Chuqi Wang

Irvine, CA 92617 | (949) 849-3189 | chuqiw4@uci.edu

[GitHub](#) | [LinkedIn](#) | [Website](#)

RESEARCH INTERESTS

Statistical methods and applications, machine learning, causal inference, experimental design, biostatistics and applications in health-related problems, such as mental health issues and public health concerns.

EDUCATION

University of California, Irvine   *Irvine, CA, USA*
Master of Data Science   *Sept. 2023 – Expected Dec. 2024*

- GPA: 3.97/4.0
- Relevant Courses: Databases & Data Management, Artificial Intelligence, Probability & Statistical Theory, Statistical Methods, Bayesian Data Analysis, Machine Learning & Data Mining

McGill University   *Montreal, QC, Canada*
Bachelor of Science in Statistics, minor in Computer Science   *Sept. 2018 – May 2022*

- GPA: 3.56/4.0, Major GPA: 3.76/4.0
- Relevant Courses: Algorithm & Data Structures, Advanced Calculus, Algebra and Analysis, Statistical Learning, Mathematical Statistics, Generalized Linear Models

PUBLICATION

- Wang, C. (2023, January). A REVIEW on 3D convolutional neural network. In 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA) (pp. 1204-1208). IEEE. [[Link](#)]

EXPERIENCE

Olivares Lab, UCI Civil & Environmental Engineering   *Jun. 2024 – Present*
Data Analyst – Research Team member (Advisor: Prof. Christopher Olivares & Jialin Dong)

- Collected, cleaned and processed PFAS contamination data from 50 U.S. states, including UCMR 3&5 data, using Pandas and Camelot to extract and cleaned raw data from various sources such as PDF reports, resulting in a final dataset of 1.4 million rows focused on key contamination in drinking water.
- Conducted in-depth exploratory data analysis (EDA) using Matplotlib, Plotly, and Seaborn to visualize concentrations of six major PFAS contaminants across U.S. states, including interactive maps and bar charts.

Financial Multimodal Large Language Model Research   *Jun. 2024 – Present*
Graduate Student Researcher (Advisor: Prof. Weining Shen)

- Conducted weekly literature reviews of 5 papers on financial multimodal LLMs and collected financial datasets to support model evaluation. Presented one selected paper in team meetings to highlight key findings and advancements.
- Evaluated 5 different large language models using FinEval dataset through API calls to assess their performance across various financial tasks under four scenarios: zero-shot, zero-shot CoT, few-shot, and few-shot CoT. Compiled and presented results in detailed comparison tables to support model selection and improvement decisions.

Pelvic Floor Disorders Research Lab, UCI Health   *Jun. 2024 – Aug. 2024*
Data Analyst (Advisor: Dr. Olivia Chang)

- Utilized Pandas to clean, filter, and merge datasets containing over 4 million patient records from the 2019-2022 NSQIP database and conducted comparative statistical analysis of patients who underwent "Vaginoplasty with peritoneal pull-through" versus "Vaginoplasty alone".
- Developed interactive dashboards using Tableau to present data visualizations for researchers. Performed logistic regression analysis with stepwise selection to predict composite outcomes of patients who underwent transgender surgeries, achieving 94.89% accuracy and an AUC of 0.865, with a cross-validation error of 0.056.

PROJECTS

Stroke Prediction Using Bayesian Logistics Regression   *Feb. 2024 – Mar. 2024*

- Developed a Bayesian logistic regression model using the rstan package in R, fitted with 2000 iterations and 4 Markov chains via Markov Chain Monte Carlo (MCMC), to predict stroke occurrence based on 5110 patients' demographic, medical, and lifestyle data. Achieved a 95.2% test accuracy and improved model sensitivity through decision threshold adjustment.

- Performed data preprocessing and exploratory data analysis (EDA) using dplyr and ggplot, and applied diagnostic tools like Bulk ESS, Tail ESS, and trace plots to ensure model convergence and reliability.

Midwifery Services Database Application   *Jan. 2022 – Apr. 2022*
- Designed an Entity-Relationship (ER) diagram and a relational database model for the Quebec Ministry of Health to efficiently manage midwifery services. Established and executed a database schema using DB2.
- Authored comprehensive SQL queries for data population, maintenance, and updates. Engineered a user-friendly database application tailored for midwives via Java Database Connectivity (JDBC).

SKILLS

Programming Languages	Python, R, Java, SQL, MATLAB
Frameworks & Tools	Jupyter, SciKit-Learn, TensorFlow, Tableau, Git, LaTeX