

Fei He

(573)818-4738 | <https://coffee19850519.github.io> | hefe@umsystem.edu

EDUCATION

University of Missouri-Columbia Columbia, MO
Ph.D. candidate in Engineering January 2023 – December 2025
Department of Electrical Engineering and Computer Science, College of Engineering Advisor: [Dong Xu](#)
Emphasis: Artificial Intelligence

Jilin University Changchun, Jilin, China
Ph.D. in Engineering September 2010 – June 2015
Department of Bioinformatics, College of Computer Science and Technology Advisor: Yuanning Liu
Emphasis: Biometrics

Jilin University Changchun, Jilin, China
Master of Engineering September 2007 – June 2010
Department of Bioinformatics, College of Computer Science and Technology Advisor: Yuanning Liu
Emphasis: Biometrics

Jilin University Changchun, Jilin, China
Bachelor of Engineering September 2003 – June 2007
College of Computer Science and Technology

PROFESSIONAL EXPERIENCE

University of Missouri-Columbia Columbia, MO
Department of Electrical Engineering and Computer Science, College of Engineering
Graduate Research January 2023 – December 2025

Stowers Institute Kansas City, MO
Laboratory Assistant Intern June 2025 – July 2025
Goal: Survey Recent Advancements in single-cell Large Language Models Supervisor: [Hua Li](#)

Northeast Normal University Changchun, Jilin, China
Department of Computer Science, School of Information Science and Technology
Associate Professor July 2022 – November 2022

University of Missouri-Columbia Columbia, MO
Bond Life Science Center
Visiting Scholar October 2018 – January 2021

Northeast Normal University Changchun, Jilin, China
Department of Computer Science, School of Information Science and Technology
Lecturer August 2015 – June 2022

CAREER HIGHLIGHTS

-
- Published over 50 peer-reviewed papers in reputed journals and conferences in venues such as *Nature Machine Intelligence*, *Nucleic Acids Research*, *IEEE Journal of Biomedical and Health Informatics*, *International Conference on Bioinformatics and Biomedicine*, etc. (h-index:20, i10-index: 32, citation > 1400 according to Google Scholar)

- Secured ¥545,000 in sponsored research funding as Principal Investigator across 5 projects, supported by the National Natural Science Foundation of China, Jilin Provincial Natural Science Foundation, and others.
- Developed or co-developed 7 tools/web-services/databases serving single cell, protein, and plant research communities with a total of more than 5k users.
- Presented research through talks and posters at over 10 worldwide Bioinformatics and Biomedicine conferences such as *ISMB*, *AMIA*, *BIBM*, etc.
- Taught over 6 undergraduate and graduate courses, including bilingual instruction for medium-to-large class sizes (~20-50 students), with highly positive feedback (average teaching evaluation score: 90.6/100)
- Mentored or co-mentored 2 PhD students and 16 master's Graduate Students.
- Served as a reviewer for over 10 major journals/conferences such as *Nature Communications*, *Pattern Recognition*, *Briefings in Bioinformatics*, *Neurocomputing*, etc, spanning the fields of Bioinformatics, Biomedicine, Machine Learning, Natural Language Processing, and Computer Vision.

RESEARCH INTERESTS

- Large Language Model (LLM) and Deep Learning (DL) have been making waves in diverse science and they are transforming diverse fields across science and technology. My recent research focuses on developing and adapting LLMs and DL models for a wide range of biological and biomedical applications, with the goals of uncovering underlying molecular programs from single-cell data, elucidating functional mechanisms in protein analysis, and facilitating knowledge curation in biomedical research.
- We are actively developing computational tools and databases to characterize the composition, interactions, and functions of single-cell and protein data, and to investigate their connections with human diseases, drug target discovery, and plant biology. Our efforts integrate data mining, computational analysis, and network-based methods. Additionally, we are also exploring strategies to enhance data privacy and security in the development and deployment of these computational resources.

