Walter Gerych
Framingham, MA wgerych@mit.edu

RESEARCH INTERESTS

Machine Learning, Deep Learning, Semi-Supervised Learning, Correcting Biased Data, Generative Modeling, Human Activity Recognition, Machine Learning for Healthcare

EDUCATION

Worcester Polytechnic Institute, *Worcester, MA* 2017 - 2023
PhD, Data Science 2023
MS, Data Science; GPA: 4.0/4.0 2019

SUNY Geneseo, *Geneseo, NY* 2013 - 2017
BS, Mathematics
Edgar Fellows Honors Program

RESEARCH EXPERIENCE

MIT CSAIL, Postdoctoral Associate 10/2023 – Current

- Working under Prof. Ghassemi in CSAIL's HealthyML group.
- Researching methods to debias pretrained models, metric learning for diabetes subtyping, modeling uncertainty in survivorship analysis, and on verifying generative models

Worcester Polytechnic Institute, Research Assistant 06/2018 – 10/2023
Advisors: Elke Rundensteiner, Emmanuel Agu

- Working on DARPA WASH project
  - Team of five students, two professors, and four medical collaborators
  - Developed methods to detect physical activity using mobile sensors
  - Created deep-learning classifiers to detect symptoms of influenza from sensor data
  - Proposed generative model to synthesize user-specific activity data
  - Developed generative methods for activity data augmentation
  - Developed method for detecting depression from changes in activity data

- Developing methods for stabilizing GAN training
- Proposed first study of and algorithms for identifiable biased positive-unlabeled learning
- Created Bayesian Network-based recurrent approach for multi-label learning
- Developed first deep-learning formulation of Maximum Variance Unfolding
- Proposed method to perform classification from biased sequential partially labeled data
- Developed deep anomaly detection technique based on Wasserstein autoencoders
- Collaborated on the following research projects:
  - Knowledge amalgamation for multi-label and multi-task learning
  - Mathematically rigorous theoretical comparison of fair-ranking metrics
  - Conditional generative models for text

- Early classification of irregularly sampled time series
- Few-shot learning for human activity data
- Open-set classification using likelihood-based generative models
- Visualization tools for human activity data and health

IBM Research, Research Scientist Intern 05/2023 –08/2023

Mentor: Yara Rizk, Manager: Vatche Isahagian

- Conducted research into measuring confidence of large language models (LLMs)
- Developed unsupervised system to select the optimal LLM and prompt template for each user query
- Submitted paper at AAAI 2024 on research carried out during internship *(accepted for publication)*
- Submitted patent disclosure on system developed during internship
  - *Title: "A Method For Selecting Foundation Model-Prompt Pairs Using Estimated Confidences"*
  - *Inventors: Yara Rizk, Walter Gerych, Vinod Muthusamy, Vatche Isahagian, Praveen Venkateswaran, Evelyn Duesterwald*

MIT Lincoln Lab, Intern 05/2022 –10/2023

Advisor: John Moores

- Developed deep models to forecast optical turbulence timeseries
- Created novel multi-stream Seq2Seq network to incorporate asynchronous exogenous variable streams for optical turbulence prediction
- Designed next-frame prediction network for optical turbulence phase screen prediction
- Performed robust analysis on scintillation for multiple wavelengths of real-world optical link

Kansas State University, Research Intern 05/2015 - 08/2015

PI: Jeremy LeCrone

- Developed cellular automata method for modeling mean curvature flow

SELECTED HONORS & AWARDS

DARPA Riser, *DARPA* 2022
DARPA award for early-career scientists identified as up-and-coming in their field

Data Science Leadership Award, *WPI,* 2021, 2020

Best Poster, *Graduate Research Innovation and Exchange, WPI* 2022, 2021

Best Poster Finalist, *Graduate Research Innovation and Exchange, WPI* 2020

V. Ambujamma Memorial Scholarship, *SUNY Geneseo* 2016
For outstanding achievement as a student of mathematics

PUBLICATIONS

I have first-author publications in NeurIPS, AAAI, SDM, CIKM, AIES, Big Data, ICMLA, and ICSC.

