

Syllabus

CSCI6410/CSCI4148/EPAH6410

Applied Research in Health Data Science

Summer (A) 2024

COURSE INFORMATION

Instructor Information

Lecturer:	Finlay Maguire
E-mail:	finlay.maguire@dal.ca
Class Meeting Time:	MWF 1305-1455 in Mona Campbell 1201
Course Homepage:	https://maguire-lab.github.io/health_data_science_research_2024
Teaching Assistant:	David Mahoney (david.mahoney@dal.ca)
Office Hours:	Wednesday 1500-1600 (following class)

Important Dates

- Classes begin: **May 6th**
- Victoria Day - University Closed: **May 20th**
- Classes end: **June 17th**
- Submission Deadlines: 4 assignments due May 26th, June 2nd, June 9th, June 16th and project proposal due June 21st

Course Description

Health data science is a rapidly growing research field across academia, government, and industry. It relates to the application of statistical and machine learning approaches to analyse large complex medical datasets including electronic medical records, radiological imaging, physiological sensor data, and longitudinal databases. This course combines an overview of these key types of medical data, hands-on introduction to their principal analysis methods, and training in how to apply them in interdisciplinary research contexts. Using a combination of lectures, R-based exercises, student-driven tutorials, and collaborative development of a research proposal, students will gain the skills necessary to plan and conduct effective health data science research.

Course Learning Outcomes

The aim of this course is to provide students with the skills and knowledge required to plan effective research in the application of data science approaches to medical data. Specifically, by the end of the course students will:

1. Understand the 4 principal sources and data types of medical data: longitudinal databases (tabular), electronic medical records (structured, semi-structured, and unstructured text), radiological imaging (image), and physiological (signal and time-series).
2. Identify and apply appropriate type of method to the analysis of each data type
3. Gain the technical skills necessary for effective health data science research including data management, reproducibility, and version control.
4. Understand the key collaborative, legal, ethical, and knowledge translation concepts required in interdisciplinary health data science research.

5. Critically appraise research literature in health data science.
6. Combine these skills to develop high-quality collaborative health data science research proposals

Course Rationale

This course is designed as an elective course for graduate students in Community Health & Epidemiology and graduate/advanced undergraduate students in the Faculty of Computer Science interested in working at the intersection of medicine and computer science. The primary goals of this course are to (i) provide an overview of the main types of medical data, (ii) introduce key analysis methods for each data, and (iii) build skills necessary for effective interdisciplinary research in this area. This will complement existing non-cross listed/co-located CS and CH&E courses by providing students with an introduction to a wide-range of concepts across those courses or an opportunity to apply those skills within a growing interdisciplinary research context. Within FCS, it will complement existing courses focused on specific analysis methods (e.g., CSCI 6504/6505/6509/6515/6612) and research skills focused courses (e.g., CSCI 6055/6061). Similarly, within CH&E, it will complement technical-skill focused courses (e.g., CH&E 6054/6056) as well as research training (e.g., HINF 6020.03/CH&E 8040) by supporting specific training at their intersection.

Class Format

This class is formatted around a mixture of didactic lectures (typically Monday/Wednesday with day change occurring due to Victoria Day), assessed in-class practical exercises (typically Friday/Monday), and tutorials around primary literature (typically Wednesday/Friday). These tutorials will involve a rotation of students presenting papers and class-based discussion of the strengths/weaknesses and key methodological take-aways of the presented work. The main assessment output for the course will be a collaboratively developed research proposal supported by in-class time and proposal/research-skills related tutorials (typically Wednesday/Friday).

Minimal Technical Requirements

This course will require access to an internet-enabled computer capable of installing and running Rstudio (>1.0). Rstudio is freely available and can be install from here: <https://www.rstudio.com/products/rstudio/>

Course Pre-requisites, Co-requisites, Exclusions and/or other Restrictions

Students should either have some previous programming experience (ideally with R) and a knowledge of basic machine learning and/or statistical methods. This includes the ability to parse and manipulate tabular datasets, generate simple visualisations such as scatter plots, and apply basic statistical modelling methods such as regression. For graduate students this will be self-certified and relevant additional training material can be found in course readings. For undergraduate students pre-requisites of CSCI2110 and either STAT2060 or CSCI2360 are required.