RECENT RESEARCH

Large Language Models for Single-Cell Analysis

- Developed *scPEFT* to efficiently adapt multiple single-cell Large Language Models (e.g., *scGPT*, *Geneformer*, and *scBERT*) for out-of-distribution application scenarios (e.g., disease-specific datasets, cross-species comparisons, and under-characterized cell groups)
- Reduced parameter tuning by at least 96% and lowered GPU memory costs by more than 50%
- Key developer in NIH-R35 Grant *Multi-view Self-Supervised Deep Learning for Biological Sequences and Beyond* (PI: Dong Xu, 2024-2028)
- Led collaborative research with Prof. [Qin Ma lab](#) from Ohio State University, resulting in a 1st author submission under **revision** at *Nature Machine Intelligence*.

Natural Language Processing for Health Informatics

- Developed multi-modal text-image understanding tool pathCLIP to extract gene interactions from literature pathway figures
- Designed prompt engineering strategies to mine biological events from biomedical publications with large language models
- Key developer in NIH-R01 Grant *Image-guided Bio-curation of Disease Pathways from Scientific Literature* (PI: Mihail Popescu, 2020-2023)
- Led collaborative research with Prof. [Mihail Popescu](#) lab from MU School of Medicine, resulting in 10+ publications in *IEEE Journal of Biomedical and Health Informatics*, *IEEE International Conference on Bioinformatics and Biomedicine*, etc.

Domain-Specific Algorithm Development for Diverse Computational Biology Applications

- Leveraged protein language models and deep learning models for protein sequence applications such as protein modification, interaction, and binding site predictions
- Introduced point cloud and graph neural networks for protein structure analyses, such as protein docking and functional predictions
- Constructed expression spectrum visualizations and image segmentation models for discovering tissue architecture and functional regions with spatial transcriptomic data
- Led (co-led) in 10+ publications in *Nucleic Acids Research*, *Computational and structural biotechnology journal*, *BMC Bioinformatics*, *BMC Genomics*, etc.

GRANTS

Jilin Provincial Natural Science Foundation, China <i>General fund</i>	PI, ¥ 100,000 2021-2022
<ul style="list-style-type: none">• Title: Protein Docking Conformation Ranking Based on Capsule Networks	
National Natural Science Foundation of China <i>Young Scientists Fund</i>	PI, ¥ 260,000 2019-2021
<ul style="list-style-type: none">• Title: Target Protein–Ligand Complex Activity Prediction Integrating Multi-view 3D Depth Descriptors	
Jilin Provincial Department of Education, China <i>Scientific Research Project</i>	PI, ¥ 25,000 2019-2020
<ul style="list-style-type: none">• Title: Hybrid Deep Learning Model for Dynamic Protein Interaction Networks	
Jilin Provincial Natural Science Foundation, China <i>Outstanding Youth Fund</i>	PI, ¥ 80,000 2017-2018
<ul style="list-style-type: none">• Title: Deep Learning Network with Adaptive and Self-Learning Characteristics for Texture Analysis	

PUBLICATIONS

As of January 1, 2026, h-index:20, i10-index: 32, citation: 1446 according to [Google Scholar](#)

Preprints/Manuscripts Under Review and In Press

1. He, F., Wang, X., & Xu, D (2025). Reusability Report: Pan-Peptide Meta Learning for T-cell receptor–antigen binding recognition. *Nature Machine Intelligence*. Under revision.
2. He, F., Fei, R., Gao, M., Su, L., Zhang, X., & Xu, D. (2024). Parameter-Efficient Fine-Tuning Enhances Adaptation of Single Cell Large Language Model for Cell Type Identification. *bioRxiv*.
3. Manshour, N., He, F., Wang, D., & Xu, D. (2024). Integrating protein structure prediction and Bayesian optimization for peptide design. *Research Square*, 3, 4045284.

