Zhengqiao **Zhao, PhD**

PRINCIPAL DATA SCIENTIST, CAPITAL ONE

Philadelphia, PA 19104

 z2e2.github.io | z2e2 | zhengqiaozhao

Research Summary

I'm a data scientist and bioinformatician with key talents in research, manuscript writing, and mentoring. I have successfully developed efficient machine learning algorithms to uncover hidden signals in genomic data and text data over the past 6 years. I am experienced in meeting tight deadlines, cross-disciplinary collaboration, and communicating data-driven solutions to a wide variety of stakeholders.

In my graduate studies, I've taken on various leadership roles, including instructing undergraduate courses/laboratories, mentoring undergraduate/graduate research assistants, and coordinating cross-disciplinary research projects. I was awarded the Research Excellence Award and Teaching Assistant Excellence Award by the Drexel Graduate College for my exemplary commitments to innovative research and student learning.

Education

Drexel University   *Philadelphia, PA*

DOCTOR OF PHILOSOPHY, ELECTRICAL & COMPUTER ENGINEERING   *2016 - 2020*

- Mentor: Gail L. Rosen PhD
- Lab: Ecological and Evolutionary Signal-Processing and Informatics
- Thesis: Improvements in machine learning for predicting taxon, phenotype and function from genetic sequences

Drexel University   *Philadelphia, PA*

MASTER OF SCIENCE, ELECTRICAL ENGINEERING   *2014 - 2016*

Northwestern Polytechnical University   *Xi'an, China*

BACHELOR OF ENGINEERING, ELECTRONICS AND INFORMATION ENGINEERING   *2010 - 2014*

Working Experience

2021.3-	**Principal Data Scientist, Risk Management**, Capital One	*McLean, VA*
2019.6-8	**Senior Data Scientist, Analyst Internship Program (AIP)**, Capital One	*Richmond, VA*
2015.6-9	**Electrical Engineering Intern**, Kulicke & Soffa	*Fort Washington, PA*

Awards & Honors

2020	**Research Excellence Awards**, The Graduate College of Drexel University	*Philadelphia, PA*
2019	**Teaching Assistant Excellence Award**, The Graduate College of Drexel University	*Philadelphia, PA*
2018	**Finalist, Teaching Assistant Excellence Award**, The Graduate College of Drexel University	*Philadelphia, PA*
2018	**Lerey L. Resser Endowed Fellowship**, The Drexel University College of Engineering	*Philadelphia, PA*
2017	**Lee Smith Traveling Fellowship**, The Graduate College of Drexel University	*Philadelphia, PA*
2016-18	**Graduate College Fellowship**, The Graduate College of Drexel University	*Philadelphia, PA*
2016	**Outstanding Promising Award (Honorable Mention)**, The Graduate College of Drexel University	*Philadelphia, PA*
2014-16	**Dean's Fellowship**, The Drexel University College of Engineering	*Philadelphia, PA*

Professional Affiliations

2016.6-	**IEEE Eta Kappa Nu (HKN)**
2016.1-	**Institute of Electrical and Electronics Engineers**

Scientific Contributions

Articles

- **Zhengqiao Zhao**, Bahrad A. Sokhansanj, Charvi Malhotra, Kitty Zheng, Gail Rosen, (2020). Genetic grouping of SARS-CoV-2 coronavirus sequences using informative subtype markers for pandemic spread visualization. *PLoS Computational Biology* [PLoS]
- **Zhengqiao Zhao**, Alexandru Cristian, Gail Rosen, (2020). Keeping up with the genomes: efficient learning of our increasing knowledge of the tree of life. *BMC Bioinformatics* [BMC]

