Shihui Zhu

Email: sz3029@cumc.columbia.edu
Linkedin: https://www.linkedin.com/in/shihui-zhu-319266173/
Mobile: +1(858)-736-5630
Github: github.com/sz3029

Education

- **University of California, San Diego** — San Diego, CA
  Bachelor of Science - Mathematics-CS and Bioengineering:Bioinformatics; GPA: 3.7 — *Sep. 2017 - June 2021*
  Courses: *Data Structures, Algorithms, Machine Learning, Databases, Bioinformatics, NGS*
- **Columbia University** — New York, NY
  Master of Science - Biostatistics (Theory and Method Track); GPA: 3.9 — *Sep. 2021 - Present*
  Courses: *Statistics Inference, Machine Learning, Deep Learning, Data Mining, Databases, RCT, Longitudinal and Survival Analysis*

Skills Summary

- **Languages**: Python, R, SAS, SQL, C++, Bash, JAVA
- **Frameworks**: Scikit, Pytorch, TensorFlow, Pandas, Caret, Tidyverse, Flask
- **Tools and IDEs**: GIT, MySQL, Pycharm, DataGrip
- **Platforms**: Linux, Web, Windows, Raspberry, AWS, Azure

Experience

- **Research Assistant at Columbia University Medical Center** — New York, NY
  Clinical Research (ML) (On Campus, Part-time) — *Oct. 2021 - Present*
  - **Data Collection and EDA**: Data Cleaning, Feature Selection and AB testing for 1M rows of Clinical Data (30-day mortality) from Anesthesiology Department. Bootstrapping to accommodate Data Imbalanceness.
  - **Modeling and Model Evaluation**: Data Normalization and Standardization. Evaluation via CV for traditionally used subset selection and Tree Models, ML models Random Forest, XGBoosting, Logistic Regression, and DNN model.
  - **Model Interpretation**: Interpreted model prediction via SHAP values. Interpreted how much each feature contributed differently among different patients.
- **Yrobot Inc.** — Remote
  Data Analyst and Machine Learning (Internship) — *May 2022 - Sep. 2022*
  - **Libraries for walking data analysis**: Created python libraries that formulate the raw IMU sensor data, and design plots for patients' walking data analysis via plotly
  - **DNN-based model adaptation for 3D Human Motion Estimation**: Adapted the model for the company's experimental data.

Projects

- **ML - Improving Automated CLL/ALL Diagnosis via Machine Learning and CNN models on Flow Cytometry Dataset**: High-dimensional Flow Cytometry Data Classification for Leukemia patients via the UMAP and CNN models. Tech: Python, Scikit, matlibplot (June '21). Project Website: https://sites.google.com/ucsd.edu/group232021leukemiadiagnosis
- **ML - Training and Comparing Machine Learning Models for HCV diagnosis**: Use linear regression, logistic regression, GLM, discriminant analysis and tree models to improve the traditional HCV diagnosis. (Oct. '21). Tech: R. Project Website: https://github.com/MefiMefi/P8106_Final/blob/main/P8106_Final_Report.pdf
- **DeepL - Improving Estimate of Prior in Variational Nonparametric Empirical Bayes (Work in Progress)**: Use CNN model to improve variational autoencoder training with self-defined loss functions. Tech: Python (Nov. '21)
- **Web Development, Database, Machine Learning - Clinical Registration Suggestion System (Work in Progress)**: Search engine for patient to search their symptoms and choose the registration type. Tech: Flask, SQL, Python (September '22)
- **Data Analysis and Visualization - Investigate the Influence of COVID-19 Pandemic on Personal Expenditure in the U.S.**: EDA and regression analysis on impact of COVID-19 U.S. cases on personal income and expenditure. Project Website: http://sz3029.github.io/final_project/

Publications

- **Differential Translation Elongation in Yeast Cells**: Co-authored, published by RNA Biology, Volume 19, Issue 1. Tech: Bash, Python, IGV (November '21)
- **Interpretable machine learning models improve clinical understanding of factors with the greatest impact on perioperative mortality risk**: Work in Progress, to be published by International Anesthesia Research Society in late 2022. Tech: TensorFlow, Python, R (Oct. '21)

Awards and Certificates

- Awarded Second Prize in Clinical Research at Department of Anesthesiology, CUIMC - May, 2022
- SAS 9.4 Base Programming Certified - Jan, 2022
- College Honor and Provost Honor at UCSD - 2017 - 2021

Volunteer Experience

- **Mentor at UCSD Bioinformatics Club** — San Diego, CA
  Mentoring new undergrad Bioinformatics students — *Jan 2018 - Jan 2021*
- **Maps building for disaster responses** — Remote
  Teams organization, completing assigned projects from the MissingMap project website. — *Jan 2021*