Journal Articles (*co-first author, †co-corresponding author)

1. He, F., Fei, R., Krull, J. E., Zhang, X., Gao, M., Su, L., Chen, Y., Yu, Y., Li, J., Jin, B., Chang, Y., Ma, A., Ma, Q., & Xu, D. (2025). Harnessing the power of single-cell large language models with parameter-efficient fine-tuning using scPEFT. *Nature Machine Intelligence*. 1-16. <https://doi.org/10.1038/s42256-025-01170-z>

2. Han, Y., **He, F.**, Shao, Q., Wang, D., & Xu, D. (2025). MTPrompt-PTM: A Multi-Task Method for Post-Translational Modification Prediction Using Prompt Tuning on a Structure-Aware Protein Language Model. *Biomolecules*, 15(6), 843. **Cover paper**
3. Essien, C.^{*}, Wang, N.^{*}, Yu, Y., Alqarghuli, S., Qin, Y., Manshour, N., **He, F.**[†], & Xu, D.[†]. (2025). Predicting the location of coordinated metal ion-ligand binding sites using geometry-aware graph neural networks. *Computational and Structural Biotechnology Journal*, 27, 137-148.
4. Xu, C., Shaw, T., Choppararu, S. A., Lu, Y., Farooq, S. N., Qin, Y., Hudson, M., Weekley, B., Fisher, M., **He, F.**, Da Silva Nascimento, J. R., Wergeles, N., Joshi, T., Bates, P. D., Koo, A. J., Allen, D. K., Cahoon, E. B., Thelen, J. J., & Xu, D. (2024). FatPlants: a comprehensive information system for lipid-related genes and metabolic pathways in plants. *Database*, 2024, baac074.
5. He, Y., Huang, R., Zhang, R., **He, F.**[†], Han, L.[†], & Han, W.[†]. (2024). PredCoffee: A binary classification approach specifically for coffee odor. *Isience*, 27(6).
6. Song, R., Liu, K., He, Q., **He, F.**[†], & Han, W.[†]. (2024). Exploring bitter and sweet: the application of large language models in molecular taste prediction. *Journal of Chemical Information and Modeling*, 64(10), 4102-4111.
7. Lin, Y., Liang, Y., Wang, D., Chang, Y., Ma, Q., Wang, Y.[†], **He, F.**[†], & Xu, D.[†]. (2024). A contrastive learning approach to integrate spatial transcriptomics and histological images. *Computational and Structural Biotechnology Journal*, 23, 1786-1795.
8. **He, F.**, Liu, K., Yang, Z., Chen, Y., Hammer, R. D., Xu, D., & Popescu, M. (2024). pathclip: Detection of genes and gene relations from biological pathway figures through image-text contrastive learning. *IEEE Journal of Biomedical and Health Informatics*, 28(8), 5007-5019.
9. **He, F.**, Liu, K., Yang, Z., Hannink, M., Hammer, R. D., Popescu, M., & Xu, D. (2023). Applications of cutting-edge artificial intelligence technologies in biomedical literature and document mining. *Medical Review*, 3(3), 200-204.
10. Wang, D., **He, F.**, Yu, Y., & Xu, D. (2023). Meta-learning for T cell receptor binding specificity and beyond. *Nature machine intelligence*, 5(4), 337-339.
11. Han, Y., Zhang, S., & **He, F.** (2023). A Point Cloud-Based Deep Learning Model for Protein Docking Decoys Evaluation. *Mathematics*, 11(8), 1817.
12. Azam, M., Arowolo, M. O., **He, F.**, Popescu, M., & Xu, D. (2023). Recognition of Gene Names from Gene Pathway Figures Using Siamese Network. *International Journal of Bioengineering and Life Sciences*, 23-30.
13. Gong, J., Zhao, Y., Heng, X., Chen, Y., Sun, P.[†], **He, F.**[†], Ma, Z.[†], & Ren, Z.[†] (2023). Deciphering and identifying pan-cancer RAS pathway activation based on graph autoencoder and Classifier Chain. *Electronic Research Archive*, 31(8), 4951-4967.
14. Chen, Y., Qin, W., Liu, T., Li, R., **He, F.**[†], Han, Y.[†], Ma, Z.[†], & Ren, Z.[†] (2023). MTNA: A deep learning-based predictor for identifying multiple types of N-terminal protein acetylated sites. *Electronic Research Archive*, 31(9), 5442-5456.
15. Lin, Y., Wang, Y.[†], Liang, Y., Yu, Y., Li, J., Ma, Q., **He, F.**[†], & Xu, D.[†] (2022). Sampling and ranking spatial transcriptomics data embeddings to identify tissue architecture. *Frontiers in Genetics*, 2007.
16. Chang, Y.^{*}, **He, F.**^{*}, Wang, J.^{*}, Chen, S., Li, J., Liu, J., Yu, Y., Su, L., Ma, A., Allen, C., Lin, Y., Sun, S., Liu, B., Otero J., Chung D., Fu, H., Li, Z.[†], Xu, D.[†], Ma, Q.[†]. (2022). Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning. *Computational and structural biotechnology journal*, 20, 4600-4617.
17. Tang, B., He, F., Liu, D., **He, F.**, Wu, T., Fang, M., Niu, Z., Wu, Z., & Xu, D. (2022). AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *Biomolecules*, 12(6), 746.
18. Han, Y., Chen, Y., **He, F.**, Qin, W., Yu, H.[†], & Xu, D.[†] (2021). Quality Assessment of Protein Docking Models Based on Graph Neural Network. *Frontiers in Bioinformatics*, 1, 693211.
19. Ji, J., Li, R., Pang, W.[†], **He, F.**, Feng, G., & Zhao, X.[†] (2021). A multi-view clustering algorithm for mixed numeric and categorical data. *IEEE Access*, 9, 24913-24924.

