

# Yi Han

677 Huntington Ave, Boston, MA 02115 | 607-327-7183  
yhan@hsph.harvard.edu | yiiihan.github.io

## EDUCATION

### Shanghai Jiaotong University (SJTU), Shanghai, China

Sept. 2019 - Jun. 2023 (Expected)

*Bachelor of Science in Statistics*

GPA: 3.93/4.3 ; Rank: 2/48

- **Relevant Coursework:** Probability Theory (A+), Multivariate Statistics (A+), Stochastic Processes (A+), Optimization (A), Stochastic Simulation (A), Mathematical Analysis (A+), Numerical Analysis (A), Math Finance (A), PDEs (A), ODEs (A), Complex Analysis (A)
- **Major Honors:** Fan Hsu-chi Scholars (awarded to 10 undergraduates each year), Outstanding student (top 5%)

### Cornell University, Ithaca, NY

Jan. 2022 - May. 2022

*Exchange Student, Department of Statistics and Data Science*

GPA: 3.957/4.3

- **Relevant Coursework:** Machine Learning (A), Time Series Analysis (A), Undergraduate Study in Statistics (A+)

### Harvard T.H. Chan School of Public Health, Boston, MA

Sep. 2022 - Feb.2023

*Visiting Student, Department of Biostatistics*

## PUBLICATIONS

- [1] Gu T, **Han Y**, Duan R.(2022) A transfer learning approach based on random forest with application to breast cancer prediction in underrepresented populations.*Proceedings of Pacific Symposium on Biocomputing* [\[LINK\]](#)
- [2] Gu T, **Han Y**, Duan R.(2022) Robust angle-based transfer learning built on ridge regression. *ArXiv* [\[LINK\]](#)

## EXPERIENCE

### Research Assistant

Department of Biostatistics, Harvard T.H.Chan School of Public Health

*Supervised by Dr. Rui Duan*

Jun. 2022 - Present

- Exploit transfer learning(TL) methods based on machine learning algorithms and application on biomedical datasets
  - Proposed a random forest-based transfer learning framework to incorporate risk prediction models trained in a source population to improve the prediction performance in a target underrepresented population with limited sample size.
  - Designed a robust angle-based transfer learning approach that leverages the concordance between the source and the target model parameters, which unifies several benchmark methods by construction.
  - Reproduced existing DNN and CNN transfer learning algorithms with Pytorch to biomedical data.
  - Developed Rpackage **multiTL** for multiple transfer learning methods [\[LINK\]](#)
- Designed federated learning method in mix-typed data integration
  - Proposed a ranking-based polygenic scores(PRS) ensemble method and generated PRSs using multiple summary statistics and PRS estimation method (LDpred, PRSCS, SDPR, Sbayes, Lassosum).
  - Designed a federated clustering method with summary level data.
- Conducted statistical analysis using Wilcoxon signed rank test and GLMM on survey data in Guinea Epilepsy Project.

### How does mask mandate effect online learning? From a regression discontinuity perspective

Cornell University

*Independent Study Advised by Dr. Yang Ning*

Feb. 2022 - May 2022

- Constructed a regression discontinuity framework including sharp regression with discontinuity design and constant treatment effect model and discovered negative treatment effects on the face mask mandate had on online learning.
- Evaluated the influence of the COVID-19 pandemic on education inequity by comparing the effects of face masks across school districts varying in percentage of minorities and per pupil total expenditure.

### Selecting Hyper-parameters for Options Pricing Model

Financial Engineering Research Center, SJTU

*Independent Study Advised by Dr. Yingda Song*

Jun. 2021 - Mar. 2022

- Conducted simulations on five continuous Markov chain grids in the pricing of European double barrier options.
- Analyzed the applicability and features of these grid design methods under the choice of underlying asset model.
- Designed adaptive grids by iterating continuous Markov chain to simulate strike prices at expiration using Monte Carlo.

### Multi-factor Stock Selection Model Based on Regression Model

MS&E, Stanford University

*Independent Study Advised by Dr. Chenru Liu*

Jan. 2021 - Apr. 2021

- Preprocessed stock data to construct market and financial indicators and validate factors using single-factor analysis.
- Performed PCA tests to select factors and reduce multicollinearity.

### Data Science Intern

Shanghai Fields Technology

*Intern at the Data & Algorithm Team*

Jul. 2021 - Aug. 2021

- Performed front-end configuration to regional carbon rating related data using SQL in investment advisory system.
- Utilized Python module to crawl statistical contents on web pages.

## ADDITIONAL INFORMATION

- **Programming Skills:** R, Python, Matlab, Latex
- **Language:** Chinese (native), English (fluent), Korean (beginner)
  - TOEFL iBT: 110/120 (Reading 30, Listening 30, Speaking 23, Writing 27)
  - GRE General: Verbal 161/170, Quantitative 170/170, Analytical Writing 4.0/6.0
- **Leadership:** President of Student Union, School of Mathematics, SJTU