University of Florida
College of Public Health & Health Professions Syllabus
Course Number: PHC 7083
Title: ***Computational Epidemiology in Population Science*** (2 credit hours)
Semester: Summer A; Year: 2020
(May 11 to June 19 2020; Tue from 11am – 1:45pm and from 2:00pm – 4:45pm)
Delivery Format: On-Campus
E-Learning: https://ufl.instructure.com/courses

Instructor Name: Dr. Mattia Prosperi
Room Number: 4234 (office)
Phone Number: 352-273-5860
Email Address: m.prosperi@ufl.edu
Office Hours: *On appointment* (Mon-Fri 9:00-5:00)
Canvas url: http://elearning.ufl.edu/;
Teaching Assistants: *TBA*
Preferred Course Communications: *e-mail*

Prerequisites PHC 6000 Epidemiology Research Methods I

PURPOSE AND OUTCOME

Course Overview. This course blends methodological, practical, and translational aspects of computational epidemiology. The course is not intended to provide statistical training, but rather to teach students to recognize suitable computational data mining approaches, and to become acquainted with machine learning software capable of processing big data.

Relation to Program Outcomes. This course covers new and interdisciplinary aspects of epidemiology in response to the growth of big data repositories, integrated multi-centric studies/cohorts, merges of heterogeneous data sources (genetics, clinical, surveillance, laboratory, sensors), and new data types (such as those from social media). This course brings our Epi program into the 'next-generation' era of epidemiological research and data science, in compliance to up-to-date accreditation standards, and with translational relevance to clinical and professional practice.

Course Objectives and/or Goals. This course has the objective to provide the students training in computational models expressively for epidemiology, aiming at enabling them to formulate and test appropriate modelling hypotheses for large scale and heterogeneous study designs. This training is valuable for a PhD student who is interested into the 'next-generation' data science of epidemiology and biomedical modelling. Upon successful completion of the course, students will be able to:

- Identify data characteristics and issues by design, e.g. how to query an electronic medical record data base.
- Formulate (multiple) modelling hypothesis, e.g. how to perform large-scale exploratory analysis on genomics data.
- Design data analysis plans, e.g. how to compare a linear regression prediction model with a decision tree or a neural network.
- Parallelize analyses to reduce complexity, e.g. how to prepare statistical scripts to be run on a computer cluster.
- Interpret prediction models and generalize findings, e.g. understand the actionable variables in a risk score and identify strategies to reduce risks.

Knowledge-based goals according to Bloom's taxonomy of educational objectives:

1. *Knowledge.* Recognition of computational learning techniques and health informatics terms/procedures, e.g. "What is a decision tree?"
2. *Comprehension.* Ability to extrapolate the functional value of computational models, e.g. "Is a decision tree a nonlinear classifier?"

3. *Application.* Ability to use a computational method in a specific context, e.g. “Can you extract data from the health record system following the study design and fit a random forest on the study outcome using a parallelized statistical software library?”
4. *Analysis.* Ability to test hypotheses using the data and different modelling approaches, e.g. “What happens if the prior probability is changed when we fit a Bayesian model?”
5. *Synthesis.* Ability to combine different computational models on the basis of a problem of interest, e.g. “For this problem, for which we found evidence of nonlinearity, we decide to use the alternating decision tree model to achieve better prediction performance yet maintaining an acceptable level of interpretability.”
6. *Evaluation.* Ability to formulate new evidence-based research questions; ability to evaluate the generalizability and translational importance of findings, e.g. “What information domains shall we look at when designing a prediction system for dengue infection risk in Haiti, and what would be the impact of a prediction model that includes modifiable variables for which an intervention could reduce incidence?”

Instructional Methods. Face-to-face lectures (using PPT/PDF presentations) divided in three parts: 1) methodological introduction, 2) applied/translational exemplification, 3) students’ feedback (Q&A, reflections/commentaries, et cetera); homework; usage of printed/PDF textbook(s) and critical reading of scientific papers. Teaching material will be posted online. All course slides will be made available online for download, some accompanied by audio recording of what it has been taught in the class (up to 5 hrs). The online material (including this syllabus) will be processed through SensusAccess according to Federal, State and University’s accessibility policies and governance.