20. **He, F.**, Li, J., Wang, R., Zhao, X., & Han, Y. (2021). an ensemble deep learning-based predictor for simultaneously identifying protein ubiquitylation and SUMOylation sites. *BMC bioinformatics*, 22, 15-Jan.
21. Li, Y., Pu, F., Wang, J., Zhou, Z., Zhang, C., **He, F.**[†], Ma, Z.[†], & Zhang, J.[†] (2021). Machine Learning Methods in Prediction of Protein Palmitoylation Sites: A Brief Review. *Current Pharmaceutical Design*, 2021(27), 2189-2198.
22. Khan, S. M., **He, F.**, Wang, D., Chen, Y., & Xu, D. (2020). MU-PseUDeep: a deep learning method for prediction of pseudouridine sites. *Computational and Structural Biotechnology Journal*, 18, 1877-1883.
23. Wang, D., Liu, D., Yuchi, J., **He, F.**, Jiang, Y., Cai, S., Li, J., & Xu, D. (2020). MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Research*, 48(W1), W140-W146.
24. Ji, J., Pang, W.[†], Li, Z., **He, F.**, Feng, G., & Zhao, X.[†] (2020). Clustering mixed numeric and categorical data with cuckoo search. *IEEE Access*, 8, 30988-31003.
25. Sun, P. P., Chen, Y. B., Liu, B., Gao, Y. X., Han, Y., **He, F.**[†], & Ji, J. C.[†] (2019). DeepMRMP: A new predictor for multiple types of RNA modification sites using deep learning. *Math. Biosci. Eng.*, 16(6), 6231-6241.
26. Gong, J., Chen, Y., Pu, F., Sun, P., **He, F.**, Zhang, L., Li, Y.[†], Ma, Z.[†], & Wang, H.[†] (2019). Understanding membrane protein drug targets in computational perspective. *Current Drug Targets*, 20(5), 551-564.
27. Ji, J., Chen, Y., Feng, G., Zhao, X.[†], & **He, F.**[†] (2019). Clustering mixed numeric and categorical data with artificial bee colony strategy. *Journal of Intelligent & Fuzzy Systems*, 36(2), 1521-1530.
28. Lu, C., Liu, Z., Zhang, E., **He, F.**[†], Ma, Z.[†], & Wang, H.[†] (2019). MPLs-Pred: predicting membrane protein-ligand binding sites using hybrid sequence-based features and ligand-specific models. *International journal of molecular sciences*, 20(13), 3120.
29. Liu, Y., Wang, D., **He, F.**, Wang, J., Joshi, T., & Xu, D. (2019). Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean. *Frontiers in genetics*, 10, 1091.
30. Han, Y., **He, F.**, Chen, Y., Liu, Y., & Yu, H. (2018). SiRNA silencing efficacy prediction based on a deep architecture. *BMC genomics*, 19, 59-65.
31. Zhao, X., Li, J., Wang, R., **He, F.**, Yue, L., & Yin, M. (2018). General and species-specific lysine acetylation site prediction using a bi-modal deep architecture. *IEEE Access*, 6, 63560-63569.
32. **He, F.**, Wang, R., Li, J., Bao, L., Xu, D., & Zhao, X. (2018). Large-scale prediction of protein ubiquitination sites using a multimodal deep architecture. *BMC systems biology*, 12, 81-90.
33. **He, F.**, Han, Y., Gong, J., Song, J., Wang, H., & Li, Y. (2017). Predicting siRNA efficacy based on multiple selective siRNA representations and their combination at score level. *Scientific Reports*, 7(1), 44836.
34. Han, Y., Liu, Y., Zhang, H., **He, F.**, Shu, C., & Dong, L. (2017). Utilizing Selected Di-and Trinucleotides of siRNA to Predict RNAi Activity. *Computational and Mathematical Methods in Medicine*, 2017(1), 5043984.
35. **He, F.**, Han, Y., Wang, H., Ji, J., Liu, Y., & Ma, Z. (2017). Deep learning architecture for iris recognition based on optimal Gabor filters and deep belief network. *Journal of Electronic Imaging*, 26(2), 023005-023005.
36. Liu, Y., **He, F.**, Zhu, X., Liu, Z., Chen, Y., Han, Y., & Yu, L. (2015). The improved characteristics of bionic Gabor representations by combining with SIFT key-points for iris recognition. *Journal of Bionic Engineering*, 12(3), 504-517.
37. Huo, G., Liu, Y., Zhu, X., Dong, H., & **He, F.** (2015). Face-iris multimodal biometric scheme based on feature level fusion. *Journal of Electronic Imaging*, 24(6), 063020-063020.