Basic pre-course familiarisation with R is recommended for all students e.g., completion of the Harvard-Chan School Bioinformatics training module: <https://hbctraining.github.io/Training-modules/IntroR/>

Exclusions: CSCI6401, CSC4148, and EPAH6410 are mutually exclusive of one another. Having previously taken one of these co-located courses precludes taking any of the others.

Required Text(s)

- R for Data Science by Wickham & Grolemund (freely available from authors: <https://r4ds.had.co.nz/>)
- Hands on Machine learning for R by Boehmke & Greenwell (freely available from authors: <https://bradleyboehmke.github.io/HOML/>)
- Text Mining with R by Julia Slige & Davin Robinson (freely available from authors: <https://www.tidytextmining.com/>)

Course Schedule

(Subject to minor adjustment as the course proceeds)

Module	Date	Day	Topic	Format
1	May 6th	Mon	Reproducible Research	Lecture: Introduction to Health Data Science
	May 8th	Wed		Tutorial: Reproducibility in research review papers Proposal Class: Introduction to proposal and group formation
	May 10th	Fri		Practical 0: Use of rstudio, rmarkdown, git, and exploratory data analysis, (formative non-graded)
2	May 13th	Mon	Medical Databases	Lecture: Medical Databases (ontologies/codes, survey weights, manifold learning, boosted decision trees)
	May 15th	Wed		Tutorial: research papers using manifold learning and boosted decision trees with databases. Proposal Class: Considerations of accessing datasets
	May 17th	Fri		Practical 1: Using medical databases
	May 20th	Mon	University Closed	Victoria Day
3	May 22nd	Wed	Electronic Medical Records	Lecture: Electronic Medical Records (Structured, semi-structured, unstructured text, NLP, Language Models)
	May 24th	Fri		Tutorial: research papers using NLP and ontologies to analyse EMR text. Proposal Class: Proposing ethical research
	May 27th	Mon		Practical 2: NLP analysis of EMR text data
4	May 29th	Wed	Radiological Imaging	Lecture: Medical imaging, data formats, classical computer vision vs deep learning, extensions for video

	May 31st	Fri		Tutorial: pathology and radiology research papers with automated image/MRI analysis methods. Proposal Class: Intermediate check-in and troubleshooting.
	June 3rd	Mon		Practical 3: Predictive categorisation of radiological/pathological image data
5	June 5th	Wed	Physiological Signal Data	Lecture: Signal data and time-series, transformation, fourier transform, de-trending, LSTMs.
	June 7th	Fri		Tutorial: Examples of research literature using EEG and wearable monitor datasets Proposal: Effective knowledge translations plans
	June 10th	Mon		Practical 4: Predictive analysis of EEG data
6	June 12th	Wed	Proposal	Proposal Presentations
	June 14th	Fri		Proposal Presentations
	June 17th	Mon		Proposal Presentations

Course Assessments

Assessment	Date Due	Weight
Practical 1: Medical Databases	May 26th	10%
Practical 2: Electronic Medical Records	June 2nd	10%
Practical 3: Radiological Imaging	June 9th	10%
Practical 4: Physiological Time-Series	June 16th	10%
Proposal Presentation	June 12-17th	25%
Proposal Write-Up	June 21st	15%
Tutorial Paper Presentation	Throughout	10%
Participation	Throughout	10%

Grading: This course uses the standard Dalhousie grading scheme: (https://www.dal.ca/campus_life/academic-support/grades-and-student-records/grade-scale-and-definitions.html)

Note: for graduate students the minimum passing grade is B- so any mark <70% will be converted to an F

Submission: Assignments must be submitted via Brightspace and also include a link to a git repository containing the analyses. These can be hosted on github.com, gitlab.com, or Dalhousie's internal git.cs.dal.ca (requires a CSID).

Late Policy: Late assignments will be penalised at 20% per day. Assignments submitted more than 5 days late can still be evaluated for feedback but the final grade will be 0. Late submissions without penalty will be considered **only if** the appropriate channels (student declaration of absence, doctor's note) are used.

