CS 5970 : Bioinformatics

Marmar Moussa, PhD

Fall 2024

Course Description:

This course offers a comprehensive introduction to the fundamental mathematical models and computational techniques in bioinformatics. Topics include sequence analysis, functional genomics, single cell omics, and machine learning in computational biology.

1 Introduction

Recent technological advances are providing unprecedented opportunities to analyse the complexities of biological systems at this minute level. High-throughput analyses of the genomes, transcriptomes and proteomes are providing novel and important insights into diverse processes such as development, gene-expression dynamics, tissue heterogeneity and disease pathogenesis. Computer Science and especially bioinformatics are at the heart of this new frontier. This course covers the algorithmic and machine learning foundations of bioinformatics and computational biology combining theory with practice. We will cover foundational topics, current research frontiers, and work directly with current large-scale biological datasets.

Examples of the topics that will be covered:

- Definition and scope of bioinformatics. (Historical perspective and evolution of the field, importance and applications in biological research, examples of Biological Databases).
- Sequence and Genome Analysis.
  - DNA, RNA, and protein sequence analysis.
  - Sequence mapping, alignment, and SNP calling.
  - Markov chains and hidden Markov models.
- Phylogenetics (Tree construction algorithms)
- Structural Bioinformatics (Protein structure prediction and modeling)
- Next-Generation Sequencing (NGS)

- Single Cell Technologies: (What they are, when to use them, input libraries and interpreting outputs)
  - Single Cell Gene Expression (RNA-Seq)
  - Single Cell DNA-Seq
  - Other Single Cell Modalities (Single Cell Epigenomics (ATAC-Seq and Multiomics cytometry), Immune Receptor Mapping (single cell TCR/BCR-Seq), High Definition spatially resolved transcriptomics, Single Cell Proteomics, Metagenomics, CRISPR Screening.
- Functional Genomics:
  - Primary vs. Secondary pipelines
  - Computational methods for single cell data
  - Computational challenges in single-cell transcriptomics
- Visualization techniques for biological data.
- Machine Learning and Algorithms for omics data
  - Type vs. State identification
  - Differential and Enrichment Analysis
  - Trajectory inference
  - Interaction Maps
  - Multimodal integration
- Practical applications of bioinformatics in research projects: (Areas: Precision/personalized medicine, Immunology, Oncology, Neuroscience, Developmental Biology, Drug Resistance, Genetic Disease)
- Ethical considerations in bioinformatics research. (Data sharing, privacy, and security.)
- Hands-on Scientific Programming Workshops
  - Primer to R programming language
  - Introduction to python notebooks
- Recent Advances and Emerging Trends
- Research Projects and Case Studies

2 Course Information and Material

Instructor's Office Hours: W 15:00 - 17:00, in DEH 253 *and/or* Zoom link here by appointment.
Course Texts and Material (Textbooks are optional): An Introduction to Bioinformatics Algorithms. A **Classic** Book by Neil Jones and Pavel Pevzner.[1]
Bioinformatics Algorithms: An Active Learning Approach. 3rd edition(2018), book by Phillip Compeau.[2]
Assigned readings ahead of class (Typically up to 2 papers will be shared on Canvas for each unit/topic: One classic paper or review and one recent (state-of-the-art) methods.
Time: August 19 - December 6, 2024. TR 4:30 - 5:45 pm
Location: Carson Engr Ctr 0121
Scheduled final exam: December 9, 2024. M 4:30 - 6:30 pm. Carson Engr Ctr 0121.
Learning Activities, Assignments, and Assessment: Choose 3 out of 4 Assignments/Small Projects, due 2 weeks from release date, each of the top three worth 10% of final grade (30% total). These could also be multiple small programming challenges assigned through Rosalind. Extra Credit opportunities up to 5% of total course grade (e.g. for using Latex).
Projects, Midterms and Final: Two journal-club-style presentations (i.e. presenting a reviewed paper or paper collection on a topic) each 15% (30% total); Final Project is 30%.
Class Participation: 10% total ($\approx$ 5 small in-class surveys or quizzes each worth 2%).
Grading Scale: Letter grades will be assigned based on the scale below. The scale already includes the rounding up allowed. $90\% \leq A \leq 100\%$; $80\% \leq B < 90\%$; $70\% \leq C < 80\%$; $60\% \leq D < 70\%$; $C < 60\%$.

