

Haibin Guan

35-43 84th ST APT 321 Jackson Heights,NY | 7868326130 | hguan003@fiu.edu | www.linkedin.com/in/haibin-guan

EDUCATION

Florida International University
MS in Data Science (1/2017–8/2018)

Miami, FL, United States
transferred to MS in Computer Science (8/2018-5/2019)

Wenzhou Medical University
Bachelor in Preventive Medicine

Wenzhou, Zhejiang, China
9/2011–7/2016

PROJECTS

TReNDS Neuroimaging Kaggle Competition
Participant

4/2020-6/2020
Kaggle, the United States

- Applied PCA on 2 sets of given preprocessed features: static FNC correlation features and sMRI SBM loadings to reduce the dimension.
- Trained and optimized 3 baseline regression models (SVR, KNN and ElasticNetCV) with the above top 500 PCA components to predict the corresponding age and assessment value.
- Added a simple blending model with optimized blending weights on top of those 3 models in the end of my ML pipeline boosted the performance (feature-weighted, normalized absolute errors on 10-fold CV) by 0.0127.
- The best model among my total 76 submission got a final local CV 0.15986 and a final LB 0.15833. Ended up getting ranked 27/1047, a solo silver.

Metabolomics Data Integration System
Biomedical Data Analyst Intern

1/2018-7/2018
INFOTECHSoft, Miami, FL, the United States

- Assisted in a NIH funded research project aiming to develop a Metabolomics Data Integration System.
- Integrated multiple metabolite data across various laboratory platforms.
- Performed a complete QIIME 2 workflow included demultiplex, denoise/cluster, taxonomy classification, alignment and diversity analysis with the FASTQ data.
- Worked collaboratively with the bioinformatics researcher and software engineers.

Multiple Myeloma DREAM Challenge
Participant

7/2017-12/2017
Biorg Lab, Florida International University

- Used an algorithm that creates datasets with balanced distributions by combining oversampling by SMOTE and undersampling by Tomek due to the given four training data sets in the competition were known to have distinct subject profiles. The R Bioconductor limma package was used for assessing differential gene expression and returned a ranking of the genes. Utilized SVM-RFECV and MRMR approaches to further perform feature selection.
- The selected genes were then used to train six baseline classifiers: Support Vector Machine (SVM), Neural Network (NN), Random Forest (RF), Gradient Boosting Machine (GBM), Learning Vector Quantization (LVQ) and Generalized Linear Model (GLM). Applied stacking strategy in the end by training ensemble classifiers to combine the results of the above mentioned baseline classifiers (At the 2nd round of competition, this model achieved the first place out of 40 teams).

Modeling of influenza like illness prediction based on Elman neural network
CDC Disease Surveillance Intern

1/2016-6/2016
Zhejiang Provincial CDC, Hangzhou, Zhejiang, China

- 10 highly correlated weather factors were selected by examining the correlation between two time series variables contemporaneously and at 7 lagged values in R.
- Generated a predictive model by implementing Elman Neural Network (a 3-layer RNN) in MATLAB to provide an early detection of influenza pandemic in Zhejiang Province (the optimal model with NN structure 10-15-1 obtained 10.58 % mean error rate and 0.8767 nonlinear correlation coefficient).

TECHNICAL SKILLS

Programming Languages:	R, Python, MATLAB, Linux, XML/HTML, D3.js, CSS, LaTeX, SPSS
Frameworks:	TensorFlow, PyTorch
Development Tools:	Git, Docker, AWS
Database:	MySQL, PostgreSQL
Bioinformatical analysis Tools:	Bioconductor, QIIME, XCMS, CAMERA, PICRUSt, Kraken

PUBLICATIONS

- Pedro Soto, Ilia Ilmer, **Haibin Guan**, Jun Li, “Lightweight Projective Derivative Codes for Compressed Asynchronous Gradient Descent”. Available: <https://arxiv.org/abs/2201.12990v1>
- Pedro Soto, **Haibin Guan**, Jun Li. Locally Random P-adic Alloy Codes with Channel Coding Theorems for Distributed Coded Tensors. Available: <https://arxiv.org/abs/2202.03469v2>
- Ruiz-Perez, D., **Guan, H.**, Madhivanan, P. et al. So you think you can PLS-DA?. BMC Bioinformatics 21, 2 (2020). <https://doi.org/10.1186/s12859-019-3310-7>
- D. RuizPerez, **H. Guan**, P. Madhivanan, K. Mathee, and G. Narasimhan, “So you think you can plsda?” in2018IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), Oct 2018,pp. 1–1.
- T. Zhang, **H. Guan**, F.Li, and F.He, “Modeling of influenzalike illness prediction based on elman neural network,”Preventive Medicine, vol. 31, no. 2, p. 113, 2019. [Online]. Available: <http://www.zjfyxzz.com/CN/Y2019/V31/I2/113>

EXTRA-CURRICULAR

YOUTH ASSEMBLY OF THE UNITED NATIONS	New York, the United States
Youth Assembly Delegate	2/2016
Summer Work Travel Program	Estes Park,Colorado, the United States
Exchange Visitor	6/2015-9/2015
International Study in Gachon University	Seoul, Korea
Exchange Visitor	1/2015-2/2015