Undergraduate Criteria: will be assessed using the same criteria as graduate students with the exception of having their worst practical assignment score dropped (resulting in 13.33% weighting for remaining 3 practical assignments) and not having to provide answers to graduate-only questions in the practical assignments. Undergraduates can also receive a minimum passing grade of C- compared to B- for graduate students.

EVALUATION CRITERIA

Practical Assignments: each of the practical assignments draws on material from the related module of the course and applies those methods and to problem data sets. This will start with a worked step-by-step example data set (introduced and explained in the practical session by the instructor) with short knowledge and understanding questions. Following this, students will then complete a short report applying the full analysis approach to a new dataset. These assignments are to be completed during the practical session or outside of class time before the next practical (1 week's time) and must be submitted via Brightspace and including a link to a matching git repository (github.com, gitlab.com, or Dalhousie's internal git.cs.dal.ca). Collaboration is **not** permitted on these assignments.

Tutorials: students will be assigned a relevant paper from the health data science literature and will present the paper to the class and lead subsequent discussion and appraisal of the methods and results of the research. Presentation assignment will take place on the first day of class.

Component	Very Good to Excellent (4-5)	Good (3)	Fair to Poor (≤ 2)	Score
Presentation of Paper	Demonstrates good preparation and comprehension of the paper(s) presented. Able to critically appraise most of the material and identify key strengths and limitations.	Demonstrates adequate preparation and comprehension of the paper(s) presented. Offers straightforward information and some insight into key strengths and limitations.	Demonstrates minimal preparation and very limited comprehension of the paper(s) presented. Offers only straightforward information and has no understanding of strengths and limitations.	/10
Leading Discussion	Effectively led discussion of the paper(s) strengths and weaknesses and facilitated involvement of the whole class.	Facilitated adequate discussion of the paper(s) strengths and weaknesses but with more limited involvement of the class.	Led only minimal conversation of the paper(s) strengths and weaknesses involving very few members of the class.	/10
Total				/20

Participation: students will be expected to participate in the tutorial discussions and in-class (worth 10% of final grade).

Component	Very Good to Excellent	Good (3)	Fair to Poor (≤ 2)	Score
-----------	------------------------	----------	---------------------------	-------

	(4-5)			
Preparation	Arrives fully prepared.	Arrives mostly, though not full, prepared.	Arrives with limited preparation.	/5
Engagement	Actively supports, engages, and listens to peers.	Makes a sincere effort to interact with peers.	Limited interaction with peers.	/5
Initiative	Plays an active role in discussions or actively asks questions.	Participates constructively in discussions or lectures.	Follows the discussion groups or the class.	/5
Quality	Questions or comments reflect analysis, synthesis, and evaluation; level of discussion is consistently better because of the student's presence.	Questions or comments are relevant to the topic discussed; level of discussion is occasionally better because of the student's presence.	Questions or comments are generally vague or too far from the topic discussed; level of discussion is not affected by the student's presence.	/5
Total				/20

Research Proposal: students will be assigned to groups (if class size is limited then individual projects may be possible) in the first week of class and throughout the course will collaboratively develop a research proposal. There will be weekly class time dedicated to training in the specifics of developing research proposals and for group work on the proposal. However, the groups will be expected to coordinate with one another and work on the proposal outside of class time as well. The final reports will be presented by the groups to the rest of the class during the last week of the course with Q&A from the other course participants. This will be followed by submission of a group written report including a statement of contribution that clearly indicates each team member's contribution to the proposal. The follow rubric summarises how this will be graded:

Component	Very Good to Excellent 4-5/5 7-10/10	Good 3/5 4-6/10	Fair to Poor 0-2/5 0-3/10	Score
Teamwork	Contributions statement and presentation show signs of strong equitable teamwork. Clear indication of a high level of mutual respect and collaboration within the group.	Contributions statement and presentation show signs of generally strong and equitable teamwork. Indication of a good level of mutual respect and collaboration within the group.	Contributions statement and presentation show little indication of strong or equitable teamwork. Signs of low levels of mutual respect and collaboration within the group.	/5
Research Question	Engaging and clearly defined interdisciplinary research question that solves an important extant issue in the field of health data science.	Relatively well defined interdisciplinary research question that addresses a relatively important extant issue in the field of health data science.	Poorly defined question with limitations in scope, importance, or interdisciplinarity.	/5

