

University of Florida
College of Public Health & Health Professions Syllabus
PHC 7083: Computational Data Science for Epidemiology (2 credit hours)

Semester: A 2023
 Delivery Format: On-Campus (HyFlex option)

Instructor Names:	Simone Marini, Mattia Prospero
Room Number:	HPNP G-301A
Phone Number:	3522948951
Email Address:	{simone.marini, m.prosperi}@ufl.edu
Office Hours:	On appointment (Mon-Fri, 9:00am-5:00pm)
Teaching Assistants:	TBA
Preferred Course Communications:	email

Prerequisites PHC 6000 Epidemiology Research Methods I

PURPOSE AND OUTCOME

Course Overview. This course blends methodological, practical, and translational aspects of computational epidemiology with emphasis on new data science methods. 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 data science expressively for epidemiology, aiming at enabling them to formulate and test appropriate modeling 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 modeling. 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 database
- Formulate (multiple) modeling 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:

Knowledge. Recognition of computational learning techniques and health informatics terms/procedures, e.g. “What is a decision tree?”

Comprehension. Ability to extrapolate the functional value of computational models, e.g. “Is a decision tree a nonlinear classifier?”

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?”

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?”

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.”

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); usage of printed/PDF textbook(s) and critical reading of scientific papers, including critical paper discussion among students. Teaching material will be posted online. All course slides will be made available online for download. 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, once a week (based on current class availability, two classes will be given every Tuesday starting on May 16 (until June 20) from 10am. The software practice session will be carried out on the same day in the afternoon (1:30pm).

Week 1

Lecture 1 – Principles of data science for epidemiology: beyond prediction toward intervention with an eye on causality.

Lecture 2 - Genetic determinants of diseases and disorders.

Practice - Introduction to data formatting and machine learning software suites, e.g., Weka, Orange, R and Python machine learning frameworks.

Readings:

Prosperi, M., Min, J.S., Bian, J. et al. Big data hurdles in precision medicine and precision public health. *BMC Med Inform Decis Mak* 18, 139 (2018) doi:10.1186/s12911-018-0719-2.

Whalen, Sean, et al. "Navigating the pitfalls of applying machine learning in genomics." *Nature Reviews Genetics* 23.3 (2022): 169-181.

Uitterlinden AG. An Introduction to Genome-Wide Association Studies: GWAS for Dummies.. *Semin Reprod Med*. 2016. PMID: 27513020.

MacEachern, Sarah J., and Nils D. Forkert. "Machine learning for precision medicine." *Genome* 64.4 (2021): 416-425.

Friedman NP, Banich MT, Keller MC. Twin studies to GWAS: there and back again. *Trends in cognitive sciences*. 2021 Oct 1;25(10):855-69.

Week 2

Lecture 1 - Data science and machine learning methods for epidemiology I.

Lecture 2 - Data science and machine learning methods for epidemiology II.

Practice – A primer on prediction model design, implementation, and evaluation

Readings:

Qayyum, Adnan, et al. "Secure and robust machine learning for healthcare: A survey." *IEEE Reviews in Biomedical Engineering* 14 (2020): 156-180.

Zhang, Angela, et al. "Shifting machine learning for healthcare from development to deployment and from models to data." *Nature Biomedical Engineering* (2022): 1-16.

Badillo, Solveig, et al. "An introduction to machine learning." *Clinical pharmacology & therapeutics* 107.4 (2020): 871-885.

Week 3

Lecture 1 & 2 – Molecular epidemiology: tracing outbreaks and their dynamics (phylogenetics/phylogenomics, equations, and simulations).

Practice – Feature extraction, clustering, and classification from scratch

Readings:

Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prospero 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.

Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*. 2004 Jan 16;303(5656):327-32. doi: 10.1126/science.1090727.

Week 4

Lecture 1 & 2 - Deep learning and explainable AI.

Practice – Other machine learning methods, e.g., support vector machines, (deep) neural networks.

Readings: Papers agreed for the journal discussion.

Week 5

Lecture 1 & 2 – Journal discussions (student presentations).

Practice – General Machine Learning practice.

Readings:

Week 6

Final assessment (exam).

Readings:

Prospero, M., Bian, J. Is it time to rethink institutional review boards for the era of big data?. *Nat Mach Intell* 1, 260 (2019) doi:10.1038/s42256-019-0059-7.

Course Materials and Technology. Course slides will be provided by the teachers 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. Recommended journal papers are listed in the previous section, already divided by topic (lecture). Students might need to install R <https://www.r-project.org/> (links to an external site), Orange <https://orangedatamining.com/> (links to an external site), and Weka <https://www.cs.waikato.ac.nz/ml/weka/> (links to an external site) on their laptop for the practice sessions. All software is free.

