

Yaowen Gu

yg3191@nyu.edu | [Github](#) | [Google Scholar](#) | [Web Page](#)

EDUCATION

New York University	2023-Present
Ph.D. in Chemistry	
Chinese Academy of Medical Science (CAMS)& Peking Union Medical	2020-2023
M.Med. in Medical Informatics (Advised by Prof. Jiao Li)	
Peking University	2016-2020
B.S. in Medical Experimental Technology (Advised by Prof. Zhengwei Xie)	

RESEARCH PROJECTS

Molecular Graph Learning (Advised by Prof. [Jiao Li](#) and Prof. [Liang Li](#) in CAMS)

- Proposed a curriculum learning training strategy called CurrMG for molecular graph learning, evaluating the performance of CurrMG on 5 GNN models and 8 molecular property prediction benchmarks (overall relative improvement 4.08%). (Publication #7 and 9)
- Designed advanced GNN framework for LBVS-based bioactivity prediction, anti-mycobacterium inhibitor VS, and drug ADMET prediction. (Publication #2, 4, and 10)

Computational Drug Repositioning (Advised by Prof. [Jiao Li](#) and Prof. [Rui Jiang](#) in THU)

- Organized a drug repositioning benchmark including 41k nodes and 1m edges with 5 biological entities; Proposed a heterogeneous GNN and attention mechanism-based method called REDDA for the drug-disease association (DDA) prediction. (Publication #6)
- Proposed a multi-instance learning-based heterogeneous GNN called MilGNet for DDA prediction, containing a pseudo meta-path generator, a bidirectional translating embedding method, and a multi-scale interpretable joint predictor. (Publication #1 and 5)

Gene-disease Prioritization (Advised by Prof. [Jiao Li](#))

- Established a multi-graph representation learning-based ensemble learning method for gene-disease association prediction, simultaneously utilizing genetic, therapeutic, and network topological features and has achieved AUC score of 0.924. (Publication #3)

Machine Learning Modeling on EHR Data (Advised by Prof. [Jiao Li](#))

- Constructed ensemble learning models/ Siamese network to predict the urinary tract infection risk, achieving desirable performances (AUC > 0.8). (Publication #16 and 17)
- Established multiple machine learning models methods for waiting time predictions in pediatric emergency department (R-square > 0.7). (Publication #12)

INTERSHIP EXPERIENCE

Xtalpi AI Research Center **Nov. 2020–Sept. 2021**

Algorithm intern (6 months) & Data science intern (3 months)

- Recurrent and deployed pretrained molecular graph models, and tested the performances of multiple GNN models on MoleculeNet and other ADMET datasets. (Patent 2); Proposed a curriculum learning methods for molecular graph learning and assembled it with Contextpred. (Patent 1)
- Designed a normalized data collecting, cleaning, and preprocessing workflow and other standards. Then integrated and preprocessed more than 100,000 molecular ADMET data collected from Reaxys, PubChem, ChEMBL for DL and AutoML model training.

SILEXON **Sept. 2022–Dec. 2022**

Algorithm intern (Advised by Prof. [Jianyang Zeng](#) in SILEXON and THU)

- Designed deep learning model for compound-protein interaction (CPI) prediction.

