

[SHAOPENG] [GU]

☎ [605-691-1562] | ✉ [shaopenggu1990@gmail.com] | 🏠 [1428 5TH ST S, BROOKINGS, SD 57006]

Knowledgeable computational scientist with over 5 years of hands-on experience with programming languages such as Python, SAS, R, MATLAB, C++, Neural Network, BI tools, SQL and statistical and network modeling. Experienced with standard and advanced statistical analysis and machine learning methods, including multivariate linear and non-linear regression, normalization and regularization, SVM, random forest, Mutual Information, Lasso, KNN, xgboost, lightGBM, ElasticNet, and regression trees, among many other techniques. Vast experience with data mining and analysis to extract relevant and useful information from various types of data, including genomic/transcriptomic, healthcare/EMR, and engineering data, with capable skills to quickly adapt to new methods and data types.

EXPERIENCE

Research Assistant

[2018-05 — present]

Bioinformatics and Mathematical Biosciences Lab, South Dakota State University

1) Using machine learning methods to analyze and predict the outcomes of EMR data; 2) develop a feature selection Python package for DNA, RNA and Protein sequence data; 3) research and develop RNA-seq pipeline tools (**GeneQC**), specifically related to the quality control of read alignment process and end-stage interpretations; 4) develop a sequence data simulation Python tool

Research Assistant

[2017-08 — 2018-05]

Department of Electrical Engineering, South Dakota State University

1) Develop machine learning methods to analyze electricity consuming data; 2) develop computational tool to identify complex attack in smart grid security

IT Technician

[2017-08 — 2018-05]

Department of Instructional Design Service, South Dakota State University

1) Instructional services designing; 2) equipment maintenance and installing

Research Assistant

[2013-01 — 2016-01]

Organic Electronic Laboratory, South Dakota State University

1) Fabricate organic small molecule and polymer solar cells 2) performance measurements

EDUCATION

MS in Mathematics - Statistics Specialization, South Dakota State University

[2017-05 — present]

Coursework in Statistical Inference, Computational Intelligence, Statistical Programming, Measure & Probability Theory, Advanced Calculus

BS in Electrical Engineering, South Dakota State University

[2012-01 — 2017-05]

Coursework in C, C++, Data Structure, Computer Organization, Differential Equation, Calculus, Digital Image Processing, Signal, Probabilistic Methods, Microcontroller

PUBLICATIONS

- Agricultural Factors in Chronic Kidney Disease*. American Journal of Medicine. doi: 10.1016/j.amjmed.2019.03.036 [2019-03]
- Predicting outcomes of chronic kidney disease from EMR data based on Random Forest Regression*. Mathematical Biosciences. doi: 10.1016/j.mbs.2019.02.001 [2019-02]
- A new machine learning-based framework for mapping uncertainty analysis in RNA-Seq read alignment and gene expression estimation*. Frontiers in Genetics. doi: 10.3389/fgene.2018.00313 [2018-09]
- Solution processed pristine PDPP3T polymer as hole transport layer for efficient perovskite solar cells with slower degradation*. Solar Energy Materials and Solar Cells. Volume 145 [2016-02]
- Crystallization of a perovskite film for higher performance solar cells by controlling water concentration in methyl ammonium iodide precursor solution*. Nanoscale. Issue 5, 2016 [2016-06]
- Shelf life stability comparison in air for solution processed pristine PDPP3T polymer and doped spiro-OMeTAD as hole transport layer for perovskite solar cell*. Data in Brief. Volume 7, 2016 [2016-06]

COMMUNITY SERVICE

President, Chinese Student & Scholar Association, South Dakota State University [2013-01 — 2014-09]

SKILLS

- 5+ years working of programming languages (**Python**, R, SQL, MATLAB, SAS, C++, Perl)
- 2+ years of large-scale data mining and analysis with statistical & network modeling and machine learning methods
- 2 years of mathematical & statistical algorithm development
- 1 year of next-generation sequencing data analysis
- Strong written and oral communication
- Capabilities writing production-level code
- Ability to work independently and as a member of team