# Zongchao Liu

zl2860@cumc.columbia.edu | 646-249-6941 | 100 Haven Ave, Tower 3, New York, NY 10032

## EDUCATION

**Mailman School of Public Health, Columbia University**      *New York, USA*
*Master of Science* in Biostatistics, GPA: 3.92/4.3      *Expected May.2021*

**School of Public Health, Shandong University**      *Jinan, China*
*Bachelor of Medical Science in* Preventive Medicine, GPA: 3.96/4.0      *Jun. 2019*

**School of Population and Public Health, University of British Columbia** *Vancouver, Canada*
*Exchange student*      *Jul. 2017 -Aug. 2017*

## SKILLS

**Knowledge:** Biostatistical Methods, Statistical Inference, Epidemiology, Computational Statistics, Machine Learning, Deep Learning(DNN,VAE-GANs, CNN), Basic & Clinical Medicine, Public Health

**Technical:** R, SAS, SPSS, Neo4j, Octave, EpiData, Python, MATLAB, LaTeX, Git version control

**Language:** Fluently both orally and written in English, Cantonese and Mandarin

## RESEARCH EXPERIENCE

**Columbia University Irving Medical Center**      *Mar.2020 – present*

- Constructed a Variational Auto-Encoder using PyTorch for generating Model-X knockoffs, which is a novel way to conduct variable selection among high-dimensional data while controlling False Discovery Rate(FDR)

**Department of Oncology, Sixth Affiliated Hospital of SYSU**      *Jun.2019 – Jan.2020*

- Applied Laplacian of Gaussian and wavelet filters in Python to process CT images and extracted 1219 first order statistics, texture and shape features from Gray Level Co-occurrence Matrix and Gray Level Run Length Matrix of the transformed and original images

- Selected the robust features that can be used for cancer prediction via LASSO regression and finally identified nearly 22 robust features by using R packages

- Calculated the rad-score for each patient based on the selected features and conducted a survival analysis among patients with high, medium and low levels of rad-scores, verifying the predictive value of the selected features

- Constructed predictive models including random forest, boosting, logistic regression to

predict the chances of being diseased based on the characteristics of patients, yielding a high value in overall accuracy.

**Department of Biostatistics, Shandong University**                    *Jan.2018 – Jun.2019*

- Built an improved gray model(1,1) with high accuracy by Python to predict the incidence rates of cervical cancer and endometrial carcinoma from 2018 to 2020 in Shandong, verifying other previous prediction of the incidence rates

- Conducted an epidemiology study by presenting the crude, age-standardized and urban(rural) incidence rates of cervical cancer and endometrial carcinoma in Shandong Province, 2013~2017

- Designed and constructed a database for nearly 9,900 published articles about GWAS of cancer and cardiovascular, documenting specific SNPs, statistical methods, adjusted covariates, interval validation results and other parameters of each study

- Collected data using a multistage sampling method and standardized the dataset on healthcare services and drug usage in Weihai from the 1980s, improving the validity of 136 datasets with more than 10,000 variables of approximately 5 million people in Shandong Province

- Designed and constructed Diabetes Knowledge Graph using Neo4j by coding specific nodes and relationships including the complete process for screening, diagnosis, treatment, and education

**Center for Hygienic Analysis and Detection, Shandong University**     *Apr. 2017 - Mar. 2018*

- Measured the TFA content in 69 types of semi-solid lipid and liquid milk teas via gas chromatography, following the national standard (GB5009.257-2016)

- Conducted quantitative and qualitative research on GC-2010 Gas Chromatography to validate the nutrient contents of samples

- Analyzed the collected data and drafted a research report for customers, presenting the risk index on TFA intake and improving the accuracy of nutrition labels

**School of Dentistry, Shandong University**                          *Sept. 2016 - Mar. 2018*

- Investigated the role of Periostin (OSF-2) in bone tissue regeneration and discussed potential strategies for nano-based drug delivery, such as nephroblastoma overexpressed (CCN3/NOV) and human laminin$\gamma$2 chain, in a mini-review

- Compared the feasibility and delivery efficiency of different types of nanoparticles carrying miRNA-21 to promote bone tissue regeneration

- Proposed the significance of phage display methodology in seeking peptide motifs of translocating nanoparticles

## RELEVANT PROJECTS

**Fragility Index for Clinical Trials**

- Designed and constructed a dataset containing general information of more than 320000 clinical trials based on the sources from clinicaltrials.gov
- Constructed a shiny-app visualizing the spatial distribution of the clinical trials in the United States
- Assisted in developing a fragility index calculator for meta analyses with different methods, measures and random effects

**A Simulation Study to Compare Two Bootstrapping Methods for propensity-score matching**

- Generated 15 scenarios of epidemiological confounding data with weak, medium and strong confounding relationship between covariates and continuous & binary outcomes
- Randomly Generated 1000 datasets for each scenarios for calculating the true effect and variance
- Conducted the propensity score matching method via both complex and simple bootstrap to calculate and compare the variability of the average treatment effects with true variance

**Implementation and optimization of algorithms on cancer diagnosis dataset**

- Built a predictive model based on Logistic Regression to facilitate cancer diagnosis
- Trained Logistic Regression models with Newton Raphson, Gradient Decent algorithms from scratch in R (No package used)
- Decreased the misclassification rate by 4%, via implementing a Logistic-LASSO Regression model with Path-wise Coordinate Descent

Evaluation of the spread of COVID-19

**Analyses of daily COVID-19 cases across nations**

- Develop the Newton-Raphson algorithm to fit a logistic curve to each region
- Apply K-mean and Gaussian mixture model with EM algorithm to cluster the fitted parameters

# RELEVANT WORK EXPERIENCES

**Qingdao Center for Disease Control and Prevention**                    *Qingdao, China*
*Staff Member*                                                          *Feb. 2019 – Jun. 2019*

- Managed health records of the citizens in Tsingdao and assisted in occupational disease assessment meetings
- Conducted multivariate logistic regression to analyze the relationship between occupational factors and women depression and to analyze the relationship between tobacco use and occupational diseases

**Shandong Qianfoshan Hospital**                                       *Jinan, China*
*Intern Physician*                                                     *May. 2017 - Jul. 2017*

- Monitored nutritional and health status of patients and collaborated with surgeons in surgeries in obstetrics and urology
- Input and standardized data on infants' health status and children's bone density using EpiData

## HONORS & AWARDS

| | |
|---|---:|
| Outstanding Graduates | *2019* |
| Excellent Student Scholarship (2017~2018) | *2019* |
| Excellent Student Scholarship (2016~2017) | *2018* |
| Member of the Elite Class, Research Center for Eco-Environmental Sciences of Chinese Academy of Sciences | *2018* |
| First Prize, Shenzhen Cup Mathematical Modeling Competition | *2016* |
| Bronze Award, Information Technology and Information Innovation and Entrepreneurship Competition in Shandong Province | *2017* |
| Excellent Student Scholarship (2015~2016) | *2016* |