DESCRIPTION OF COURSE CONTENT

Topical Outline/Course Schedule

The course is divided into 12 classes of 75 minutes each, for a total of 900 minutes. The lessons are accompanied by PPT/PDF slides and/or research papers. Usually the last part of the lecture classes is dedicated to interactive questions & answers and discussion on the topics presented, whilst the last part of the practice sessions let the students use the programs by themselves after following the instructor’s tutorial.

Time frame for classes: Summer A, from May 12 to June 16 2020, once a week. Based on current class availability, two classes will be given every Tuesday 11:00am-1:45pm, with ~15 minutes pause in between; the software practice session will be carried out on the same day in the afternoon (2:00pm-4:45pm).

- Week 1
  - Lecture 1 - Precision medicine: molecular medicine & multi-domain inference in healthcare.
  - Lecture 2 - Genetic determinants of diseases and disorders.
  - Practice - Introduction to software suites Weka and R.
  - Readings:
    - Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics. 2015 Jun 27;8:33.
    - Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012 Jan 13;90(1):7-24.
- Week 2
  - Lecture 1 - Eagle’s view on computational epidemiology and biomedical modelling.
  - Lecture 2 - Identifying diagnostic pathways: a study on age-related macular degeneration.
  - Practice - Linear regression models in R (set up of link functions, AIC/BIC stepwise selection, LASSO, packages glm, MASS, glmnet).
  - Readings:
    - Borman N, Gillblad, D. Learning machines for computational epidemiology. IEEE International Conference on Big Data, 27-30 Oct. 2014, Washington, DC, pp. 1-5.
    - Fraccaro P, Nicolo M, Bonetto M, Giacomini M, Weller P, Traverso CE, Prosperi M, OSullivan D. Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach. BMC Ophthalmol. 2015 Jan 27;15:10.
- Week 3
  - Assignment #1 due
  - Lecture 1 - Tracing outbreaks and predicting their dynamics trees.
  - Lecture 2 - Tracing outbreaks and predicting their dynamics: networks and agents.
  - Practice - Decision trees and random forests in R (packages rpart, randomForest, party).

    - Readings:
        - Pybus OG, Tatem AJ, Lemey P. Virus evolution and transmission in an ever more connected world. Proc Biol Sci. 2015 Dec 22;282(1821).
        - Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prosperi M. Phylodynamic applications in 21 st century global infectious disease research. Global Health Research and Policy 2017; 2:13. DOI: 10.1186/s41256-017-0034-y
        - Patlolla P, Gunupudi V, Mikler AR, Jacob Rt. Agent-Based Simulation Tools in Computational Epidemiology. Innovative Internet Community Systems. Volume 3473 of the series Lecture Notes in Computer Science pp 212-223.
- Week 4
    - Combined Lecture 1&2 - Computational Epidemiology Journal discussion.
    - Practice - Machine learning methods in Weka.
    - Readings: Paper(s) agreed for the journal discussion.
- Week 5
    - Assignment #2 due
    - Lecture 1 - Practical insights on big data in healthcare, electronic medical records, ontologies, and big –omics data.
    - Lecture 2 - Critical review of methods and findings: Discussion of assignment #1.
    - Practice - Cross-validation and bootstrap in Weka/R, performance functions (e.g. sensitivity, specificity, AUROC).
    - Readings:
        - Pesquita C, Ferreira JD, Couto FM and Silva MJ. The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources. Journal of Biomedical Semantics20145:4. DOI: 10.1186/2041-1480-5-4.
- Week 6
    - Lecture 1 - Conclusive remarks (research integrity, ethics).
    - Final assessment (exam)
    - Readings: None.

Note: Journal discussions are chaired by a volunteer or randomly selected student (or two students if they wish to do it as a collaboration) who will make a short presentation (10-15 minutes) illustrating the content of the article of choice and posing arguments and questions to be discussed in the following roundtable session with all the other students. The article will be circulated a few days prior the class so that the other students can read it. The teacher will facilitate the session and will encourage students to have a critical group discussion. A good presentation may result in a 1-2 bonus points for the final grading. No penalty points will incur in any case.

Course Materials and Technology

Course slides. Provided by the teacher and posted online.

Textbook(s): None mandatory (the course material and the suggested/assigned papers will be sufficient). Students may ask the teacher for a selection of interesting textbooks in computational epidemiology.

