Yan Chak (Richard) Li

BIOINFORMATICIAN

richardpokard@gmail.com | huhrichard.github.io/ | github.com/huhrichard | linkedin.com/in/yan-chak-li-865b6b124/ | @rlychuh

Personal Statement

My research interest and experience are in developing and applying machine learning techniques to biomedical data to aid tasks such as clinical diagnosis, biomedical knowledge discovery etc. In this direction, I have led the development and evaluation of novel automated methods for the processing of mass spectrometry data and diagnosis of vertebral fractures from X-ray images. I am now working with Dr. Gaurav Pandey on multi-modal data integration methods to address biological prediction and knowledge discovery problems, such as protein function predictions, and disease outcomes. Besides that, I am also working on web applications to enable public sharing and data visualization.

Education

The Hong Kong University of Science and Technology *Clear Water Bay, Hong Kong*

M.Phil. in Bioengineering *Sept 2017 - Aug 2019*

- Thesis: Deep Learning Enables Instance Edge Detection of Vertebral Bodies on X-ray Images
- Teaching Assistant of IELM/IEDA 2100E
- Courses: Computer Vision, Mathematical Foundations of Imaging, Topological and Geometric Data Reduction and Visualization etc.

The Hong Kong University of Science and Technology *Clear Water Bay, Hong Kong*

B.Eng. in Computer Engineering *Sept 2013 - Aug 2017*

- Undergradudate Research Project: Improving the Efficiency of Spectral Library Searching in Mass Spectrometric Data Analysis
- Final Year Project: Digitizing Receipts
- Courses: Introduction to Bioinformatics Algorithms, Medical Imaging, Heterogeneous Parallel Programming etc.

Work Experience

Icahn School of Medicine at Mount Sinai *New York City, U.S.A.*

Bioinformatician *Nov 2019 - Current*

- Develop ensemble machine learning methods for multimodal biomedical data
- Analyze clinical cohorts by machine learning techniques
- Teaching Assistant of 'Machine Learning for Biomedical Data Science' (Spring 2020 & 2021)

The Hong Kong University of Science and Technology *Clear Water Bay, Hong Kong*

Consultant *Sept 2019 - Oct 2019*

- Integrate stereo & thermal camera for Smart Fever Screening System & localize fever people

Hong Kong Telecommunication Limited *Quarry Bay, Hong Kong*

Summer Internship *Jun 2015 - Aug 2015*

- Summarize up-to-date anti-DDoS solution & audit Data by Microsoft Excel

Pigeon City Creative Computer Centre *Prince Edward, Hong Kong*

Part-time Tutor *Feb 2015 - May 2015*

- Teach students to build their own programmable LEGO and mini-games.

SkyWare Technologies Limited *Tsuen Wan, Hong Kong*

Technical Support *May 2013 - Aug 2013*

- Test new network firmwares and hardwares.

Publications

eipy: An Open-Source Python Package for Multi-modal Data Integration using Heterogeneous Ensembles *arXiv*

Bennett JJR, **Li YC**, Pandey G *Jan 2024*

Link: eipy package, preprint

KinAce: a web portal for exploring kinase-substrate interactions *bioRxiv*

Sekar JAP, **Li YC**, Schlessinger A, Pandey G *Dec 2023*

Link: KinAce - web portal, preprint

Exploring the Druggable Conformational Space of Protein Kinases Using AI-Generated Structures — *bioRxiv*
Herrington NB, Stein D, **Li YC**, Pandey G, Schlessinger A — *Sep 2023*
Link: Preprint

Facilitating youth diabetes studies with the most comprehensive epidemiological dataset available through a public web portal — *medRxiv*
McDonough C, **Li YC**, Vangeepuram N., Liu B., Pandey G. — *Aug 2023*
Link: POND - web portal, preprint

Developing better digital health measures of Parkinson's disease using free living data and a crowdsourced data analysis challenge — *PLOS Digital Health*
Sieberts SK, Borzymowski H, Guan Y, Huang Y, Matzner A, Page A, … , **Li YC**, … , Stanescu A, … , Pandey G, Shawen N, Synder P, Omberg L — *Apr 2023*
link: paper

Integrating multimodal data through interpretable heterogeneous ensembles — *Bioinformatics Advances*
Li YC, Wang L, Law JN, Murali TM, Pandey G — *Sep 2022*
link: Paper, github repository

Machine learning-driven identification of early-life air toxic combinations associated with childhood asthma outcomes — *Journal of Clinical Investigation*
Li YC, Hsu HL, Chun Y, Chiu PH, Arditi Z, Claudio L, Pandey G, Bunyavanich S — *Nov 2021*
Link: paper, github repository

Clinical features of COVID-19 mortality: development and validation of a clinical prediction model — *Lancet Digital Health*
Yadaw AS, **Li YC**, Bose S, Iyengar R, Bunyavanich S, Pandey G — *Oct 2020*
Link: paper, github repository

Presentations

Integrating multimodal data through interpretable heterogeneous ensembles — *Madison, Wisconsin, U.S.A.*
Li YC, Wang L, Law J, Murali TM, Pandey G — *Jul 2022*
Oral and poster present at The 30th Conference on Intelligent Systems for Molecular Biology (ISMB)

Automatic Instance-edge Detection Network (AID-Net) - Vertebral Edge Detection by Deep Learning — *Coimbra, Portugal*
Li RYC, Chin NJW, Wang Y, So RHY — *May 2019*
Oral present at European Society for Clinical Investigation Congress (ESCI Congress) 2019

Fast Similarity Measure of SWATH-MS by Cosine Similarity of Random Pairs (CS-RP) — *Biopolis, Singapore*
Li YC, Wu L, Lam H — *Dec 2017*
Oral present at Asia Oceania Mass Spectrometry Conference (AOMSC) 2017

Skills

Data science Pandas, NumPy, Scikit-learn, PyTorch, OpenCV, Keras etc. on Python; statistical analyses, data visualization on R, basic SQL
Web development Shiny application on R and R Markdown; Vue.js
Other computing skills High-performance computing, administrator of cloud computing on Oracle Cloud and Google Cloud, Firebase, Linux, LaTeX, Git, Google Analytics, CUDA C, C, C++, Java

Awards

2017 **Young Scientist Travel Award**, Asia Oceania Mass Spectrometry Conference 2017 — *Singapore*
2013 **Dean of Engineering Scholarship**, HKUST — *Hong Kong*

Languages

English Professional proficiency
Cantonese Native proficiency