CSCI 1851: Machine Learning for Biology and Health

Spring 2026

Instructor:	Ritambhara Singh	**Time:**	TTh 10:30 – 11:50 AM
Email:	ritambhara@brown.edu	**Location:**	CIT 241

Course Description:

Can machine learning models that have defeated world champions in games or surpassed humans in image recognition also help us understand biology and improve human health? How far can these computational approaches take us toward diagnosing disease, discovering new therapies, and personalizing medicine? In an era of rapidly expanding biological, clinical, and health data, ranging from genomics and imaging to electronic health records, machine learning methods are becoming essential tools for extracting insight and guiding decision-making.

Despite their promise, applying machine learning to biological and health data presents unique challenges, including data heterogeneity, limited labels, bias, interpretability, and ethical considerations. As researchers working at the intersection of machine learning, biology, and health, we must think carefully about when and how to use these methods. Is machine learning appropriate for a given biological or clinical problem? Which models are best suited for different data types? How can we ensure that our approaches lead not only to accurate predictions but also to meaningful biological or medical insight?

In this course, you will explore these questions through the study of machine learning concepts and hands-on practice of real-world applications. You will learn about a range of biological and health-related tasks, core machine learning models (ranging from linear regression to transformers), and how they fit together in practice. The course emphasizes critical thinking, interdisciplinary collaboration, and the thoughtful application of machine learning to advance our understanding of biology and improve health outcomes.

This course is an undergraduate-level course. Enrollment limited to 40.

Course Objectives: By the end of this course, you will be able to:

1. Connect different machine learning models from linear regression to transformers to applications in biology and health domain.
2. Understand key machine learning concepts when solving homework assignments.
3. Think critically about using a machine learning method for a task - what works, what doesn't work, and how a particular model may or may not be appropriate for the task.
4. Collaborate with classmates on a team project to apply machine learning models to a domain task.
5. Communicate your findings (both positive and negative results are encouraged) clearly by writing a report and through oral presentations.

Prerequisites: This course requires some basic understanding of Artificial Intelligence (AI) concepts through courses like CSCI 0410, 1411.

Instructor's Office Hours: Fridays 3:30-5:30 PM in Data Science Institute (DSI) Room #313 (164 Angell St., 3rd Floor) or over video call. Book an appointment slot here. (Note that the office hour slots may move during the weeks the instructor is traveling for work).

Teaching Assistants (TA) Information:

- Head TAs: Falak Pabari and Kyle Yeh

Email: cs1851headtas@lists.brown.edu

In the aftermath of December 13 shootings:

While the rest of this syllabus has a business-as-usual tone, I want to remain mindful that this semester will be hard from time to time for students. I will try to be as flexible as possible to accommodate your needs, so you should feel free to communicate them.

Undergraduate students who are concerned and feel they need academic flexibility or special accommodations beyond those spelled out on the syllabus below, should first consult the appropriate academic dean. Graduate students can e-mail me directly. It is always okay to reach out to Campus life, Counseling and Psychological Services, and/or support in our larger Rhode Island community.

The university has provided a list of resources that you can use if you would like to learn more or connect with psychological services. You can find all of them on the Ever True website.

Course websites: We will be using the following websites for smooth running of the online course:

1. Course Canvas Site: https://canvas.brown.edu/courses/1102089
   - **Central site to access all course-related information.** This includes EdStem, links to the course website etc.
   - **Instructor announcements**. The instructor will use Canvas for weekly announcements and updates during the semester.
   - **Project report submission**. Project report will be submitted via the Assignments section on Canvas.
2. EdStem (Accessed through Canvas):
   - **HTA announcements:** EdStem will be used by HTAs to provide information and updates regarding the course.
   - **Out-of-class discussions:** Students are encouraged to ask questions and conduct course-related discussions outside class times.
3. Course Website: https://csci1851.github.io
   - **Publishing assignments:** All course assignments will be published on the course website
   - **Lecture slides:** All lecture slides will be posted on the course website
   - **Extra resources:** Any extra resource useful for the class will be made available on the course website
4. Gradescope (Accessed through Canvas):
   - **Homework assignments.** Students will be able to submit their homework assignments on Gradescope.

