Jennifer Williams

✉ jlw1@alumni.cmu.edu
○ http://github.com/jennwilliams
⌂ jennwilliams.github.io

Highlights

- Deep expertise in applying machine learning and statistics to high dimensional data, including natural language processing (NLP), signal processing, classification, regression, clustering, and deep learning
- Collaborated with interdisciplinary teams to manipulate, analyze, and visualize healthcare data (i.e., neuroimages including sensor data, psychological surveys, histology images, genomic data)
- Proposed, implemented, and validated novel machine learning frameworks to analyze neuroimaging timeseries data (i.e., fMRI, EEG, MEG)

Experience

Senior Data Scientist, Machine Learning (ML) Platform, *CVS Health*, since April 2023

- Orchestrated and led a self-organizing agile team responsible for developing a productionalization tool for ranking models on Google Cloud Platform (GCP), resulting in a remarkable reduction in time-to-production to just 20 minutes
- Delivered engaging and informative technical demonstrations of ML Platform tools to diverse audiences, spanning from junior colleagues to Senior VPs, in small groups to gatherings of up to 1,200 colleagues
- Designed and developed a tool to streamline productionalization of large-scale TensorFlow personalization models, notably enhancing scalability within our ML infrastructure

Education

PhD in Computational Biology, *Carnegie Mellon University (CMU)*, 2016 - 2022
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* (in submission)
- **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*

Computing Skills

- Python (Scikit-learn, Pandas, NumPy, SciPy, Matplotlib), MATLAB, R, Java, Cloud Platforms, Git, Container Platforms, Linux/Bash, High Performance Computing (Slurm, PBS schedulers)

Relevant Courses

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

Leadership and Service

- Co-founder ML Platform Lunch and Learn Series
- Reviewer for: ML4H Conference, Learning from Time Series for Health and WiML NeurIPS Workshops, New in ML NeurIPS and ICML Workshops, Nature Scientific Reports, ECCB
- 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)
- CMU Provost Conference Award

Selected Research Projects

Deep learning to integrate imaging and genomics data

- Question: Similarly to clinicians, can sample-specific models integrate multi-modal health data (i.e., whole-slide 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)

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 novel 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., videos) do not affect all parts of the language network similarly.
- Paper: CLeaR 2022 (arXiv 2202.10376)
- Code: github.com/brainML/stim-effect

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 computational approach 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

Fellowships