1. ***Who Knows the Answer? Finding the Best Model & Prompt Using Confidence-Based Search.***
   Walter Gerych, Yara Rizk, Vatche Isahagian, Vinod Muthusamy, Evelyn Duesterwald, Praveen Venkateswaran. **AAAI,** 2024.

2. *Amalgamating Multi-Task Models with Heteogeneous Architectures.*
Jidapa Thadajarassiri, **Walter Gerych**, Xiangnan Kong, Elke Rundensteiner. **AAAI,** 2024.

3. *Debiasing Pretrained Generative Models By Uniformly Sampling Semantic Attributes.*
Walter Gerych, Kevin Hickey, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Emmanuel Agu, Elke Rundensteiner. **NeurIPS**, 2023

4. *Stabilizing Adversarial Training for Generative Networks.*
Walter Gerych*, Kevin Hickey* (Joint First Author), Thomas Hartvigsen, Luke Buquicchio, Abdulaziz Alajaji, Kavin Chandrasekaran, Hamid Mansoor, Emmanuel Agu, Elke Rundensteiner. **IEEE Big Data MLDB**, 2023.

5. *Population-Level Visual Analytics of Smartphone Sensed Health Using Community Phenotypes.*
Hamid Mansoor, **Walter Gerych**, Abdulaziz Alajaji, Luke Buquicchio, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **IEEE ICHI,** 2023.

6. *Adversarial Human Context Recognition: Evasion Attacks and Defenses.*
Abdulaziz Alajaji, **Walter Gerych**, Kavin Chandrasekaran, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. **IEEE COMPSAC**, 2023.

7. *Knowledge Amalgamation for Multi-Label Classification via Label Dependency Transfer.*
Jidapa Thadajarassiri, Thomas Hartvigsen, **Walter Gerych**, Xiangnan Kong, Elke Rundensteiner. **AAAI**, 2023. **Oral spotlight.**

8. *Domain Adaptation Methods for Lab-to-Field Human Context Recognition.*
Abdulaziz Alajaji, **Walter Gerych**, Luke Buquicchio, Kavin Chandrasekaran, Hamid Mansoor, Emmanuel Agu, Elke Rundensteiner. **Sensors 23(6)**, 2023.

9. *INPHOVIS: Interactive Visual Analytics For Smartphone-Based Digital Phenotyping.*
Hamid Mansoor, **Walter Gerych**, Abdulaziz Alajaji, Luke Buquicchio, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner, Angela Incollingo Rodriguez. **Visual Informatics**, 2023.

10. *HAR-CTGAN: A Mobile Sensor Data Generation Tool for Human Activity Recognition.*
Joshua DeOliveira, **Walter Gerych**, Aruzhan Koshkarova, Elke Rundensteiner, Emmanuel Agu. **IEEE Big Data 4th Special Session on HealthCare Data**, 2022.

11. *Text Generation to Aid Depression Detection: A Comparative Study of Conditional Sequence GANs*
ML Tlachac, **Walter Gerych**, Kratika Agrawal, Benjamin Litterer, Nicholas Jurovich, Saitheeraj Thatigotla, Jidapa Thadajarassiri, Elke Rundensteiner. **IEEE Big Data 4th Special Session on HealthCare Data**, 2022.

12. *Positive Unlabeled Learning with a Sequential Selection Bias.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. **SDM**, 2022.

13. *Robust Recurrent Classifier Chains for Multi-Label Learning with Missing Labels.*
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. **CIKM,** 2022.

14. *Stop&Hop: Early Classification of Irregular Time Series*

Thomas Hartvigsen, **Walter Gerych**, Jidapa Thadajarassiri, Xiangnan Kong, Elke Rundensteiner. **CIKM,** 2022.

15. Recovering The Propensity Score From Biased Positive Unlabeled Data.
Walter Gerych, Thomas Hartvigsen, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. **AAAI**, 2022.
Oral spotlight.