- **Zhengqiao Zhao**, Stephen Woloszynek, Felix Agbavor, Joshua Chang Mell, Bahrad A. Sokhansanj, Gail Rosen, (2020). Learning, Visualizing and Exploring 16S rRNA Structure Using an Attention-based Deep Neural Network. *Cold Spring Harbor Laboratory* [bioRxiv]
- Taha Valizadehaslani, **Zhengqiao Zhao**, Bahrad A. Sokhansanj, Gail Rosen, (2020). Amino Acid k-mer Feature Extraction for Quantitative Antimicrobial Resistance (AMR) Prediction by Machine Learning and Model Interpretation for Biological Insights. *Biology* [MDPI]
- Stephen Woloszynek, **Zhengqiao Zhao**, Jian Chen, and Gail L. Rosen, (2019). 16S rRNA sequence embeddings: Meaningful numeric feature representations of nucleotide sequences that are convenient for downstream analyses. *PLOS Computational Biology* [PLOS]
- Stephen Woloszynek, Joshua Chang Mell, **Zhengqiao Zhao**, Gideon Simpson, Michael P. O'Connor, and Gail L. Rosen, (2019). Exploring thematic structure and predicted functionality of 16S rRNA amplicon data. *PLoS ONE* [PLOS]

Chapters and Reviews

- Cullen, Chad M. and Aneja, K. and Beyhan, S. and Cho, C. and Woloszynek, S. and Convertino, M. and McCoy, S. and Zhang, Y. and Anderson, M. and Alvarez-Ponce, D. and Smirnova, E. and Karstens, L. and Dorrestein, P. and Li, H. and Sen G. and Cheung, K. and Powers, J. and **Zhao, Z.** and Rosen, G., (2020). Emerging Priorities for Microbiome Research. *Frontiers in Microbiology* [Frontiers]
- S. Woloszynek, **Z. Zhao**, G. Ditzler, J. Price, E. Reichenberger, Y. Lan, J. Chen, J. Earl, S. Langroodi, G. Ehrlich, G. Rosen, (2018). Analysis Methods for Shotgun Metagenomics. Book chapter for *Theoretical and Applied Aspects of Systems Biology*, Editors: F. A. Silva, N. Carels, F. P. Silva Junior. [Springer]

Conference Papers and Patent applications

- **Zhengqiao Zhao**, Gail Rosen, (2018). Multi-temporal Information Object Incremental Learning Software System. *The United States Patent and Trademark Office Patent Application: us15/910845* [USPTO]
- **Zhengqiao Zhao**, Jason Rollins, Linge Bai, Gail Rosen, (2017). Incremental Author Name Disambiguation in Scientific Citation Dataset. *4th IEEE international conference on Data Science and Advanced Analytics* [IEEE]
- Marco A Janko, **Zhengqiao Zhao**, Moshe Kam, Yon Visell, (2018). A partial contact frictional force model for finger-surface interactions. *IEEE Haptics Symposium 2018* [IEEE]

Invited Talks

- **Zhengqiao Zhao**, Gail Rosen, (2020). Visualizing and Annotating Protein Sequences using Deep Neural Network. Invited oral presentation at *54th Annual Asilomar Conference on Signals, Systems, and Computers* [IEEE]
- **Zhengqiao Zhao**, Stephen Woloszynek, Felix Agbavor, Joshua Chang Mell, Gail Rosen, (2019). Learning, Visualizing and Exploring 16S rRNA Structure Using an Attention-Based Deep Neural Network. Invited Keynote speaker at *2019 IMMID Workshop on Applications of Artificial Intelligence to Microbial Genomics in Health and Disease* [Drexel University, College of Medicine]
- **Zhengqiao Zhao**, Jason Rollins, Linge Bai, Gail Rosen, (2017). Incremental Author Name Disambiguation in Scientific Citation Dataset. Invited oral presentation and poster at *4th IEEE international conference on Data Science and Advanced Analytics* [IEEE]
- **Zhengqiao Zhao**. (2017). Machine Learning in an Evolving Domain. Presented at *Scholars Share: Conversations on Graduate Student Research* [Drexel University, Graduate College]

