CSCI 1850: Deep Learning in Genomics

Spring 2021

Instructor:	Ritambhara Singh	**Time:**	TTh 10:30 – 11:50 AM
Email:	ritambhara@brown.edu	**Format:**	Online (Synchronous)

Course Description:

Can deep learning models that have defeated gamers or recognized images better than humans also help us understand genomics? How far will this interdisciplinary research take us on our quest to cure cancer? In an era with faster-than-Moore's-Law exponential growth of the genomics data (Berger et al. 2016), deep learning methods are finally able to assist in solving essential problems in the field. However, these exciting developments also face challenges that are unique to working with data from our DNA.

As researchers trying to combine deep learning and genomics, we have to think carefully about applying these models effectively to genomics tasks. Is it appropriate to use deep learning for our application? What model should we use? Will our approach improve our understanding of the data or the problem? In this course, you will answer these questions by our coverage of recent research literature in the class. You will learn about different genomics tasks, deep learning models, and how they fit together. The course is designed to enable critical thinking and allows students to work together to apply these models.

This course is an undergraduate-level course. Enrollment limited to 30.

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

1. Connect different state-of-the-art models like Convolutional Neural Networks, Recurrent Neural Networks, etc. to applications in genomics
2. Extract key ideas from research papers when solving homework assignments.
3. Think critically about using a deep-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 deep learning models to a genomics 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 an understanding of machine learning and deep learning concepts. Knowledge about genomics is useful but not necessary.

Instructor's Office Hours: Thursdays 12PM – 1 PM, or by appointment (will be held over Zoom).

Teaching Assistants Information:

Office hours will be held over zoom

- HTA: Daniel Ben-Isvy: Wednesdays 4 PM – 6 PM
- UTA: Giselle Garcia : Mondays 8 PM – 10 PM

Email: cs1850tas@lists.brown.edu

Online (Synchronous) format: What does that entail?

All classes will be conducted online over zoom and recorded for future viewing. Active student participation during the class is highly encouraged. Students anticipating difficulties in attending classes at the assigned time are recommended to email the instructor by January 19,2021, so that accommodations could be made accordingly.

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

1. Course Canvas Site: https://canvas.brown.edu/courses/1084361
   - **Zoom recordings:** Classes will be conducted via Canvas Zoom interface that will allow the recordings to be available immediately after class.
2. Piazza: https://piazza.com/brown/spring2021/csci1850
   - **Class announcements:** Piazza will be used to provide information and regular updates regarding the course.
   - **Out-of-class discussions:** Students are encouraged to ask questions and conduct course-related discussions outside class times using Piazza
3. Gradescope: https://www.gradescope.com/courses/228828
   - **Assignment submissions:** All course assignments will be submitted via Gradescope
4. Course Website: http://cs.brown.edu/courses/csci1850
   - **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

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)

(60%) Final course project: With 1 other student, you will work on a course project that will apply a deep learning model to a particular genomics 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 deep learning framework to solve that task using the released dataset. For mid-term evaluations, the teams will submit their trained models, present their methods, and write a report. 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 and discussions during the mid-term presentations, 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 announce prizes for the top-performing teams (during both mid-term and finals). (Details/guidelines will be given in class)

- **(20%) Mid-term Project presentations:** Each team will get a 15-minute slot (10 min for presentation + 5 min for questions) to present their method and results.

- **(10%) Mid-term Report:** The teams will submit a report that will be reviewed by the instructor and the TAs. You will include the 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.
- **(20%) Final Project presentations:** Each team will get a 15-minute slot (10 min for presentation + 5 min for questions) to present their project at the end of the semester.
- **(10%) Final Report:** 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.

(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 project	40
Final project	40
Total	180

Tentative Course Calendar:

Note: All assignments are due at 11:59 PM EST on the specified day.

Date, Day	**Agenda**
Introduction	
Jan 21, Thursday	**Welcome to CSCI 1850: Deep Learning in Genomics** Discussion of the course schedule, assignments, evaluation, and brief overview of genomics tasks that will be covered in the course.
Section I: Predicting from DNA sequences	
Continued on next page	

Jan 26, Tuesday	**Transcription Factor Binding Prediction:** Task, biology, previous works
Jan 27, Wednesday	Course setup form due (link)
Jan 28, Thursday	**Transcription Factor Binding Prediction:** Deep learning models (DeepBind)
Feb 02, Tuesday	**Chromatin Accessibility Prediction:** Task, biology, previous works, deep learning models (Basset) **Homework 1 released**
Feb 04, Thursday	**Multi-omics prediction:** Task, biology, previous works
Feb 09, Tuesday	**Multi-omics prediction:** Deep learning models (DeepSEA, DANQ)
Section II: Predicting from other genomic signals	
Feb 11, Thursday	**Gene expression prediction:** Task, biology, previous works
Feb 16, Tuesday	**No Class**: Long weekend **Homework 1 due**
Feb 18, Thursday	**Gene expression prediction:** Deep learning models (DeepChrome) **Homework 2 released**
Feb 23, Tuesday	**Back-propagation and Improving HiC Resolution:** Task, biology, previous works
Feb 25, Thursday	**Improving HiC Resolution:** Deep learning models using CNNs
Mar 02, Tuesday	**Improving HiC Resolution:** Deep learning models using GANs
Section III: Imputation	
Mar 04, Thursday	**Imputation of genomic signals:** Task, biology, previous works **Homework 2 due**
Mar 09, Tuesday	**Imputation of genomic signals:** Deep learning models (Avocado) and **Working with single cells:** Task, biology, previous works **Homework 3 released**
Mar 11, Thursday	**Working with single cells:** Deep learning models
Mar 16, Tuesday	**Mid-term project presentations (Slot 1)**
Mar 18, Thursday	**Mid-term project presentations (Slot 2)**
Section IV: Interpretation of Deep Learning models	
Continued on next page	

Mar 23, Tuesday	Perturbation based methods and Saliency maps **Mid-term project reports due**
Mar 25, Thursday	Class based optimization and Attention based methods **Homework 3 due**
Mar 30, Tuesday	Other gradient based methods **Homework 4 released**
Section V: Other interesting applications	
Apr 01, Thursday	Graph neural networks
Apr 06, Tuesday	Manifold Alignments in single cells
Apr 08, Thursday	Working with Single cell microscopy data
Final Course Project	
Apr 13, Tuesday	**Final project presentations (Slot 1)** **Homework 4 due**
April 15, Thursday	**Final project presentations (Slot 2)**
Apr 20, 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.
- 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 3-day extension for at most 3 deadlines without penalty. Excluding the scenario mentioned above, 20% of the total points will be deducted for late submissions and missed submissions won't 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.

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.