
Course title and number ECEN 766 Algorithms in Structural Bioinformatics
Term Spring 2020
Meeting times & location TR 12:45-2:00pm HRBB 204

Course Description and Prerequisites

This course introduces fundamental concepts, modeling techniques, and computational algorithms in structural bioinformatics especially for students interested in algorithmic development and application for computational challenges arising from the data-rich field. With a focus on algorithms involving molecular modeling, systems simulation, optimization, machine learning, and deep learning, the course provides essential knowledge for students without prior background in the application domain, addresses their learning barriers, and helps them make unique contributions to the field.

Applications of these algorithms are centered on how to analyze, predict, and engineer biomolecules and biomolecular systems: protein sequence, structure and function; genotype-phenotype association; computer-aided/AI-empowered drug discovery; and biomolecular systems modeling and engineering. Algorithmic solutions to these applications can provide case studies for algorithmic thinking and innovation. Students interested in practical problem-solving skills for specific applications as well as undergraduate students interested in exploring new fields are also welcome to attend (seven undergraduate students attended before).

The course will involve literature-based presentations, case studies, short projects in homework, and a main final project, in addition to regular lectures.

Prerequisites: Basic knowledge in algorithms and programming. No prior knowledge in biomolecules or biomolecular systems is required.

Learning Outcomes

By taking the course, students are expected

1. to gain knowledge about fundamental concepts, pressing challenges, and rich opportunities in developing and applying algorithms for structural bioinformatics and healthcare;
2. to apply and to strengthen engineering principles, data-science skills, and algorithmic thinking to the emerging applications of structural bioinformatics, AI/healthcare, and other fields; and
3. to develop practical skills in computational approaches to analyze, predict, and engineer biomolecules and biomolecular systems.

Instructor Information

Name Yang Shen
Telephone number 979-862-1694
Email address yshen@tamu.edu
Office hours T 3:00pm-4:30pm (except for M 3:00-4:30pm on the week of Feb. 24 and Mar. 3)
or by appointment
Office location Wisenbaker Engineering Building 215I

Textbook and Resource Material

No textbook is required. Notes and additional materials such as tutorials, papers and book chapters will be provided to the class.

Other References:

(Concise Bioinformatics)

- **[Do]** B Donald. [Algorithms in Structural Molecular Biology](#). MIT Press. 2011.

(Comprehensive Bioinformatics)

- **[GB]** J Gu and PE Bourne. [Structural Bioinformatics](#). Wiley-Blackwell. 2009.

(Python Programming)

- **[VdP]** Jake VanderPlas. [Python Data Science Handbook](#). O'Reilly. 2016.

[VdP Jupyter Notebooks](#) (with colab links).

(Machine Learning and Deep Learning Practice)

- **[Ge]** Aurélien Géron. [Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems](#). O'Reilly. 2017.

[Ge Jupyter Notebooks](#) (DL based on the TensorFlow interface).

(More on Deep Learning)

- **[Go]** Ian Goodfellow, Yoshua Bengio, and Aaron Courville. [Deep Learning](#). MIT Press. 2016.
- **[Zh]** Aston Zhang, Zachary Lipton, Mu Li, and Alexander Smola. [Dive into Deep Learning](#). 2020. (Interactive book with math, figures and code. NumPy interface).

(Other books for DL with codes based on Keras or PyTorch)

- **[Ch]** Francois Chollet. Deep Learning with Python. Manning Publications. 2017.

[Keras-based Jupyter Notebooks](#).

- **[SAV]** Eli Stevens, Luca Antiga and Thomas Viehmann. Deep Learning with PyTorch. Manning Publications. 2020.

[PyTorch-based Jupyter Notebooks](#).

Grading Policies

Weights towards final grades for undergraduate students

- 50% Homework
- 10% 1 journal-club presentation
- 40% Final Project: Short Proposal (10%), Progress Update (10%), & Final Report and Presentation (20%)

Weights towards final grades for graduate students

- 40% Homework
- 10% Mini Project
- 10% 1 journal-club presentation
- 40% Final Project: Short Proposal (10%), Progress Update (10%), & Final Report and Presentation (20%)

Please note that that the Proposal and the Update essentially are early versions of Introduction and Methods/Results sections for the Final Report.

