

Xiao (Katrina) Liu

 [katrina-liu](#) |  [Katrina Liu](#) |  xliu@hms.harvard.edu |  +01.412.315.5028

EDUCATION

Aug. 2022 - present	MBI, Biomedical Informatics, Harvard Medical School	(GPA: 3.75/4.0)
	Expected Graduate Date: Mar. 2024	
Aug. 2018 - May 2022	BS, Computer Science, Carnegie Mellon University	(GPA: 3.86/4.0)
	Minor: Mathematical Sciences, Computational Biology	

RESEARCH EXPERIENCE

Harvard Medical School, Boston, MA Sep. 2022 - present

- Worked as a graduate research assistant in the Gehlenborg Lab and supervised by Dr. Nils Gehlenborg in the Department of Biomedical Informatics at Harvard Medical School.
- Worked on a project to realize automatic interpretation and generation of genomic visualizations
- Built a JavaScript program and Python program to generate training images and extract feature information.
- Built Python program to reconstruct visualization structure using the predicted information.
- Built the front-end web application to allow users to upload their own images, view the generated output, and modify the output image.

Carnegie Mellon University, Pittsburgh, PA Feb 2021 - Dec. 2022

- Worked as a research assistant in the Metabolomics and Metagenomics Lab and was supervised by Dr. Hosein Mohimani in the Computational Biology Department at Carnegie Mellon University.
- Mostly worked on the project Type II Polyketide Structure Prediction while helping other projects in the lab with illustrations, data processing, and testing.
- Collected data of each type II polyketide biosynthesis pathway from research papers and online databases.
- Preprocessed the genomic sequences of related domains and build machine learning models to predict features of the synthesis process of type II polyketides.
- Built program to construct the type II polyketide structures from the predicted features.
- Benchmarked with other state-of-the-art structure predictions methods.

Acuity Diagnostics, Pittsburgh, PA Jun. 2022 - Sep. 2022

- Worked as a machine learning research assistant and was supervised by Dr. Yongxin Zhao.
- Improved the understanding and ability to read cancer cell slides by hand and annotate regions of tumors.
- Worked on a project to improve the explainability of deep neural networks by predicting certain cancer pathological features.
- Built Python program to produce slide patches, normalized slide tiles, and conducted semantic segmentation and nucleus segmentation.
- Completed the pipeline to process whole slide images to detect dead cells and tumor-infiltrating lymphocytes.
- Worked on detection of muscle invading tumor from whole slide images.

AWARDS AND FELLOWSHIPS

Summer 2021	Summer Undergraduate Research Fellowship, Carnegie Mellon University
Summer 2020	Summer Experience for Mathematical Science, Carnegie Mellon University
Fall 2019-Spring 2022	Dean's List, Carnegie Mellon University
	No Dean's List awarded for Spring 2020 due to pandemic
Fall 2018-Spring 2019	Dean's List of High Honors, Carnegie Mellon University

RESEARCH PROJECTS

Analysis of TCGA-MESO Samples Based on Mutation Variants and Gene Expression

Oct. 2022, Harvard Medical School

The TCGA dataset is comprised of a variety of patient sample information collected by many project. The project TCGA-MESO in particular focused on mesothelioma, a type of cancer which mostly arises in the mesothelial surfaces of tissues in the pleura. In this project, sample information, gene expression level, and mutation variants information associated with the disease are collected by the TCGA-MESO project, providing the foundation of this study. We conducted principle component analysis and clustering to see if there are correlation between the gene expression data and mutational variants on the general level.xxxx

Predicting Biological Process Membership of Proteins from PPI Using LMM

May 2022, Carnegie Mellon University

We used the latent mixed-membership model to predict the biological process membership of proteins from the Protein-Protein Interaction data. We implement an Estimation- Maximization (EM) algorithm to approximate the model to calculate the degree of membership of each protein in each biological process group. Our result proves a noticeable improvement in prediction accuracy of latent mixed membership model to mixed membership stochastic blockmodel.

Comparison of Active Learning and Traditional Supervised Classification Methods on High-Volume Cancer Gene Expression Data

May 2022, Carnegie Mellon University

We use the Random Forest and Support Vector Machine to serve as baseline classifiers for the comparison between active learning strategies and traditional supervised strategies. We implemented different active learning strategies and we measured the accuracy of each classification strategy with 5-fold cross-validation, recording the standard deviation mean prediction accuracy for the unobserved datasets during the learning process.

Comparison of TAD Callers for Structural Analysis of Cancer and Normal Cells

Dec. 2021, Carnegie Mellon University

We compare the effects of different computational methods in identifying structurally similar regions of TAD structures between normal and cancer cell types. We explored quantitative and qualitative analysis of the results of each method with a set of metrics: accuracy, run time, resolution, and adaptability.

