CSCI2952-G: Deep Learning in Genomics

Fall 2023

Instructor:	Ritambhara Singh	**Time:**	TTh 1:00 – 2:20 PM
Email:	ritambhara@brown.edu	**Location:**	CIT 477

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 reading recent research literature and discussing it during 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 a graduate-level seminar. Enrollment limited to 20. Instructor's permission required.

Course Objectives: When you complete 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 and **discuss** them with your peers.
3. **Think** critically about using a deep-learning method for a new 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 genomics datasets of your interest.
5. **Communicate** your findings (both positive and negative results are encouraged) clearly by writing a research paper 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.

Website: http://cs.brown.edu/courses/csci2952g/
Canvas site: https://canvas.brown.edu/courses/1092952

We will be using the course Canvas site for the following activities:

- Class announcements: Canvas Announcements 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 Canvas Discussions.
- Assignment submissions: All course assignments will be submitted via Canvas Assignments.

Instructor's Office Hours: Thursdays 2:30 – 3:30 PM in CCMB Room #313 (164 Angell St.), or by appointment.

Graduate TA: Atishay Jain
Graduate TA Email: atishay_jain@brown.edu
TA office hours: TBD.

Assessment of learning:

(20%) Paper presentations:
This assignment will help you in developing your ability to distill and clearly present the ideas from a research paper. Each student will sign-up for one paper (link to sign-up sheet) to present during the class in a 40-minute presentation. You are free to choose your medium; however, using slides is recommended to gain practice for presenting in conferences. (Other details/guidelines will be given in class)

(10%) Pre-class reading: Everyone (except the presenter) will be required to read the assigned paper few days before the class and submit a short paragraph summarizing the article as well as listing 4 points that were convincing in the paper and 4 points that were confusing. These points will be provided to the presenter a day before the class, so they can address them during the presentation or raise them during the discussion.

(60%) Final course project: With 1-2 other students, you will work on a course project that will apply a deep learning model to a genomics task of your interest. If you cannot think of an application, you can pick one from a paper discussed in the class. However, you will then apply a model (or models) that is different from the chosen article and compare and contrast your work with that of the paper. At the end of the semester, you will present your completed project as a presentation to the class as well as submit a 4-page workshop paper describing the work. For fruitful project collaboration, teams would be required to setup project repositories on GitHub or Bitbucket. (Other details/guidelines will be given in class)

- **(10%) Mid-term Project presentations:** Each team will give a 5-minute presentation pitching its project idea to the class to get feedback.
- **(10%) Mid-term Research paper:** The teams will submit an initial draft of the research paper (including a literature survey) that will be reviewed by the instructor and the students. You may include preliminary results (if available) or describe the expected results.
- **(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.
- **(20%) Final Research paper:** The teams will be required to submit their research paper drafts 1-week prior to the project presentations. This draft will be also be reviewed by the instructor and the students. The teams may incorporate the suggestions from the reviews and presentations to improve their papers. The revised final research papers (one per team) will be due after a week.

(10%) Course participation: This course aims to promote engagement and exchange of research ideas. Thus, your level of participation in online and offline discussions will count for 10% of your final grade. Your participation may include asking questions, providing insights during discussions, giving feedback to your peers, contributing questions or answers on the discussion boards 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
Reading papers	60
Writing reviews	20
Final project	60
Total	180

Tentative Course Calendar:

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

Date, Day	**Agenda**
	Introduction
Sep 07, Thursday	**Welcome to CSCI2952-G: Deep Learning in Genomics** Discussion of the course schedule, assignments, evaluation, etc.
Sep 10, Sunday	**Summary due:** Deep Learning: new computational modeling techniques for genomics
Sep 12, Tuesday	A primer on biology **Summary due:** Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning
	Section I: Predicting from DNA Sequences
Sep 14, Thursday	**Presentation:** Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning
Sep 17, Sunday	**Summary due:** Predicting effects of noncoding variants with deep learning-based sequence model
Sep 19, Tuesday	**Presentation:** Predicting effects of noncoding variants with deep learning-based sequence model **Summary due:** DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences.
Sep 21, Thursday	**Presentation:** DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences.
Sep 22, Friday	**Course Project: Project team plans due**
Continued on next page	

