

Teaching Statement

I am passionate about understanding how students learn, how to design curricula that ignite students' passions, and how to build a student's capacity and enthusiasm for future innovations. As a computer scientist and machine learning researcher whose research interest lies in machine learning for drug discovery and development, I have diverse interests in multiple disciplines such as chemistry, biology, pharmacy, health science, etc. I have established my teaching interests through designing and teaching undergraduate and graduate courses, as well as through mentoring and cross-disciplinary collaborations.

Teaching Interests

As artificial intelligence (AI)/machine learning (ML)/deep learning (DL) becomes more ubiquitous and is widely used in many applications, e.g., computer vision (CV), natural language processing (NLP), drug discovery and development, universities are increasingly responsible for training broad-minded machine learning engineers, as well as people from other disciplines, e.g., bioinformatics, electrical and computer engineering.

I am eager and qualified to take on the challenges of designing and teaching

- **undergraduate** courses: Introduction to AI/ML/DL/CV/NLP/Algorithms/Computer Science/Python Programming/C++ Programming.
- **graduate** courses: AI/ML/DL for Drug Discovery and Development/Healthcare/Scientific Discovery, AI/ML/DL and its Application, Pattern Recognition, Computational Data Analysis, Advanced Algorithm, Design and Analysis of Algorithms.

Further, I am more than willing to design **seminars** and research projects that allow students to explore AI for drug discovery and development, healthcare and scientific discovery, etc. The target students of the graduate courses and seminars are graduate (e.g., MS and Ph.D. students) or advanced undergraduate students majoring in computer science, engineering, medicine, chemistry, biology, medical informatics, biostatistics, etc.

Teaching Experience

I have been the TA for Prof. Jimeng Sun's course "Big Data Analytics for Healthcare" three times (2019 Fall, 2020 Spring, 2020 Fall) at Georgia Institute of Technology. The course is very popular at Georgia Institute of Technology, and it attracted more than 100 graduate students every semester. I design the slides for the lectures on machine learning applications in medicine, which cover some popular topics, including how to leverage machine learning models for designing novel drug molecules with desirable pharmaceutical properties, pharmaceutical property prediction for drug molecules, drug adverse effect detection, clinical trial outcome prediction, etc. Also, following the "learning by doing" principle, I design some supporting projects (e.g., jupyter notebooks using publicly available datasets and codes) to help students have a deep intuition of what they are creating. Further, in these courses, I was responsible for the final projects and host weekly office hours to help students with final projects.

Mentoring

I am particularly excited to begin forming lifelong connections with new students, to nurture their interests and fan their passions, and to shepherd their growth over the course of their degree and beyond. I have worked with several students (including both junior Ph.D. students and master students) in Prof. Jimeng Sun's lab. For example, currently I am working with Chufan (Andy) Gao (a first-year Ph.D.

student) on clinical trial outcome prediction specifically for biologics-related clinical trials, where biologics belong to a specific kind of drug. We focus on extracting the biologics' features from DrugBank, including amino acid sequence, numeric features about biologics' pharmacodynamic properties, and text descriptions from the relevant literature. Andy is highly self-motivated and very organized. For example, we schedule a weekly meeting and before each meeting, he would write down all the experimental results, the next steps we should explore, and potential questions in a document, which makes the meeting very efficient. We have got significant progress in the project. I am working with Weijiang Li (a first-year master student) on drug adverse effect detection. Weijiang is also self-motivated and independent in research. We focus on utilizing machine learning models to learn the drugs' adverse effect, which is a very important task in drug development. During the first meeting, I described several interesting projects we might work on together and shared several representative papers for each project. After two weeks, when we met again, she has already prepared a short survey draft and a google sheet. For each direction, she investigated the existing literature and put them into a table, she characterized each paper from various aspects, including dataset, problem setup, code repository, method, experiment sketch, etc. It gave me great pleasure to supervise early-career students and guide them in research projects.

Textbook

My research interest lies in ML for drug discovery and development. Despite the rapid growth of this area, educational resources are still lacking and are in high demand. To meet the needs in educational resources, I, Dr. Cao Xiao, and Prof. Jimeng Sun are writing the textbook entitled "Machine Learning for Drug Discovery and Development"¹ [1]. The book is intended for a graduate course on "AI/ML for drug discovery and development". It can also serve the graduate course "AI/ML and its application". The target audiences are graduate (e.g., MS and Ph.D. students) or advanced undergraduate students majoring in computer science, engineering, medicine, chemistry, biology, medical informatics, biostatistics, etc. Our book covers a comprehensive overview of most of the essential topics in AI for drug discovery & development and starts from basic concepts in machine learning and pharmaceutical science. Thus it is more friendly to readers who do not have sufficient background knowledge. Also, for most of the ML for discovery and development tasks, we provide hands-on examples (e.g., tutorials in terms of jupyter notebook). The book is expected to be published by Springer Press in May 2023.

Leadership

I co-founded the series of AI for Science (AI4Science) workshops² at leading AI venues (NeurIPS 2021, ICML 2022, and NeurIPS 2022 upcoming) to bring researchers working on AI4Science topics together and consolidate the fast-growing area of AI4Science into a recognized field. For example, in the 3rd AI4Science workshop co-located with NeurIPS upcoming in December 2022, we received over 140 high-quality submissions.

References

- [1] Fu, Tianfan, C. Xiao, and J. Sun. Machine learning for drug discovery and development. *Springer (to appear)*, 2023.

¹<https://ml4drug-book.github.io/>

²<https://ai4sciencecommunity.github.io/>