PUBLICATIONS

1. **Gu, Y.**, Zheng S., Zhang, B., Kang, H., Jiang, R., & Li, L. (2023). Meta-Path-Based Deep Multiple Instance Learning with Heterogeneous Graph Neural Network for Drug-disease Association Prediction. *IEEE Journal of Biomedical and Health Informatics*. (Under review round 1)
2. **Gu, Y.**, Li, J., Kang, H., Zhang, B., & Zheng, S. (2023). Employing Molecular Conformations for Ligand-based Virtual Screening with Equivariant Graph Neural Network and Deep Multiple Instance Learning. *Molecules*. (IF=4.6)
3. Wang Z.#, **Gu, Y.#**, Zheng, S., Yang, L., & Li, J. (2023). MGREL: A multi-graph representation learning-based ensemble learning method for gene-disease association prediction. *Computers in biology and medicine*. (IF=7.7)
4. **Gu, Y.**, Zheng, S., & Li, J. (2023). GNN-MTB: An Anti-Mycobacterium Drug Virtual Screening Method Based on Graph Neural Network. *Data Analysis and Knowledge Discovery*. (Chinese)
5. **Gu, Y.**, Zheng, S., Zhang, B., Kang H., & Li, J. (2022). MilGNet: A Multi-instance Learning-based Heterogeneous Graph Network for Drug repositioning. *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*.
6. **Gu, Y.**, Zheng, S., Yin Q., Jiang R., & Li, J. (2022). REDDA: integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction. *Computers in biology and medicine*. (IF=7.7)
7. **Gu, Y.**, Zheng, S., Xu, Z., Yin, Q., Li, L., & Li, J. (2022). An efficient curriculum learning-based strategy for molecular graph learning. *Briefings in Bioinformatics*. (IF=9.5)
8. **Gu, Y.**, & Li, J. (2022). Research Progress of Electronic Health Record Data Mining Based on Unsupervised Deep Learning. *Journal of Medical Informatics*. (Chinese)
9. **Gu, Y.**, Zheng, S., & Li, J. (2021). CurrMG: A Curriculum Learning Approach for Graph Based Molecular Property Prediction. *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*.
10. **Gu, Y.**, Zheng, S., Yang, F., & Li, J. (2021). Predicting Drug ADMET Properties Based on Graph Attention Network. *Data Analysis and Knowledge Discovery*. (Chinese)
11. Xu, Z., **Gu, Y.**, Xu, X., Topaz, M., Guo, Z., Sun, L., & Li, J. (2023). Developing a personalized meal recommendation system for Chinese elderly population: a knowledge-driven and community-based approach. *Journal of Medical Internet Research*. (Under review round 1)
12. Guo, L., Li, J., **Gu, Y.**, Wang, J., Cui, Y., Qian, Q., Chen, T., Jiang, R., Guo, L., & Zheng, S. (2023). Characteristics and Admission Preferences of Pediatric Emergency Patients and Their Waiting Time Prediction Using Electronic Medical Record Data: A Retrospective Comparative Analysis *Journal of Medical Internet Research*. (IF=7.4, Accepted)
13. Zhou, X., Zhuang, Y., Liu, X., **Gu, Y.**, Wang, J., Shi, Y., Zhang, L., Li, R., Zhao, Y., Chen, H., Li, J., Yao, H., Li, L. (2023). Study on tumor cell-derived hybrid exosomes as dasatinib nanocarriers for pancreatic cancer therapy. *Artificial Cells, Nanomedicine and Biotechnology*. (IF=5.8, Accepted)
14. Yi, M., Cao, Y., Wang, L., **Gu, Y.**, Zheng X., Wang, J., Chen, W., Wei, L., Zhou, Y, Shi, C., & Cao, Y. (2023). Prediction of Medical Disputes Between Health Care Workers and Patients in Terms of Hospital Legal Construction Using Machine Learning Techniques: Externally Validated Cross-Sectional Study. *Journal of Medical Internet Research*. (IF=7.4)
15. Kang, H., Hou, L., **Gu, Y.**, Lu, X., Li, J., & Li, Q. (2023). Drug–disease association prediction with literature based multi-feature fusion. *Frontiers in Pharmacology*. (IF=5.6)
16. Wang, Z., Lan, Y., Xu, Z., **Gu, Y.**, & Li, J. (2022). A Comparison Study of Mortality Predictive Models of Sepsis Patients Based on Machine Learning. *Chinese Medical Sciences Journal*.
17. Xu, Z., Zhu, C., **Gu, Y.**, Zheng, S., Sun, X., Cao, J., Song, B., Jin J., Liu Y., Wen X., Cheng S., Li, J., & Wu, X. (2022). Prediction of Poststroke Urinary Tract Infection Risk in Immobile Patients Using Machine Learning: a observational cohort study. *Journal of Hospital Infection*. (IF=6.9)
18. Xu, Z., Zhu, C., **Gu, Y.**, Zheng, S., Sun, X., Cao, J., Wu, X., & Li J. (2021). Developing a Siamese

Network for UTIs Risk Prediction in Immobile Patients. *MedInfo 2021*.

PATENTS

1. Gu, Y., Zhai K., Zhang, B., Wu Z., Ma, S., Data processing methods, devices and model training methods, devices and electronic equipment. CN114842926A. (initiative for examination as to substance)
2. Zhai, K., Zhang, B., Wu Z., Gu, Y., Li, P., Wang, Y., Ma, S., Wang Z., Molecular Property Prediction Methods and Systems, Devices, Storage Media, and Processors. CN114067928A. (granted)
3. Li, J., Xu, X., Gu, Y., Methods and Devices for Constructing Knowledge Representation Models for Clinical Practice Guidelines. CN113421657A. (granted)
4. Li, J., Xu, X., Gu, Y., A method and device for constructing computerized clinical guidelines based on graphical representation. CN114023462A. (initiative for examination as to substance)

FUNDINGS&AWARDS

MacCraken Fellowship 2023

President's Award of PUMC (Top20 at PUMC) 2023

China National Scholarship (20k CNH, Top38 at PUMC) 2022 –2023

Fundamental Research Funds for the Central Universities (30k CNH, 171/1450) 2022 –2023

“The construction and application study in graph neural network-based drug virtual screening

9th “Sharing Cup” Innovation Competition of Science and Technology 2021 –2022

Resources Sharing Service

Third Class Prize (Building a Bayesian network for the mortality risk prediction of sepsis patients.)

Third Class Prize (Construction of graph neural network based on multi-omics heterogeneous network for drug-disease association prediction.)

8th “Sharing Cup” Innovation Competition of Science and Technology 2020 –2021

Resources Sharing Service

First Class of Pharmaceutical Prize (Building drug ADMET prediction model based on machine learning and graph neural network.)

Second Class Prize (Adopting ensemble-learning method to construct a prediction model for acute respiratory distress syndrome (ARDS) identification among sepsis patients in ICU.)

Learning Excellent Award of Peking University (Twice) 2016-2018

Third Class Prize of Student Scholarship of Peking University (Twice) 2016-2018

ACADEMIC SERVICES

- Program Committee member for *IEEE BIBM 2023*.
- Journal reviewer for *Briefings in Bioinformatics*, *Journal of Medical Internet Research*, *Infectious Disease Modeling*, *Heliyon*, *JMIR Formative Research*, and *Frontiers in Endocrinology*.
- Conference reviewer for *AMIA 2023* and *IEEE BIBM 2023*.

SKILLS

- Foreign language: English
- Programming language: Python, R, MATLAB
- Frameworks and packages: Pytorch, Tensorflow, DGL, PyG, Sklearn, RDkit
- Dry-lab technologies: molecular docking, metabolism network construction, prognostic model construction
- Wet-lab technologies: qPCR, cell survival assay, self-administration for rat