

# CHENWEI WU

chenweiwu@g.harvard.edu ♦ [Personal Site](#) ♦ GitHub: [ChenweWu](#)

## EDUCATION

**Harvard University**, Cambridge, MA 09/2021-05/2023  
M.S. in Data Science GPA: 4.00/4.00  
**Focuses:** Self-supervised learning (SSL), Computer Vision, Out-of-domain (OOD) Generalization  
**TF:** CS109A Introduction to Data Science  
**University of Rochester**, Rochester, NY 09/2016-05/2020  
B.S. in Data Science (highest distinctions), B.A. in Financial Economics (distinctions), Minor in Japanese  
**Honors:** Department Top 3, Phi Beta Kappa, Deans' List 2016-2020 GPA: 3.92/4.00  
**TF:** CS Department Tutor, CSC 265 Intermediate Statistical Methods, CSC 240 Data Mining, ECON 268 Economics of Globalization

## PUBLICATIONS & CONFERENCE PRESENTATIONS

[BenchMD: A Benchmark for Modality-Agnostic Learning on Medical Images and Sensors](#)  
First Author Submitted to CVPR 2023  
**A Case Study of the Challenges of Applied Machine Learning in Assisted Reproductive Technology**  
First Author In preparation for ICML 2023  
**Sensor-based Gaze and Pose Tracking Analysis of Student Group Behavior in the Makerspace**  
First Author In preparation for CSCW 2023  
[Interpretable Multimodal Deep Learning to Predict Breast Cancer Stage](#)  
First Author Submitted to Nightingale Contest & IEEE Transactions on Medical Imaging  
**De-identification and Obfuscation of Sensitive Attributes From Retinal Scans**  
First Author In preparation for Nature Communications  
[Social affective forecasting and social anhedonia in schizophrenia-spectrum disorders: a daily diary study](#)  
Co-Author Nature, Schizophrenia, 2022  
[Effective Clustering of Nursing Homes Using Unsupervised Machine Learning Focusing on Dementia and Mental Illness](#)  
Co-Author Academy Health ARM 2020  
[Using Optical Flow to Quantify Movement Differences in Responses to Emotional Stimuli Among People with Schizophrenia](#)  
Co-Author Cognitive Neuroscience Society 2020  
[Does functional connectivity within the DMN predict individual differences in social pleasure in schizophrenia?](#)  
Co-Author Cognitive Neuroscience Society 2020, Society for Research in Psychopathology 2019

## TALKS

**Johnson & Johnson Data Science R&D Symposium**, New Brunswick, NJ 08/2022  
"Self-supervised Learning for Survival Rate Prediction of Histopathology Images"  
**Harvard Medical AI Talk, Harvard Medical School**, Boston, MA 08/2022  
"[Remote Working on a Cluster](#)"  
**Harvard Medical AI Talk, Harvard Medical School**, Boston, MA 04/2022  
"Model Soup"  
**Harvard Medical AI Talk, Harvard Medical School**, Boston, MA 03/2022  
"[Ensemble of Averages: Improving Model Selection](#)"  
**Harvard Medical AI Talk, Harvard Medical School**, Boston, MA 01/2022  
"[Docker for ML Practitioners](#)"

## RESEARCH EXPERIENCES

**Medical AI Lab, Harvard Medical School** 11/2021 – Present  
*Graduate Researcher* Advisor: Dr. Pranav Rajpurkar

- Viewmakers: Learning Neural Augmentations for Electrocardiograms in Self-supervised Learning
  - Developed generative diversity-viewmaker networks budgeted by stochastic L1 boundaries to adversarially learn SSL augmentations on 12-lead electrocardiogram; Viewmakers eliminate the

manual expert augmentation process and perform spurious feature suppression.

- Developed self-distillation with no labels algorithms to improve performance for the CNN and ViT encoders.
- **BenchMD: A Benchmark for Modality-Agnostic Learning on Medical Images and Sensors**
  - Designed BenchMD, a modality-agnostic benchmark that tests how different architectures and training techniques (SSL & SL) perform on domain-shift medical tasks; This benchmark covers 19 publicly available datasets for 7 diverse medical modalities, ranging from 1D sensor data, 2D images, to 3D volumetric scans.
  - Designed three SSL techniques (Emix, Shed, Agnostic MAE) and evaluated their few-shot and zero-shot performance on OOD medical data; beat SOTA AUROC in EEG and Dermatology.

#### **Learning, Information & Technology Lab, Harvard University**

10/2021 – Present

*Graduate Researcher*

*Advisor: Dr. Bertrand Schneider*

- **Multimodal Learning Analytics (MMLA) for Makerspaces**
  - Developed computer vision 3D gaze detection and facial segmentation pipelines for student collaborative learning behaviors in the Harvard Makerspace using Pytorch.
  - Performed gaze and pose data re-projections into the 3D simulated lab space to generate better visualizations.
  - Collect and analyze student behavior data in Makerspace by leveraging multimodal pipelines to understand social learning aspects like student collaboration and self-efficacy.

