JUNZHE SHAO

Columbia University, Mailman School of Public Health
(+1)858-766-1048 ⋄ js5959@columbia.edu

EDUCATION

Columbia University, Mailman School of Public Health Sep 2021 - Present
M.S. in Biostatistics

Peking University, School of the Life Sciences Sep 2016 - June 2021
B.S. in Biological Science
Minor in Physics

University of California, San Diego Sep 2019 - Sep 2020
Research Assistant at Jain Lab

RESEARCH EXPERIENCE

Causal Inference for Non-Stationary Time Series Data in Digital Psychiatry. (On going)
November 2021 - Present
Advisor: Professor Linda Valeri, Professor David Blei

- A novel parametric generalization of the synthetic control approach when the invariance assumption is violated.
- Demonstrate the model's flexibility by using a state-space model, a Kalman filter/smoother estimator for a time-varying weight.
- Using a generalized synthetic control approach to estimate heterogeneous treatment effect for non-stationary time series of COVID-19 with digital psychiatry data.
- Developing an R package SSMimpute implementing the missing data imputation approach for "State space model multiple imputation for missing data in non-stationary multivariate time series".

Integrative High-throughput Metabolomics Analysis of Pulmonary Arterial Hypertension Phenotypes and Outcomes
Sep 2019 - June 2021
Advisor: Professor Mohit Jain. Co-Advisor: Tao Long, Head of Bioinformatics.

Co-first authored manuscript: *Bioactive Metabolomic Profiles of Scleroderma-PAH are different than idiopathic PAH and associated with worse clinical outcomes*, submitted to *Chest*, under revision.
`https://www.medrxiv.org/content/10.1101/2021.07.10.21259355v1`

- Built a regularized regression statistical model based on high throughput mass spectrum of bioactive lipids in plasma sample to do mortality prediction of Pulmonary Arterial Hypertension (PAH).
- The model outperformed traditional clinical variants in the metric of AUC(0.78). Gave out a quick and non-invasive mortality risk score to be practically utilized.
- Further studied the properties of metabolites by molecular networking and the causal inference by Mendelian randomization.
- Adapted the model on the subtype prediction of PAH type I, using scleroderma and IPAH as the new response.

Image Based Age and Life Expectancy Prediction of C.elegans
Mar 2021 - Present
Advisor: Professor Jingdong Jackie Han

- Developed a deep learning method based on Inception-ResNet-V2 of image processing to identify the movement of *Caenorhabditis elegans* model from microscopic video of different levels of resolution.
- Proposed a novel Multi-task learning approach for both prediction of mortality and life expectancy to get generalization ability.
- Defined features of vitality index and calculated frailty score for predicting the chronological age and life expectancy of worms.
- Achieve a Mean Absolute Error of 1.8 days for age results and 2.6 days for lifespan results.

RELATED COURSES

Mathematics Courses: Mathematical Analysis, Linear Algebra, Mathematical Methods for Physics (Complex Analysis and PDE)

Statistics Courses: Probability and Mathematical Statistics, Applied Regression Analysis, Casual Inference (PhD-Level), Applied Causality (PhD-Level) , Data Science I (Data Analysis Using R), Data Science II (Statistical Learning), Biostatistical Method I (Statistical Inference and Linear Regression), Biostatistical Method II (Generalized Linear Regression and Longitudinal Data Analysis), Applied Stochastic Process, Data Analysis of Genomics, Mathematical Modeling in the Life Sciences

Physics Courses: Theoretical Mechanics, Electrodynamics, Equilibrium Statistical Physics, Quantum Physics, Solid State Physics

EXTRACURRICULAR ACTIVITIES

Piano Club, PKU June 2016 - Present
Beijing, China

Played Lindraja by Debussy in Duo Piano Concert of Peking University, 2017

SKILLS AND TECHNICAL STRENGTHS

Programming Languages	R, Python, MATLAB, C
Software & Tools	LaTeX, Microsoft Office, Adobe Photoshop
Languages	Mandarian(Native), English(Proficient)