Junyi Chen

Research Assistant Professor
Hong Kong SAR, China
(852) 53058149, (86) 15521319231 | jerry-junyi.chen@polyu.edu.hk | juychen

Biography

Junyi Chen is currently a research assistant professor at the Department of Health Technology and Informatics in The Hong Kong Polytechnic University. He finished his PhD in Computer Science at the City University of Hong Kong. Junyi's research focuses on the application of artificial intelligence in bioinformatics, including single-cell RNA-seq analysis, and sequence analysis. He published first-author papers in reputed journals such as Nature Communications, GigaScience, and Briefings in Bioinformatics.

Position

Research Assistant Professor — *Hong Kong SAR, China*
The Hong Kong Polytechnic University — *Oct 2024 to present*
Department of Health Technology and Informatics

Postdoctoral Fellowship — *Hong Kong SAR, China*
Laboratory of Data Discovery for Health ($D^2$4H) — *Dec 2020 to Oct 2024*
AI-Assisted Scalable Data Analytics for Global Health Protection led by Dr. Joshua Ho

Education

Doctor of Philosophy (PhD) in Computer Science — *Hong Kong SAR, China*
City University of Hong Kong — *Sep 2017 to Aug 2020*
In East Asian Bioinformatics and Computational Biology (Hong Kong) supervised by Dr. Ka-Chun Wong

Master of Science (MSc) in Computer Science — *Hong Kong SAR, China*
City University of Hong Kong — *Sep 2016 to Aug 2017*

Bachelor of Engineering (BEng) in Software Engineering — *Guangzhou, Guangdong, China*
Guangzhou University — *Sep 2011 to Aug 2015*

Experiences

Visiting Scholar — *Columbus, OH, US*
The Ohio State University — *May 2019 to Nov 2019*
In Bioinformatics and Mathematical Biosciences Lab supervised by Dr. Qin Ma

Research Assistance — *Guangzhou, Guangdong, China*
Sun Yat-sen University — *Jun 2016 to Sep 2016*

Skills

Research Interests	Bioinformatics, sing-cell RNA-Seq, Pharmacogenomics, Machine learning, Deep learning, Active learning.
Teaching Assistance Courses	Software Design, Fundamentals of Data Science, Software Engineering Practice, Computer Networks.
Core Courses	Machine Learning, Data Warehousing & Data Mining, Big Data Algorithms and Tech.
Programming Languages	Python, R, Java.
Languages	Mandarin, Cantonese, English.

Publications

Journal Papers

- **Chen, J.**, Yin, D., Wong, H. Y., Duan, X., Yu, K. H., & Ho, J. W. (2024). Vulture: cloud-enabled scalable mining of microbial reads in public scRNA-seq data. GigaScience, 13, giad117. (Impact Factor: 11.8, Ranking: MULTIDISCIPLINARY SCIENCES 10/134, First author)
- Ma, S., **Chen, J.**, & Ho, J. W. (2024). An edge-device-compatible algorithm for valvular heart diseases screening using phonocardiogram signals with a lightweight convolutional neural network and self-supervised learning. Computer Methods and Programs in Biomedicine, 243, 107906. (Impact Factor: 4.9, Ranking: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS 31/169, Second author)
- Zheng, Z., **Chen, J.**, Chen, X., Huang, L., Xie, W., Lin, Q., ... and Wong, K. C. (2023). Enabling Single-Cell Drug Response Annotations from Bulk RNA-Seq Using SCAD. Advanced Science, 2204113. (Impact Factor: 14.3, Ranking: CHEMISTRY, MULTIDISCIPLINARY 18/230, Second author)
- **Chen, J.**, Wang, X., Ma, A., Wang, Q. E., Liu, B., Li, L., ... and Ma, Q. (2022). Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. Nature Communications, 13(1), 6494. (Impact Factor: 14.7, Ranking: MULTIDISCIPLINARY SCIENCES 8/135, First author)
- Yin, D., Cao, Y., **Chen, J.**, Mak, C. L., Yu, K. H., Lin, Y., ... & Yang, J. Y. (2022). Scope+: An open source generalizable architecture for single-cell atlases at sample and cell levels (pre-print).
- **Chen, J.** and Wong, K. C. (2020). RNCE: Network Integration with Reciprocal Neighbors Contextual Encoding for Multi-modal Drug Community Study. Briefings in Bioinformatics. (Impact Factor: 6.8, Ranking: BIOCHEMICAL RESEARCH METHODS 3/85, First author)
- **Chen, J.** and Wong, K. C. (2020). Categorical Matrix Completion with Active Learning for High-throughput Screening. IEEE/ACM transactions on computational biology and bioinformatics.(Impact Factor: 3.6, Ranking: BIOCHEMICAL RESEARCH METHODS 25/85, First author)
- Lin, J., Zhang, Z., Zhang, S., **Chen, J.**, & Wong, K. C. (2020). CRISPR-Net: A Recurrent Convolutional Network quantifies CRISPR Off-target Activities with Mismatches and Indels. Advanced Science. (Impact Factor: 14.3, Ranking: CHEMISTRY, MULTIDISCIPLINARY 18/230, Third author)
- Zhao, J., Hinton, P., **Chen, J.**, & Jiang, J. (2020). Causal inference for the effect of environmental chemicals on chronic kidney disease. Computational and Structural Biotechnology Journal, 18, 93-99. (Impact Factor: 4.4, Ranking: BIOCHEMISTRY & MOLECULAR BIOLOGY 82/313, Third author)
- Wong, K. C., **Chen, J.**, Zhang, J., Lin, J., Yan, S., Zhang, S., ... & Kwong, S. (2019). Early Cancer Detection from Multianalyte Blood Test Results. iScience, 15, 332-341. (Impact Factor: 4.6, Ranking: MULTIDISCIPLINARY SCIENCES 27/135, Second author)
- **Chen, J.**, Yan, S., & Wong, K. C. (2018). Verbal aggression detection on Twitter comments: Convolutional neural network for short-text sentiment analysis. Neural Computing and Applications, 1-10. (Impact Factor: 4.5, Ranking: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE, 52/197, First author)