Abstract	A concise, informative, and well-written summary of the proposal targeted to an expert audience.	A relatively well written summary of the proposal targeted to an expert audience.	A poorly written summary that only conveys limited details about the proposal.	/5
Lay Summary	A concise, well-written plain-language summary aimed at the general public summary which articulates the problem, proposed solution, and relevance in an accessible and clear manner.	A generally clear plain-language summary aimed at the general public summary which largely articulates the problem, proposed solution, and relevance in a mostly accessible way.	Poorly written plain-language summary that doesn't convey the problem, proposed solution, or relevance in an accessible manner.	/5
Introduction	A detailed and well-written technical overview that positions the research idea within the wider research and social context, articulates the knowledge gap the project will address, and introduces relevant methods	Generally detailed and well-written technical overview that positions the research idea within the wider research and social context, articulates the knowledge gap the project will address, and introduces relevant methods	A detailed and well-written technical overview that positions the research idea within the wider research and social context, articulates the knowledge gap the project will address, and introduces relevant methods	/10
Literature Review	Critically appraises and summarises a broad sampling of relevant literature that fully justifies the research question and proposed methodology.	Critically appraises and summarises a reasonably broad sampling of relevant literature that mostly justifies the research question and proposed methodology.	Reports a limited range of relevant literature with limited critical appraisal and doesn't justify the research question and proposed methodology.	/10
Methodology	Clearly communicates a reasonable, appropriate, and fully-developed methodology (including data gathering/access) that solves the research question and is well supported by the literature review.	Communicates a reasonable, appropriate, and relatively developed methodology (including data gathering/access) that solves the research question and is reasonably well supported by the literature review.	Limited methodology with significant gaps in, disconnection from the key research question, and limited literature support.	/10
Ethics	Thoughtfully and robustly explores the research ethics, potential obstacles/risks, and the risk/benefits of the research question, methodology, and impact of the proposed work.	Relatively strong exploration of the research ethics, potential obstacles/risks, and the risk/benefits of the research question, methodology, and impact of the proposed work.	Limited treatment of the ethics, potential obstacles/risks, and the risk/benefits of the research question, methodology, and impact of the proposed work.	/10
Discussion	Clearly addresses potential	Addresses most potential	Only addresses a small	/10

	limitations in the proposed project, identifies, explores potential future directions/extensions, and evaluates implications of a reasonable range of potential results.	limitations in the proposed project, identifies some future directions/extensions, and explores some implications of a reasonable range of results.	range of limitations, implications, and/or future research directions.	
Knowledge Translation	Develops a robust and impactful plan to effectively mobilise knowledge gained from the proposed research across a range of sectors/settings.	Develops a relatively complete plan to mobilise knowledge gained from the proposed research across a range of sectors/settings.	Develops a limited plan that doesn't successfully mobilise potential findings and/or does so across a limited range of sectors/settings.	/10
Formatting	Proposal is clearly formatted with consistent citations and use of section headings, figure legends, and appendices where appropriate.	Proposal is mostly clearly formatted with some minor formatting errors e.g., malformed citations, inconsistency, missing sections.	Poorly formatted proposal with numerous issues with use of sections, citations, and presentation of figures/tables.	/5
Presentation	Well crafted presentation, with good use of slides to clearly convey components of the proposal.	Reasonably clear presentation, that mostly conveys key components of the proposal.	Less well organised presentation that doesn't convey key components of the proposed work.	/10
Presentation: Q&A	Well-prepared team that comprehensively handles questions that arise about their work.	Well-prepared team that handles questions that arise about their work relatively well.	Relatively poor team-wide ability to answer questions about their proposed work.	/5
Total				/100

COURSE POLICIES

Academic Standards

Failure to properly attribute sources in your work will be treated as an academic standards issue and points may be deducted for not following citation requirements. For example, forgetting to quote text taken from other sources, failure to include in-text citations, or a failure to include required information in the citations or references. Please see the resources on proper citation provided by the Dalhousie Writing Center (<https://dal.ca/libguides.com/c.php?g=257176&p=5001261>).