38. Chen, Y., Liu, Y., Zhu, X., Chen, H., **He, F.**, & Pang, Y. (2014). Novel approaches to improve iris recognition system performance based on local quality evaluation and feature fusion. *The Scientific World Journal*, 2014(1), 670934.
39. **He, F.**, Liu, Y., Zhu, X., Huang, C., Han, Y., & Chen, Y. (2014). Score level fusion scheme based on adaptive local Gabor features for face-iris-fingerprint multimodal biometric. *Journal of Electronic Imaging*, 23(3), 033019-033019.
40. Chen, Y., Liu, Y., Zhu, X., **He, F.**, Wang, H., & Deng, N. (2014). Efficient iris recognition based on optimal subfeature selection and weighted subregion fusion. *The Scientific World Journal*, 2014(1), 157173.
41. Liu, Y., **He, F.**, Zhu, X., Chen, Y., Han, Y., & Fu, Y. (2014). Video sequence-based iris recognition inspired by human cognition manner. *Journal of Bionic Engineering*, 11(3), 481-489.
42. **He, F.**, Liu, Y., Zhu, X., Huang, C., Han, Y., & Dong, H. (2014). Multiple local feature representations and their fusion based on an SVR model for iris recognition using optimized Gabor filters. *EURASIP Journal on Advances in Signal Processing*, 2014, 17-Jan.

