

# YIRU GONG

100 Haven Avenue, New York, NY 10032 | (917) 742-0036 | yg2832@cumc.columbia.edu

## EDUCATION

### Columbia University Mailman School of Public Health

New York, NY

Master of Science (MS), Biostatistics

09/2021 – 05/2023 (Expected)

- Public Health Data Science Track; GPA: 3.85
- Relevant Courses: *Deep Learning, Introduction to Database, Data Science II, Biostatistical Methods II*

### University of Edinburgh - Zhejiang University Joint Institute

Edinburgh, UK & Zhejiang, China

Bachelor of Science with Honors in Biomedical Science, Dual degree program of UoE & ZJU

09/2017 – 06/2021

- GPA: 3.82
- Relevant courses: Applied Biomedical Science (Statistics); Introductory Data Science with Python and Tableau, Data Analytics for Customer Insights (NUS summer program)

## SKILLS AND CERTIFICATES

- Proficient in R (tidyverse, Shiny), Python (Numpy, Pandas, sklearn, Pytorch, Keras, Flask), SQL, and MATLAB; familiar with Linux Shell, JavaScripts, SAS, C language, and VBA
- Good command of MS Office software; Familiar with data visualization in Tableau, R Shiny
- Familiar with cloud computing platforms such as Google Cloud Platform (GCP)
- Languages: Mandarin (Native), English (Proficient)

## RELEVANT EXPERIENCE

### Elevance Health, *Health Data Analytics Intern, AIM Specialty Health*

Remote, US, 06/2022-08/2022

- Carried forward the project on “Clinical AI & UM Automation” using Natural Language Processing (NLP) Model and Predictive Machine Learning Models
  - Built a machine learning model on claim data to predict UM case approval status by using Python (LightGBM)
  - Pulled raw claim data by merging multiple datasets in the Microsoft SQL Server Management Studio (SSMS)
  - The model achieved an 86% precision score and is proposed to reduce the number of reviewed cases by 10%
  - Refined the NLP model on document scanning by communicating with Medical Doctors for user interface improvement and false positive cases reviewing

### Competition: Covid-19 Infection Percentage Estimation (CodaLab)

01/2022-03/2022

- Established a Convolutional Neural Network (CNN) [model](#) to estimate the Covid-19 infection percentage from the CT scans of Covid-19 patients.
- Achieved 7.2 MAE score and ranked in the top 20% of the leaderboard.

### GlaxoSmithKline (GSK), *Digital Analyst Intern, R&D Tech*

Shanghai, China, 04/2021-07/2021

- Managed a project on “Establishment of a Database to Track Differences in Medication Treatment Response by Ethnicity”; designed drug data collection and analysis methods, established computational workflow using R and Python to support drug pharmacokinetics analysis.
  - Reduced manual workload from 2 months to 2 days by using both R and Python
  - Tidied semi-structured JSON data of > 400k FDA clinical trials by Elasticsearch in Python
  - Expanded information on the use of 56 chemical drugs and 10 biotech drugs by ethnicity
  - Identified three ethnic difference-related genes and signaling pathways
- Facilitated project on “Natural Language Processing (NLP)-based Medical Knowledge Graph Establishment.”
  - Presented researchers the principle of Natural Language Processing and AI to extract disease-related information from > 10 million research articles and official documents
  - Communicated with 6-8 Physicians and Clinicians for technical issues and designed big data-based solutions
  - Applied NLP to clinical trials to help colleagues identify potential competitor drugs and trials
  - Applied Deep Learning (NER, RE) to carry out auto-revision of medical writing and reduced time and effort on medical document translation and revision by 50%

### DuPont Danisco Nutrition & Bioscience, *Sales/Marketing Intern, Intellifresh™*

Shanghai, China, 03/2021-07/2021

- Initialized the Wechat Digital Marketing Platform for Danisco, localized and digitalized the Global marketing sheets into styles favored by Chinese customers
- Expanded potential target customers from 200 to 2,000 by effectively improving the product exposure and customer management method, significantly increasing customer orders
- Established an automatic-filling labeling application document using VBA in Word and Excel. Significantly reduced the time for format editing and content filling

