Course Change Request

New Course Proposal

Date Submitted: 10/08/24 7:05 pm

Viewing: BINF 761 : Artificial Intelligence and Deep Learning in Bioinformatics

Last edit: 10/08/24 7:05 pm

Changes proposed by: clockha2

In Workflow

1. **BINF Representative**
2. **SC Curriculum Committee**
3. SC Assistant Dean
4. Assoc Provost-Graduate
5. Registrar-Courses
6. Banner

Approval Path

1. 10/28/24 11:34 am Iosif Vaisman (ivaisman): Approved for BINF Representative

Are you completing this form on someone else's behalf?

No

Effective Term: Spring 2025

Subject Code: BINF - Bioinformatics

Course Number: 761

Bundled Courses:

Is this course replacing another course? No

Equivalent Courses:

Catalog Title: Artificial Intelligence and Deep Learning in Bioinformatics

Banner Title: AI & Deep Learning in Bioinf

Will section titles vary by semester? No

Credits: 3

Schedule Type: Lecture

Hours of Lecture or Seminar per week: 3

Repeatable: May be only taken once for credit, limited to 3 attempts (N3)

Max Allowable Credits: 9

Default Grade Mode: Graduate Regular

Recommended Prerequisite(s):

Equivalent of 631 (molecular biology) and BINF 634 (bioinformatics programming) or with permission of the instructor.

Recommended Corequisite(s):

Required Prerequisite(s) / Corequisite(s) (Updates only):

Registrar's Office Use Only - Required Prerequisite(s)/Corequisite(s):

And/Or	(	Course/Test Code	Min Grade/Score	Academic Level	)	Concurrency?

Registration Restrictions (Updates only):

Registrar's Office Use Only - Registration Restrictions:

Field(s) of Study:

Class(es):

Level(s):

Degree(s):

School(s):

Catalog Description:

This course explores the use of artificial intelligence (AI) and deep learning in bioinformatics. Students will gain hands-on experience training deep learning models on biological datasets, with a focus on sequence data, medical images, omics data, and biomolecular structures related to human health, cancer, and infectious diseases. The models that will be covered include convolutional neural networks (CNNs), language models, variational autoencoders (VAEs), and generative adversarial networks (GANs).

Additionally, the course will focus on existing predictive tools such as AlphaFold. Students will present papers from literature to survey recent applications of deep learning in bioinformatics and complete a deep learning project.

Justification:

What: Offering a new course on the use of AI and deep learning in bioinformatics.

Why: AI and deep learning have been increasing applied to biological datasets, yet there is not currently a course offering that introduces these approaches to BINF graduate (master's and PhD) students despite student interest in this subject and its relevance to current bioinformatics research. This course will provide students in the School of Systems Biology with explicit experience in the theoretical underpinnings of deep learning algorithms, training deep learning on biological data, leveraging existing AI and deep learning tools such as AlphaFold, and surveying bioinformatics literature where deep learning is applied. Therefore, this course is important for BINF students to gain practical experience with these techniques and emerging AI technologies.

Does this course cover material which crosses into another department? Yes

Impacted Departments:

Department
CS - Computer Science

Learning Outcomes:

By the end of this course, students will be able to:

1. Discuss various deep learning algorithms and when they are appropriate to use
2. Train deep learning models on biological datasets
3. Leverage existing AI and deep learning bioinformatics tools
4. Appraise recent literature in bioinformatics that apply AI and deep learning

Will this course be scheduled as a cross-level cross listed section? No

Attach Syllabus

BINF 761 - AI and Deep Learning in Bioinformatics - Syllabus.pdf

Additional Attachments

Staffing:

Christopher Lockhart, PhD

Relationship to Existing Programs:

The proposed BINF 761 course will be included as an elective in the PhD and MS Bioinformatics and Computational Biology programs. It may also be taken as elective by the students from Biology MS and Biosciences PhD programs.

Relationship to Existing Courses:

SSB offers Machine Learning for Bioinformatics (BINF 760), which covers supervised and unsupervised predictive algorithms such as Naïve Bayes, logistic regression, and support vector machines. Deep learning is not presented in this course and cannot be included due to the specialized focus of deep learning. CS offers Deep Learning (CS 747) and Generative Deep Learning (CS 757), but the focus of these courses is not on biological datasets. To be relevant to SSB students, biological datasets and their analysis is fundamental to the proposed course.

Additional Comments:

Reviewer Comments

Key: 18784

BINF 761
Artificial Intelligence and Deep Learning in Bioinformatics

Spring 2025

School of Systems Biology
George Mason University
Manassas, VA

Instructor: Chris Lockhart
Email: clockha2@gmu.edu (preferred)
Phone: Microsoft Teams
Office: Virtual
Office Hours: By appointment (Microsoft Bookings)

Course Modality: Online or in person
Course Website: Canvas

Credits: 3

Course Description

This course explores the use of artificial intelligence (AI) and deep learning in bioinformatics. Students will gain hands-on experience training deep learning models on biological datasets, with a focus on sequence data, medical images, omics data, and biomolecular structures related to human health, cancer, and infectious diseases. The models that will be covered include convolutional neural networks (CNNs), language models, variational autoencoders (VAEs), and generative adversarial networks (GANs). Additionally, the course will focus on existing predictive tools such as AlphaFold. Students will present papers from literature to survey recent applications of deep learning in bioinformatics and complete a deep learning project.

