Course Information

Course Number:	ECEN 766
Course Title:	Algorithms in Structural Bioinformatics
Section:	600
Time:	TR 11:10am-12:25pm
Location:	ETB 1003
Credit Hours:	3

Instructor Details

Instructor:	Yang Shen
Office:	WEB 215I
Phone:	(979)862-1694
E-Mail:	yshen@tamu.edu
Office Hours:	TBA

Course Description

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.

Course Prerequisites

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

Special Course Designation

NA

Course Learning Outcomes

By taking the course, students are expected

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

Textbook and/or Resource Materials

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. 2023. (Interactive book with math, figures and codes based on PyTorch, TensorFlow, MXNet, and JAX).
- **[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 Policy

Weights towards final grades:

- 45% Homework
- 10% 1 journal-club presentation
- 45% Final Project: Short Proposal (10%), Progress Update (10%), & Final Report and Presentation (25%)

Please note that the Proposal and the Update are essentially 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 when they choose final project topics.

Late Work Policy

Late submission, i.e. submitting a deliverable after the established deadline, would incur a penalty of 10% each calendar day after the due time, up to a maximum of 7 calendar days. For instance, submissions late within 1 day of the due time receive 90% of their scores, those late 1-2 days receive 80%, and those late for > 7 days are not accepted.

Work submitted by a student as makeup work for an excused absence is not considered late work and is exempted from the late work policy (Student Rule 7).

Course Schedule

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

Week	Topic	Related Reading
1-2	Introduction to biomolecules and structural bioinformatics	**GB**1-3
2-3	Data visualization and analysis: protein sequence, structure, and function	**GB**9
3-5	Structure prediction from sequence – Classics	
	A. template-based homology modeling and threading	**GB**30,31

(Optimization fundamentals; Convex optimization; Algorithm fundamentals; Complexity; Sequence alignment as dynamic programming & database search; Threading as linear programming and

machine learning)

B. ab initio methods **GB**8,32
(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)

C. [Protein docking] (optional) **Do**20-23, **GB**24-27
(Dimensionality reduction and optimization algorithms revisited)

5-6 Protein function prediction from sequence, structure, and big data **Go**5,**GB**21,22
(*Machine learning, classification, kernels, data integration, regression, diagnostics)*

7-9 Revisiting sequence-structure-function – New Waves

A. Learning protein sequences as texts **Go**10,15/**Zh**8,10,14 & Literature
(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: "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences" (Rives et al. 2021, PNAS) & "Evolutionary-scale prediction of atomic-level protein structure with a language model" (Lin et al. 2023, Science)

B. Predicting protein properties (structures, interactions and mutational effects) based on learned sequence representations
(Unsupervised/supervised/semi-supervised/self-supervised learning; Pre-training and fine-tuning; End-to-end learning; Transfer learning; Case Study: "Learning the protein language: Evolution, structure, and function" (Bepler and Berger 2021, Cell Systems) & "Embeddings from protein language models predict conservation and variant effects" (Marquet et al. 2022, Human Genetics))

C. Predicting contact/distance maps in structures as images **Go**6-9/**Zh**6,7,13 & Literature
(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)

D. Predicting protein 3D structures from sequence: The revolution has arrived!
(Multimodal approach involving texts, graphs, and point clouds; Graph neural networks; Roto-translationally equivariant neural networks; Case study: DeepMind's AlphaFold 2 and Meta AI's ESM-fold)

E. [Learning protein functions or protein-protein interactions as graphs] (optional)

10-11 Drug Discovery: from computer-aided to AI-empowered molecular design
Go20/**Zh**16 & 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
(Special texts and graphs for small-molecule drugs; inverse design from desired property to identity text/graph; generative models such as GAN, VAE, flow-based and diffusion models; Case Study: Various generative models for text/graph-based drug/protein/RNA design including ProGen2 (2023, Cell Systems) and Chroma (2023, Nature))

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) Literature
(Topology, steady states & dynamics; New AI waves)

13-14 Final project presentations

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

University Policies

This section outlines the university level policies that must be included in each course syllabus. The TAMU Faculty Senate established the wording of these policies.

***NOTE**: Faculty members should not change the written statements. A faculty member may add separate paragraphs if additional information is needed.*

Attendance Policy

The university views class attendance and participation as an individual student responsibility. Students are expected to attend class and to complete all assignments.

Please refer to Student Rule 7 in its entirety for information about excused absences, including definitions, and related documentation and timelines.

Makeup Work Policy

Students will be excused from attending class on the day of a graded activity or when attendance contributes to a student's grade, for the reasons stated in Student Rule 7, or other reason deemed appropriate by the instructor.

Please refer to Student Rule 7 in its entirety for information about makeup work, including definitions, and related documentation and timelines.

Absences related to Title IX of the Education Amendments of 1972 may necessitate a period of more than 30 days for make-up work, and the timeframe for make-up work should be agreed upon by the student and instructor" (Student Rule 7, Section 7.4.1).

"The instructor is under no obligation to provide an opportunity for the student to make up work missed because of an unexcused absence" (Student Rule 7, Section 7.4.2).