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

- helpdesk@ufl.edu
- (352) 392-HELP - select option 2
- <https://helpdesk.ufl.edu/>

Additional Academic Resources

[Career Connections Center](#): Reitz Union Suite 1300, 352-392-1601. Career assistance and counseling services.

Library Support: Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center: Broward Hall, 352-392-2010 or to make an appointment 352- 392-6420. General study skills and tutoring.

Writing Studio: 2215 Turlington Hall, 352-846-1138. Help brainstorming, formatting, and writing papers.

Student Complaints On-Campus: [Visit the Student Honor Code and Student Conduct Code webpage for more information.](#)

On-Line Students Complaints: [View the Distance Learning Student Complaint Process.](#)

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 facilitates their study.

Assignments. Students will present to the class a critical review of a scientific paper addressing computational modelling and large data in epidemiology, population science, public health, or biomedical sciences. While the instructors will provide a list of papers to review, students are encouraged to propose papers they find interesting. Papers needs to be approved by the instructors in advance. The presentation counts for 40% of the total final mark; points' scale is in grade letters (see Grading). Comments and questions to the presentations of other students are highly encouraged.

Grading

Requirement	Due date	Points or % of final grade (% must sum to 100%)
Presentation	05/13	40%
Exam	05/20	40%
Attendance and participation	N/A	20%

Point system used (i.e., how do course points translate into letter grades). The presentations and the final exam will be assigned a grade each (see table below). The final points-grade will be calculated as the weighted average of the assignments and the exam, i.e. $0.4 * \text{Presentation} + 0.4 * \text{Exame} + 0.2 * \text{Presence}$. 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.

Grade letters and grade points

LetterGrade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	E
	WF	I	NG	S-U								
GradePoints	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.								

Please be aware that a C- is not an acceptable grade for graduate students. The GPA for graduate students must be 3.0 based on 5000 level courses and above to graduate. A grade of C counts toward a graduate degree only if based on credits in courses numbered 5000 or higher that have been earned with a B+ or higher.

More information on UF grading policy may be found at:

<http://gradcatalog.ufl.edu/content.php?catoid=10&navoid=2020#grades>

Exam Policy. 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. There will be 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.

Tentative exam location(s)/dates(s)/times(s): HPNP G-301A, Tue 06/20, 10:00 am. Exam duration will be approximately 50 minutes. Note: the exact location, date, and time of the exam might vary to accommodate the students' schedule (other exams, conference attendance, etc.).

Policy Related to Make up Exams or Other Work. Please note: Any requests for make-ups due to technical issues MUST be accompanied by the UF Computing help desk (<http://helpdesk.ufl.edu/>) correspondence. You MUST e-mail me within 24 hours of the technical difficulty if you wish to request a make-up.

Policy Related to Required Class Attendance. Excused absences must be consistent with university policies in the Graduate Catalog (<https://catalog.ufl.edu/graduate/regulations/#text>). Additional information can be found here: <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.php.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.

Recording Within the Course:

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A “class lecture” is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To “publish” means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

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: <https://phhp.ufl.edu/policy-classroom-guests-of-students/>

Online Faculty Course Evaluation Process

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.

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.

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.
- **U 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>
- **University Police Department:** [Visit UF Police Department website](#) or call 352-392-1111 (or 9-1-1 for emergencies).
- **UF Health Shands Emergency Room / Trauma Center:** For immediate medical care call 352-733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville, FL 32608; [Visit the UF Health Emergency Room and Trauma Center website.](#)

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.

Inclusive Learning Environment

Public health and health professions are based on the belief in human dignity and on respect for the individual. As we share our personal beliefs inside or outside of the classroom, it is always with the understanding that we value and respect diversity of background, experience, and opinion, where every individual feels valued. We believe in, and promote, openness and tolerance of differences in ethnicity and culture, and we respect differing personal, spiritual, religious and political values. We further believe that celebrating such diversity enriches the quality of the educational experiences we provide our students and enhances our own personal and professional relationships. We embrace The University of Florida's Non-Discrimination Policy, which reads, "The University shall actively promote equal opportunity policies and practices conforming to laws against discrimination. The University is committed to non-discrimination with respect to race, creed, color, religion, age, disability, sex, sexual orientation, gender identity and expression, marital status, national origin, political opinions or affiliations, genetic information and veteran status as protected under the Vietnam Era Veterans' Readjustment Assistance Act." If you have questions or concerns about your rights and responsibilities for inclusive learning environment, please see your instructor or refer to the Office of Multicultural & Diversity Affairs website: www.multicultural.ufl.edu