List of journal papers for Computational Epidemiology journal discussion:

- Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics. 2015 Jun 27;8:33. (for week 1)
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012 Jan 13;90(1):7-24. (for week 1)
- Fraccaro P, Nicolo M, Bonetto M, Giacomini M, Weller P, Traverso CE, Prosperi M, OSullivan D. Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach. BMC Ophthalmol. 2015 Jan 27;15:10. (for week 2)
- Borman N, Gillblad, D. Learning machines for computational epidemiology. IEEE International Conference on Big Data, 27-30 Oct. 2014, Washington, DC, pp. 1-5. (for week 2)
- Pybus OG, Tatem AJ, Lemey P. Virus evolution and transmission in an ever more connected world. Proc Biol Sci. 2015 Dec 22;282(1821). (for week 3)
- Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prosperi M. Phylodynamic applications in 21 st century global infectious disease research. Global Health Research and Policy 2017; 2:13. DOI: 10.1186/s41256-017-0034-y (for week 3)
- Patlolla P, Gunupudi V, Mikler AR, Jacob Rt. Agent-Based Simulation Tools in Computational Epidemiology. Innovative Internet Community Systems. Volume 3473 of the series Lecture Notes in Computer Science pp 212-223. (for week 3)

- Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. Journal Of Big Data 2014; 1:2. DOI: 10.1186/2196-1115-1-2 (for week 5)
- Pesquita C, Ferreira JD, Couto FM and Silva MJ. The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources. Journal of Biomedical Semantics20145:4. DOI: 10.1186/2041-1480-5-4. (for week 5)
- For assignments, discussion classes and deepening:
    - Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018 Apr 27. doi: 10.1093/jamia/ocy032.
    - Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. Eur J Clin Invest. 2015 Feb;45(2):204-14.
    - Ferrucci D, Levas A, Bagchi S, Gondek D, Mueller ET. Watson: Beyond Jeopardy! Artificial Intelligence 2013 199-200:93-105.
    - Su C, Andrew A, Karagas MR, Borsuk ME. Using Bayesian networks to discover relations between genes, environment, and disease. BioData Min. 2013 Mar 21;6(1):6.
    - Marshall BD, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. Am J Epidemiol. 2015 Jan 15;181(2):92-9.
    - Binder H, Blettner M. Big data in medical science -a biostatistical view. Dtsch Arztebl Int. 2015 Feb 27;112(9):137-42;
    - Hauser J, Rybakowski J. Three clusters of male alcoholics. Drug Alcohol Depend. 1997 Dec 15;48(3):243-50;
    - Dasgupta A, Sun YV, König IR, Bailey-Wilson JE, Malley JD. Brief Review of Regression-Based and Machine Learning Methods in Genetic Epidemiology: The Genetic Analysis Workshop 17 Experience. Genet Epidemiol. 2011; 35(Suppl 1): S5–11.
    - Mate S, Köpcke F, Toddenroth D, Martin M, Prokosch H-U, Bürkle T, et al. (2015) Ontology-Based Data Integration between Clinical and Research Systems. PLoS ONE 10(1): e0116656. https://doi.org/10.1371/journal.pone.0116656
    - And a good selection here: https://compepi.cs.uiowa.edu/index.php/Publications/Main

Students need to install R (https://www.r-project.org/) and Weka (https://www.cs.waikato.ac.nz/ml/weka/) on their laptop for the practice sessions.

For technical support for this class, please contact the UF Help Desk at:

- Learning-support@ufl.edu
- (352) 392-HELP - select option 2
- https://lss.at.ufl.edu/help.shtml

ACADEMIC REQUIREMENTS AND GRADING

Homework

Students are supposed to review course material as suggested by the teacher. The teacher will clearly explain the pathway to acquire the knowledge, develop critical understanding, and explain the requirements for the Assignment(s)/Exam. The reading of all suggested papers (besides those assigned) is not enforced: a student may prefer to study on other texts or educational media if this facilitated their study.

