CINDY ZHANG

Seattle, Washington cnmnzhang@gmail.com 320-266-0887 cnmnzhang.github.io/me

EDUCATION

Johns Hopkins University 2019 – 2023
BS in Computer Science with Honors, BS Biomedical Engineering with Honors
Minor in Applied Mathematics and Statistics
Member of Habitat for Humanity, Society of Women Engineers
GPA: 3.59/4.0

EXPERIENCE

Research Associate June 2024 – Present
University of Washington Medical Center *Seattle, WA*

- Investigated the utility of intra-patient hematologic setpoints from laboratory markers for predicting patient outcomes using Kaplan-Meier analysis and COX proportional-hazard modeling.
- Developed Bayesian framework to adaptively model personalized reference ranges from time-series laboratory markers data.
- Parallelized pipelines to analyze gigabytes of diverse laboratory markers and optimize hyperparameters for hidden Markov models, state-space models, and classical Bayesian inferencing models.

Senior Programmer Analyst June 2023 – June 2024
Johns Hopkins Health System *Baltimore, MD*

- InBasket Clinical Message Triage Model Support
  - Deployed BERT model to classify 100,000+ messages, reducing cognitive load on physicians and administrative staff.
  - Evaluated model misclassification rates across different clinical departments.
  - Explored BERTopic to improve model training by supplying a minimum set of representative department-specific examples.
  - Conducted prompt engineering experiments and cost analysis for labeling messages using GPT models (GPT-4-32K, GPT-4 Turbo, LLaMA2) to reduce human labeling efforts.
- Monitored ETL pipelines for the Precision Medicine Analytics Platform using Azure Data Factory and Databricks.
- Supported the centralized development of the Precision Medicine Analytics Platform, focusing on electronic health record data quality standards and system integration requirements. Including:
  - Refactored and validated WHO Clinical Progression Scale status assignment logic with physicians, reducing run-time from 3 hours to 11 minutes.
  - Scripted a Python library for automated extraction of high-volume perinatal monitoring waveform data using REST API methods and Delta Live Tables in Databricks.
  - Led an initiative to create a ground truth configuration table in SQL Server.
  - Created a master Azure ADF pipeline for table updates and bulk data refreshes.
  - Addressed internal needs for high-level key performance indicators by analyzing ETL pipeline statistics and projection database usage from SQL Server audit logs, and visualized the analytics in Splunk.
  - Facilitated schema changes for EPIC upgrades and CRMS to OnCore transition.

Undergraduate Researcher Aug 2022 - May 2023
Translational Informatics Research and Innovation Lab *Baltimore, MD*

- Assessing Associations Between COVID-19 Symptomology and Adverse Outcomes After Piloting Crowdsourced Data Collection: Cross-sectional Survey Study
  - Crowdsourced responses for a COVID-19 symptomology Qualtrics study using Amazon Mechanical Turk.
  - Developed a strategy to filter responses and ensure data reliability using attention checks within the survey.
  - Performed chi-squared analyses to assess differences in sociodemographic and hospitalization outcomes.
- Designed a prototype Android app to assist pregnant women in seeking reliable health information.
- Developed an electronic health record converter between HL7 and Eventflow formats.
- Investigated MyChart logs for patterns of continuity of care and assessed association with sociodemographics and divergent health outcomes.

Systems Engineer May 2022 - Aug 2022
Johnson & Johnson *Santa Clara, CA*

- Executed C++ test suites in a Linux environment to characterize a drift in robotic surgical arm joints
- Drove system shutdown requirements to prevent excessive drift.
- Verified design documents and requirements across engineering teams.
- Developed and documented scripts to link and populate Airtable databases, replacing manual processes with automated solutions.
- Refactored JAMA REST API methods and designed an automated error-logging strategy using Python.

Researcher and Software Developer Nov 2021 - May 2022
Applied Physics Laboratory *Laurel, MD*

- Proofread AI-annotated neural connectomes to verify image segmentation of neurons and synapses.
- Conceptualized game-ification of connectome proofreading to facilitate human reinforcement learning.
- Wrote C# code to generate 3-D large-scale neuron meshes for the Unity game.

ACADEMIC RESEARCH

Senior Design Project: A machine learning approach to predict emergent right ventricular pressure-volume loop phenotypes Aug 2022 - May 2023

- Collaborated with a clinician on a weekly basis to devise a clinically meaningful paradigm for pulmonary hypertension risk stratification informed by the pathophysiology of the right ventricle.
- Clustered pressure-volume loop measurements to obtain three patient groups with distinct right heart functional phenotypes and statistically significantly different times to clinical worsening.
- Integrated right heart catheterization and magnetic resonance imaging measurements to train a random forest model to predict the pressure-volume loop patient group (AUC = 0.94).
- Identified the most informative features with feature importance and SHAP analysis.
- Contributed to the literature review, methodology, and results of the JAMA submission manuscript.

Delineo Disease Modeling Jan 2021 - Nov 2021

- Implemented a Wells-Riley model to simulate the spread of diseases in a parameterizable "anytown".
- Managed the simulation team by leading meetings, distributing tasks, and providing mentorship.
- Decreased simulation time for 6 months of infection by 59% by code characterization using CPython.

PUBLICATIONS AND PRESENTATIONS

"Haematologic Setpoints Are a Stable and Patient-Specific Deep Phenotype"
Brody H Foy, Rachel Petherbridge, Maxwell Roth, Cindy Zhang, Christopher Mow, Hasmukh R Patel, Chhaya H Patel, Samantha N Ho, Evie Lam, Konrad J Karczewski, Veronica Tozzo, John M Higgins
Dec 2023, DOI: 10.1101/2023.09.26.23296146
PMID: 37808854

"Assessing Associations Between COVID-19 Symptomology and Adverse Outcomes After Piloting Crowdsourced Data Collection: Cross-sectional Survey Study"
Natalie Flaks-Manov, Jiawei Bai, **Cindy Zhang**, Anand Malpani, Stuart C Ray, Casey Overby Taylor
Nov 2022, DOI: 10.2196/37507
PMID: 36343205

"OPTIC: Optimizing Patient-Provider Triaging & Improving Communications in Clinical Operations using GPT-4 Data Labeling and Model Distillation".
Working Paper

"Machine Learning Approach to Predict Emergent Right Ventricular Pressure-Volume Loop Phenotypes in Pulmonary Hypertension".
Design Day, Featured Presentation. Johns Hopkins University, Baltimore, MD, May 2023

SKILLS

Languages: Python (scikit-learn, PyTorch), MATLAB, C, C++, SQL, Java, JavaScript, TypeScript, HTML LaTeX, Apache Spark, Scala, Microsoft SQL Server, PostgreSQL,

Technologies: Git, JupyterLab, MongoDB, ExpressJS, ReactJS, NodeJS, Docker, Django, Azure Data Factory, Databricks

Interests: yoga, running, swimming, climbing, art, piano, Mandarin, reading!