Relevant Textbooks (free online access):

- A course in machine learning, by Hal Daume III
- Dive into Deep Learning, by Aston Zhang, Zack C. Lipton, Mu Li, and Alex J. Smola.

Assessment of learning:

(30%) Homework Assignments: The assignments will help you in developing your ability to distill ideas from a research paper. We will hand out five homework assignments through the duration of the course. Each assignment will consist of 4-5 conceptual questions from the papers related to the topics covered in the class and 1 programming assignment. The programming assignment (stencil code provided) will provide a practical understanding of working with genomics data and deep learning models. (Details/guidelines will be given in class)

(15%) Mid-term exam: This exam will take place during class time and the goal will be to assess the conceptual understanding of the material until mid semester.

(15%) Final exam: This exam will take place during exam period and the goal will be to assess the conceptual understanding of the material after mid semester.

(30%) Final course project: With 2-3 other students, you will work on a course project that will apply a machine learning model to a particular biology and health domain task. This project is modeled like a Kaggle competition, where we will describe the task and provide the related dataset. Each team will then develop and apply a machine learning framework to solve that task using the released dataset. We will test the models on an unreleased dataset and publish a scoreboard ranking their performances.

For the next round, we will release the dataset that the mid-term models were tested on and update the task. Based on the performance during the mid-term mark, each team will now modify their models accordingly. At the end of the semester, you will once again submit your final model, present the updated method as well as submit a report describing the work. All the submitted models will be tested on another unreleased dataset to generate the final scoreboard. We will also the top-performing teams (during both mid-term and finals). Details/guidelines will be given in class.

Note: Equal participation is an important requirement for the final project. If there is any indication (through feedback forms) that a student did not contribute equally to the project, then the instructor will deduct points from the final project tally for that student.

- **(5%) Mid-term model evaluation:** This will just be a check-point to see how the initial model implementations are performing before updating the task. Scored on just completion.
- **(15%) Final Project presentations:** Each team will get a 10-minute slot (8 min for presentation + 2 min for questions) to present their project at the end of the semester.
- **(10%) Final Report and Code:** The teams will submit the final report that will be reviewed by the instructor. You will include the updated details of the task, data, model, training, experiments, and results. The reports will be due after the presentations. You may incorporate the suggestions from the presentations to improve it. We will also check quality and correctness the code implementation (including documentation)

(10%) In-class participation: This course aims to promote engagement and exchange of research ideas. Thus, your level of participation in class discussions will count for 10% of your final grade. Your participation may include asking questions, providing personal insights during discussions, giving feedback to your peers, etc.

Semester Hours:

Total time spent in and out of class for this course is estimated at 180 hours. During the semester you will spend approximately following number of hours for in-class and out-of-class course work:

Task	Hours Spent on Task
Class Time	40
Homework Assignments	60
Mid-term exam prep	15
Mid-term project	25
Final exam prep	15
Final project	25
Total	180

Tentative Course Calendar:

Note: All assignments are due at 5 PM EST on the specified day.

Date, Day	Agenda
Introduction	
Jan 22, Thursday	**Welcome to CSCI 1851** Discussion of the course logistics, assignments, evaluation, and brief overview of machine learning and health and biology tasks.
Section I: Linear Models	
Jan 27, Tuesday	Linear classification (Perceptron)
Jan 29, Thursday	Logistic Regression, introduction to heart disease prediction task
Feb 03, Tuesday	Linear Regression, introduction to the aging prediction task **Homework 1 released**
Feb 05, Thursday	Overfitting and regularization
Section II: Decision trees	
Feb 10, Tuesday	Decision Trees and Random Forests
Feb 12, Thursday	Gradient Boosting, introduction to Cancer prediction task
Feb 13, Friday	**Homework 1 due** **Homework 2 released**

