Haowen Guan

Jersey City, NJ, 07306 | (253)-230-2080 | Website | haowen@nyu.edu

Education

New York University | NYC, NY **Sep. 2021–May 2023 (Expected)**

M.S. in Data Science, Center of Data Science | GPA (In progress): **3.7/4.0**

Relevant Coursework: Statistical Method (T-test, Regression, ANOVA, etc.), Natural Language Processing, Linear Algebra for Data Science, Machine Learning, Big Data, Database, Distributed system, etc.

University of Washington | Seattle, WA **Sep. 2018–Jun. 2021**

B.S. in Applied Physics, Department of Physics

B.S. in Applied and Computational Mathematical Sciences (ACMS) – Data Science and Statistics, Interdisciplinary Program

GPA: 3.61/4.0 (Cumulative), **3.67/4.0** (Physics), **3.71/4.0** (ACMS-DS&STAT)

Relevant Coursework: Computer Programing, Statistical software, Database System, Data Structure and Algorithm, Algorithms and Computational Complexity, Calculus, Linear Algebra, Differential Equations, Modern Physics, Numerical Methods, Discrete/Continuous Mathematical Modeling, Probabilistic Theories, etc.

Honors: 5 consecutive Dean Lists

Coursera **Jun. 2020-Aug.2020**

Deep Learning Specialization

Relevant Coursework: Neural Networks and Deep Learning, Convolutional Neural Networks, Sequence Models, etc.

Work Experience

DeepTracer, University of Washington **Jun. 2020-Aug. 2022**

Senior Backend Algorithm Engineer

- Backend developer working on Deep Tracer (https://deeptracer.uw.edu) – A application that do de novo protein complex structure prediction from Cryo-EM Maps, focused on training the Deep Learning model, Python code refinement (numpy), and model performance testing.
- Built a 3D Cryo-EM density map denoising Deep Learning model in Python, which adopted a novel UNet architecture based on convolutional neural network, and achieved an overall 95% accuracy in identifying noise pixels and low FP rates – a break-through in the domain, summarized the model & findings for publication.
- Researched ML approaches for cryogenic electron microscopy (Cryo-EM) Imaging, 3D reconstruction, and protein structure modeling and mastered academic writing skills by summarizing the insights in a review paper titled “Machine Learning and Artificial Intelligence Advances for De Novo Molecular Structure Modeling in Cryo-EM”.

Dais Group, University of Washington - Bothell **Sep. 2020-Nov. 2020**

- ***Research Assistant***
- Researched prevalent Machine Learning Algorithms for Cybersecurity, reviewed key techniques and use cases from publications within the past 10 years, evaluated pros & cons of popular Cybersecurity defense mechanisms.
- Investigated ML techniques like KNN, SVM and Bayesian Networks to Deep learning methods like CNN, RNN, GAN & autoencoders, their advantages and key applications in Cybersecurity.
- Collaborated with T-Mobile, on specific cybersecurity research targeted on 5G Telecom security, summarized the research findings in a detailed report and led a presentation on the research.

Pierce College Tutoring Center **Mar. 2017-Jun. 2018**

Teaching Assistant / Science Tutor

- Supported the professor in the facilitation of 7 freshman and sophomore level Mathematics & Physics courses.
- Provided drop-in, 1-on-1, and group academic guidance to 100+ students with varied backgrounds & skill levels, responded to their queries and recommended memory techniques to grasp complex concepts.

Project Experience

FIFA Players Value Analysis | *Python* **Dec. 2021**

- Cleansed and transformed ~20k records of Player values with 110 related variables, imputed missing data and treated outliers using Pandas and NumPy in Python.
- Performed correlation analysis and hypothesis tests for 7 highly correlated features to understand if they significantly impact player value and visualized them with box plots and trend lines using Matplotlib.
- Applied Kmeans clustering to understand if player positions influence their scores and built linear & lasso regression models using Sci-kit learn to predict player value based on 80+ features, achieved lowest MSE of ~0.08 with Lasso.

HIV Infections Modeling and Analysis | *R* **Aug. 2020**

- Reviewed existing research on HIV models, analyzed past 10-year infection data to identify most relevant models.
- Applied Murray's model described in the book Mathematical Biology, and another non-linear ODE model in R from "Mathematical Model for an Effective Management of HIV Infection" publication.
- Compared the model analysis and results and summarized the outcomes in a detailed project report.

Research Experience

Predicting the Best Sequence of Treatment for Cancer Patients with Big Data **Jan. 2020-May 2020**

- Manipulated, filtered and transformed 200k+ patient records from a database using custom SQL queries.
- Cleansed & preprocessed the clinical dataset using R, identified and treated outliers & missing values, performed exploratory data analysis and applied Hypothesis Tests to determine significant trends & relationships in the data,
- Built logistic regression, random forest and LGBM ML models using R to predict the next best treatment sequence for cancer patients, evaluated the models using confusion matrix and achieved an F1-score of 0.74 with LGBM model.

Skills & Interests

Programming Skills: Python, R, Java, SQL, MATLAB, LaTeX, Git.
Professional Skills: Microsoft Office (Word, Excel, etc.); Windows, Mac, and Linux Operating System
Interests: Physics, Artificial intelligence, Deep Learning, Computational and Biomedical Informatics, Blockchain, Stock Trading.

Publication

Si, D., Nakamura, A., Tang, R., Guan, H., Hou, J., Firozi, A., Cao, R., Hippe, K., & Zhao, M. (2021). Artificial intelligence advances for de novo molecular structure modeling in cryo-electron microscopy. WIREs Computational Molecular Science. Published. https://doi.org/10.1002/wcms.1542