Assignments

Assignments:

1. (Theory) Conduct a critical review, in writing of no more than two pages (normal page margins, 1.5 line spacing, Palatino Linotype font 11pt), of a scientific paper which addresses computational modelling and large data in epidemiology, population science, public health, or biomedical sciences. The instructor can provide a list of papers to review or the student can propose one. This assignment counts for 25% of the total; points' scale is in 30 units (see Grading). Submission is exclusively via e-mail to the instructor.
2. (Practice) Write and execute properly a machine learning pipeline in R or Weka. The pipeline is intended to be an implementation of the analytics part upon a study design: data are assumed to be already formatted, and the student has to concentrate on the correct pipelining of commands to

perform data analysis. Examples include model selection through bootstrap or cross validation, parameter optimization, external validation, comparison of different techniques and ranking by complexity, model ensemble, et cetera. The delivery is either by e-mail as a fully working script, including a brief report, or in person during one of the practice sessions. This assignment counts for 25% of the total; points' scale is in 30 units (see Grading).

Exam

Five questions, six points each. There will be two or three technical/methodological questions and two or three applied/discussion questions. The instructor will give examples of mock-up exams during classes to prepare the students.

Exam location(s)/dates(s)/times(s): *Same location of course lectures, same time; exam duration will be approximately 50 minutes.*

Grading

Requirement	Due date	Points or % of final grade (% must sum to 100%)
Assignment #1 (theory)	May 26 2020	25%
Assignment #2 (practice)	June 9 2020	25%
Exam	June 16 2019	30%
Attendance and participation in lectures and practical sessions	N/A	20%

Point system used (i.e., how do course points translate into letter grades): Assignment #1 , Assignment #2 and the final exam will be assigned up to 30 points each. The final points-grade will be calculated as the weighted average of the assignments and the exam, i.e. $0.25*A_1+0.25*A_2+0.3*E+0.2*P$ where A_1 and A_2 are the points received for the Assignments,*E* are the points earned in the Final Exam, and P are the points earned by attending and participating in the lectures and the practical sessions. The final letter grade will be obtained by converting the final points-grade in accordance to the table below. Decimals will be rounded to the nearest integer.

Points Earned	**30-29**	**28-27**	**26-25**	**24-23**	**22-21**	**20-19**	**18-17**	**16-15**	**14-13**	**12-11**	**10-9**	**9-0**
Letter Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	E

Please be aware that a C- is not an acceptable grade for graduate students. The GPA for graduate students must be 3.0. in all 5000 level courses and above to graduate. A grade of C counts toward a graduate degree only if a sufficient number of credits in courses numbered 5000 or higher have been earned with a B+ or higher. In addition, the Bachelor of Health Science Program does not use C- grades.

Letter Grade	**A**	**A-**	**B+**	**B**	**B-**	**C+**	**C**	**C-**	**D+**	**D**	**D-**	**E**	**WF**	**I**	**NG**	**S-U**
Grade Points	4.0	3.67	3.33	3.0	2.67	2.33	2.0	1.67	1.33	1.0	0.67	0.0	0.0	0.0	0.0	0.0

For greater detail on the meaning of letter grades and university policies related to them, see the Registrar's Grade Policy regulations at: http://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx

Exam Policy

Written exam. The student may not bring any didactic material, nor have an open computer or phone on their desk. The student must write on paper sheets that the instructor will provide. Five questions, six points each. The exam will cover material from the didactic and the practical sessions. There will be two or three technical/methodological questions and two or three applied/discussion questions. The instructor will give examples of mock-up exams during classes to prepare the students.

Exam location(s)/dates(s)/times(s): *Same location of course lectures, same time; exam duration will be approximately 50 minutes.*

Policy Related to Make up Exams or Other Work

Any requests for make-ups due to technical issues MUST be accompanied by the ticket number received from LSS when the problem was reported to them. The ticket number will document the time and date of the problem. You MUST e-mail me within 24 hours of the technical difficulty if you wish to request a make-up.

Late submissions will not be generally accepted. Exceptions will be handled on an individual basis and may result in a 2 points penalty.

Policy Related to Required Class Attendance

Please note all faculty are bound by the UF policy for excused absences. For information regarding the UF Attendance Policy see the Registrar website for additional details:
https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx
Attendance will be verified by signature sheets available during the class. Excused absences can be reported to the instructor prior to or on the day of absence, not later. Absence is defined as not showing up at class or being late more than 15 minutes.

