XIANGYU GUO

xiangyu0925@gmail.com

Summary

Passionate in the insight of data and have strong ability to learn. Eager to solve practical problems in both academic and industrial. With interdisciplinary background and logical thinking ability, I am looking for opportunity or position in data science field.

Education

University of St Andrews, the UK 09/2017 – 12/2018

- MSc in Data Intensive Analysis grade: 14.7/20

University College Dublin, Ireland 09/2016 – 05/2017

- Computer science (Full-year exchange student) and bioinformatics related modules

South China Agricultural University, China 09/2013 – 06/2016

- Bachelor of Science, Biological Science grade: 87.45/100

Academic Achievements & Publication

- Yong Bai, **XY Guo,** K Liu, Q Luo, …, Xin Jin. Analysis of spatial transcriptomics at varying resolution levels using the unified framework of SpaSEG. (First author, **BioRxiv**)
- Li LS, **Guo XY**, Sun K. Recent advances in blood-based and artificial intelligence-enhanced approaches for gastrointestinal cancer diagnosis. ***World J Gastroenterol*** 2021; 27(34): 5666-5681 **PubMed**
- T Zhang, X Jiang, **XY Guo**, …, Xin Jin, Yong Bai. A machine learning based INDEL prediction algorithm that utilizes functional difference of INDELs to improves prediction of pathogenicity and aid in interpretation clinical genome ultra-fast. (**Finished**, to be submitted to ***the American Journal of Human Genetics***)
- R Zhou, T Zhang, **XY Guo**, …, Yong Bai. **scDualGN,** fast, robust and accurate clustering of large-scale single-cell RNA-seq data using dual generative network (**Finished**, to be submitted to ***Nature Computer Science***)
- International Genetically Engineered Machine Competition ***(iGEM),*** MIT, USA 10/2016

  project link: http://2016.igem.org/Team:SCAU-China

  Awards: • Gold Medal in iGEM 2016 Giant Jamboree (The best achievement among universities in mainland of China)
  - Second Runner-Up • Best Plant Synthesis Biology
  - Best New Application Project • Best Education and Public Engagement

Research Experience

Research associate & Algorithm Engineer ***BGI-Research*** 10/2020 – Present

Institute of Cancer Research 08/2022 — Present

Project in progression: **1.** Prediction the seeding and metastatic potential to Lymph node of intrahepatic cholangiocarcinoma(ICC) in tumor tissue and profiling the molecular characteristics of metastasis of ICC by integrating spatial transcriptomics and scRNA-seq data. **2.** Investigating the difference of MIA and IA in lung cancer by using spatial transcriptomics. **3.** Spatial transcriptomics data batch correction in tumor sample slice.

Institute of Precision Health 06/2021 – 08/2022

Research Topic: ***1.*** An Unsupervised Deep Learning algorithm for Spatial transcriptomics data Clustering and downstream analysis, **finished. 2.** Prediction the gene expression and prognosis from **HE** images in SRT of breast cancer, **in progression**.

Research Assistant 10/2020 – 05/2021

Institute of Cancer Research, Shenzhen Bay Laboratory(SZBL) joint with BGI-Research, Shenzhen

Research Topic: Deep Learning based approach to build a pan-cancer classifier for tumor status classification and tissue origin classification. The public data TCGA was utilized to build machine learning models(LR, LightGBM, ANN) for the two tasks. For the tumor status prediction, our best model could achieve 97.95% accuracy and 0.9877 for the AUC. As the tissue origin classification, the best model gives us 98.21% overall accuracy.

MSc Dissertation 06/2018-09/2018

University of St. Andrews

Topic: Machine Learning in Glucose Level Prediction

Background: This research topic came from KDH 2018 (The 3rd International Workshop on Knowledge Discovery in Healthcare Data) and the dataset contains six real world clinic diabetes' physiological data through continuous glucose monitoring.

Objectives: 1. Data extraction, cleaning and transformation. Knowledge-drive approach on information retrieval and explanatory analysis on patients' healthcare data.

2. Building benchmark for human glucose level prediction based on statistical modelling techniques (LMs & GLS) and evaluated different machine learning algorithms' performance (SVM, RF, xgBoost) on the problem.

Supervisor: Prof. Ognjen Arandjelovic

Ref: https://sites.google.com/view/kdhd-2018/bglp-challenge

Research Assistant – Wet Lab 06/2014-06/2016

The State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources

Topics: 1. Genetically modifying the metabolism pathway by utilizing Cas9 gene editing system to produce astaxanthin.

2. research on transportation-related gene NtAN9 of anthocyania in tobacoo.

Supervisor: Prof. Letian Chen, Prof. Qinlong Zhu. Ref: https://doi.org/10.1016/j.molp.2018.09.007

Skills

- **Data Analysis & Statistical Modelling & Data Visualization**
  Quantitative data analysis, data manipulation and statistical modelling with GLMs, GAMs, GEEs, random intercept and random coefficient models.
- **Machine learning & Deep Learning**
  Decision tree, Gradient Boosting tree, Random Forest(RF), Support Vector Machines(SVM), CNN, RNN
- **Programming**
  R, Java and Python, Tensorflow, PyTorch Basic: C++, SQL, JavaScript

Certificates and awards

- ***China Scholarship Council*** ***9000€*** 2016
- ***South China Agriculture University***

First prize scholarship and the honor of **outstanding graduate** in SCAU 2017

The Merit Student and The Second Prize Scholarships. 2014

Working Experience and Social Activity

Data Mining Engineer 01/2019-08/2020

The Research & Development Center, PingAn Insurance Group, China

- Developed and deployed ML and Deep Learning models for our 1.8 million insurance agents and clients.
- Built big data modelling platform (real-time prediction) for our team.

Team Leader 09/2014-06/2016

Table Tennis Team of College of Life Science, South China Agricultural University, China

Security Maintenance & Merchant Receptionist

The 2015 Guangzhou International Vertical Marathon 04/2015

The 114th China Import and Export Fair (Canton Fair) 04/2015