Recommended Prerequisites: Equivalent of BINF 631 (molecular biology) and BINF 634 (bioinformatics programming) or with permission of the instructor.

Learning Outcomes

By the end of this course, students will be able to:

1. Discuss various deep learning algorithms and when they are appropriate to use
2. Train deep learning models on biological datasets
3. Leverage existing AI and deep learning bioinformatics tools
4. Appraise recent literature in bioinformatics that apply AI and deep learning

Course Textbook

Izadkhah, H. (2022). *Deep Learning in Bioinformatics: Techniques and Applications in Practice* (1st ed.). Elsevier Science & Technology.

Tentative Course Topics

Week	Topic	Reading
1	Introduction to machine learning, artificial intelligence, and deep learning; Introduction to PyTorch through logistic regression; Classification of Wisconsin breast cancer dataset	Izadkhah, Ch. 2, 7
2	Multilayer perceptrons (MLP), backpropagation, activation functions; Classification of RNA-seq data related to HIV	Izadkhah, Ch. 4-5
3	Convolutional neural networks (CNNs) and their application to biomedical images; Classification of breast cancer histopathology image dataset	Izadkhah, Ch. 8
4	Recurrent neural networks (RNNs) and their application to time series and sequential data; Predicting protein secondary structure from primary sequence with the CB513 dataset	Izadkhah, Ch. 12
5	Attention mechanisms and transformers for sequence-to-sequence translation; Predicting gene function with the ENCODE dataset	Zhang et al. (2023) Bioinf. Adv. 3(1), vbad001
6	Language models; Named entity recognition with the CORD-19 (COVID-19 article) dataset using BERT	Liu et al. (2024) arXiv 2401.04155
7	Graph neural networks and the application to interconnected data; Protein-protein interaction prediction with STRING Viruses dataset	Zhang et al. (2021) Front. Genet. 12, 690049
8	Generative AI; AlphaFold, RoseTTAFold	Krishna et al. (2024) Science 384, 6693
9	Encoder-decoder architectures; Repurposing ESM embeddings for viral protein annotation using the Influenza Research Database	Izadkhah, Ch. 11
10	Variational autoencoders (VAEs); Generating novel antimicrobial peptides with VAE	Wei et al. (2020) IEEE Access 9, 4939-4956
11	Generative adversarial networks (GANs); Generating novel antimicrobial peptides with GAN	Lan et al. (2020) Front. Public Helth 8, 164
12	Diffusion models; RFdiffusion	Watson et al. (2023) Nature 620, 1089-1100
13	Boltzmann generators for biomolecules	Noé et al. (2019) Science 365, 6457
14	Future of AI and its use in personalized medicine and drug development	Izadkhah, Ch. 13

Each lecture is a 2½ hour presentation with a 10-minute break.

Course Policies

Grading scale (points): A+ (≥97), A (94-96.9), A- (90-93.9), B+ (87-89.9), B (84-86.9), B- (80-83.9), C (60-79.9), F (<60).

Grading policy: Students will be graded on homework (30%), journal club (30%), and a final project (40%).

- There will be weekly homework assignments, which will feature theoretical questions and programming exercises related to that week's lecture. Each assignment will be equally weighted.
- Journal club will be met by (1) once during the semester finding a recent peer-reviewed article that applies deep learning to biological data, writing 1-2 paragraphs on the article, and posting this information to Canvas for discussion and (2) asking critical questions of the papers posted by other students. To earn full credit, students must contribute to the discussion of 4 articles throughout the semester.
- The final project will require students to complete their own deep learning project from topics prepared by the instructor. Students may pursue their own topics with permission. The project will be presented in a written report.

Late assignments and make-up work: In case of illness or quarantine, please contact the instructor to set up a plan for make-up work. Late assignments will not be accepted unless due to emergency, illness, quarantine, work-related, or other documented reasons.

Course recordings: All lectures in this class will be recorded to provide necessary information for students in this class. Recordings will be stored on Canvas and will only be accessible to students taking this course during this semester.

Other considerations: If there are any schedule issues related to religious holidays, please inform the instructor the first week of class.

Course Logistics

Content distribution: The course uses Canvas for distributing lecture materials, submission of homework, and grading. Canvas can be accessed by visiting https://canvas.gmu.edu/ and logging in with your MasonID and password.

Communication: I will use Mason email to distribute class updates and communicate with students (see Email section in Student Responsibilities). If you wish, please share your name and gender pronouns with me and how best to address you in class and via email. I use he/him for myself, and you may address me as Chris or Dr. Lockhart in email and verbally. Communication over email is largely preferred, and I will respond to student emails promptly within 48 hours.

