Bao Hoang

hoangbao@msu.edu | Google Scholar | GitHub | Personal Website

Research interest

My research interest lies in trustworthy machine learning, focusing on robustness, privacy, and explainability, with applications in healthcare settings.

Education

Michigan State University East Lansing, MI

Bachelor of Science, Computer Science and Advanced Mathematics, Honors College *Expected May 2025*

- GPA: 4.0/4.0
- Graduate-level Coursework: Machine Learning, Data Mining, Computational Foundations in Artificial Intelligence, Algorithmic Graph Theory.
- Honors and Awards: Honors College Excellence Scholarship, Professorial Assistantships, R.E. Phillips Memorial Scholarship, EnSURE Funding 2023 and 2024, Dean List all semesters.

Publications

[1] **Bao Hoang**, Yijiang Pang, Siqi Liang, Liang Zhan, Paul Thompson, Jiayu Zhou.*"Distributed Harmonization: Federated Clustered Batch Effect Adjustment and Generalization"*. KDD 2024 [Paper][Code]

[2] **Bao Hoang**, Yijiang Pang, Hiroko H. Dodge, and Jiayu Zhou.*"Translingual Language Markers for Cognitive Assessment from Spontaneous Speech"*. InterSpeech 2024 [Paper][Code]

[3] **Bao Hoang**, Yijiang Pang, Hiroko H. Dodge, and Jiayu Zhou.*"Subject Harmonization of Digital Biomarkers: Improved Detection of Mild Cognitive Impairment from Language Markers"*. PSB 2024 [Paper][Code]

[4] Yijiang Pang, **Bao Hoang**, Jiayu Zhou.*"Cross-modality debiasing: using language to mitigate sub-population shifts in imaging"*. arXiv 2024 [Paper]

[5] Yijiang Pang, Shuyang Yu, **Bao Hoang**, Jiayu Zhou.*"Towards Stability of Parameter-free Optimization"*. submitted to ICLR 2025 [Paper]

Experience

Research Assistant August 2021 – Present

ILLIDAN Lab *East Lansing, MI*

- Project Contributions:
- 1. Improved Detection of Mild Cognitive Impairment (MCI) via Spontaneous Conversations:
  - Led multimodal analysis of linguistic, facial, and acoustic features from patient-doctor conversations to identify MCI digital biomarkers.
  - Applied Natural Language Processing techniques to analyze lexical, semantic, and syntactic features.
  - Utilized the VGG-Face Convolutional Neural Network to extract patients' facial emotion embeddings.
  - Extracted Mel-frequency Cepstral Coefficients as paralinguistics features from patient speech signals.
  - Developed a novel deep learning-based subject harmonization approach to map original digital biomarkers into generalizable subject-invariant representations, achieving a 5% increase in the AUC performance of MCI classifiers.
- 2. Multi-site Brain Imaging Efficient Distributed Harmonization:

    - Enhanced the ComBat harmonization technique for brain imaging by integrating a K-means Clustering algorithm, allowing for the harmonization of new sites' data without requiring retraining, thus doubling inference time while maintaining similar performance.
    - Extended the proposed harmonization algorithm to a distributed framework utilizing the Federated Averaging algorithm, thus protecting the privacy of sensitive health information.
- 3. Machine Learning Research:
    - Developed innovative ensemble method combining multiple sentence transformers' embeddings using Kullback–Leibler Divergence loss, resulting in improved In-Context Learning performance for Large Language Models.
    - Conducted theoretical analysis on the convergence rate of AdaGrad-norm optimization in non-convex settings to identify the optimal step size, which minimizes the upper bound of gradient norm, leading to the design of the proposed parameter-free Adam optimization (AdamG).
    - Improved the worst-case accuracy of the Contrastive Language-Image Pre-training (CLIP) model by developing effective debiased prompts for the proposed Language-Distributional Robustness Optimization (L-DRO) algorithm.

Machine Learning Engineer Intern May 2022 – August 2022
FPT Software AI Center *Hanoi, Vietnam*

- Project Contributions:
- 1. Healthcare Chatbot Development:
    - Developed a multilingual healthcare support chatbot in Python using the Bot Framework SDK v4.
    - Integrated APIs from Google Cloud and Microsoft Azure to enhance intent recognition capabilities, support multilingual conversations, and provide location-based services to find the nearest hospital.
    - Implemented Dijkstra's algorithm for efficient indoor hospital navigation.
- 2. Smart Workplace Alert System:
    - Engineered a real-time monitoring system utilizing Streamlit, FastAPI, and pre-trained YOLOv5 to detect and alert on non-compliant behaviors, such as cell phone usage during work hours.

Machine Learning Engineer January 2021 – August 2021
Projectube *Hanoi, Vietnam*

- Developed a machine learning server using Flask for Projectube, an interactive platform enabling Vietnamese high school students to discover and explore extracurricular opportunities.
- Designed a personalized recommendation system leveraging User-Based Collaborative Filtering algorithm to suggest activities aligned with users' interests, based on real-time browsing behavior analysis.
- Integrated PhoBERT model using PyTorch to efficiently detect and filter Vietnamese toxic comments.

Programming Competitions

- Ranked 14th, 26th, and 21st in The 2023, 2022, and 2021 ICPC East Central NA Regional Contest (Profile)
- Ranked 260th, 196th, 252nd, 281st in Google Kick Start Round A, B, G, H 2022 (Certificate)
- Ranked 1597th and 928th in Meta Hacker Cup Round 2 2022 (Certificate) and 2023 (Certificate)
- Top 0.3% contestants in LeetCode contest platform (Profile)

Technical Skills

Programming Languages: Python, C++, SQL, JavaScript
Backend Frameworks: Flask, FastAPI, NodeJS
Machine Learning Frameworks: TensorFlow, PyTorch, Scikit-learn
Cloud Platform: Microsoft Azure, Google Cloud
Miscellaneous: Git, LaTeX