#### **DtaK Lab, Harvard University**

02/2022 – Present

*Graduate Researcher*

*Advisor: Dr. Weiwei Pan*

- **A Case Study of the Challenges of Applied Machine Learning in Assisted Reproductive Technology**
  - Proposed and implemented two improved methods to overcome the existing limitations of machine learning application in the In Vitro Fertilization domain.
    - Explored limitations in current literature including zero external validation, data leakage, heterogeneity and lack of timeliness.
    - Developed Phase-by-phase model to provide interpretable and progressive assistance for clinicians at different IVF stages, and developed Subgroup model to cope with data heterogeneity.
    - Collaborated with clinicians to outline a standardized data selection, preprocessing and modeling pipeline.
  - Used mixed effects models and mixtures of regressions to interpret predictors contributing to successful pregnancies and live births.

#### **SANA Lab, MIT**

09/2022 – Present

*Graduate Researcher for Capstone Project*

*Advisor: Dr. Leo Anthony G. Celi*

- **De-identifying Retinal Fundus Images for MIMIC**
  - Develop multi-objective algorithms to de-identify retinal fundus images without hurting down-stream disease classification performance; Integrate a new Brazilian retinal fundus dataset into MIMIC following HIPAA privacy regulations.
    - Successfully decreased the gender identification accuracy from 81% to 64%, while maintaining the diabetic retinopathy classification accuracy at 95%.
- **Interpretable multimodal deep learning to predict breast cancer stage**
  - Develop multi-modal models to combine attention-based multiple-instance learning on biopsy images and self-normalized networks on structured clinical metadata to predict breast cancer staging. Achieved Cohen Kappa of 71% and AUROC of 80% over 5-fold cross validation.

#### **Social Cognition & Psychopathology Lab, University of Rochester**

04/2019-04/2022

*Undergraduate Researcher*

*Advisor: Dr. David Dodell-Feder*

- **Social Brain Study**
  - Performed sentiment and topic analysis on text behavioral data in a functional neuroimaging study to investigate the relationship between brain functional connectivity and social anhedonia.

- Movement Study
  - Applied semi-supervised CNNs, optical flow and OpenFace on patient videos to analyze group differences in nonverbal social synchrony between schizophrenia patients and controls.
  - Studied the patterns of combinations of facial activation units to evaluate the effect of oxytocin on patients' social abilities.

#### **Geriatric Oncology, University of Rochester Medical Center**

05/2019 - 08/2021

*Undergraduate Researcher*

*Advisor: Dr. Huiwen Xu*

- Nursing Home Clustering
  - Utilized unsupervised algorithms to cluster nursing homes based on the percentage of residents with dementia, depression, and serious mental illness; Detected previously unknown patterns of resident case-mix and staffing in nursing homes; Predicted deficiency scores of nursing homes on a longitudinal basis using supervised learning algorithms.
- Caregiver Health Forecasting
  - Developed text association and pattern mining algorithms to classify cancer therapies.
  - Combined synthetic minority oversampling Technique (SMOTE) with supervised learning techniques to process imbalanced cancer caregiver EHR datasets; Implemented Local Interpretable Model-Agnostic Explanations (LIME) to help clinicians identify mental and physical health risk factors.

### **WORK EXPERIENCES**

**Janssen**, Titusville, NJ

06/2022 - 08/2022

*CV Research Intern*

- Developed self-supervised learning pipeline for prostate cancer histopathology WSI's feature extraction and downstream survival prediction.

**Credit Suisse**, New York, NY

06/2020 - 06/2021

*Full-time Analyst*

- Served as an IB analyst and data scientist for the Investment Banking Emerging Markets Division.
- Built Airflow automated trading data ETL pipelines and constructed a centralized Azure cloud data platform for bonds and credit default swaps.
- Served as project manager to create a firm-wide chat platform that leverages NLP to assist sales & trading teams to gain a competitive edge.
- Collected big data streams and performed deep learning time series predictions on stock trends.

### **OUTREACH**

**Rochester Data Science Society, University of Rochester**

10/2018 - 05/2020

*President*

- Formed a Data Science Colloquia Series; Invited renowned speakers from various fields of data science to share SOTA topics and methods.
- Connected Data Science students with faculty and alumni to help them gain internship and research opportunities.

**The Commonwealth of Massachusetts**, Boston, MA

06/2018-08/2018

*Data Analyst Intern*

- Analyzed certification and accounting data via R and Excel VBA to identify opportunities and make strategies for the recruitment, engagement, and retention processes for vendors of minorities, including women, LGBTQ, physically disadvantaged, and veterans.
- Attended in-person events and made calls to promote government assistance programs and establish connections with disadvantaged vendors. Utilized BI reporting tool (Sugar) to collect, analyze, and visualize government vendor business data for state government budgeting.

### **SKILLS**

**Computer Languages & Tools::** Python, Java, MATLAB, R, SQL, HTML, Airflow, Docker, Excel, Tableau

**Machine Learning:** PyTorch, Pytorch-lightning, Tensorflow, Keras, Sklearn, OpenCV

**Computing Platforms:** SLURM, AWS, GCP, Azure Databricks

**Languages:** Chinese (Native), English (Proficient), Japanese (Intermediate), Hu Dialect (Proficient)

**Personal Interests:** Rowing (UR Varsity Team), Powerlifting (Bench Press 345 lbs, Deadlift 565 lbs, Squat 535 lbs), Diving (Open Water Permit), Valorant: First Person Shooter Game (Peak Top 135 North America)