

# Yaowen Gu

[gu.yaowen@imicams.ac.cn](mailto:gu.yaowen@imicams.ac.cn) | No.3 Yabao Rd, Chaoyang Dist, Beijing, CHN | [Github](#) | [Google Scholar](#)

## EDUCATION

**Chinese Academy of Medical Science (CAMS)& Peking Union Medical College** 2020-2023

M.Med. in Medical Informatics (Advised by Prof. [Jiao Li](#))

**Peking University**

2016-2020

B.A. in Experimental Medicine (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, and tested the performance of CurrMG on 5 GNN models and 8 molecular property prediction benchmarks (overall relative improvement 4.08%). (Publication #1 and 2)
- Built graph neural network for drug ADMET prediction and anti-mycobacterium inhibitor virtual screening. (Publication #3 and 4)

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

- Collected and integrated a drug repositioning benchmark including 41,100 nodes and 1,008,258 edges with 5 biological entities; Proposed a heterogeneous graph neural network and attention mechanism-based method called REDDA for the drug-disease association (DDA) prediction. (Publication #5)
- Proposed a multi-instance learning-based heterogeneous graph neural network called MilGNet for drug-disease association prediction, containing a pseudo meta-path generator, a bidirectional translating embedding method, and a multi-scale interpretable joint predictor. (Publication #6)

**Gene-disease Association Prediction** (Advised by Prof. [Jiao Li](#) in CAMS)

- Established a multi-graph representation learning-based ensemble learning method for gene-disease association prediction, which simultaneously utilizes genetic, therapeutic, and network topological features and has achieved AUC score of 0.924. (Publication #7)

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

- Established 6 machine learning models and 2 rolling average methods for waiting time predictions in pediatric emergency department (R-square > 0.7).
- Constructed ensemble learning models/ Siamese network to predict the urinary tract infection risk, achieving desirable performances (AUC > 0.8). (Publication #8 and 9)

## INTERNSHIP EXPERIENCE

**Xtalpi AI Research Center**

Nov. 2020–Sept. 2021

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

- Recurred 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–Now

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

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

## PUBLICATIONS

- 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=13.99)
- 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)*.
- Gu, Y.**, Zheng, S., Yang, F., & Li, J. (2021). [Predicting Drug ADMET Properties Based on Graph Attention Network](#). *Data Analysis and Knowledge Discovery*. (Chinese)
- Gu, Y.**, Zheng, S., & Li, J. (2022). [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., 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=6.69)
6. **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)*. (Accepted)
7. Wang Z.#, **Gu, Y.#**, Zheng, S., Yang, L., & Li, J. (2022). [MGREL: A multi-graph representation learning-based ensemble learning method for gene-disease association prediction](#). *Computers in biology and medicine*. (Under review round 1)
8. Zhu, C., Xu, Z., **Gu, Y.**, Zheng, S., Sun, X., Cao, 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= 8.94)
9. Xu, Z., Zhu, C., **Gu, Y.**, Zheng, S., Sun, X., Cao, J., ... & Wu, X. (2021). [Developing a Siamese Network for UTIs Risk Prediction in Immobile Patients](#). *MedInfo 2021*.
10. 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*.
11. **Gu, Y.**, & Li, J. (2022). Research Progress of Electronic Health Record Data Mining Based on Unsupervised Deep Learning. *Journal of Medical Informatics*. (Chinese)
12. Zheng, S., **Gu, Y.**, & Li, J. (2021) [Data Mining Methods for Precision Medicine](#), *Q&A for Precision Medicine*, Tianjin Science and Technology Press (In press, chapter)

## 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.**, [A method and device for constructing computerized clinical guidelines based on graphical representation](#). CN114023462A. (initiative for examination as to substance)
4. Li, J., Xu, X., **Gu, Y.**, [Methods and Devices for Constructing Knowledge Representation Models for Clinical Practice Guidelines](#). CN113421657A. (initiative for examination as to substance)

## FUNDINGS&AWARDS

<b>China National Scholarship</b> (20k CNH, Top38 at PUMC)	<b>2022 –2023</b>
<b>The Fundamental Research Funds for the Central Universities</b> (30k CNH, 171/1450)	<b>2022 –2023</b>
“The construction and application study in graph neural network-based drug virtual screening model.”	
<b>9<sup>th</sup> “Sharing Cup” Innovation Competition of Science and Technology Resources Sharing Service</b>	<b>2021 –2022</b>
<i>Third Class Prize</i> (Building a Bayesian network for the mortality risk prediction of sepsis patients.)	
<i>Third Class Prize</i> (Construction of graph neural network based on multi-omics heterogeneous network for drug-disease association prediction.)	
<b>8<sup>th</sup> “Sharing Cup” Innovation Competition of Science and Technology Resources Sharing Service</b>	<b>2020 –2021</b>
<i>First Class of Pharmaceutic Prize</i> (Building drug ADMET prediction model based on machine learning and graph neural network.)	
<i>Second Class Prize</i> (Adopting ensemble-learning method to construct a prediction model for acute respiratory distress syndrome (ARDS) identification among sepsis patients in ICU.)	
<b>Learning Excellent Award of Peking University</b> (Twice)	<b>2016-2018</b>
<b>Third Class Prize of Student Scholarship of Peking University</b> (Twice)	<b>2016-2018</b>

## ACADEMIC SERVICES

- Journal reviewer for *Briefings in Bioinformatics* and *Infectious Disease Modeling*.

## SKILLS

- Foreign language: English (TOFEL 91)
- Programming language: Python, R, MATLAB
- Frameworks and packages: Pytorch, DGL, PyG, Sklearn, Networkx, RDkit
- Dry-lab technologies: metabolism network construction, prognostic model construction
- Wet-lab technologies: qPCR, cell survival assay, self-administration for rat