Conference Papers

- **Chen, J.**, Yan, S., & Wong, K. C. (2017, March). Aggressivity detection on social network comments. In Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (pp. 103-107).

Book Chapters

- **Chen, J.** and Wong, K. C. (2020). Analyzing High-order Epistasis from Genotype-Phenotype Maps using 'Epistasis' Package. Epistasis in Methods in Molecular Biology.

Teaching experiences

Taking CityU course SG8001: Teaching Students: First Steps

City University of Hong Kong

Student

- Learned the framework of Outcomes-based Teaching and Learning (OBTL)
- Designed a course by justifying Intended Learning Outcomes (ILOs)
- Targeting ILOs by setting up Teaching and Learning Activities (TLAs)
- Validated the ILOs for students by setting up Assessment Tasks (ATs)

Software Engineering Practice

City University of Hong Kong

Teaching assistant

- **Intentended learning outcomes 1**: Properly apply the principles and techniques of requirements specification and analysis, software design, implementation, testing, delivery, and maintenance to create software applications
- **Intentended learning outcomes 2**: Present projects effectively
- **Intentended learning outcomes 3**: Write technical documentation in a clear and concise manner
- **Intentended learning outcomes 4**: Work effectively in a team environment.
- **Teaching activity 1**: A team building event (the straw tower challenge) to enhance engagement of students (week 1)
- **Teaching activity 2**: in-class coding tutorials weekly (week 2,3,4,7,8,11)
- **Accessment task 1**: (40%) two in-class mini-quizzes with coding (week 5,10) and the continues performances in the tutorial
- **Accessment task 2**: (30%) a final group project with presentation and report (week 13)
- **Accessment task 3**: (30%) a final examination (week 12)

Data Warehousing and Data Mining

City University of Hong Kong

Teaching assistant

- **Intentended learning outcomes 1**: Identify the main characteristics of different data warehousing and data mining techniques through observation of their operations.
- **Intentended learning outcomes 2**: Perform a critical assessment of current data warehousing and data mining techniques.
- **Intentended learning outcomes 3**: Implement the main algorithms in data warehousing and data mining in a computationally efficient way.
- **Intentended learning outcomes 4**: Propose new solutions for data warehousing and data mining problems by improving and combining current techniques.
- **Teaching activity 1**: A lecture introducing the concepts of Data Warehousing and Data Mining.
- **Teaching activity 2**: Two programming assignments. One implementing the data mining algorithms, the second using data warehousing and data mining techniques to real world problems.
- **Accessment task 1**: (30%) programming assignments
- **Accessment task 2**: (20%) in-class quiz
- **Accessment task 3**: (50%) a final examination

Research Projects

Large language model-based anti-sense oligonucleotides drugs discovery

Leading project *2023 to now*

- Collecting anti-sense oligonucleotides drug data from Google patent and scholar.
- Developing the deep learning methods for anti-sense oligonucleotides drug potency prediction.

Single cell analysis for pelvic pain from endometriosis patients

Leading project *2021 to now*

- Performing pre-processing and downstream analysis for the BDrhapsody and inDrops3 pipeline.

