

# Chengshuai Zhao

No.1, Shizishan Street, Wuhan, P.R.China | cszhaoapplication@gmail.com

## EDUCATION

---

Huazhong Agricultural University, College of Informatics

Sep. 2019 – Jun. 2023 (expected)

Bachelor of Engineering, Data Science and Big Data Technology

Overall GPA: 3.97/4.0    Ranking: 1/63

Huazhong Agricultural University, College of Plant Science & Technology

Sep. 2018 – Jun. 2019

Plant Production

Overall GPA: 3.87/4.0    Ranking: 2/364

## PUBLICATIONS

---

X Xu, S Liu, Z Yang, X Zhao, Y Deng, G Zhang, J Pang, **C Zhao**, W Zhang. "A systematic review of computational methods for predicting long noncoding RNAs." *Briefings in functional genomics* 20.3 (2021): 162-173.

**C Zhao**, Y Qiu, S Zhou, S Liu, W Zhang, Y Niu. "Graph embedding ensemble methods based on the heterogeneous network for lncRNA-miRNA interaction prediction." *BMC genomics* 21.13 (2020): 1-12.

**C Zhao**, S Liu, F Huang, S Liu, W Zhang. "CSGNN: Contrastive self-supervised graph neural network for molecular interaction prediction." *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, Online. 2021.

## ACADEMIC EXPERIENCE

---

Undergraduate Researcher | BioMedical Big Data Mining Lab, Wuhan, P.R.China

**Molecular Interaction Inference**

Aug. 2020 – Aug. 2021

- Collected, processed, and constructed homogeneous and heterogeneous networks with **40,00+** nodes and **500,00+** edges
- Proposed a novel **graph neural network (GNN)** framework based on **self-supervised learning** and **contrastive learning**, which generated pseudo-labeled data to facilitate representation learning from the limited samples
- Implemented the algorithm with **PyTorch**, achieved a maximum **20%** improvement in **AUC** values and **35%** improvement in **AUPR** values compared to state-of-the-art methods (e.g., node2vec, Line, and GAT) on 7 publicly available datasets
- Analyzed experimental results and pipelined the framework with the **t-SNE** algorithm, **Matplotlib** and **MS Visio** tools
- Summarized the ideas as a manuscript and published it at a top-tier conference

**LncRNA-MiRNA Interaction Prediction**

Feb. 2020 – Dec. 2020

- Developed a high-accuracy algorithm entailing **ensemble learning** and **attention mechanism** based on the heterogeneous network to predict lncRNA-miRNA interactions and prompt to reveal potential genetic disease
- Architected the model with **Scikit-learn** and **TensorFlow** framework, conducted 5-fold **cross-validation (CV)** experiments, and advanced the predictive capacity (**5.12%-562.84%** improvement in **F1 score**)
- Design and conducted **ablation study** and **case study** to validate proposed methods both theoretically and practically

## Long Noncoding RNAs Identification

Oct. 2019 – May. 2020

- Reviewed previous studies, and developed a comprehensive Python package ezLncPred which integrated 9 machine learning based models and a convenient command-line method to assist researchers who intended to identify lncRNAs

## COURSE PROJECTS

---

### Machine Learning Algorithm Replications

Oct. 2021 – Dec. 2021

- Implemented **Linear Regression, Logistic Regression, Decision Tree, Naive Bayes, Principal Component Analysis, and K-means Clustering** algorithms with Python package NumPy and SciPy

### Drug Analysis System

Jul. 2021 – Aug. 2021

- Designed and built a lightweight Python-based drug analysis system that allowed drug interactions real-time search, online acquisition, group-based statistics, network visualization, and similarity-based filter to promote drug development
- Utilized **Selenium** based **multi-threading** web crawler to extract drug structure, attribution, and interaction information from dynamic webpages, collected and processed **1,350,000+** drug interactions with Xpath location technique
- Implemented backend API endpoints with PyMongo and pass on queries to the **MongoDB** database, developed a friendly frontend Graph User Interface and implemented with TkInter and PyQt5
- Modeled a comprehensive algorithm that combined **Jaccard similarity coefficient** based structure and **node2vec** embedding base topology to filter candidate drugs and predict potential interactions via multi-view similarity

### Mutation Point of Gene Expression Detection

Jun. 2021

- Designed and architected **Markov chain Monte Carlo (MCMC)** based methods with **PyMC3** to precisely detect the mutation point of ENSG00000000419 gene expression in the human body

## ACTIVITIES

---

### Yunshu Club Academic Salon

Aug. 2020 – Aug. 2021

*Chief Director* | Huazhong Agricultural University, Wuhan, P.R.China

- Led and managed an academic organization including 20 undergraduate students, where weekly academic seminars and regular salons were conducted to brainstorm and share literature, ideas, or projects

## HONORS & AWARDS

---

Software Copyright (Aug. 2021): Drug Interaction Data Collection and Analysis System V1.0

*Grant by: National Copyright Administration, P.R.C*

Honorable Mention (Apr. 2021): Mathematical Contest in Modeling

*Awarded by: COMAP (the Consortium for Mathematics and Its Application)*

Bronze Prize (Dec. 2020): Mechanisms of Action (MoA) Prediction

*Awarded by: Kaggle*

Third Prize (Jul. 2020): 13rd Chinese College Students Computer Design Contest

*Awarded by: Chinese College Students Computer Design Contest Committee*

Advanced Individual (Oct. 2019): 2018 - 2019 Around-Lake Marathon

*Awarded by: Huazhong Agricultural University*

## **SCHOLARSHIP**

---

Huazhong Agricultural University:

- Outstanding Study Scholarship (Oct. 2021) (ratio: less than 1%)
- Merit Student Scholarship (Oct. 2020) (ratio: 20%)
- Study Scholarship (Oct. 2020) (ratio: 10%)
- Advanced Study Scholarship (Mar. 2020; Oct. 2019; Mar. 2019) (ratio: 3%)

## **TECHNICAL SKILLS**

---

**Programming Languages:** C, C++, Java, Python, R, Go, Scala, SQL, Verilog HDL

**Operating Systems:** MS Windows, Linux

**Database:** Oracle, MySQL, SQLite, MongoDB

**Frameworks:** Scikit-learn, Keras, TensorFlow, PyTorch, Hadoop, Spark

**Tools:** Git, Docker