Please note that if it appears that the error was made with intent to claim other people's work as your own such as a lack of both citations and references, an allegation of plagiarism will be submitted to the Faculty Academic Integrity Officer, which could result in consequences such as a course failure.

Responsible Computing Policy

Usage of all computing resources in the Faculty of Computer Science must be within the Dalhousie Acceptable Use Policies (https://www.dal.ca/dept/university_secretariat/policies/information-management-and-technology/acceptable-use-policy-.html) and the Faculty of Computer Science Responsible Computing Policy. For more information please see

https://www.dal.ca/content/dam/dalhousie/pdf/faculty/computerscience/policies-procedures/fcs_policy_local.pdf

Use of Plagiarism Detection Software

All submitted code may be passed through a plagiarism detection software, such as the plagiarism detector embedded in Codio, the Moss (<https://theory.stanford.edu/~aiken/moss/>) Software Similarity Detection System, or similar systems. If a student does not wish to have their assignments passed through plagiarism detection software, they should contact the instructor for an alternative. Please note, that code not passed through plagiarism detection software will necessarily receive closer scrutiny. https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf

Use of Artificial Intelligence Tools

You may use AI-driven tools to assist you in learning but remember that your objective is to understand, achieve, and apply the course competencies and outcomes. While you may use tools for learning, specific assessments in this course will disallow the use of AI-driven tools to assert that you have attained course learning outcomes. This is because a graduate must be able to analyze, assess and produce work unassisted by AI technology. Where tools are allowed: you must acknowledge all tools used to assist you. If applicable, you must provide links to chat logs. Using AI-driven tools where prohibited constitutes an academic offense.

Student Health and Wellness

Taking care of your health is important. As a Dalhousie student, you have access to a wide range of resources to support your health and wellbeing. Students looking to access physical or mental health & wellness services at Dalhousie can go to the Student Health & Wellness Centre in the LeMarchant Building. The team includes: registered nurses, doctors, counsellors and a social worker. Visit dal.ca/studenthealth to learn more and book an appointment today. Students also have access to a variety of online mental health resources, including telephone/texting counselling and workshops/training programs. Learn more and access these resources at dal.ca/mentalhealth.

Culture of Respect¹

Every person has a right to respect and safety. We believe inclusiveness is fundamental to education and learning. Misogyny and other disrespectful behaviour in our classrooms, on our campus, on social media, and in our community is unacceptable. As a community, we must stand for equality and hold ourselves to a higher standard.

What we all need to do:

1. **Be Ready to Act:** This starts with promising yourself to speak up to help prevent it from happening again. Whatever it takes, summon your courage to address the issue. Try to approach the issue with open-ended questions like “Why did you say that?” or “How did you develop that belief?”
2. **Identify the Behaviour:** Use reflective listening and avoid labeling, name-calling, or assigning blame to the person. Focus the conversation on the behaviour, not on the person. For example, “The comment you just made sounded racist, is that what you intended?” is a better approach than “You’re a racist if you make comments like that.”
3. **Appeal to Principles:** This can work well if the person is known to you, like a friend, sibling, or co-worker. For example, “I have always thought of you as a fair-minded person, so it shocks me when I hear you say something like that.”
4. **Set Limits:** You cannot control another person’s actions, but you can control what happens in your space. Do not be afraid to ask someone “Please do not tell racist jokes in my presence anymore” or state “This classroom is not a place where I allow homophobia to occur.” After you have set that expectation, make sure you consistently maintain it.
5. **Find or be an Ally:** Seek out like-minded people that support your views, and help support others in their challenges. Leading by example can be a powerful way to inspire others to do the same.
6. **Be Vigilant:** Change can happen slowly, but do not let this deter you. Stay prepared, keep speaking up, and do not let yourself be silenced.

¹ Source: Speak Up! © 2005 Southern Poverty Law Center. First Printing. This publication was produced by Teaching Tolerance, a project of the Southern Poverty Law Center. Full “Speak Up” document found at: <http://www.dal.ca/dept/dalrespect.html>. Revised by Susan Holmes from a document provided April 2015 by Lyndsay Anderson, Manager, Student Dispute Resolution, Dalhousie University, 902.494.4140, lyndsay.anderson@dal.ca www.dal.ca/think.

