Jifan Gao

Github: github.com/GGGGFan

Email: jifan.gao@wisc.edu
Mobile: +1-(206)618-7236

Research Interests

Ontology-based interpretable machine learning; Complementarity and redundancy of multimodality biomedical data; Fairness of machine learning models.

Education

- **University of Wisconsin-Madison** — Madison, Wisconsin, USA
  Ph.D. in Biomedical Data Science — *2026 (expected)*
- **University of Wisconsin-Madison** — Madison, Wisconsin, USA
  M.S. in Computer Science — *May 2023*
- **University of Wisconsin-Madison** — Madison, Wisconsin, USA
  M.S. in Biomedical Data Science — *May 2020*
- **University of British Columbia** — Vancouver, British Columbia, Canada
  M.Eng. in Biomedical Engineering — *November 2016*
- **Southeast University** — Nanjing, Jiangsu, China
  B.Eng. in Bioengineering — *June 2014*

Data Science Competitions

- **The NIH Long COVID Computational Challenge (L3C)** 2023
  - Built a LightGBM-based model with a tailored ontology rollup strategy on structured EHR data to identify which patients infected with COVID-19 have a high likelihood of developing PASC/Long COVID.
  - Achieved 4th place overall in judging and won **$16,666 prize**.
- **Preterm Birth Prediction - Microbiome DREAM Challenge** 2022
  - Built a LightGBM-based model for preterm birth prediction with an ensemble strategy tailored for **vaginal microbiome data collected from multiple studies**.
  - Achieved **1st place** in Task 1 - Preterm Birth Prediction.
- **2022 National NLP Clinical Challenges Track 3 - Progress Note Understanding** 2022
  - Incorporate **medical ontology** into a **transformer-based workflow** for understanding the relation between assessment and plan subsections in clinical notes.
  - Achieved a macro F1 of 81.19 that ranked at **3rd place** among all participants.
- **Pediatric COVID-19 Data Challenge Task 1 - Predicting Need for Hospitalization** 2021
  - Competition is hosted by the Biomedical Advanced Research and Development Authority (BARDA), an office of the **U.S. Department of Health and Human Services (HHS)**.
  - Used **standardized vocabularies** and **ontology rollup** to **reduce dimensionality**, mitigated **sample size bias** by customized training/tuning strategy, used LightGBM for prediction, used **SHAP values** to enhance interpretability.
  - Achieved **best performance** and won **$100,000 prize** to the home university (University of Wisconsin-Madison).
- **COVID-19 DREAM Challenge Q1 (COVID-19 Diagnosis Prediction)** 2021
  - Used handcrafted features and LightGBM models for prediction.
  - Achieved **First Runner-up** among all participants.
- **First EHR DREAM Challenge – Patient Mortality Prediction** 2020
  - Used **standardized vocabularies** and **ontology rollup** to **reduce dimensionality**, used time-binning to capture **longitudinal properties**, customized training/tuning strategy to encourage the model's **"future-proof" ability**.
  - Achieved **best performance** among all participants.

Publications

Gao J, Chen G, O'Rourke AP, Caskey J, Carey K, Oguss M, Stey A, Dligach D, Miller T, Mayampurath A, Churpek MM. Automated stratification of trauma injury severity across multiple body regions using multi-modal, multi-class machine learning models. Accepted by Journal of the American Medical Informatics Association.

Golob JL, Oskotsky TT, Tang A, Roldan A, Chung V, Ha CW, Wong R, Flynn K, [et al, including **Gao J**]. Microbiome Preterm Birth DREAM Challenge: Crowdsourcing Machine Learning Approaches to Advance Preterm Birth Research. Cell Reports Medicine. 2024 Jan 16;5(1).

Bergquist T, Schaffter T, Yan Y, Yu T, Prosser J, **Gao J**, Chen G, Charzewski Ł, Nawalany Z, Brugere I, Retkute R. Evaluation of crowdsourced mortality prediction models as a framework for assessing AI in medicine. Journal of the American Medical Informatics Association. 2024 Jan 1;31(1):35-44.

Gao J, He S, Hu J, Chen G. A hybrid system to understand the relations between assessments and plans in progress notes. Journal of Biomedical Informatics. 2023 May 1;141:104363.

Bergquist T, Wax M, Bennett TD, Moffitt R, **Gao J**, Chen G, Telenti A, Maher MC, Bartha I, Walker L, Orwoll B. A framework for future national pediatric pandemic respiratory disease severity triage: The HHS pediatric COVID-19 data challenge. Journal of Clinical and Translational Science. 2023:1-29.

Gao J, Mar PL, Chen G. More Generalizable Models For Sepsis Detection Under Covariate Shift. AMIA Summits on Translational Science Proceedings. 2021;2021:220.

Yan Y, Schaffter T, Bergquist T, Yu T, Prosser J, Aydin Z, Jabeer A, Brugere I, **Gao J**, Chen G, Causey J. A Continuously Benchmarked and Crowdsourced Challenge for Rapid Development and Evaluation of Models to Predict COVID-19 Diagnosis and Hospitalization. JAMA network open. 2021 Oct 1;4(10):e2124946-.

Professional Service

- **Student Editor, Journal of the American Medical Informatics Association (JAMIA)** 2024-2025
  - One of the 6 student editors for the 2024-2025 term
  - Duties include reviewing 6 manuscripts annually and moderating the monthly online JAMIA Journal Club.