Tentative Schedule: Table 1 and potential updates on Canvas.
Prerequisites: C S 2413 , and C S 2813. or permission of instructor. If you have not mastered any of the course prerequisites, consider reaching out to me or a teaching assistant and we will provide resources.
Notes: Canvas - Please check the course on Canvas for announcements, updates to schedule, zoom links to office hours, etc.

Assignments and Exams - I am interested in your learning and your approach to problem solving, therefore, partial credit will be given when you have solved parts of the problem correctly. If you feel you are falling behind or stressed about the material, please come see me or a teaching assistant, we will help.

Attendance - I do not take attendance in class, however there can be graded surveys and quizzes counted towards class participation. I recommend that all students attend every lecture and review, regardless of previous performance. All students will learn something new in class, receive tips and best practices to solving problems, learn about potential pitfalls and frequently made mistakes and attending class is one of the best ways to learn the concepts and improve your skills with minimal stress.

Additional Support - Consider additional services and support offered by OU, e.g. free walk-in tutoring sessions through Action Tutoring or support from the Accessibility and Disability Resource Center.

Week	Topics	Assignments Released	Notes
1	Syllabus + Introduction		
2	Next-Generation Sequencing (NGS) and High-Throughput Data		Potentially Asynchronous
3	Sequence and Genome Analysis		hands-on[1]
4	Markov chains and hidden Markov models	Assignment 1	
5	Phylogenetic tree algorithms		journal-club-style[2]
6	Single Cell Technologies		journal-club-style
7	Computational methods for single cell data		journal-club-style
8	Muli-omics Computational Challenges		journal-club-style - Hybrid
9	Visualization techniques for biological data	Assignment 2	hands-on
10	Machine Learning and Algorithms for omics data		journal-club-style, hands-on
11	Biostatistical methods for Differential and Enrichment Analysis		hands-on
12	Applications of bioinformatics algorithms in medical research		journal-club-style
13	Advanced Algorithms: e.g. Trajectory Inference and Shared Space Embeddings (Multi-modal integration)	Assignment 3	
14	Final Projects Presentations / Ethical considerations in bioinformatics research		
15	Final Projects Presentations		Thanksgiving
16	Pre-Finals Week		
17	Finals Week - Final Projects Presentations Continued		

Table 1: ***Tentative*** *schedule for Bioinformatics Fall 2024. See Canvas for most recent updates.*

[1] Can include hands-on components.

[2] Suitable for journal-club-style presentation topics.

3 Course Policies

3.1 Submissions

Typeset, electronic versions (preferably in PDF format) of assignments and tests should be submitted to Canvas by the due dates. Late submissions of homework assignments might be accepted with a per day 10% penalty. No late submissions/uploads of tests will be accepted.

3.2 Typesetting

You are required to typeset your assignments using a word processor or other tools with an equation editor or -preferably but **optional**- using LaTeX [3] which is a software system for document preparation that is especially well-suited for documents with complex mathematical equations, scientific notation, and technical content. LaTeX excels in typesetting mathematical equations and symbols. It provides extensive support for various mathematical notations, making it the preferred choice for writing papers in fields like mathematics, physics, computer science, and engineering. LaTeX also simplifies the process of cross-referencing sections, equations, figures, and tables within your document. It also seamlessly integrates with BibTeX or BibLaTeX for managing and formatting citations and bibliographies. LaTeX and related material and resources will be discussed in some detail during the first week of class. Examples and templates as well as cheat sheets will also be provided. We highly encourage using Overleaf, MiKTeX or other free LaTeX tools.

3.3 References and Literature Citations

Any statement or answer in any submitted work (assignments, tests, exams, projects, etc.) that relies on external sources of information (i.e. not the student's own new ideas or findings or solutions or general knowledge) should use a citation.

Citations are allowed in any format including but not limited to : APA, MLA, Chicago, and Harvard. You are free to choose which style to use as long as you use it consistently within an assignment or submitted work. You are encouraged to explore styles and use a citation generator.

