Jennifer Williams

✉ jlw1@cs.cmu.edu

⌂ http://github.com/jennwilliams

⌂ jennwilliams.github.io

Highlights
- Deep expertise in applying machine learning and statistics to high dimensional data, including dimensionality reduction, classification, regression, natural language processing, and deep learning
- Collaborated with cross-functional teams to analyze human health data (i.e., neuroimages including sensor data, psychological surveys, histology images, genomic data) to test scientific hypotheses
- Proposed, implemented, and evaluated new machine learning frameworks to analyze neuroimaging timeseries data (i.e., fMRI, EEG, MEG)
- Conceptualized and implemented causal inference tools (i.e., Bayesian networks) for scientific discovery

Education

PhD in Computational Biology, *Carnegie Mellon University*, 2016 - 2022 (Expected)
Advisor: Dr. Leila Wehbe (Machine Learning Department)
Thesis: Modeling individual differences in the brain in the presence and absence of disease

Cancer and Systems Biology EU-USA Atlantis Dual Degree Program, 2013 - 2016
MS in Natural Science, *Roswell Park, University at Buffalo*
MSc in Integrated Systems Biology**,** *University of Luxembourg*

BS in Biology**,** *Canisius College,* Magna Cum Laude*, 2009-2013*

Publications
- **Same cause; different effects in the brain.**
  M. Toneva*, J. Williams* (co-first), A. Bollu, C. Dann, L. Wehbe*. Causal Learning and Reasoning (CLeaR) 2022*
- **Behavior measures are predicted by how information is encoded in an individual's brain.**
  J. Williams, L. Wehbe*. arXiv 2021* (in submission)
- **Discriminative subtyping of lung cancers from histopathology images via contextual deep learning.**
  B.J. Lengerich, M. Al-Shedivat, A. Alavi, J. Williams, S. Labbaki, E.P. Xing*. medRxiv 2020*
- **LSD1 dual function in mediating epigenetic corruption of the vitamin D signaling in prostate cancer.**
  S. Battaglia, E. Karasik, B. Gillard, J. Williams, T. Winchester, M.T. Moser, D.J. Smiraglia, B.A. Foster. *Clinical Epigenetics 2017*

Fellowships
- **Digital Health Fellowship** – Center for Machine Learning and Health (2020 - 2021)
  Full tuition and stipend for 12 months and $3,000 for research-related expenses
- **NIH T32 Training Grant** – National Institute of Biomedical Imaging and Bioengineering (2017 - 2019)
  Full tuition and stipend for 2 years and $6,000 for research-related expenses
- **CanSys MS Scholarship** – Atlantis EU-USA Training Program (2013 - 2015)
  Stipend for 12 months

Computing Skills
- Python (Scikit-learn, Pandas, NumPy, SciPy, Matplotlib), R, MATLAB, Java, Linux/Bash, High Performance Computing (Slurm, PBS), LaTeX, Git, Container Platforms (i.e., Singularity)

Relevant PhD Courses

- Machine Learning
- Computational Medicine
- Intermediate Statistics
- Probabilistic Graphical Models
- Cognitive Neuroscience
- ABCDE of Statistical Methods in Machine Learning

Leadership and Service

- Co-organizer brAIn Seminar
- President PhD Program Graduate Student Assoc.
- Reviewer for: ML4H Conference, New in ML NeurIPS and ICML Workshops, Nature Scientific Reports, ECCB, Bioinformatics Advances
- Student Representative PhD Program Steering Committee and Admissions Committee
- Mentored Undergraduate Computer Science student (currently PhD student at Princeton)

Awards

- Top 10 Reviewer Machine Learning for Health (ML4H)
- International Conference on Machine Learning Travel Award
- Invited to Machine Learning Summer School (MLSS)
- PhD Program Student Service Award
- Canisius Excellence in Science Education Research Award

Selected Research Projects

Modeling individual differences in the brain to predict behavior

- Question: Do individual differences in how information is encoded in the brain predict behavior?
- Method Contribution: Built on insights from two sub-fields of neuroscience (brain mapping and behavioral neuroscience), to create the first machine learning framework to identify individual differences in how information is encoded in the brain and test if these differences predict behavior.
- Scientific Insights:
  - Individual differences in brain encoding can predict variability in cognitive behavior.
  - Advise researchers interested in predicting behavior to optimize their choice of neuroimaging task and feature-space for their behavior of interest.
- Paper: arXiv 2112.06048
- Code: github.com/brainML/great-apes

Disambiguating scientific inferences with causality

- Question: Why are large parts of the brain well predicted by the same neural network derived stimulus feature-space? Do these brain zones process the features similarly or differently?
- Method Contribution: Developed a causal inference framework, that includes two new metrics, to provide insights beyond current brain mapping techniques. Specifically, the framework enables researchers to infer if a complex (multivariate and high dimensional) stimulus affects two brain zones similarly.
- Scientific Insight: Real-world language stimuli (i.e., language spoken in videos) do not affect all parts of the language network similarly.
- Paper: CLeaR 2022 (arXiv 2202.10376)
- Code: github.com/brainML/stim-effect

Deep learning to integrate imaging and genomics data

- Question: Similarly to clinicians, can sample-specific models integrate multi-modal health data (i.e., histology images and transcriptomic data) to obtain accurate and interpretable lung cancer classification?
- Method Contribution: Inspired by contextual deep learning, adapted Contextual Explanation Networks (CENs) (Al-Shedivat et al., JMLR 2020) to the problem of lung cancer classification.
- Scientific Insight: Sample-specific multimodal models increase the accuracy of classification and captured the heterogeneity of biological processes underlying lung cancer.
- Paper: medRxiv (DOI: 10.1101/2020.06.25.20140053)