Yina Hou

yhou@my.tnstate.edu | yhou | YinaHou | Google Scholar
YinaHou.github.io | Nashville, Tennessee

Education

Tennessee State University Tennessee, USA
M.Sc. Candidate, Computer Science (Data Science Specialization) (GPA: 4.0/4.0) *Jan'2024 – Dec'2025*

- **Thesis**: Causal Explainability of Machine Learning in Heart Failure Prediction from Electronic Health Records

HeBei University HeBei, China
B.Sc. in Information and Computing Science (GPA: 3.25/4.0) *Sep'2019 – June'2023*

- **Thesis**: A Study on Data Augmentation to Improve the Performance of Deep Classification Models

Professional Experience

Graduate Research Assistant Spring 2024, Fall 2025
CIDA Lab, Tennessee State University

- Developed and optimized deep learning models for processing and analyzing electronic health records (EHR)
- Designed a novel imputation framework using sample- and feature-level attention to address missing data
- Investigated the impact of imputation-induced data distribution shifts on downstream predictive performance
- Processed and curated real-world healthcare datasets (e.g., heart failure, breast cancer) from the All of Us research workbench

Research Assistant May 2025 – Aug 2025
Biomedical Data Science Department, Meharry Medical College

- Analyzed and fine-tuned large language models (LLMs) on medical data sets
- Evaluated LLM performance on downstream tasks including diagnosis prediction and medical code generation
- Assisted in setting up experiments for retrieval-augmented generation (RAG) for extracting medical concepts

Graduate Teaching Assistant Sep 2024 – May 2025
Computer Science Department, Tennessee State University

- Supported COMP3000 (Computer Programming) and COMP4830 (Introduction to Data Science) through student consultations and tutoring
- Proctored and graded assignments and quizzes, maintaining academic standards and timely feedback

Peer Reviewed Publications

1. **Hou, Y.**, Rabbani, S. B., Hong, L., Diawara, N., & Samad, M. D. (2025). *Causal Explainability of Machine Learning in Heart Failure Prediction from Electronic Health Records*. *In Proceedings of the 26th IEEE International Conference on Information Reuse and Integration for Data Science (IRI 2025)*.

2. Kowsar, I., Rabbani, S. B., **Hou, Y.**, & Samad, M. D. (2025). *DeepIFSAC: Deep imputation of missing values using feature and sample attention within contrastive framework*. *Knowledge-Based Systems*, 113506. Elsevier.

Technical Skills

Programming Languages: Python, C++, SQL
Deep Learning Frameworks: PyTorch, TensorFlow, Keras
ML/NLP Libraries: Scikit-learn, Hugging Face Transformers, OpenCV, SciPy, Pandas, NumPy, Seaborn, Matplotlib
Tools & Platforms: Jupyter, Git, Docker, Google Cloud Platform (GCP), Weights & Biases (WandB), SQLAlchemy
Databases: MySQL, MongoDB
Other Software: MATLAB, LaTeX

Projects

- **OMOP Common Data Model for EHR Data (COMP5400)** (GitHub)
  - Extracted synthetic EHR records using ICD codes from the All of Us workbench to simulate real-world structure
  - Designed and implemented a relational database following OMOP-CDM standards using SQL and RDBMS principles
  - Validated relationships across clinical entities using realistic privacy-preserving patient data
- **Feature Importance and Causal Structure Analysis on Real-World EHR Data** (GitHub)
  - Extracted and curated datasets from the All of Us workbench using ICD codes for heart failure and fatty liver
  - Designed the study to analyze gender- and age-specific patterns in medical feature behavior
  - Evaluated causal discovery models and feature importance techniques to uncover key risk factors for heart failure
- **A Novel Imputation Method Using Sample and Feature Attention for EHR Data** (GitHub)
  - Extracted ICD-based EHR data from the All of Us workbench for simulation of real-world missingness
  - Developed and refined an attention-based imputation pipeline leveraging feature and sample dependencies
  - Validated imputation effectiveness through downstream classification tasks and reconstruction error analysis

Certificates & Awards

- Third Prize, 13th National College Student Math Competition (Hebei Province) 2022
- Outstanding Board Member, University Student Association, Hebei University 2021
- Performance-Based Scholarship, Hebei University 2020–2021