Policy Related to Guests Attending Class

Only registered students are permitted to attend class. However, we recognize that students who are caretakers may face occasional unexpected challenges creating attendance barriers. Therefore, by exception, a department chair or his or her designee (e.g., instructors) may grant a student permission to bring a guest(s) for a total of two class sessions per semester. This is two sessions total across all courses. No further extensions will be granted. Please note that guests are not permitted to attend either cadaver or wet labs. Students are responsible for course material regardless of attendance. For additional information, please review the Classroom Guests of Students policy in its entirety. Link to full policy:
http://facstaff.phhp.ufl.edu/services/resourceguide/getstarted.htm

STUDENT EXPECTATIONS, ROLES, AND OPPORTUNITIES FOR INPUT

Expectations Regarding Course Behavior

Keep cell phones silenced in class. Ask permission to teacher and students for using audio/video/image recording devices. Be educated and polite.

Communication Guidelines

Follow netiquette for online communications http://teach.ufl.edu/wp-content/uploads/2012/08/NetiquetteGuideforOnlineCourses.pdf

Academic Integrity

Students are expected to act in accordance with the University of Florida policy on academic integrity. As a student at the University of Florida, you have committed yourself to uphold the Honor Code, which includes the following pledge:

"**We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity**."

You are expected to exhibit behavior consistent with this commitment to the UF academic community, and on all work submitted for credit at the University of Florida, the following pledge is either required or implied:

"On my honor, I have neither given nor received unauthorized aid in doing this assignment."

It is your individual responsibility to know and comply with all university policies and procedures regarding academic integrity and the Student Honor Code. Violations of the Honor Code at the University of Florida will not be tolerated. Violations will be reported to the Dean of Students Office for consideration of disciplinary action. For additional information regarding Academic Integrity, please see Student Conduct and Honor Code or the Graduate Student Website for additional details:
https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/
http://gradschool.ufl.edu/students/introduction.html

Please remember cheating, lying, misrepresentation, or plagiarism in any form is unacceptable and inexcusable behavior.

Online Faculty Course Evaluation Process
Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at https://evaluations.ufl.edu. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at https://evaluations.ufl.edu/results/.
Students' evaluations are important (e.g. for assessing instructors' performance on their teaching activities in relation to promotion) and they will be used to improve the course topics, materials, assignments, exams, and the instructor's teaching style.

COUNSELING AND STUDENT HEALTH

Students sometimes experience stress from academic expectations and/or personal and interpersonal issues that may interfere with their academic performance. If you find yourself facing issues that have the potential to or are already negatively affecting your coursework, you are encouraged to talk with an instructor and/or seek help through University resources available to you.

- The Counseling and Wellness Center 352-392-1575 offers a variety of support services such as psychological assessment and intervention and assistance for math and test anxiety. Visit their web site for more information: http://www.counseling.ufl.edu. On line and in person assistance is available.
- You Matter We Care website: http://www.umatter.ufl.edu/. If you are feeling overwhelmed or stressed, you can reach out for help through the You Matter We Care website, which is staffed by Dean of Students and Counseling Center personnel.
- The Student Health Care Center at Shands is a satellite clinic of the main Student Health Care Center located on Fletcher Drive on campus. Student Health at Shands offers a variety of clinical services. The clinic is located on the second floor of the Dental Tower in the Health Science Center. For more information, contact the clinic at 392-0627 or check out the web site at: https://shcc.ufl.edu/
- Crisis intervention is always available 24/7 from:
  Alachua County Crisis Center:
  (352) 264-6789
  http://www.alachuacounty.us/DEPTS/CSS/CRISISCENTER/Pages/CrisisCenter.aspx

Do not wait until you reach a crisis to come in and talk with us. We have helped many students through stressful situations impacting their academic performance. You are not alone so do not be afraid to ask for assistance.

Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

SUPPORT SERVICES

Accommodations for Students with Disabilities
If you require classroom accommodation because of a disability, it is strongly recommended you register with the Dean of Students Office http://www.dso.ufl.edu within the first week of class or as soon as you believe you might be eligible for accommodations. The Dean of Students Office will provide documentation of accommodations to you, which you must then give to me as the instructor of the course to receive accommodations. Please do this as soon as possible after you receive the letter. Students with disabilities should follow this procedure as early as possible in the semester. The College is committed to providing reasonable accommodations to assist students in their coursework.