ADAM CATTO

agocatto@gmail.com • adamcatto.github.io • New York, NY, USA

Data Scientist & Machine Learning Researcher. See Google Scholar for technical papers: https://scholar.google.com/citations?user=wuJLVSYAAAAJ&hl=en

EXPERIENCE

Data Scientist, Icahn School of Medicine at Mount Sinai **Jun 2022 – Present**

- Designed, implemented, and trained a deep learning model to detect spontaneous postsynaptic currents in current-clamp neuronal electrophysiology signals. Trained the model to an F1 score of 81%. My model saves dozens of hours of manual annotation per experiment.
- Designed, implemented, and trained a deep learning model to forecast behavioral events in laboratory animals, using computer vision and sequence modeling on behavioral neuroscience videos. Published a preprint detailing results.
- Finetuned and trained from scratch multiple DNA-sequence large language models (LLMs) on the order of 100M-2B parameters using PyTorch Fully Sharded Data Parallel. Improved existing LLM architecture's parallelism by 4x.
- Refactored a torchvision-based medical image analysis pipeline to be hardware-agnostic (GPU & TPU).
- Led analysis of spatial transcriptomics, single-cell sequencing, whole-brain microscopy, and electrophysiology data.

Research Data Scientist Intern, Johnson & Johnson **Jun 2021 – Oct 2021**

- Discovered nine digital biomarkers of four autoimmune disorders from wearable sensor data.
- Implemented distributed pipeline for terabyte-scale accelerometer data. Improved running time > 100x.
- Implemented signal processing algorithms to detect interpretable sleep and activity patterns in accelerometer data.
- Trained statistical machine learning models to classify disease vs. healthy controls from interpretable time-series features. Presented all results to project stakeholders.

Graduate Research Assistant, Research Foundation of CUNY **Aug 2020 – Jun 2022**

- Built machine learning models to predict adverse outcomes in pregnancy.
- Contributed to data preprocessing, model building, and visualization/analysis that won a national data challenge hosted by the National Institutes of Health: https://www.nichd.nih.gov/research/supported/challenges/decoding-maternal-morbidity. Wrote multiple technical machine learning preprints documenting the methods and results.
- Conceptualized, implemented, and validated two novel ensemble learning techniques for handling large amounts of missing data in tabular datasets, with and without imputation.

SKILLS

- Languages: Python (Advanced), Shell Scripting (Beginner), R (Beginner)
- Machine Learning / Data Analysis: PyTorch, PyTorch-Lightning, PyTorch-XLA, PyTorch Fully Sharded Data Parallel, TorchVision, PyTorch-Forecasting, Pandas, Scikit-Learn, HuggingFace, Seaborn
- Miscellaneous: Linux, Google Cloud Platform, Multi-GPU programming, Object-Oriented Design Principles

EDUCATION

CUNY Graduate Center, MS Data Science **2020 – 2022**

- (GPA 3.90/4.00) Graduate Coursework: Machine Learning, Data Visualization, Artificial Intelligence, Data Mining, Computational Biology, Digital Image Processing, Deep Learning for Genomics, Big Data Analytics

Stony Brook University, BS Applied Mathematics & Statistics, Philosophy **2015 – 2019**

PROJECTS

Parkinson's Freezing of Gait Prediction **2023**

Deep learning to segment bouts of freezing of gait in Parkinson's patients from hip-worn accelerometer signals.

- Placed in the 21st percentile of global Kaggle competition.

Computer Vision and Image Processing for Intelligent Transportation Systems **2021**

Built an efficient prior- and motion-based image processing architecture for real-time multi-object tracking in noisy tunnel traffic video feeds. Developed an image processing library from scratch in Python: github.com/adamcatto/dippy