Conference Papers

1. Azam, M., **He, F.**, Santosh, K., Xu, D., & Popescu, M. (2024). Evaluation and Integration of Advanced AI Chatbots for Biological Pathway Curation. *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 6342-6348.
2. Arowolo, M. O., Azam, M., **He, F.**, Popescu, M., & Xu, D. (2024). Gene Name Recognition in Gene Pathway Figures Using Siamese Networks. *2024 IEEE International Conference on Medical Artificial Intelligence (MedAI)*, 218-230.
3. **He, F.**, Yang, Z., Gao, M., Poudel, B., Dhas, N. S. E. S., Gyawali, R., Dhakal, A., Cheng, J., & Xu, D. (2024). Adapting segment anything model (SAM) through prompt-based learning for enhanced protein identification in cryo-EM micrographs. *2024 IEEE International Conference on Medical Artificial Intelligence (MedAI)*, 20-Sep.
4. Bao, Y., Wang, W., Dong, M., **He, F.**, & Wang, H. (2021). Discover the Binding Domain of Transmembrane Proteins Based on Structural Universality. *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 11-May.
5. **He, F.**, Thompson, J., Mao, Z., Ren, Y., Nussbaum, Y., Kholod, O., Shin, D., Hannink, M., Popescu, M., & Xu, D. (2021). Identifying genes and their interactions from pathway figures and text in biomedical articles. *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 398-405.
6. Thompson, J., Dong, H., Liu, K., **He, F.**, Popescu, M., & Xu, D. (2021). A rule-based approach for generating synthetic biological pathways. *International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*, 105-116.
7. Ren, Y., **He, F.**, Qu, J., Li, Y., Thompson, J., Hannink, M., Popescu, M., & Xu, D. (2021). Text Mining Enhancements for Image Recognition of Gene Names and Gene Relations. *International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*, 132-142.
8. **He, F.**, Wang, D., Innokenteva, Y., Kholod, O., Shin, D., & Xu, D. (2019). Extracting molecular entities and their interactions from pathway figures based on deep learning. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 397-404.
9. **He, F.**, Wang, R., Gao, Y., Wang, D., Yu, Y., Xu, D., & Zhao, X. (2019). Protein ubiquitylation and sumoylation site prediction based on ensemble and transfer learning. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 117-123.
10. Han, Y., **He, F.**, Tan, X., & Yu, H. (2017). Effective small interfering RNA design based on convolutional neural network. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 16-21.
11. **He, F.**, Bao, L., Wang, R., Li, J., Xu, D., & Zhao, X. (2017). A multimodal deep architecture for large-scale protein ubiquitylation site prediction. *2017 IEEE international conference on bioinformatics and biomedicine (BIBM)*, 108-113.

12. Huo, G., Liu, Y., Zhu, X., Huang, C., **He, F.**, Gao, S., Dong, H., Yu, L., & Yang, X. (2015). An iris recognition method based on annule-energy feature. *Biometric Recognition: 10th Chinese Conference, CCBR 2015, Tianjin, China, November 13-15, 2015, Proceedings 10*, 341-348.
13. Huo, G., Liu, Y., Zhu, X., Wang, H., Yu, L., **He, F.**, Gao, S., & Dong, H. (2014). An effective iris recognition method based on scale invariant feature transformation. *Biometric Recognition: 9th Chinese Conference, CCBR 2014, Shenyang, China, November 7-9, 2014. Proceedings 9*, 301-310.
14. **He, F.**, Liu, Y., Zhu, X., Deng, W., Zhang, X., & Huo, G. (2013). The affection of Gabor parameters to iris recognition and their optimization. *Biometric Recognition: 8th Chinese Conference, CCBR 2013, Jinan, China, November 16-17, 2013. Proceedings*, 330-337.

Posters

1. **He, F.**, Fei, R., Krull, J., Zhang, X., Chang, Y., Gao, M., Su, L., Chen, Y., Yu, Y., Ma, Q., & Xu, D. (2024). Harnessing the power of single cell large language models with parameter efficient fine-tuning. *2024 International Conference on Intelligent Biology and Medicine, ICIBM 2024, Houston TX*.
2. **He, F.**, Chen, Y., Petruc, M., Xu, D., & Popescu, M. (2023). Ten-year pathway figure mining and analysis for non-small cell lung cancer. *American Medical Informatics Association 2023 Annual Symposium, New Orleans, LA*.
3. **He, F.**, Chen, Y., Petruc, M., Xu, D., & Popescu, M. (2023). Ten-year pathway figure mining and analysis for non-small cell lung cancer. *The 8th Workshop on Biostatistics and Bioinformatics, Atlanta, GA*.

