Yan Chak (Richard) Li

BIOINFORMATICIAN

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

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 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
- 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

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.

Projects

SunBEAm-ABC web portal

Link: public version: public version, beta version

The SunBEAm Analysis & Bioinformatics Center (SunBEAm-ABC) is assaying biosamples using omics and will apply integrative systems biology to identify novel determinants of food allergy and atopic dermatitis. The web portal aims to provide data visualization of multi-omics data via different types of plots, such as boxplot, barplot, heatmap, etc. On top of it, we also build network visualization (https://sunbeamtest.net/network-vis/index.html), which allows users to upload & download network data to explore their interests.

KiNet - Kinase-Substrate Interaction Network Visualization

Link: bioRxiv, KiNet website

The KiNet web portal aggregates and visualizes the network of interactions between protein-kinases and their substrates in the human genome. Each tab provides different ways to select proteins and display the known kinase-substrate interactions between them. We also provided detailed information on interactions by selecting the edges.

POND - Prediabetes/diabetes youth ONline Dashboard

Link on shinyapps.io / hpc.mssm.edu

POND is an interactive dashboard for exploring factors associated with prediabetes and diabetes mellitus (preDM/DM) among youth (aged 12-19 years) in the United States. Raw data were obtained from the National Health and Nutrition Examination Survey (NHANES) and processed into a multi-domain dataset that is the foundation of our study and this portal.

Ensemble Integration - multimodal machine learning

Link: Paper / eipy python package documentation

Ensemble Integration (EI, ensemble-integration/eipy as a python package) is a multimodal machine learning package for generating diverse ensembles of heterogeneous classifiers, as well as the accompanying metadata needed for ensemble learning approaches utilizing ensemble diversity for improved performance.

Data-driven ExposurE Profile (DEEP) - feature combination extraction from tree-based models

Link: Paper / github repository

DEEP uses the XGBoost algorithm to identify air toxic combinations associated with health outcomes. The combinations identified using XGBoost were then adjusted for potential confounders to identify early-life multi-air toxic combinations.

Identifying clinical features of COVID-19 mortality

Link: Paper / github repository

We developed a machine learning model to predict COVID-19 mortality using clinical data from a large cohort of patients treated at Mount Sinai Health System. The model trained on data from 3,841 patients, achieved high accuracy (AUC=0.91) in predicting mortality when tested on retrospective and prospective datasets. The model relies on just three clinical features: patient age, minimum oxygen saturation during their medical encounter, and type of patient encounter (inpatient vs outpatient/telehealth).

Publications

A web portal for exploring kinase-substrate interactions

npj Systems Biology and Applications

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

Oct 2023

Link: KinAce - web portal, paper, github repository

A comprehensive exploration of the druggable conformational space of protein kinases using AI-predicted structures

PLoS Comput. Biol.

Herrington NB, **Li YC**, Stein D, Pandey G, Schlessinger A

Jul 2024

Link: paper

A comprehensive youth diabetes epidemiological dataset and web portal: Resource Development and Case Studies

JMIR Public Health Surveill

McDonough C, **Li YC**, Vangeepuram N., Liu B., Pandey G.

Jul 2024

Link: POND - web portal, paper

Machine learning-driven identification of air toxic combinations associated with asthma symptoms among elementary school children in Spokane, Washington, USA

Science of The Total Environment

Amiri S, **Li YC**, Buchwald D, Pandey G

Feb 2024

Link: paper

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

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, joblib, graphviz etc. on Python; statistical analyses, ggplot2, plotly on R, basic SQL
Web development	R Shiny application; Vue.js, JavaScript, HTML, Flask, Django
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*