

Feng Jiang

+1 (682) 208 3748 - alanfengjiang@gmail.com - UTA 3rd Year PhD - Arlington, TX, USA

EDUCATION

The University of Texas at Arlington
GPA: 4.0/4.0

Ph.D. in Computer Science
Fall 2022 - Present, Arlington, TX

The University of Texas at Arlington
GPA: 4.0/4.0

M.S. in Software Engineering
Fall 2021 - Fall 2022, Arlington, TX

Xi'an Jiaotong University City College
GPA: 3.8/4.0

B.E. in Software Engineering
Fall 2015 - Fall 2019, Xi'an, China

RECENT HONORS

Best Ph.D

John S. Schuchman Outstanding Doctoral Student

The University of Texas at Arlington
Apr 2023

RESEARCH INTERESTS

- Large Foundation Model & Multimodal Large Language Model
- Graph Neural Networks
- Deep AUC Maximization
- Long-tailed Learning

TECHNICAL SKILLS

Programming Languages: Python, PHP, Java, SQL

Libraries and Tools: PyTorch, Sklearn, Pandas, Numpy, Hugging Face, Git, Docker, Laravel

ML Architectures: CNN, GNN, BERT, LSTM, LoRA, BLIP2, LLaVA, Distributed Data Parallel

RESEARCH EXPERIENCE

- **(2024.2-Now) Multimodal Large Language Model.**
 - LLaPA: We leverage multimodal approaches using large language models to enhance protein function prediction. Our method integrates protein sequences and functional descriptions from UniProt with contrastive learning techniques to predict the potential functions of unknown sequences. The model demonstrates superior performance in protein-text retrieval tasks and downstream predictions, while also showing reliable capabilities in both protein sequence and functional description generation.
 - AlphaEpi: We combine sequence and structure information by using the protein language model ESM-2 for sequence data and AlphaFold 3 for structural data. A dynamic alignment module merges these modalities, leading to exceptional performance in B-cell epitope prediction. *This work has been accepted to the ACM BCB 2024 Conference.*
 - Toxmm: We developed a multi-modal molecular embedding model for toxicity prediction by integrating SMILES structures and their textual descriptions. We curated a comprehensive dataset of over 50,000 molecular SMILES-text pairs from PubChem database for pre-training. Our approach achieved state-of-the-art performance on the toxicity dataset.
- **(2023.7-2024.3) A Graph Framework for TCR and Epitopes Binding Prediction.**
 - Developed a novel graph-based framework that models TCR-epitope binding as a topology learning problem, utilizing residue interaction patterns and structural information to capture complex binding mechanisms.

- Enhanced model performance through deep AUC maximization and dynamic graph attention mechanisms, achieving state-of-the-art performance on multiple benchmark datasets.
- Our method significantly outperforms existing approaches in both accuracy and efficiency for TCR-epitope binding prediction. *Published in Briefings in Bioinformatics.*
- **(2023.3-2024.1) SAM-based Auto-prompting Fine-tuning Framework.**
 - Segment Any Cell is an innovative framework designed to enhance the Segmentation Anything Model (SAM) for nuclei segmentation.
 - Low-Rank Adaptation (LoRA) is integrated within the attention layer of the Transformer to improve the fine-tuning process. An auto prompt generator is introduced to guide the segmentation process.
- **(2022.3-2023.4) LLM in Protein Structure Prediction.**
 - Developed a novel teacher-student framework to address the challenge of limited structural data in protein research. The Teacher model, trained on over 20,000 high-quality labeled structures, generates reliable pseudo labels to guide the training of a Student protein language model across millions of sequences.
 - Published in BCB'22 and accepted by the Journal of Computational Biology 2024.

PUBLICATION

- **Jiang, Feng**, et al. "GTE: a graph learning framework for prediction of T-cell receptors and epitopes binding specificity." *Briefings in Bioinformatics* 25.4 (2024).
- Weizhi An, Wenliang Zhong, **Feng Jiang**, Hehuan Ma and Junzhou Huang, "Causal Subgraphs and Information Bottlenecks: Redefining OOD Robustness in Graph Neural Networks", In Proc. of the 18th European Conference on Computer Vision, ECCV'24, Milan, Italy, October 2024.
- Hehuan Ma, **Feng Jiang**, Yuzhi Guo and Junzhou Huang, "Towards Robust Self-training for Molecular Biology Prediction Tasks, *Journal of Computational Biology*, Volume 31, Issue 3, pp. 213-228, March 2024.
- Na, S., Guo, Y., **Jiang, F.**, Ma, H., & Huang, J. (2024). Segment Any Cell: A SAM-based Auto-prompting Fine-tuning Framework for Nuclei Segmentation. *arXiv preprint arXiv:2401.13220*.
- Ma, H., **Jiang, F.**, Rong, Y., Guo, Y., & Huang, J. (2022, August). Robust self-training strategy for various molecular biology prediction tasks. In *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 1-5).

TEACHING ASSISTANT EXPERIENCE

- 2248-CSE-6324-980-ADV TOPS SOFTWARE ENGINEERING *Fall 2024*
- 2245-CSE-3314-002-PROFESSIONAL PRACTICES *Summer 2024*
- 2242-CSE-6324-001-ADV TOPS SOFTWARE ENGINEERING *Spring 2024*
- 2238-CSE-6324-980-ADV TOPS SOFTWARE ENGINEERING *Fall 2023*
- 2232-CSE-6324-001-ADV TOPS SOFTWARE ENGINEERING *Spring 2023*
- 2228-CSE-6324-980-ADV TOPS SOFTWARE ENGINEERING *Fall 2022*
- 2228-CSE-5325-003-SFWR ENG II: MGMT, MAIN, & QA *Fall 2022*