Sep 24, Sunday	**Summary due:** Effective gene expression prediction from sequence by integrating long-range interactions
Sep 26, Tuesday	**Presentation:** Effective gene expression prediction from sequence by integrating long-range interactions **Summary due:** DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome
Sep 28, Thursday	**Presentation:** DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome
Section II: Predicting from other genomic signals	
Oct 01, Sunday	**Summary due:** DeepChrome: deep-learning for predicting gene expression from histone modifications
Oct 03, Tuesday	**Presentation:** DeepChrome: deep-learning for predicting gene expression from histone modifications **Summary due:** Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus
Oct 05, Thursday	**Presentation:** Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus
Oct 06, Friday	**Course Project: Literature review due**
Oct 8, Sunday	**Summary due:** hicGAN infers super resolution Hi-C data with generative adversarial networks
Oct 10, Tuesday	**Presentation:** hicGAN infers super resolution Hi-C data with generative adversarial networks **Summary due:** Epiphany: predicting Hi-C contact maps from 1D epigenomic signals
Oct 12, Thursday	**Presentation:** Epiphany: predicting Hi-C contact maps from 1D epigenomic signals
Oct 15, Sunday	**Summary due:** Avocado: Multi-scale Deep Tensor Factorization Learns a Latent Representation of the Human Epigenome
Oct 17, Tuesday	**Presentation:** Avocado: Multi-scale Deep Tensor Factorization Learns a Latent Representation of the Human Epigenome **Summary due:** Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks
Section III: Interpretation of deep learning models	
Oct 19, Thursday	**Presentation:** Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks

Continued on next page

Oct 20, Friday	**Course Project: First draft of research paper due**
Oct 22, Sunday	**Summary due:** Attend and Predict: Understanding Gene Regulation by Selective Attention on Chromatin
Oct 24, Tuesday	**Presentation:** Attend and Predict: Understanding Gene Regulation by Selective Attention on Chromatin **Summary due:** Learning Important Features Through Propagating Activation Differences
Oct 26, Thursday	**Presentation:** Learning Important Features Through Propagating Activation Differences
Oct 29, Sunday	**Course Project: Reviews due for first draft**
Section IV: Deep learning and single-cells	
Oct 31, Tuesday	**Course Project: Project Idea Presentations** **Summary due:** DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning
Nov 02, Thursday	**Presentation:** DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning
Nov 05, Sunday	**Summary due:** Single-cell RNA-seq denoising using a deep count autoencoder
Nov 07, Tuesday	**Presentation:** Single-cell RNA-seq denoising using a deep count autoencoder **Summary due:** MAGAN: Aligning Biological Manifolds
Nov 09, Thursday	**Presentation:** MAGAN: Aligning Biological Manifolds
Nov 12, Sunday	**Summary due:** Multi-domain translation between single-cell imaging and sequencing data using autoencoders
Nov 14, Tuesday	**Presentation:** Multi-domain translation between single-cell imaging and sequencing data using autoencoders **Summary due:** Mapping single-cell data to reference atlases by transfer learning
Nov 16, Thursday	**Presentation:** Mapping single-cell data to reference atlases by transfer learning
Section V: Interesting applications of Graph Neural Networks	
Nov 21, Tuesday	**Course Project: Project check-in with the teams** **Summary due:** Integrating Long-Range Regulatory Interactions to Predict Gene Expression Using Graph Convolutional Networks
Nov 23, Thursday	**No class:** Thanksgiving Holiday

Continued on next page

Nov 28, Tuesday	**Presentation:** Integrating Long-Range Regulatory Interactions to Predict Gene Expression Using Graph Convolutional Networks **Summary due:** Single-cell classification using graph convolutional networks
Nov 30, Thursday	**Presentation:** Single-cell classification using graph convolutional networks
Dec 03, Sunday	**Summary due:** Connecting high-resolution 3D chromatin organization with epigenomics
Dec 05, Tuesday	**Presentation:** Connecting high-resolution 3D chromatin organization with epigenomics **Summary due:** Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens
Dec 07, Thursday	**Presentation:** Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens
Final Course Project	
Dec 08, Friday	**Course Project: Second draft due**
Dec 12, Tuesday	**No class: Reading period** **Course Project: Reviews due for second drafts**
Dec 14, Thursday	**Final project presentations**
Dec 17, Thursday	**Final project papers due**

How can you do well? The course-load for this class is relatively low and and you can ensure your success in it by doing the following:

- Regularly attending classes and actively participating in class discussions.
- Reading the assigned paper prior to coming to class and providing useful feedback to your peers.
- Preparing well for your paper presentation and addressing the questions raised by your peers.
- Completing and turning in all assignments on time.
- Equally contributing to the project assignment and clearly presenting your project idea and results in the final presentation and research paper.

Missed assignments (including late assignments): You can skip submitting up to 3 pre-class reading summaries during the semester without any explanation or penalty. For all other submissions – like research paper, drafts, reviews – you can get 1-day extension for at most 2 deadlines without penalty. Excluding the two scenarios mentioned above, 20% of the total points will be deducted for late submissions and missed submissions won't be assigned any score. 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.