

Hanyu Wang

Email : hanyv.wang@yale.edu

GitHub: <https://github.com/hanyvwang>

Homepage: <https://hanyvwang.github.io/>

OVERVIEW

Highly quantitative MSc student in **Epidemiology of Infectious Diseases (Quantitative Specialization)** at Yale University, with a strong background in Statistics and Quantitative Biology. Research focuses on **applying advanced statistical modeling and machine learning** to understand how environmental factors influence infectious disease dynamics and public health outcomes. Seeking a PhD position to integrate **epidemiological modeling, environmental data, and advanced computation** for actionable disease prevention and risk mitigation.

EDUCATION

- Yale University | Master of Science** New Haven, CT, USA
Epidemiology of Infectious Diseases - Quantitative Specialization *Expected May 2026*
Relevant Coursework: *Quantitative Methods for Infectious Disease (R), Global Change and Infectious Disease, Advanced Regression Model (R), Transmission Dynamic Models for Infectious Diseases (R), Machine Learning & Causal Inference (R)*
- University of Toronto | Honours Bachelor of Science** Toronto, Canada
Statistics Major, Quantitative Biology Major *Jun 2025*
Relevant Coursework: *Time Series Analysis (R), Methods for Multivariate Data (R), Machine Learning (R & Python), Experimental Analysis (R), Population Ecology (R), Computer Programming (Python)*

WORK EXPERIENCE

- Multifactorial Compartmental Model for Infectious Disease Dynamics** York University
Research Project (Supervisor: Prof. Jianhong Wu) *May 2024 - Aug 2024*
 - Developed an optimized **SEIRS compartmental model** to simulate infectious disease transmission, incorporating demographic and biological factors.
 - Integrated **stochasticity and spatial dynamics**, explicitly modeling birth, death, and immune decay via computational simulations.
 - Collaborated with a team of four to refine model accuracy; enhanced expertise in **mathematical modeling and public health informatics**.
- Single-Cell Transcriptome Sequencing Data Analysis** Xibei Hospital
Research Project *Sep 2024 - Dec 2024*
 - Analyzed scRNA-seq data from enteritis patients to study gene expression heterogeneity and identify therapeutic cell populations.
 - Built **R pipelines** using **Seurat**, implementing PCA, Louvain/Leiden clustering, and visualization (t-SNE, UMAP).
 - Identified **10 cell subtypes from 19 clusters**, advancing skills in biomedical data analysis and machine learning.

ACADEMIC PROJECT

- Identifying Autism-Specific Brain Connectivity Patterns | University of Toronto | Jan - Apr 2025** [view]
 - Analyzed ABIDE fMRI data using multivariate statistical methods including **correlation matrices, permutation testing, effect-size estimation, and FDR correction** to identify ASD-related connectivity differences.
 - Found robust hyperconnectivity centered on the Left Middle Temporal Gyrus (posterior division), refining understanding of ASD neural mechanisms.
- Prediction for Alzheimer's Disease Risk and Progression | University of Toronto | Sep – Dec 2024** [view]
 - Designed and compared **Logistic Regression, Random Forest, and Naive Bayes** models, achieving **94.33% accuracy**.
 - Enhanced robustness via **grid search hyperparameter tuning, k-fold cross-validation, and SMOTE** for data balancing.

TECHNICAL & RESEARCH SKILLS

Programming & Software

R (Expert): Custom pipelines, **Seurat**, time series, multivariate analysis, experimental design

Python (Proficient): **scikit-learn**, Scanpy, machine learning

Statistical Modeling & Machine Learning

Probabilistic Modeling: **Gaussian Processes** (self-study/projects), **Gaussian Mixture Models**

Algorithms: Logistic Regression, Random Forest, LDA, Naive Bayes, PCA, UMAP, t-SNE

Techniques: Grid search, k-fold CV, SMOTE

Research Domains

Epidemiology: **SEIRS modeling**, infectious disease dynamics, public health analytics

Biomedical Informatics: scRNA-seq, gene expression, computational neuroscience