16. On Detecting COVID-Risky Behavior from Smartphones.
Thomas Hartvigsen*, **Walter Gerych* (Joint First Author)**, Marzyeh Ghassemi. **Workshop on Epidemiology meets Data Mining and Knowledge Discovery, KDD,** 2022.

17. Triplet-based Domain Adaptation (Triple-DARE) for Lab-to-field Human Context Recognition.
Abdulaziz Alajaji, **Walter Gerych**, Luke Buquicchio, Kavin Chandrasekaran, Hamid Mansoor, Emmanuel Agu, Elke Rundensteiner. **IEEE PerCom Industry Track**, 2022.

18. Recurrent Bayesian Classifier Chains for Exact Multi-Label Classification.
Walter Gerych, Tom Hartvigsen, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. **NeurIPS**, 2021.

19. GAN For Generating User-Specific Human Activity Data From An Incomplete Training Corpus.
Walter Gerych, Harrison Kim, Joshua DeOliveira, MaryClare Martin, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Emmanuel Agu, Elke Rundensteiner. **IEEE Big Data** 4th Special Session on HealthCare Data, 2021.

20. Variational Open Set Recognition.
Luke Buquicchio, **Walter Gerych**, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Thomas Hartvigsen, Elke Rundensteiner, Emmanuel Agu. **IEEE Big Data**, 2021.

21. Local Geometry Preserving Deep Networks For Featurizing High-Dimensional Datasets.
Walter Gerych, Jessica Bader, Declan Nelson, Thalia Chai-Zhang, Luke Buquicchio, Abdulaziz Alajaji, Kevin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **IEEE ICMLA**, 2021.

22. Few-Shot Classification for Human Context Recognition Using Smartphone Data Traces.
Luke Buquicchio, **Walter Gerych**, Abdulaziz Alajaji, Kavin Chandrasekaran, Hamid Mansoor, Emmanuel Agu, and Elke Rundensteiner. **IEEE ICMLA**, 2021.

23. Visual Analytics of SmartphoneSensed Human Behavior and Health.
Hamid Mansoor, **Walter Gerych**, Abdulaziz Alajaji, Luke Buquicchio, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **IEEE Computer Graphics and Applications**, 2021.

24. Smartphone Health Biomarkers: Positive Unlabeled Learning of In-the-Wild Contexts.
Abdulaziz Alajaji, **Walter Gerych**, Luke Buquicchio, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **Pervasive Computing**, 2021.

25. Measuring Group Advantage: A Comparative Study of Fair Ranking Metrics.
Caitlin Kuhlman*, **Walter Gerych* (Joint First Author)**, Elke A. Rundensteiner. **AIES**, 2021.

26. ***PLEADES: Population Level Observation of Smartphone Sensed Symptoms for In-the-wild Data***
Hamid Mansoor, **Walter Gerych**, Abdulaziz Alajaji, Luke Buquicchio, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **VISIGRAPP**, 2021.

27. ***Complex Activity Recognition Using Topic Models for Feature Generation from Wearable Sensor Data.***
Kavin Chandrasekaran, **Walter Gerych**, Luke Buquicchio, Abdulaziz Alajaji, Emmanuel Agu, Elke Rundensteiner. **SMARTCOMP**, 2021.

28. ***BurstPU: Classification of Weakly Labeled Datasets with Sequential Bias.***
Walter Gerych, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Aidan Murphy, Emmanuel Agu, Elke Rundensteiner. **IEEE Big Data**, 2020.

29. ***INTOSIS: Interactive Observation of Smartphone Inferred Symptoms for In-The-Wild Data.***
Hamid Mansoor, **Walter Gerych**, Luke Buquicchio, Abdulaziz Alajaji, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **IEEE Big Data**, 2020.

30. ***DeepContext: Parameterized Compatibility-Based Attention CNN for Human Context Recognition.***
Abdulaziz Alajaji, **Walter Gerych**, Kavin Chandrasekaran, Luke Buquicchio, Emmanuel Agu, Elke Rundensteiner. **ICSC**, 2020.