Continued on next page

Feb 17, Tuesday	**No Class**: Long weekend
Section III: Support Vector Machines	
Feb 19, Thursday	Kernels
Feb 24, Tuesday	Kernels
Feb 26, Thursday	Support Vector Machines (SVMs)
Feb 27, Friday	**Homework 2 due**
Mar 03, Tuesday	Kernalized SVMs **Homework 3 released**
Section IV: Neural Networks and CNNs	
Mar 05, Thursday	Fully connected neural networks
Mar 10, Tuesday	Neural networks continued, predicting from X-rays **Homework 3 released**
Mar 12, Thursday	Convolutional Neural Networks (CNNs)
Mar 13, Friday	**Homework 3 due**
Mar 17, Tuesday	Deep architectures, U-Nets **Homework 4 released**
Mar 19, Thursday	**Mid-term examination**
Mar 24, Tuesday	**No Class**: Spring Break
Mar 26, Thursday	**No Class**: Spring Break
Section IV: Unsupervised Learning	
Mar 31, Tuesday	Unsupervised learning, introduction to single-cell gene expression
April 02, Thursday	Autoencoders
April 03, Friday	**Homework 4 due**

Continued on next page

April 07, Tuesday	Variational Autoencoders **Homework 5 released**
	Section V: Graph Neural Networks
Apr 09, Thursday	Graph neural networks (GNNs)
Apr 14, Tuesday	GNNs continued, introduction to drug-based prediction task
	Section VI: Sequential Modeling
Apr 16, Thursday	Recurrent Neural Networks and LSTMs, predicting from DNA sequences
Apr 17, Friday	**Homework 5 due**
Apr 21, Tuesday	Attention and Transformers **Homework 6 released**
April 23, Thursday	Transformers contd., single-cell foundation models
Apr 28, Tuesday	**No Class**: Reading Period
April 30, Thursday	**No Class**: Reading Period
May 01, Friday	**Homework 6 due**
May 05, Tuesday	**Final project presentations**
May 08, Friday	**Final examination**
May 12, Tuesday	**Final project reports due**

How can you do well? This class has a medium-level course load and you can ensure your success in it by doing the following:

- Regularly attending classes, asking questions, and actively participating in the class discussions.
- Completing and turning in all assignments on time.
- Preparing well for the course exams.
- Equally contributing to the project assignment and clearly presenting your project and its results in the presentations and reports.

Collaboration Policy: Discussion of material with your classmates is both permitted and encouraged. However, showing, copying or other sharing of answers to written questions and actual code on homework and projects is forbidden, unless specified. This includes publishing projects publicly on Github or any other public platform. In addition, reusing code or pre-trained models from another student or any public platform is forbidden, unless specified.

Missed assignments (including late assignments): For all assignment submissions, you can get a 72 hours extension for at most 4 homework deadlines without penalty. Excluding the scenario mentioned above, 20% of the total points will be deducted for late submissions and missed submissions will not be assigned any score. For project presentations, if you are unable to present on a particular day, please exchange your slot with another student/team and inform the instructor. No-show on the day of your assigned presentation will be treated as a missed assignment.

Policy on the use of AI-powered tools for course assignments: All work that students submit during the course must be their own original work and represent their own thoughts and ideas. As such, the use of AI-powered tools (such as OpenAI's ChatGPT or GitHub's CoPilot) for completing course assignments is discouraged. The use of AI-powered tools without citation will be considered academic misconduct.

If a student chooses to use these tools for course assignments, they must acknowledge and thoroughly document their use of the tool. The student must: 1) cite the tool used, 2) include an explanation of how the tool was used for the assignment, and 3) fully document the student's own contribution versus the contribution of the tool (e.g., including full ChatGPT transcripts as an appendix to your assignment). All assignments will be graded based on the student's original ideas – students risk losing credit if the documentation provided is insufficient to determine the student's original contributions.

Students with Special Needs: Brown University is committed to full inclusion of all students. Please inform me early in the term if you have a disability or other conditions that might require accommodations or modification of any of these course procedures. You may speak with me after class or during office hours. For more information, please contact Student and Employee Accessibility Services at 401-863-9588 or SEAS@brown.edu. Students in need of short-term academic advice or support can contact one of the deans in the Dean of the College office.

Diversity Statement: This course is designed to support an inclusive learning environment where diverse perspectives are recognized, respected and seen as a source of strength. It is our intent to provide materials and activities that are respectful of various levels of diversity: mathematical background, previous computing skills, gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

Multilingual Students: Brown welcomes students from around the country and the world, and their unique perspectives enrich our learning community. To support students whose primary language is not English, an array of English support services are available on campus including language and culture workshops and individual appointments. For more information, contact english-support@brown.edu or (401) 863-5672.