# Curriculum Vitae

chenweiw@umich.edu ◊ GitHub: ChenweWu

## Education

**University of Michigan,** Ann Arbor, MI — 08/2023 - Present

*Ph.D. Student* — **Advisors**: Liyue Shen; Zhongming Liu

**Harvard University,** Cambridge, MA — 09/2021-05/2023

*M.S. in Data Science* — **GPA**: 3.94/4.00

**Teaching:** CS109A Introduction to Data Science — **Advisors**: Pranav Rajpurkar; Leo Celi

**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* — **GPA**: 3.92/4.00

**Honors:** Phi Beta Kappa, Deans' List 2016-2020, Department Top 3

**Teaching:** CS Department Tutor, Intermediate Statistical Methods, Data Mining, Economics of Globalization

## Research Experiences

**SANA Lab, MIT** — 09/2022 – 08/2023

*Graduate RA* — *Advisor: Dr. Leo Anthony G. Celi*

- De-identifying Retinal Fundus Images
  - Develop two 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.
- Multi-modal Fusion of Multi-spectral Satellite Imagery for Public Health
  - Develop a framework, integrating Sentinel Hub and Google Earth Engine, facilitates efficient and generalized embeddings extraction. Also introduce a unique recursive de-noising algorithm to reduce noise anomalies and a cryptographic hash method to ensure data quality.
  - Propose a novel multimodal fusion pipeline that exclusively relies on satellite imagery and metadata. This stands in contrast to prevalent autoregressive prediction methodologies, which heavily depend on historical Dengue data often absent in real-world scenarios.

**Medical AI Lab, Harvard Medical School** — 11/2021 – 05/2023

*Graduate RA* — *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.
- 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; 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 – 08/2023

*Graduate RA* — *Advisor: Dr. Bertrand Schneider*

- Multimodal Learning Analytics for Makerspaces
  - Developed 3D gaze detection and facial segmentation pipelines to capture student collaborative learning behaviors in the Harvard Makerspace using Pytorch.
  - Performed gaze and pose data re-projections into the 3D point cloud space.

## Work Experiences

**Janssen,** Titusville, NJ — 06/2022 - 08/2022

*Computer Vision 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
*Analyst*

- Served as a business analyst and data scientist for Investment Banking Group, 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 chatbot application that leverages NLP to assist sales & trading.

**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.

## Talks

**Johnson & Johnson Data Science R&D Symposium,** Philadelphia, PA 08/2022
"Self-supervised Learning for Survival Rate Prediction of Histopathology Images "
**Harvard Medical AI Talk Series, Harvard Medical School,** Boston, MA 08/2022
"Remote Working on a Cluster"
**Harvard Medical AI Talk Series, Harvard Medical School,** Boston, MA 03/2022
"Ensemble of Averages: Improving Model Selection"
**Harvard Medical AI Talk Series, Harvard Medical School,** Boston, MA 01/2022
"Docker for ML Practitioners"

## Papers & Abstracts

**Multi-modal Fusion of Multi-spectral Satellite Imagery for Public Health**
Co-Author In Reviews Nature Scientific Reports
**Designing New Ways of Capturing Social Learning in Makerspace using Multimodal Data Streams**
First Author In Reviews Learning & Instructions
**Interpretable Multimodal Deep Learning to Predict Breast Cancer Stage**
Co-Author In Reviews IEEE Transactions on Medical Imaging
**BenchMD: A Benchmark for Modality-Agnostic Learning on Medical Images and Sensors**
First Author In Reviews NIPS 2023
**De-identification and Obfuscation of Gender Attributes From Retinal Scans**
First Author Accepted by MICCAI 2023, Fairness of AI in Medical Imaging
**Pixel Snow and Differential Privacy in Retinal fundus photos de-identification**
Co-Author ARVO 2023; Investigative Ophthalmology & Visual Science
**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

## Skills

**Computer Languages & Tools::** Python, Java, R, Docker, Tableau, Apache Kafka, Apache Airflow, Scala
**Machine Learning:** PyTorch, Pytorch-lightning, Tensorflow, Tf.keras, Sklearn, OpenCV
**Computing Platforms:** Slurm, AWS, GCP, OCI, Azure Databricks
**Languages**: Chinese (Native), English (Proficient), Japanese (Intermediate)
**Test Scores:** GRE : V163+Q170+W4.5; CFA: Level1 passed