\*, Cansu Sen, **Thomas Hartvigsen**, Elke Rundensteiner.  
IEEE MIT Undergraduate Research Technology Conference (**URTC**), 2018.
12. *CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining.*  
Cansu Sen, **Thomas Hartvigsen**, Kajal Claypool, Elke Rundensteiner.  
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (**ECML**), 2017.

\*Student under my supervision.

## HONORS AND AWARDS

---

Best Poster, HEALTHINF	2020
Graduate Student Travel Grant, WPI	2020
IMA Travel Grant, Institute for Mathematics and its Applications, U. of Minn.	2019
KDD Student Travel Grant, NSF/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  
GAANN Ph.D. Fellowship, U.S. Department of Education

2017  
2016-2021

## INVITED TALKS

---

<b>Adaptive-Halting Policy Networks for Early Classification</b>	
MITRE, Bedford, MA	2020
University of Minnesota, Minneapolis MN	2019
New England Machine Learning Day Poster, Northeastern University, Boston MA	2019
<b>Recurrent Models for Clinical Time Series</b>	2019
Lightning Talk at WPI Arts & Sciences Week	
<b>Selective Activation in Recurrent Neural Networks</b>	2019
Data Science Research Group Colloquium, WPI	
<b>Early Prediction of MRSA Infections Using Electronic Health Records</b>	2018
HEALTHINF conference, Funchal, Madeira Island, Portugal	

## OTHER TALKS

---

<b>Partial Recurrent State Updates for Irregular Multivariate Time Series</b>	2019
Graduate Research Innovation and Exchange Poster Session, WPI	
<b>Introduction to PyTorch and Deep Learning</b>	2019
PyTorch tutorial given to NSF-funded summer REU students with Walter Gerych, WPI	
<b>JokeR: A Recurrent Joke Generator</b>	2018
<i>Deep Learning</i> final project with Thanh Tran and Sanket Gujar, WPI	
<b>Adaptively-Halting RNN for Tunable Earliness in Multivariate Time Series Classification</b>	2018
Graduate Research Innovation and Exchange Poster Session, WPI	
<b>CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining</b>	2017
Graduate Research Innovation and Exchange Poster Session, WPI	

## TEACHING

---

<b>NSF REU Project Advisor, WPI.</b>	Summers of 2017-19
<i>Students:</i> L. Klindziuk, D. Johnston, L. Nazarov, J. Friend, A. Hauck, S. Kurada, S. Brownell, S. Tocci.	
<i>Outcomes:</i> One paper per summer from 2017, 2018, and 2019.	
<b>Teaching Assistant, SUNY Geneseo, Modeling Biological Systems (2x) and BioStats (1x).</b>	2015-2016
<b>Modeling Biological Systems, SUNY Geneseo</b>	2016
<i>Guest lecturer:</i> taught Percolation Models, including an in-class exercise in R.	

## SERVICE

---

<b>Graduate Student Council of Arts &amp; Sciences, WPI</b>	2018-2020
<b>Data Science Graduate Student Council, WPI</b>	2016-2019

## TECHNICAL SKILLS

---

**Programming:** Python, R,  $\text{\LaTeX}$ , SQL.  
**Frameworks:** PyTorch, TensorFlow, Scikit-learn, NumPy.