Rui Li

📞 716-449-6324 ✉ rli35@buffalo.edu

EDUCATION

University at Buffalo **Aug. 2018 − Dec. 2023 (expected)**
Phd in Computer Science and Engineering, supervised by Prof. Jing Gao *Buffalo, NY, USA*

Zhejiang University **Aug. 2015 − Sep. 2018**
Master in Information Science and Electronic Engineering *Hangzhou, Zhejiang, China*

Zhejiang University **Aug. 2011 − Sep. 2015**
Bachelor in Information Science and Electronic Engineering *Hangzhou, Zhejiang, China*

INTERNSHIP

Mayo Clinic **06 2022 − 12 2022, 04 2023 − 05 2023**
Machine Learning Engineer Intern *Rochester, MN, USA*

- Designed and implemented a **time-aware transformer** based rare disease differential diagnoses model, and F1 increases 17% and recall increases 46%.
- Designed and implemented a module combining **Meta Learning** (Meta Weight Net) and **generative adversarial network** (GAN) aiming to solve the imbalanced and positive unlabeled challenge in rare disease detection, recall increases 80% and F1 increases 25% for skewed data containing hundred thousands samples with 5% positive rate.
- Design and implemented a model combining **diffusion models** and nnPU aiming to solve the imbalanced and positive unlabeled challenge in rare disease detection, recall increases 40% and F1 increases 35% for highly skewed data containing hundred thousands samples with 1% positive rate.
- Integrating clinical narratives and diagnosis codes to design a **knowledge-guided multi-modal** rare disease early detection method, a **reinforcement learning** based agent is used to decide whether or not to halt at each time step.

Didi Chuxing **03 2017 − 07 2017**
Machine Learning Engineer Intern *Hangzhou, Zhejiang, China*

- Designed the predictive model to predict the attrition rate of the driver. We select attributes such as drivers' income, driving time, driving distance etc. to train a logistic regression model.
- Writing SQL queries and working with product managers to help the decision making process.

TECHNICAL SKILLS

Programming Languages/framework: Python,C, C++, SQL, Matlab, Pytorch, Tensorflow
Machine Learning Models: meta learning, generative adversarial network, diffusion model, multi-modal representation learning, reinforcement learning

Research Interests

My research interest mainly resides in using deep learning techniques such as meta learning, generative adversarial networks, diffusion models and multi-modal representation learning to help the prediction problem in healthcare. And I also did some research about federated learning.

PROJECTS

Multi-modal multi-label classification task in EHR data

- Consider two modalities including diagnosis codes and clinical features. Diagnosis codes are **hierarchical** and clinical features are **multivariate irregularly-sampled time series**.
- Proposed and implemented the **transformer in hyperbolic space** to model the sequential dependency among multiple admissions.
- Regarded clinical feature as a significant complement of diagnosis codes, and design a weight adjustment mechanism which **learns the correlation between two modalities**, and adjusts the weight of clinical features. This work has been published on [AMIA2021].
- Proposed and implemented the **multi-modal contrastive learning framework** to compute the inter-modal contrastive loss between two modalities. For multi-label classification task, we propose the **multi-label contrastive loss** which uses the hierarchical structure among multiple labels. AUC increases 7% and this work has been published to [ICHI 2022].

Online Federated Multitask Learning

- Proposed and implemented an **online federated multi-task learning algorithm**, which learns the model parameters for the new device without revisiting the data of existing devices.
- Our framework achieves comparable accuracy to the existing algorithms but with much smaller computation, transmission and storage cost. This work has been published on [BigData2019].

Post-Hoc Local Interpretation via Sequential Rules

- Proposed and implemented an effective local sampling method for sequential data, which utilizes the historical information of features and labels.
- Proposed and implemented the deep first search sequential rule mining algorithm that helps to illuminate the illness trajectory of the target patient and explain the transition of which feature causes the transition of the label.

MANUSCRIPTS

[1] **Rui Li**, Andrew Wen, Jing Gao, Hongfang Liu. *MLGAN: a Meta-Learning based Generative Adversarial Network adapter for Rare Disease Differentiation.*

[2] **Rui Li**, Andrew Wen, Jing Gao, Hongfang Liu. *Time-aware Transformer for Rare Disease Detection*

[3] **Rui Li**, Andrew Wen, Jing Gao, Hongfang Liu. *Positive Unlabeled Learning with Diffusion Augmentation for Rare Disease Detection*

[4] **Rui Li**, Fenglong Ma, Jing Gao. *Post-Hoc Local Interpretation via Sequential Rules in Healthcare*

PUBLICATIONS

[BigData 2019] **Rui Li**, Fenglong Ma, Wenjun Jiang, Jing Gao. *Online federated multitask learning*, IEEE International Conference on Big Data, 2019. [paper, code]

[AMIA 2021] **Rui Li**, Fenglong Ma, Jing Gao. *Integrating Multimodal Electronic Health Records for Diagnosis Prediction*, AMIA Annual Symposium Proceedings, 2021. [paper, video]

[ICHI 2022] **Rui Li**, Jing Gao. *Multi-modal Contrastive Learning for Healthcare Data Analytics*, IEEE International Conference on Healthcare Informatics, 2022.