DEVELOPED TOOLS, WEB SERVICES, AND DATABASES

scPEFT: a single-cell parameter-efficient fine-tuning framework that integrates learnable, low-dimensional adapters into scLLMs for out-of-context applications, such as disease-conditional cell type identification, conditioned bio-marker discovery, cross-species adaptation, and uncharacterized cell population detection.

- <https://github.com/coffee19850519/scPEFT>

pathCLIP: An image-text contrastive learning framework that coordinates embeddings of image snippets and text descriptions for genes and gene relation curation across diverse literature sources

- <https://github.com/yangyang-69/pathCLIP>

Prompt-SAM-cryoPPP: A protein identification framework for cryo-electron microscopy (cryo-EM) utilizing prompt-based learning strategies with Segment Anything Model (SAM), to enhance performance and reduce computational requirements in 3D protein structure construction.

- https://github.com/yangyang-69/Prompt_sam_cryoPPP

fatPlants: A comprehensive website platform for plant fat-related genes, proteins and metabolic pathways, currently providing all fatty acid related proteins and genes retrieve search for these 3 species, and most fatty acid chemical properties. Co-developed with Chunhui Xu.

- <https://www.fatplants.net>

MuSite: A web-based tool that predicts and visualizes post-translational modification (PTM) and ion-binding sites on proteins using deep neural networks. In addition, the server maintains a local database providing pre-processed PTM annotations from Uniport/Swiss-Prot. Co-developed with Dr. Duolin Wang.

- <https://www.musite.net>

RESEPT: A deep-learning framework for characterizing and visualizing tissue architecture from spatially resolved transcriptomics by learning a three-dimensional embedding mapped into an RGB image for leveraging image segmentation. Co-developed with Yuzhou Chang.

- <https://github.com/OSU-BMBL/RESEPT>

PPDocking: A graph neural network framework for scoring protein-protein docking conformations, for improving structural bioinformatics predictions. Co-developed with Dr. Ye Han.

- <https://github.com/coffee19850519/PPDocking>

INVITED TALKS AND PRESENTATIONS

Saint Louis University Bioinformatics Program St. Louis, MO
Bioinformatics and Computational Biology Colloquium Series 2025

- Invited by Prof. [Jie Hou](#)
- Title: Applications of AI and Large Language Models in Biology and Biomedicine

Shanghai Institute for Mathematics and Interdisciplinary Sciences Online
Summer Seminar Series in Shanghai 2025

- Invited by Prof. [Zhigang Yao](#)
- Title: Deep Learning Methods for Single-Cell Data Analyses

The International Conference on Intelligent Biology and Medicine (ICIBM) Houston, TX
AI Methods for Single Cell Data Analysis 2024

- Title: Enhancing adaptation of single cell large language models in single cell analyses through parameter efficient fine-tuning

IEEE International Conference on Bioinformatics and Biomedicine (BIBM) Online
Data Mining, Machine Learning, and Artificial Intelligence for Biomedicine 2021

- Title: Identifying Genes and Their Interactions from Pathway Figures and Text in Biomedical Articles

IEEE International Conference on Bioinformatics and Biomedicine (BIBM) Online
Molecular Structure, Function and Evolution 2021

- Title: Discover the Binding Domain of Transmembrane Proteins Based on Structural Universality

28th Conference on Intelligent Systems for Molecular Biology (ISMB) Online
3DSIG: Structural Bioinformatics and Computational Biophysics 2020

- Title: Quality Assessment of Protein Docking Models Based on Graph Neural Network

IEEE International Conference on Bioinformatics and Biomedicine (BIBM) San Diego, CA
Molecular Structure, Function and Evolution 2019

- Title: Protein ubiquitylation and sumoylation site prediction based on ensemble and transfer learning

10th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB) Buffalo, NY
Deep learning III 2019

- Title: Extracting molecular entities and their interactions from pathway figures based on deep learning

IEEE International Conference on Bioinformatics and Biomedicine Kansas City, MO
Molecular Structure, Function and Evolution 2017

- Title: A multimodal deep architecture for large-scale protein ubiquitylation site prediction

ACADEMIC SERVICE

Editorship

Editor, International Journal of Artificial Intelligence and Robotics Research, 2026 - Now

Journal Reviewer

Nature Communications, Pattern Recognition, Briefings in Bioinformatics, Neurocomputing, International Journal of Biological Macromolecules, Life Sciences, Computers in Biology and Medicine, BMC Medical Informatics and Decision Making, Scientific Reports, Journal of Imaging Informatics in Medicine, Engineering Applications of Artificial Intelligence, Digital Chinese Medicine, etc.