UNIVERSITY STATEMENTS

This course is governed by the academic rules and regulations set forth in the University Calendar and the Senate.

<https://academiccalendar.dal.ca/Catalog/ViewCatalog.aspx?pageid=viewcatalog&catalogid=117&loadusercredits=False>

Territorial Acknowledgement

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people. Dalhousie acknowledges the histories, contributions, and legacies of the African Nova Scotia people and communities who have been here for over 400 years.

Internationalization

At Dalhousie, 'thinking and acting globally' enhances the quality and impact of education, supporting learning that is "interdisciplinary, cross-cultural, global in reach, and orientated toward solving problems that extend across national borders." <https://www.dal.ca/about-dal/internationalization.html>

Academic Integrity

At Dalhousie University, we are guided in all of our work by the values of academic integrity: honesty, trust, fairness, responsibility and respect. As a student, you are required to demonstrate these values in all of the work you do. The University provides policies and procedures that every member of the university community is required to follow to ensure academic integrity. (read more: http://www.dal.ca/dept/university_secretariat/academic-integrity.html)

Accessibility

The Student Accessibility Centre is Dalhousie's centre of expertise for matters related to student accessibility and accommodation. If there are aspects of the design, instruction, and/or experiences within this course (online or in-person) that result in barriers to your inclusion please contact:

https://www.dal.ca/campus_life/academic-support/accessibility.html for all courses offered by Dalhousie with the exception of Truro.

Conduct in the Classroom – Culture of Respect

Substantial and constructive dialogue on challenging issues is an important part of academic inquiry and exchange. It requires willingness to listen and tolerance of opposing points of view. Consideration of individual differences and alternative viewpoints is required of all class members, towards each other, towards instructors, and towards guest speakers. While expressions of differing perspectives are welcome and encouraged, the words and language used should remain within acceptable bounds of civility and respect.

Diversity and Inclusion – Culture of Respect

Every person at Dalhousie has a right to be respected and safe. We believe inclusiveness is fundamental to education. We stand for equality. Dalhousie is strengthened in our diversity. We are a respectful and inclusive community. We are committed to being a place where everyone feels welcome and supported, which is why our Strategic Direction prioritizes fostering a culture of diversity and inclusiveness (Strategic Priority 5.2). (read more:

<http://www.dal.ca/cultureofrespect.html>)

Student Code of Conduct

Everyone at Dalhousie is expected to treat others with dignity and respect. The Code of Student Conduct allows Dalhousie to take disciplinary action if students don't follow this community expectation. When appropriate, violations of the code can be resolved in a reasonable and informal manner—perhaps through a restorative justice process. If an informal resolution can't be reached, or would be inappropriate, procedures exist for formal dispute resolution. (read more:

https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/Code%20of%20Student%20Conduct%20rev%20Sept%202021.pdf)

Fair Dealing Policy

The Dalhousie University Fair Dealing Policy provides guidance for the limited use of copyright protected material without the risk of infringement and without having to seek the permission of copyright owners. It is intended to provide a balance between the rights of creators and the rights of users at Dalhousie. (read more: https://www.dal.ca/dept/university_secretariat/policies/academic/fair-dealing-policy-.html)

Originality Checking Software

The course instructor may use Dalhousie's approved originality checking software and Google to check the originality of any work submitted for credit, in accordance with the Student Submission of Assignments and Use of Originality Checking Software Policy. Students are free, without penalty of grade, to choose an alternative method of attesting to the authenticity of their work, and must inform the instructor no later than the last day to add/drop classes of their intent to choose an alternate method. (read more: https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf)

Student Use of Course Materials

These course materials are designed for use as part of the CSCI courses at Dalhousie University and are the property of the instructor unless otherwise stated. Third party copyrighted materials (such as books, journal articles, music, videos, etc.) have either been licensed for use in this course or fall under an exception or limitation in Canadian Copyright law. Copying this course material for distribution (e.g. uploading material to a commercial third party website) may lead to a violation of Copyright law.

Learning and Support Resources

Please see https://www.dal.ca/campus_life/academic-support.html