31. ***ARGUS: Interactive Visual Analytics Framework for the Discovery of Disruptions in Bio-Behavioral Rhythms.***
Hamid Mansoor, **Walter Gerych**, Luke Buquicchio, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **EuroVis** (Short Papers), 2020.

32. ***COMEX: Identifying Mislabeled Human Behavioral Context Data Using Visual Analytics.***
Hamid Mansoor, **Walter Gerych**, Luke Buquicchio, Kavin Chandrasekaran, Emmanuel Agu, Elke Rundensteiner. **COMPSAC**, 2019.

33. ***Classifying Depression in Imbalanced Datasets Using Autoencoder-Based Anomaly Detection.***
Walter Gerych, Emmanuel Agu, Elke Rundensteiner. **ICSC**, 2019.

SUPERVISED UNDERGRADUATE PAPERS:

Human Context Recognition: A Controllable GAN Approach
Joshua DeOliveira, Harrison Kim, MaryClare Martin**, Walter Gerych**, Elke Rundensteiner. **IEEE URTC**, 2021

Positive Unlabeled Gradient Boosting
Caitlin Timmons, Andrea Boskovic, Sreeharsha Lakamsani, **Walter Gerych**, Luke Buquicchio, Elke Rundensteiner. **IEEE URTC**, 2020

Neural Network for Nonlinear Dimension Reduction Through Manifold Recovery
Jessica Bader, Declan Nelson, Thalia Chai-Zhang, **Walter Gerych**, Elke Rundensteiner. **IEEE URTC**, 2019

TEACHING/MENTORING

I have mentored teams of REU students every summer since 2019. I have also mentored three teams of students on their "Major Qualifying Projects" (year-long capstone projects). I typically give the students a research topic and direction, and then meet with them at least twice a week to resolve any issues and provide next steps.

Students Advised:

- Joshua DeOliveira, MS, WPI 05/2021–Current
- Cindy Trac, BS, WPI 08/2022–Current
- Sirut Buasai, BS, WPI 08/2022–Current
- Jason Dykstra, BS, WPI 08/2022–Current
- Dillon McCarthy, BS, WPI 08/2022–Current
- Aruzhan Koshkarova, BS, WPI 05/2022–08/2022
- Alek Lewis, BS, WPI 08/2021–05/2022
- Ryan Astor, BS, WPI 08/2021–05/2022
- Kyle Costello, BS, WPI 08/2021–05/2022
- Harrison Kim, BS, Northeastern University 05/2021–08/2021
- MaryClare Martin, BS, Holy Cross 05/2021–08/2021
- Jesse Abeyta, BS, WPI 08/2020–05/2021
- Vinay Nair, BS, WPI 08/2020–05/2021
- Nicholas Cheng, BS, WPI 08/2020–05/2021
- Bryan Gass, BS, WPI 08/2020–05/2021
- Caitlin Timmons, BS, Smith College 05/2020–08/2020
- Andrea Boskovic, BS, Amherst College 05/2020–08/2020
- Sreeharsha Lakamsani, BS, Arizona State University 05/2020–08/2020
- Jessica Bader, BS, Iowa State University 05/2019–08/2019
- Declan Nelson, BS, Georgia Institute of Technology 05/2019–08/2019
- Thalia Chao-Zhang, BS, Bard College 05/2019–08/2019

Developed workshop on Engaging Research Presentations for Graduate Students, WPI 01/2022
Lead workshop on Creating Engaging Presentations for Undergrads, WPI 06/2021
Developed workshop on Deep Learning with PyTorch for Undergrads, WPI 05/2019

Teaching assistant for Introduction to Data Science graduate class, WPI 01/2018-05/2018

SERVICE

Data Science Student Council, WPI 2019-2023
President of Council 2021-2023
Vice President 2020-2021
Judge for Graduate Qualifying Project Research Exchange, WPI 2021
Volunteer at Women in Data Science Conference @ WPI, WPI 2020, 2021
Organized Deep Learning Reading Group, WPI 2019-2020
Founder and President of Data Science Club, SUNY Geneseo 2016-2017