You are generally allowed to use any external sources of information when solving your assignments and tests or examinations that are marked as 'open-book' as long as you properly and accurately cite your references. Misrepresenting an answer as original work will constitute a violation of the academic integrity policy, for more information see section 3.5 of this syllabus.

3.4 Citing (generative) artificial intelligence (AI):

How to cite ChatGPT and similar generative AI[4] websites and tools: You can generally follow the APA (American Psychological Association) citation style for citing online software or tools. Here's an example of how you might format a citation for an online software tool (e.g. ChatGPT):

OpenAI. (Year). ChatGPT [Computer software]. Retrieved Month Day, Year, from https://www.openai.com/chatgpt/

Here's an example of how to cite *a specific conversation* with ChatGPT:

ChatGPT. (Year, Month Day). Conversation Title. [Response to prompt: "User prompt or question"]. Retrieved from [URL of the platform where the conversation took place].

Please replace "Year, Month Day" and the URL with the actual year, retrieval date, and URL as needed, e.g.:

ChatGPT. (2023, January 18). How do you cite an answer in ChatGPT. [Response to user question "How do you cite an answer in ChatGPT"]. Retrieved from https://chat.openai.com/.

3.5 Expectations for Academic Integrity

The overall goal of this course is your learning. In order to demonstrate that you have reached this goal, the work you turn in needs to be your own. This includes putting written work into your own words and citing your sources to avoid plagiarism. If you work in a group, seek assistance from a tutor, use a resource on campus, and/or use online resources (including AI software), the work you turn in must be your own, demonstrating your own understanding of the material that you have gained through the learning process. If you have questions about academic integrity or plagiarism, please ask: my aim is to foster an environment where you can learn and grow, while also maintaining academic honesty and a clear representation of your learning and ideas. Penalties for serious offenses include a zero on the assignment and egregious offenses can even result in expulsion from the university, so it is important to understand expectations.

Plagiarism as defined by the OU Integrity Office includes: • Copying words and presenting them as your own writing. • Copying words, even if you give the source, unless you also indicate that the copied words are a direct quotation • Copying words and then changing them a little, even if you give the source. • Even if you express it in your own words, it is plagiarism to use someone else's idea as your own. Visit the OU Integrity Office for more information on what constitutes plagiarism.

4 Land Acknowledgement Statement

Long before the University of Oklahoma was established, the land on which the University now resides was the traditional home of the "Hasinais" Caddo Nation and "Kirikir?i:s" Wichita & Affiliated Tribes. We acknowledge this territory once also served as a hunting ground, trade exchange point, and migration route for the Apache, Comanche, Kiowa and Osage nations. Today, 39 tribal nations dwell in the state of Oklahoma as a result of settler and colonial policies that were designed to assimilate Native people. The University of Oklahoma recognizes the historical connection our university has with its indigenous community. We acknowledge, honor and respect the diverse Indigenous peoples connected to this land. We fully recognize, support and advocate for the sovereign rights of all of Oklahoma's 39 tribal nations. This acknowledgement is aligned with our university's core value of creating a diverse and inclusive community. It is an institutional responsibility

to recognize and acknowledge the people, culture and history that make up our entire OU Community.

5 University Policies

5.1 Religious Observance

It is the policy of the University to excuse the absences of students that result from religious observances and to reschedule examinations and additional required classwork that may fall on religious holidays, without penalty. See Faculty Handbook.

5.2 Reasonable Accommodation Policy

The Accessibility and Disability Resource Center is committed to supporting students with disabilities to ensure that they are able to enjoy equal access to all components of their education. This includes your academics, housing, and community events. If you are experiencing a disability, a mental/medical health condition that has a significant impact on one or more life functions, you can receive accommodations to provide equal access. Possible disabilities include, but are not limited to, learning disabilities, AD(H)D, mental health, and chronic health. Additionally, we support students with temporary medical conditions (broken wrist, shoulder surgery, etc.) and pregnancy. To discuss potential accommodations, please contact the ADRC at 730 College Avenue, (ph.) 405.325.3852, or adrc@ou.edu.

