Chuqi Wang

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

Research Interests

Statistical methods and applications, machine learning and data mining, causal inference, biostatistics, with a focus on health-related problems, including mental health, environmental issues, food safety, public health, and other related areas. I am particularly motivated to explore statistical methods in interdisciplinary research, such as personalized medicine, and to develop AI systems for health monitoring and prevention through multimodal models.

Education

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

- GPA: 3.97/4.0
- Relevant Courses: Databases & Data Management, Big 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 & Analysis, Statistical Learning, Mathematical Statistics, Generalized Linear Models

Publications

- 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]
- Zhang, L., Cai, W., Liu, Z., Yang, Z., Dai, W., Liao, Y., ... & Chen, Y. (2023). Fineval: A chinese financial domain knowledge evaluation benchmark for large language models. arXiv preprint arXiv:2308.09975. (under review via ACL Rolling Review, submitted October 2024) [Link]

Research Experience

Survey Research on AI Applications in Sports *Oct. 2024 – Present*
Graduate Student Researcher (Advisor: Prof. Weining Shen)

- Reviewed and summarized critical findings of research papers in AI Applications in Sports, consolidated current processes, prospects, and challenges in applying computer vision, NLP, and multimodal LLMs to sports analytics, supporting the development of a comprehensive survey.
- Delivered in-depth presentations and led discussions to align project direction, provide strategic insights, and guide the team in identifying research gaps and opportunities.

Olivares Lab, UCI Civil & Environmental Engineering *Jun. 2024 – Present*
Research Team Member (Advisor: Prof. Christopher Olivares)

- Collected, cleaned and processed PFAS contamination data for drinking water from 20 public sources across 50 U.S. states. Utilized Pandas and Camelot to extract and clean raw data from various formats to build a structured dataset of 2.5 million samples for PFAS analysis.
- Conducted exploratory and spatial analyses with Matplotlib, Plotly, and Seaborn, using interactive maps to visualize PFAS concentration distributions, trends over time, and regional comparisons across states.
- Summarized and organized 40+ historical research papers, consolidated critical findings in presentation to guide the direction of research, currently applying machine learning models to predict per-and polyfluoroalkyl substances (PFAS) concentration across different states and co-authoring a research paper based on analysis results.

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

- Conducted a literature review of 50+ papers on financial multimodal LLMs, collected financial datasets to support survey research, and presented selected papers in team meetings to highlight key findings and advancements.
- Evaluated large language models on financial tasks using the FinEval benchmark via API calls, testing four scenarios: zero-shot, zero-shot CoT, five-shot, and five-shot CoT. Conducted comparative analysis to identify 19 different models' strengths and weaknesses, guiding model selection for financial applications.
- Co-authored a research paper on the FinEval benchmark with Xin Guo et al., performed data visualizations and presented findings through extensive LaTeX formatted tables and figures to enhance clarity and interpretability of benchmark questions and model results.

Pelvic Floor Disorders Research Lab, UCI Health *Jun. 2024 – Aug. 2024*

Individual Study Researcher (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", providing insights into surgical outcomes to guide clinical decision-making and support future studies on transgender healthcare.
- Developed interactive dashboards using Tableau to present data visualizations. 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.

RELEVANT PROJECTS

Climate AI-Capstone Project [GitHub] *(Industry Partner: CCEX Company)* *Sept. 2024 – Dec. 2024*

- Developed Python scripts using Selenium to scrape and preprocess approximately 2,000 verified carbon standard reports, requirements, and templates from the Verra registry. Extracted content and segmented PDF reports using pdfplumber, and generated embeddings with OpenAI's embedding models, enabling fast retrieval and efficient semantic search capabilities.
- Led the integration of LLMs via API within a Retrieval-Augmented Generation (RAG) framework and utilized Docker containers to develop Climate AI, which provided project report recommendations and generated project description documents that met Verified Carbon Standard (VCS) requirements.
- Enhanced Climate AI into a multimodal chatbot with an expanded knowledge base and follow-up question suggestions to support expert guidance, streamline project description drafting for certification, and improve customer satisfaction, cost efficiency, and risk mitigation.

Stroke Prediction Using Bayesian Logistics Regression [GitHub] *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 95.2% test accuracy and improved sensitivity by adjusting the decision threshold, enabling better stroke case identification.
- Performed data preprocessing and exploratory data analysis (EDA) using dplyr and ggplot, and applied diagnostic tools such as Bulk ESS, Tail ESS, and trace plots to ensure model convergence and reliability. Identified age and glucose levels as significant predictors, providing insights for targeted prevention and intervention strategies.

Midwifery Services Database Application [GitHub] *Jan. 2022 – Apr. 2022*

- Designed an Entity-Relationship (ER) diagram and relational database model for the Quebec Ministry of Health to manage midwifery services efficiently. Implemented and deployed the database schema using DB2.
- Developed comprehensive SQL queries for data population, maintenance, and updates. Built a user-friendly database application for midwives using Java Database Connectivity (JDBC).

SKILLS

Programming Languages:	Python, R, Java, SQL, C/C++, MATLAB
Frameworks & Tools:	Jupyter, SciKit-Learn, TensorFlow, PyTorch, Tableau, AWS, LaTeX
Databases:	MySQL, PostgreSQL, Apache Cassandra, MongoDB, Neo4J, Apache Spark