Course Technology Requirements

Software and hardware: This course uses Canvas as a learning management system available at https://canvas.gmu.edu/. Students are required to have regular, reliable access to a computer with an updated operating system (recommended: Windows 10 or Mac OS X 10.15 or higher) and a stable broadband Internet connection (cable modem, DSL, satellite broadband, etc., with a consistent 1.5 Mbps download speed or higher).

Course-specific software: This course will use Python (the Python distribution from Anaconda https://www.anaconda.com/ is recommended).

Technical help: If you have difficulty with accessing Canvas, please contact the ITS Support Center at (703) 993-8870 or support@gmu.edu. If you have trouble with using the features in Canvas, email courses@gmu.edu.

Student Responsibilities

Email: Students must use their Mason email account to receive important University information, including communications related to this class. Per University policy, I will not respond to messages sent from or send messages to a non-Mason email address.

Academic integrity: The integrity of the University community is affected by the individual choices made by each of us. Mason has an Honor Code with clear guidelines regarding academic integrity. Three fundamental and rather simple principles to follow always are that: (1) all work submitted be your own; (2) when using the work or ideas of others, including fellow students or online sites, give full credit through accurate citations; and (3) if you are uncertain about the ground rules on a particular assignment, ask for clarification. No grade is important enough to justify academic misconduct. Plagiarism means using the exact words, opinions, or information from another person without giving the person credit. Writers give credit through accepted documentation styles, such as parenthetical citation, footnotes, or endnotes. Paraphrased material must also be cited, using the appropriate format for this class. A simple listing of books or articles is not sufficient. Plagiarism is the equivalent of intellectual robbery and cannot be tolerated in the academic setting. If you have any doubts about what constitutes plagiarism, please see me, or consult the Academic Integrity website at https://oai.gmu.edu/.

Generative AI policy: Use of generative AI (e.g., ChatGPT) is strongly discouraged. If students use these tools, they must follow the fundamental principles of the Honor Code. This includes being honest about the use of generative AI for submitted work and giving credit through accurate citations.

Disability accommodations: Disability Services at George Mason University is committed to providing equitable access to learning opportunities for all students by upholding the laws that ensure equal treatment of people with disabilities. If you are

seeking accommodations for this class, please first visit http://ds.gmu.edu/ for detailed information about the Disability Services registration process. Then please discuss your approved accommodations with me. Disability Services is in Student Union Building I (SUB I), Suite 2500. Email: ods@gmu.edu | Phone: (703) 993-2474.

Student Services

University writing center: Take advantage of the Writing Center as you work on written assignments in this course. You can book a free 45-minute appointment to meet with a tutor on Zoom or to submit a draft for written feedback. Tutors will work with you on any phase of a writing project. They can help you develop your ideas, provide feedback on a draft, answer your questions, and show you strategies for brainstorming, organizing, drafting, revising, and editing. To schedule an appointment, go to writingcenter.gmu.edu, register with the center, and make an appointment using the online scheduler. Watch this short video (https://youtu.be/LA-B0Szoe28) for more detailed guidance on making an appointment and send any questions to wcenter@gmu.edu.

University Libraries: University Libraries provides resources for distance learning students (See the Library website: https://library.gmu.edu/for/online).

Counseling and psychological services: The George Mason University Counseling and Psychological Services (CAPS) staff consists of professional counseling and clinical psychologists, social workers, and counselors who offer a wide range of services (e.g., individual and group counseling, workshops and outreach programs) to enhance students' personal experience and academic performance (See the Counseling and Psychological Services website: https://caps.gmu.edu).

Family Educational Rights and Privacy Act (FERPA): The Family Educational Rights and Privacy Act of 1974 (FERPA), also known as the "Buckley Amendment," is a federal law that gives protection to student educational records and provides students with certain rights (See the Registrar's Office website: registrar.gmu.edu/ferpa/).

Course Materials and Student Privacy

Video recordings of class meetings that are shared only with the instructors and students officially enrolled in a class do not violate FERPA or any other privacy expectation. Video recordings that only include the instructor (no student names, images, voices, or identifiable texts) may be shared without violating FERPA (but see University Policies: Privacy, for some qualifications and recommendations). All course materials posted to Canvas or other course site are private to this class; by federal law, any materials that identify specific students (via their name, voice, or image) must not be shared with anyone not enrolled in this class.

Video conferencing or recordings: Video recordings – whether made by instructors or students — of class meetings that include audio, visual, or textual information from other students are private and must not be shared outside the class. Live video conference meetings (e.g., Zoom) that include audio, textual, or visual information from other students must be viewed privately and not shared with others in your household or recorded and shared outside the class.

Common Course Policies

This course adheres to the common course policies set by George Mason University, which includes polices about Academic Standards, Accommodations for Students with Disabilities, FERPA, and Title IX. These policies are described in more detail at the following link: https://stearnscenter.gmu.edu/home/gmu-common-course-policies/.