5.3 Title IX Resources and Reporting Requirement

Anyone who has been impacted by gender-based violence, including dating violence, domestic violence, stalking, harassment, and sexual assault, deserves access to resources so that they are supported personally and academically. The University of Oklahoma is committed to offering resources to those impacted, including: speaking with someone confidentially about your options, medical attention, counseling, reporting, academic support, and safety plans. If you would like to speak with someone confidentially, please contact OU Advocates (available 24/7 at 405-615-0013) or another confidential resource (see "Can I make an anonymous report?"). You may also choose to report gender-based violence and discrimination through other means, including by contacting the Institutional Equity Office(ieo@ou.edu, 405-325-3546) or police (911). Because the University of Oklahoma is committed to the safety of you and other students, I, as well as other faculty, Graduate Assistants, and Teaching Assistants, are mandatory reporters. This means that we are obligated to report gender-based violence that has been disclosed to us to the Institutional Equity Office. This includes disclosures that occur in: class discussion, writing assignments, discussion boards, emails and during Student/Office Hours. For more information, please visit the Institutional Equity Office.

5.4 Adjustments for Pregnancy / Childbirth Related Issues

Should you need modifications or adjustments to your course requirements because of documented pregnancy-related or childbirth-related issues, please contact your professor or the

Accessibility and Disability Resource Center at 405/325-3852 as soon as possible. Also, see the Institutional Equity Office FAQ on Pregnant and Parenting Students' Rights for answers to commonly asked questions.

5.5 Final Exam Preparation Period

Pre-finals week will be defined as the seven calendar days before the first day of finals. Faculty may cover new course material throughout this week. For specific provisions of the policy please refer to OU's Final Exam Preparation Period policy.

5.6 Emergency Protocol

During an emergency, there are official university procedures that will maximize your safety. Severe Weather: If you receive an OU Alert to seek refuge or hear a tornado siren that signals severe weather.

1. Look for severe weather refuge location maps located inside most OU buildings near the entrances

2. Seek refuge inside a building. Do not leave one building to seek shelter in another building that you deem safer. If outside, get into the nearest building.

3. Go to the building's severe weather refuge location. If you do not know where that is, go to the lowest level possible and seek refuge in an innermost room. Avoid outside doors and windows.

4. Get in, Get Down, Cover Up

5. Wait for official notice to resume normal activities.

Additional Weather Safety Information is available through the Department of Campus Safety.

5.7 Armed Subject/Campus Intruder:

If you receive an OU Alert to shelter-in-place due to an active shooter or armed intruder situation or you hear what you perceive to be gunshots: 1. Avoid: If you believe you can get out of the area WITHOUT encountering the armed individual, move quickly towards the nearest building exit, move away from the building, and call 911. 2. Deny: If you cannot flee, move to an area that can be locked or barricaded, turn off lights, silence devices, spread out, and formulate a plan of attack if the shooter enters the room. 3. Defend: As a last resort fight to defend yourself. For more information, visit OU's Active Shooter page. Shots Fired on Campus Procedure – Video.

5.8 Fire Alarm/General Emergency:

If you receive an OU Alert that there is danger inside or near the building, or the fire alarm inside the building activates: 1. LEAVE the building. Do not use the elevators. 2. KNOW at least two building exits 3. ASSIST those that may need help 4. PROCEED to the emergency assembly area 5 ONCE safely outside, NOTIFY first responders of anyone

that may still be inside building due to mobility issues. 6. WAIT for official notice before attempting to re-enter the building. OU Fire Safety on Campus.

5.9 Mental Health Support Services:

If you are experiencing any mental health issues that are impacting your academic performance, counseling is available at the University Counseling Center (UCC). The Center is located on the second floor of the Goddard Health Center, at 620 Elm Rm. 201, Norman, OK 73019. To schedule an appointment call (405) 325-2911. For more information, please visit University Counseling Center.

References

[1] Neil C Jones and Pavel A Pevzner. *An introduction to bioinformatics algorithms*. MIT press, 2004.

[2] Phillip Compeau. *Bioinformatics algorithms: an active learning approach*. Active Learning Publishers, 3rd edition edition.

[3] LaTeX-Project. Latex, 2023.

[4] OpenAI. ChatGPT, 2021.