Students who request an excused absence are expected to uphold the Aggie Honor Code and Student Conduct Code. (See Student Rule 24.)

Academic Integrity Statement and Policy

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

"Texas A&M University students are responsible for authenticating all work submitted to an instructor. If asked, students must be able to produce proof that the item submitted is indeed the work of that student. Students must keep appropriate records at all times. The inability to authenticate one's work, should the instructor request it, may be sufficient grounds to initiate an academic misconduct case" (Section 20.1.2.3, Student Rule 20).

You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at aggiehonor.tamu.edu.

Americans with Disabilities Act (ADA) Policy

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 the Disability Resources office on your campus (resources listed below) 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.

Disability Resources is located in the Student Services Building or at (979) 845-1637 or visit disability.tamu.edu.

Title IX and Statement on Limits to Confidentiality

Texas A&M University is committed to fostering a learning environment that is safe and productive for all. University policies and federal and state laws prohibit gender-based discrimination and sexual harassment, including sexual assault, sexual exploitation, domestic violence, dating violence, and stalking.

With the exception of some medical and mental health providers, all university employees (including full and part-time faculty, staff, paid graduate assistants, student workers, etc.) are Mandatory Reporters and

must report to the Title IX Office if the employee experiences, observes, or becomes aware of an incident that meets the following conditions (see University Rule 08.01.01.M1):

- The incident is reasonably believed to be discrimination or harassment.
- The incident is alleged to have been committed by or against a person who, at the time of the incident, was (1) a student enrolled at the University or (2) an employee of the University.

Mandatory Reporters must file a report regardless of how the information comes to their attention – including but not limited to face-to-face conversations, a written class assignment or paper, class discussion, email, text, or social media post. Although Mandatory Reporters must file a report, in most instances, a person who is subjected to the alleged conduct will be able to control how the report is handled, including whether or not to pursue a formal investigation. The University's goal is to make sure you are aware of the range of options available to you and to ensure access to the resources you need.

Students wishing to discuss concerns in a confidential setting are encouraged to make an appointment with Counseling and Psychological Services (CAPS).

Students can learn more about filing a report, accessing supportive resources, and navigating the Title IX investigation and resolution process on the University's Title IX webpage.

Statement on Mental Health and Wellness

Texas A&M University recognizes that mental health and wellness are critical factors that influence a student's academic success and overall wellbeing. Students are encouraged to engage in healthy self-care by utilizing available resources and services on your campus

Students who need someone to talk to can contact Counseling & Psychological Services (CAPS) or call the TAMU Helpline (979-845-2700) from 4:00 p.m. to 8:00 a.m. weekdays and 24 hours on weekends. 24-hour emergency help is also available through the National Suicide Prevention Hotline (800-273-8255) or at suicidepreventionlifeline.org.

Campus-Specific Policies

Texas A&M at Galveston

Classroom Access and Inclusion Statement

Texas A&M University is committed to engaged student participation in all of its programs and courses and provides an accessible academic environment for all students. This means that our classrooms, our virtual spaces, our practices and our interactions are as inclusive as possible and we work to provide a welcoming instructional climate and equal learning opportunities for everyone. If you have an instructional need, please notify me as soon as possible.

The Aggie Core values of respect, excellence, leadership, loyalty, integrity and selfless service in addition to civility, and the ability to listen and to observe others are the foundation of a welcoming instructional

climate. Active, thoughtful and respectful participation in all aspects of the course supports a more inclusive classroom environment as well as our mutual responsibilities to the campus community.

Statement on the Family Educational Rights and Privacy Act (FERPA)

FERPA is a federal law designed to protect the privacy of educational records by limiting access to these records, to establish the right of students to inspect and review their educational records and to provide guidelines for the correction of inaccurate and misleading data through informal and formal hearings. Currently enrolled students wishing to withhold any or all directory information items may do so by going to howdy.tamu.edu and clicking on the "Directory Hold Information" link in the Student Records channel on the MyRecord tab. The complete FERPA Notice to Students and the student records policy is available on the Office of the Registrar webpage.

Items that can never be identified as public information are a student's social security number, citizenship, gender, grades, GPR or class schedule. All efforts will be made in this class to protect your privacy and to ensure confidential treatment of information associated with or generated by your participation in the class.

Directory items include name, UIN, local address, permanent address, email address, local telephone number, permanent telephone number, dates of attendance, program of study (college, major, campus), classification, previous institutions attended, degrees honors and awards received, participation in officially recognized activities and sports, medical residence location and medical residence specialization.

COVID Statement

The following statement on the value of vaccinations and masking was developed by the Executive Committee of the Faculty Senate and approved by the Administration on August 18, 2021:

"To help protect Aggieland and stop the spread of COVID-19, Texas A&M University urges students to be vaccinated and to wear masks in classrooms and all other academic facilities on campus, including labs. Doing so exemplifies the Aggie Core Values of respect, leadership, integrity, and selfless service by putting community concerns above individual preferences. COVID-19 vaccines and masking — regardless of vaccination status — have been shown to be safe and effective at reducing spread to others, infection, hospitalization, and death."