Tentative Grading Scale:

A	[90%, 100%]
B	[80%, 90%)
C	[70%, 80%)
D	[60%, 70%)
F	[0%, 60%)

Final grades will be determined numerically based solely on individual standing to reflect how well

students do in homework, exams, and projects. This approach is adopted to ensure at least a fair mechanism to assess how well students learn course materials and accomplish course goals. Meanwhile, diversity in student background (engineering or science) and academic standing (undergraduate or graduate) will be respected and reflected in final project topics.

Course Topics

Here is the tentative course outline with approximately assigned lecture time:

Week	Topic	Related Reading
1-2	Introduction to biomolecules and structural bioinformatics	GB1-3
2-3	Data visualization and analysis: protein sequence, structure, and function	GB9
3-5	Structure prediction from sequence – Classics	
	A. template-based homology modeling and threading (<i>Optimization fundamentals; Convex optimization; Algorithm fundamentals; Complexity; Sequence alignment as dynamic programming & database search; Threading as linear programming and machine learning</i>)	GB30,31
	B. ab initio methods (<i>Structure prediction as energy minimization; Energy function and conformational variables; Nonconvex optimization; Gradient-based and gradient-free algorithms; Convergence and convergence rate; Great ideas for objective function, search space & constraints</i>)	GB8,32
	C. (Protein docking) (<i>Dimensionality reduction and optimization algorithms revisited</i>)	Do20-23, GB24-27
5-6	Protein function prediction from sequence, structure, and big data (<i>Machine learning, classification, kernels, data integration, regression, diagnostics</i>)	Go5,GB21,22
7-9	Revisiting sequence-structure-function – New Waves	
	A. Predicting contact/distance maps in structures as images (<i>Special images in protein structure data; Images and computer vision; Deep learning; Convolutional neural networks; Data augmentation; Optimization for deep learning; Case study: DeepMind's AlphaFold</i>)	Go6-9/Zh6,7,13 & Literature
	B. Learning protein sequences as texts (<i>Special texts in protein sequence data; Texts and natural language processing; Language models; Recurrent neural networks; Attention mechanisms; Transformers; Representation learning; Discriminative and generative models; Case Study: TAPE</i>)	Go10,15/Zh8,10,14 & Literature
	C. Predicting protein properties (structures, interactions and mutational effects) based on learned sequence representations (<i>Unsupervised/supervised/semi-supervised/self-supervised learning; Pre-training and fine-tuning; end-to-end learning; Transfer learning; Case Study: DeepAffinity & MuPIPR</i>)	
	D. (Learning protein functions or protein-protein interactions as graphs)	
10-11	Drug Discovery: from computer-aided to AI-empowered molecular design	Go20/Zh16 & Literature
	A. Drug discovery and development process; Healthcare and more data	
	B. Computer-aided drug design with energy-based combinatorial optimization	
	C. AI-empowered drug/biomolecule design with data-driven deep learning (<i>Special texts and graphs for small-molecule drugs; inverse design from desired property to identity text/graph; GAN and VAE as generative models; Case Study: Various generative models for text/graph-based drug/protein/RNA design</i>)	
11-12	Blackbox no more?	Literature
	A. Incorporating physical constraints / domain knowledge into deep learning	
	B. Demanding interpretability / explainability from deep learning.	

12 (Biomolecular system modeling)
(*Topology, steady states & dynamics; New AI waves*)

Literature

13-14 Final project presentations

Contents may subject to adjustment. Additional materials will be provided through eCampus.

Attendance and Make-up Policies

Regular and punctual attendance to the lectures facilitates the effective implementation of a systematic study plan. Please consult student rule 7 for additional information: <http://student-rules.tamu.edu/rule07>.

Americans with Disabilities Act (ADA)

Texas A&M University is committed to providing equitable access to learning opportunities for all students. If you experience barriers to your education due to a disability or think you may have a disability, please contact Disability Resources in the Student Services Building or at (979) 845-1637 or visit <http://disability.tamu.edu>. Disabilities may include, but are not limited to attentional, learning, mental health, sensory, physical, or chronic health conditions. All students are encouraged to discuss their disability related needs with Disability Resources and their instructors as soon as possible.

Academic Integrity

For additional information please visit: <http://aggiehonor.tamu.edu>

"An Aggie does not lie, cheat, or steal, or tolerate those who do."