

Yunchang Xie

Tel.: (314) 445-7899 | Email: x.yunchang@wustl.edu

EDUCATION BACKGROUND

Washington University in St. Louis	St. Louis, Missouri
<i>Ph.D in Computational and Data Science</i>	<i>August 2024 - Present</i>
Washington University in St. Louis	St. Louis, Missouri
<i>Master of Science in Biostatistics and Data Science</i>	<i>August 2022 - December 2023</i>
Courses: Biostatistics, Data Mining, Deep Learning, Data Structures and Algorithms, Bioinformatics, Survival Analysis	
Lancaster University	Lancaster, UK
<i>Bachelor of Science in Psychology</i>	<i>October 2019 - June 2021</i>
Courses: Calculus, Calculus II, Linear algebra, Statistics, Research Methods, Cognitive Neuroscience, Clinical Psychology	

RESEARCH INTERESTS

- Machine Learning, Deep Learning
- Medical Image
- Multimodal Learning

RESEARCH PROJECTS

Washington University in St. Louis, St. Louis, Missouri
Predicting Brain Age and Analyzing Age Gap in Alzheimer’s Disease Using Multi-Omics and Neuroimaging Data
<i>Advisor: Aristides Sotiras, PhD</i> May 2024 - Present
<ul style="list-style-type: none">• Harmonized MRI data from Alzheimer's Disease Neuroimaging Initiative (ADNI) and ADRC by Combat. Imputed missing values in proteomics (7009 features), and metabolomics(411 features) with MissForest.• Trained multiple machine learning models (e.g., Support Vector Machine, Gradient Boosting Decision Tree, XGBoost) on normal control subjects, then applied these models to AD patients to calculate the age gap between predicted and chronological age.• Visualized the relationship between age gap and AD pathology and identified feature importance across multimodal data.
Investigating HIV-Associated Neurocognitive Disorder Through Non-negative Matrix Factorization
<i>Advisor: Aristides Sotiras, PhD</i> July 2023 - March 2024
<ul style="list-style-type: none">• Applied non-negative matrix factorization(NMF) to decompose 1k+ cortical images into component and coefficient matrices, and determined the optimal number of components (16) by evaluating reconstruction error and reproducibility.• Assessed the association of cognitive measures with each component in R using multiple statistical methods, including multiple linear regression, polynomial regression, and mixed effect models; selected the optimal models by Akaike information criterion(AIC) and likelihood ratio test.• Performed visualizations of significant brain components using FreeSurfer, with distinct colors to represent varying effect sizes.
Applying Deep Learning To Predict Epigenetic Age With DNA Methylation Data
<i>Advisor: Lei Liu, PhD</i> May 2023 - December 2023
<ul style="list-style-type: none">• Imputed missing values with KNN. Performed variable screening by Spearman correlation between age and each CpG site, finally selecting 270 CpG sites out of 860627 CpG sites for downstream model development.• Developed the deep neural network(DNN) by TensorFlow with grid search for fine-tuning the model, employing MAE and MSE as loss functions and improving the model robustness by adversarial training.• Compared model performances among neural network, elastic net and iterative SIS regressions, where MSE and Pearson correlation serve as the evaluation metrics.

WORK EXPERIENCES

ByteDance Technology Co. Ltd.	Beijing, China
<i>Operation Analyst Intern, TouTiao Operations Team</i>	<i>April 2021- October 2021</i>
<ul style="list-style-type: none">• Utilized MySQL to process 100k+ pieces of content data, and developed Tableau dashboards to monitor campaign performance across creators with varying follower counts, presenting the bi-weekly reports to stakeholders.	

- Initiated A/B testing to determine bonus mechanism for improving retention rate, used funnel model to attribute the root cause of global low conversion rate, bringing a daily 5k+ increase in posting volume.
- Developed data pipeline by Python (e.g. Pyspark) to extract and cleanse 2TB+ post-consumption data of multiple content categories on Cloud Platforms (like Google Bigquery), providing data sources to support cross-team collaboration.

SKILLS

- **Frameworks/Libraries:** TensorFlow, PyTorch, Keras, Scikit-learn, NumPy, Pandas, SciPy, Seaborn, Matplotlib, PySpark
- **Technologies:** MySQL, Python, R, MATLAB, SAS, SPSS, AWS, Google Big Query, Git, Linux, Java, Hadoop