Prediction Model on Functions of Antibiotic Enzymes

May 2021, Carnegie Mellon University, Advised by Dr. Hosein Mohimani

Cytochrome P450 enzymes are very important in steroid biosynthesis and drug metabolism in human and natural product biosynthesis pathways. By building a classification/prediction model on the functions of P450 enzymes, we can better understand their behavior and classify them based on their functions for newly discovered P450 enzymes. This project builds classification models of the hydroxylation function of proteins in the P450 family using the k nearest model and depth-first search for their neighbors in connectivity graphs based on their pairwise alignment scores. We also examined the prediction accuracy of models and we try to add additional features that boost prediction accuracy.

Correlation between SARS-CoV-2 Spike Protein Sequences and Structures

May 2021, Carnegie Mellon University

The spike protein of SARS-CoV-2 plays a key role when the virus enters the host, and small mutations at spike protein may lead to very prevalent SARS-CoV-2 strands. Building a phylogenetic tree on the sequences of SARS-CoV-2 spike protein can serve as a classification model on SARS-CoV-2 spike protein variants. Based on the classification results, we used protein simulation to construct the structures of the spike protein variants and obtained the similarity and differences of protein structures. By analyzing the phylogenetic tree and comparing it to the results from protein simulation similarities, we found some moderate correlations between the spike protein sequence variance and its structure. However, there may be some errors in our data and measurements, so we may not reach a conclusive statement.

SKILLS

Programming Languages	Python, Java, JavaScript/Typescript, C/C++, Linux, R, OCaml, HTML/CSS
Development Skills	Django, Agile Development, Object Orient Programming, Machine Learning,
Technical Skills	L ^A T _E X, MatLab, Microsoft Office
Languages	Chinese/Mandarin (First Language), English (Professional Level)

INDUSTRY EXPERIENCE

Software Engineer Intern, Sina Corp, Beijing, China

Jun. 2019 - Aug. 2019

- Worked as a software engineer intern in the Auto Database department.
- Developed a project for enormous web pages status checking
- Familiarized with urllib, requests, HTMLParser, connection pool, os in python.
- Familiarized with debugging tools like logging in python and SMTP in email.
- Developed a project for an automatic software update.

SOFTWARE PROJECTS

Indeed for Labs

Dec. 2021, Carnegie Mellon University

We are trying to build a platform for students and academic researchers to apply to work at labs at any university. Applicants can create a profile and search for labs to apply to based on university, subject area, professor, etc. Lab professors can post their labs on the site. The primary goal is to polish up the product for use at Carnegie Mellon University

Consumer-to-Consumer eCommerce Platform

May. 2021, Carnegie Mellon University

We built a website where users can buy/sell products directly from/to users using Django and Python. All available items will be displayed on a page. Users can change the category filter to decide the kind of products they want to see. Users can search for the items based on the item name. Only registered users can post products and make purchases. Users can act as sellers and create posts for things they want to sell. These posts include item names, item prices, item descriptions, item pictures, and item reviews from other buyers. The sellers can add the product to specific categories. Users can also act as buyers on the website to purchase items, add items to wishlists and write reviews. The transactions are conducted Paypal.

Covid Analysis Framework

Apr. 2021, Carnegie Mellon University

We built a covid analysis framework with support for different data plugins and display plugins on Java. [\[Video demo link\]](#). The final product supports data plugins including different web APIs and local files. We used Java Swing Utilities for the GUI display of the framework.

Plant Pycoon

May. 2019, Carnegie Mellon University

An individual project of a multiplayer plant-growing game based on Python using Pygame and Socket.
[\[Video demo link\]](#)

TEACHING EXPERIENCE

Teaching Assistant, Carnegie Mellon University, Pittsburgh, PA Aug. 2021 - May. 2022

- Worked as a teaching assistant for course 17-214 Principles of Software Construction Objects, Design, and Concurrency, a course focusing on Java/Typescript development and software design principles. The key library used are React, Handlebars, and Nanohttpd.
- Led weekly 1-hour recitation for 40 students. Gave feedback to student presentations.
- Prepared recitation materials and code repo to distribute.
- Participated in weekly meetings and grading meetings.
- Graded homework and exams.

Grader, Carnegie Mellon University, Pittsburgh, PA Aug. 2020 - Dec. 2020

- Worked as a grader for course 21-127 Concepts of Mathematics, a course focusing on mathematical proofs and concepts.
- Graded homework and exams.

Peer Tutor, Carnegie Mellon University, Pittsburgh, PA May. 2020 - Aug. 2020

- Worked as a peer tutor for course 15-150 Functional Programming, a course emphasizing classifying expressions by types that specify their applicative behavior. The course also places a focus on well-typed functional programming, recursion, program run time, and parallel programming.
- Tutored two sections of 6 students in total for understanding course materials and homework.

Teaching Assistant, Carnegie Mellon University, Pittsburgh, PA Aug. 2019 - Dec. 2019

- Worked as a teaching assistant for course 21-120 Differential and Integral Calculus.
- Led 1-hour recitation for 20 students twice a week.
- Graded homework and exam.