Generalizable architecture for single-cell atlases at sample and cell levels

Paper pre-printed *2021 to 2022*

- Developing the database backend for the single-cell atlases.

Scalable cloud computing analysis for single-cell RNA-Seq data

Leading project *2020 to 2022*

- Developing the container-based AWS architectures for viral calling single-cell RNA-Seq analysis.
- Pre-processing and downstream analysis for the single-cell RNA-Seq analysis

Integrative drug response study from bulk to single-cell RNA-Seq data

Leading project, research proposal granted by the Center for Clinical and Translational Science Funding of The Ohio State University (CCTS) *2019 to 2020*

- Drafted the research proposal;
- Developing the single-cell RNA-Seq drug response prediction model inegrating drug-response data from the bulk level.

Transfer learning study from pharmacogenomics to single-cell RNA-Seq data

Leading project, research proposal submitted to National Natural Science Foundation of China (NSFC) *2020*

- Drafted the research proposal;
- Developing the transfer learning methods to transfer gene-drug relation to single-cell RNA-Seq data.

Network integration method for multi-modal pharmacogenomics study

Leading project, Paper accepted by briefings in bioinformatics *2019 to 2020*

- Designed a method to guide conditions-samples studies using high-throughput screening and automated robotics;
- Applied active learning with categorical matrix completion to solve the batch selection problem;
- Achieved superior the proteins compound-effect dataset than the state-of-the-art.

Smart guidance on high-throughput screening experiments

Leading project, Paper Accepted by IEEE/ACM transactions oncomputational biology and bioinformatics *2018 to 2019*

- Designed a method to guide conditions-samples studies using high-throughput screening and automated robotics;
- Applied active learning with categorical matrix completion to solve the batch selection problem;
- Tested and proved the superior performance of the propose method on the proteins compound-effect dataset than the state-of-the-art.

Alternative regulon identification from single-cell RNA-Seq data

Research proposal submitted to NIH Research Project Grant Program (R01) *2019*

- Drafted one section of the research proposal;
- To develop a deep learning based method to discover the alternative form of regulons among cell sub-types on single-cell RNA-Seq data.

Agricultural factors influencing chronic kidney disease study

Research proposal submitted to R01 and CCTS *2019*

- Assisted the drafting of the research proposal;
- Run preliminary clustering analysis for the estimated glomerular filtration rate decline pattern;
- To develop a random forest based model to discover the relation between agricultural factors, Electronic Medical Records (EMRs) and chronic kidney disease.

Early cancer detection from multianalyte blood test results

Paper Accepted by iScience *2019*

- Reviewed literature of one-dependence estimator models;
- Reviewed literature of cancer related proteins in blood;
- Assisted manuscript drafting of the method section.

Deep learning on social media text mining

Leading project, Paper published in Neural Computing and Applications *2017 to 2018*

- Built convolutional neural network (CNN) models for the task of sentiment analysis on short social media comments;
- Validate performances of different word embedding models on deep learning models.

Bio-medical text mining on disease reference

2017

- Build machine learning algorithms and cross-validate performance to classify disease references from the OMIM/ORPH dataset.

Social media text mining

Leading project, Paper published in 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence *2016 to 2017*

- Collected comments from the social media and annotate the sentiment label;
- Build machine learning algorithms to classify social comments by their sentiment labels.

Mobile application for unmanned aerial vehicle control

2016

- Developed a mobile application on Android smart phone controlling and streaming camera video from unmanned aerial vehicle.

Awards

Outstanding Academic Performance Awards (OAPA)

City University of Hong Kong *2018 to 2019*

Research Tuition Scholarship (RTS)

City University of Hong Kong *2018 to 2019*

Outstanding Poster Award for the Research Student Workshop

City University of Hong Kong *6th June 2018*

Distinction Banding for Taught Postgraduate Award Classifications

City University of Hong Kong *2016 to 2017*

First class Academic Performance Awards

Guangzhou University *2014 to 2015*

First class Academic Performance Awards

Guangzhou University *2013 to 2014*

First class Academic Performance Awards

Guangzhou University *2012 to 2013*

Review services

Journals	Nucleic Acids Research (sub-reviewer), Briefings in Bioinformatics, BioData Mining, Applied Soft Computing, Computational Biology and Chemistry.
Conferences	ISCMI 2019 (Conference Committee), BigData 2019, ICMLC 2019, ICBDE 2019, BiCOB 2019, BigData 2018, ISCMI 2018 (Conference Committee), BIBE 2018, PIC 2018, SmartData 2018,ICMLC 2018, BDET 2018,BiCOB 2018, ABPC 2018,DMBD 2018, DSS 2017, BigCom 2017, DMBD 2017, ICSI 2017,