Yanzhi Liu

315-229-2113 | yl1312@georgetown.edu | McLean, VA

Education

Georgetown University **Aug. 2020 – Dec. 2021**
M.S. in Biostatistics - GPA: 3.8/4.0 *Washington, DC*

North University of China **Sept. 2015 – Jul. 2019**
B.S. in Electrical Engineering - GPA: 3.82/4.0 *Taiyuan, China*

Employment

Georgetown University Medical Center **May. 2021 – Sept. 2021**
Research Assistant *Washington, DC*

- Implemented **Variational Bayes PCA** and **Conventional PCA** to conduct unsupervised Feature extraction
- Performed an integrated analysis of mRNA and miRNA expression from stressed mouse heart and investigated the underlying molecular biology and transcriptomic background of PTSD-mediated heart disease
- Utilized a latent class model with random effects to model the conditional dependence among multiple biomarkers. Achieved 89% of sensitivity and 92% of specificity
- Co-authored *Production of Anti-Spike Antibodies in Response to COVID Vaccine in Lymphoma Patients*

Georgetown University **Jan. 2021 – Dec. 2021**
Teaching Assistant *Washington, DC*

- Courses: BIST 510 Probability & Sampling, BIST 511 Statistical Inference, GLOH 177 Epidemiology
- Delivered and graded course materials, led and tutored weekly office hours, and improved lectures based on feedback

TCL Technology, Co. **May. 2018 – Aug. 2018**
Software Engineer Intern *Shenzhen, China*

- Participated in a collaborative project between TCL and Mozilla Thunderbird. Deployed **Knapsack** algorithm to meet the time limit requirements, analyzed different videos and made POC demo to auto-summarize educational videos
- Built a **Siamese Network** to solve the difference between laboratory condition and practical application environment based on **coarse-grained** and **fine-grained classification** and processed negative samples. Achieved 92% of accuracy
- Applied self-monitoring learning system to distinguish between constant speed and accelerated speed over 10K videos

Selected Projects

Covid-19 Blood RNA-Sequencing Analysis **Aug. 2021 – Oct. 2021**

- Constructed signature scores in blood RNA sequencing data and evaluated their diagnostic accuracy compared with the gold standard of PCR testing, by quantifying **AUROC**, sensitivities, and specificities
- Predicted the infection status by random forest model and calculated the accuracy and f1 score. Receiver operating characteristic (**ROC**) curve was used to better visualization of the models' performance

Evaluating accuracy of biomarker when gold standard is imperfect **Apr. 2020 – Aug. 2020**

- Checked data normality by Shapiro-Wilk test and homogeneity of variances by Bartlett test
- Employed T-test/Wilcoxon rank-sum test for 2 levels of variables, and Kruskal-Wallis test for higher levels of variables
- Utilized a latent class model with **Random Effects** to model the conditional dependence among multiple diagnostic tests and achieved at least **15**% improvement in sensitivity and specificity
- Contributed correlation residual plots to provide confidence bands and benchmarked the model adequacy
- Applied to the pathology review of therapeutic efficacy for pediatric high-grade gliomas and calculated progression-free survival probabilities

Article Embedding System **Sept. 2019 – Dec. 2019**

- Created document embeddings from the hidden layer of a feed-forward auto-encoder with connected weights in the hidden layer (**DocNADE**). Model was trained on an AWS GPU instance
- Designed personalized recommender systems with streaming data and employed multi-class supervised learning model. Built in **TensorFlow** and implemented with Docker on AWS EC2 instances
- Constructed an article popularity prediction model. Used **XGBoost** with batch scoring and explained the predictions from Local Interpretable Model-Agnostic Explanations (**LIME**)

Skills Summary

Languages: Python, SQL, JAVA, JavaScript, C/C++, HTML, CSS, R, SAS, MATLAB
Machine Learning framework: PyTorch, Caret, Scikit-learn, Keras, Tensorflow
Database: MySQL, Hadoop, Redis, Mybatis, MongoDB, DocumentDB, Kafka
Platform/Tool: Linux, AWS, GCP, Django, Spring MVC, Conda, Git