## OTHER PROJECTS

---

**Prof. Ali Gharavi's lab, Center for Precision Medicine and Genomics, Columbia University** 03/2022-08/2022

**Project: Establishing a Docker Image for Automated Compound Heterozygous (CH) Identification**

- Established an automated and highly efficient analyzing pipeline in R to identify the Compound Heterozygous genes from annotated VCF files and unannotated gVCF files
- Embedded the scripts into a docker image to allow large-scale analysis in the cloud computing platform

**Graduate Course: Deep Learning**

01/2022-05/2022

**Final Project: Applying Masked Token Transformer (MaskGIT) in Long Audio Generation**

- Applied the newest image transformer model in Computer Vision to audios to generate long and high-quality audio within a shorter time by using the Pytorch package in Python
- Built the [model](#) by combining the audio-pre-trained VQGAN model and the MaskGIT transformer, and trained it on the VAS video spectrum dataset on the Google Cloud Platform (GCP)
- Authored a research article and a conference-style poster as final reports, 90/100 (first class) received

**Graduate Course: Introduction to Databases**

01/2022-05/2022

**Project: Building a SQL-based Website Interface for Integrated Covid-19 Services**

- Established a [website](#) to provide user-oriented Covid-19 service appointments and nearby site finding using PostgreSQL, Flask package in Python, and Jinja in JavaScript
- Designed and built a database in PostgreSQL to store the information of users and Covid-19 service sites, and created a user-friendly website application to access and modify the database using Jinja in JavaScript
- Achieved interaction between database and website inputs by writing SQL queries-embedded python scripts with Flask package
- Perfect score (100/100) received

**Graduate Course: Data Science II**

01/2022-05/2022

**Project: Predicting Covid-19 Vaccination Status from Social-economic Variables**

- Established and compared the predictive efficiency of different machine learning models including Random Forest, glmnet, Support Vector Machine (SVM), Generalized additive model (GAM), and Neural Networks
- 92/100 (first class) received

**Graduate Course: Data Science**

09/2021-12/2021

**Final Project: Establishing a Website of World Happiness Score and Related Factors using R and Shiny App**

- Built a [GitHub Website](#) to visualize the happiness score in different countries and correlated social-economic factors
- Analyzed and visualized the yearly worldwide happiness score and GDP, COVID-19 Status, Life Pressure, and other factors by using interactive plots in ggplot, Plotly, and [shiny app](#) in R
- Established a multivariable linear regression model to explore the effect of social factors on happiness level and applied five-fold cross-validation to verify the model
- Top rank (A) received

**Prof. Roger Foo's lab, Cardiovascular Research Institute, National University of Singapore**

01/2021-5/2021

**Graduate Thesis: Investigating the role of ADAR-mediated RNA-editing in Cardiomyopathy**

- Established a computational pipeline in Linux and R for RNA-editing site identification from 324 large RNAseq data
- Performed multithread processing and core distribution by using R in the remote Linux Server
- Wrote dissertation and made thesis presentation to Department Dean and professors, top ranking received

## EXTRACURRICULAR ACTIVITIES

---

**VP for Activity Management, CUMC CSSA, Columbia University**

9/2021-7/2022

- Made the organization's yearly budget plan and activity schedule
- Launched campus-level events such as Professional Networking Series and practiced leadership by allocating work

**Outstanding Participants, 17<sup>th</sup> Qiangyin Plan for entrepreneurship, Zhejiang University**

11/2018-04/2019

- Wrote a business proposal for Anji Environment Protective Center to promote a sustainable business and public welfare model. The proposal was selected from 40 competitors to be presented on a public roadshow

**President, Residential College Student Committee, ZJU International Campus**

09/2018-06/2019

- Organized school-level events such as High table dinner (350 participants) and allocated related work to 20 staff
- Represented university to participate in the 5th Cross-Strait Forum on Education of Modern Colleges in Hong Kong
- Established connections with Residential College in more than 4 universities (Oxford University, HKU, University of Macao, SUSTC, Fudan University, etc) and launched exchange events, improving the reputation of Residential College globally.