Conference Reviewer

IEEE International Conference on Bioinformatics and Biomedicine (2019, 2023-2024, co-reviewed with Prof. Trupti Joshi).

Grant Reviewer

National Natural Science Foundation of China (2024-2025).

Thesis Reviewer

School of Information Science and Technology at Northeast Normal University, Northeast Electric Power University, Chinese Ministry of Education Degree Center's Quality Monitoring Service Platform for Doctoral and Master's Theses.

HONORS AND AWARDS

Outstanding Ph.D. Student Award	EECS MU 2025
Student and Trainee Travel Award	ICIBM 2024
Travel Fellowship Award	EECS MU 2024
Travel Award	LSC Postdoc, Graduate, and Undergraduate Association MU 2023
Travel Award	Interdisciplinary Plant Group MU 2023
Student and Trainee Travel Award	BIBM 2017
Outstanding Graduate Student	Jilin University 2014
First-class Undergraduate Scholarship	Jilin University 2007
Second-class Undergraduate Scholarship (3 times)	Jilin University 2004-2006

TEACHING EXPERIENCE

Advanced Methods in Deep Learning, University of Missouri Fall 2023, Fall 2024

- New graduate course, Assistant in Instruction for Prof. Dong Xu, ~30 students

Principles and Applications of Database, www.icourse163.org Online MOOC

- Undergraduate public MOOC, Co-Instruction with Prof. Yanling Li

Machine Learning and Deep Learning, Northeast Normal University Fall 2021

- Developed graduate-level elective course, Instructor, ~30 students

Operating Systems, Northeast Normal University Spring 2021, Spring 2022

- Undergraduate compulsory course, Instructor, ~50 students
- **Bilingual education** serving the Northeast Normal University–Kennesaw State University Global Education - Degree Program

Python Programming, Northeast Normal University Fall 2017, Fall 2018

- Undergraduate-level core course, Instructor, ~50 students

Bioinformatics, Northeast Normal University

Spring 2017, Spring 2018

- Developed graduate-level elective course, Instructor, ~20 students

C#.NET Programming, Northeast Normal University

Spring 2017, Spring 2018

- Undergraduate-level core course, Instructor, ~30 students

MENTORING EXPERIENCE

Ph.D. Students

Yongbing Chen (co-advised with Prof. Zhiguo Fu)

Northeast Normal University, 2020 - 2015

Jianting Gong (co-advised with Prof. Zhiqiang Ma)

Northeast Normal University, 2018 - 2022

M.S. Students

Xinyu Zhang, Xianyu Wang

Northeast Normal University, 2023 - present

Zhiyuan Yang, Ruixin Fei, Mingyue Gao, Ning Wang, Jialiang Zhang Northeast Normal University, 2022 - 2025

Kai Liu, Qiuyu Lyu, Yan Zhang, Lili Jiang, Chaoyang Ni

Northeast Normal University, 2021 - 2024

Wenyuan Qin

Northeast Normal University, 2020 - 2023

Jingyi Li, Yang Yu (co-advised with Dr. Pingping Sun)

Northeast Normal University, 2019 - 2022

SKILLS

Programming

Python, R, C#/C++/C

Data Analysis

NumPy, Pandas, SciPy, scikit-learn, Matplotlib, Scanpy, Seurat, BioPython, rdkit, AlphaFold, LIME, SHAP

Model Construction

PyTorch, HuggingFace Transformer, FlashAttention, TensorFlow, Keras

Languages

English (fluent), Mandarin (native), Cantonese (native)