Poster Presentations

- **Zhengqiao Zhao**, Bahrad A. Sokhansanj, Gail Rosen, (2020). Genetic Grouping of SARS-CoV-2 Coronavirus Sequences using Informative Subtype Markers (ISM) for Pandemic Spread Visualization. 4th Drexel Emerging Graduate Scholars (DEGS) Conference, Drexel University, PA, 2020.
- **Zhengqiao Zhao**, Stephen Woloszynek, Felix Agbavor, Joshua Chang Mell, Gail Rosen, (2019). Visualizing and Exploring 16S rRNA Using an Attention-based Deep Neural Network. 3rd Drexel Emerging Graduate Scholars (DEGS) Conference, Drexel University, PA, 2019.
- **Zhengqiao Zhao**, Gail Rosen. Incremental Machine Learning for Big Data. (1) DIG's 10th Annual Research Symposium, Drexel University, Philadelphia, PA, 2018, (2) 2nd Drexel Emerging Graduate Scholars (DEGS) Conference, Drexel University, PA, 2018.
- **Zhengqiao Zhao**, Jason Rollins, Linge Bai, Gail Rosen, (2017). Incremental Author Name Disambiguation in Scientific Citation Dataset. Invited oral presentation and poster at *4th IEEE international conference on Data Science and Advanced Analytics* [IEEE]

Software and Packages

- sequence_attention. **Zhengqiao Zhao**, Python package
- ncov_ism. **Zhengqiao Zhao**, command-line interface

Media Coverage

- Cover Art: Genetic grouping of SARS-CoV-2 coronavirus sequences using informative subtype markers for pandemic spread visualization. PLOS Computational Biology. 2020.
- Genetic Tracing "Barcode" Is Rapidly Revealing COVID-19's Journey and Evolution. The Drexel University College of Engineering. 2020.

Didactic

Teaching

Statistical Analysis of Metagenomes
Gail L. Rosen PhD
- Lecturer: Introduction to microbiome data science using QIIME 2

Drexel University
2020

Bioinformatics
Gail L. Rosen PhD
- Lecturer: DNA sequence coding regions identification using Fast Fourier Transform
- Final Coding Exam Developer: UPGMA algorithm and Backward algorithm for Hidden Markov model
- Final Project Instructor: Data Analysis for wastewater SARS-CoV-2 sequencing

Drexel University
2020

Multi-Disciplinary Digital Signal Processing
Gail L. Rosen PhD
- Lecturer: Introduction to Hidden Markov Models

Drexel University
2019

Computing and Control
ECE-403
- Teaching assistant

Drexel University
2019

Foundations of Electric Circuits
ECE-201
- Teaching assistant

Drexel University
2018

Programming for Engineers
ECE-203
- Teaching assistant

Drexel University
2018

Transform Methods and Filtering I
ECES-301
- Teaching assistant

Drexel University
2018

Computation Lab I
ENGR-121
- Teaching assistant

Drexel University
2018

Applied Digital Signal Processing
ECES-434
- Teaching assistant

Drexel University
2017

Multi-Disciplinary DSP
ECES-436
- Teaching assistant

Drexel University
2017

Mentoring

Melissa Gray
Undergraduate, Computer Engineering
- Projects: Incremental taxonomic classification

Drexel University
2020

Felix Agbavor
Undergraduate, Biomedical Engineering
- Projects: Host phenotype prediction

Drexel University
2017

Alexandru Cristian
Undergraduate, Computer Science
- Projects: Incremental naïve Bayes taxonomic classifier for metagenomic sequencing data

Drexel University
2017

Conferences Attended

2020	**Asilomar 2020, the 54th Annual Asilomar Conference on Signals, Systems, and Computers**	*Online*
2019	**SciPy 2019, the 18th annual Scientific Computing with Python Conference**	*Austin, TX*
2017	**DSAA 2017, the 4th IEEE international conference on Data Science and Advanced Analytics**	*Tokyo, Japan*

Skills

Programming   Python, TensorFlow, Matlab, Bash, SQL, regex, Git, cluster computing, LaTeX