BMI 5553-0010 – Autumn 2024
Predictive Analytics in Electronic Health Records

Time/Place: Asynchronous Online Learning; OSU CarmenCanvas
Course Director: Dr. Ping Zhang; 310-G LT; zhang.10631@osu.edu; http://pingzhang.net
Instructors:
- Dr. Ping Zhang; 310-G LT; zhang.10631@osu.edu; http://pingzhang.net
- Ms. Ruoqi Liu; 310-45 LT; liu.7324@osu.edu; http://ruoqi-liu.github.io/
- Ms. Seungyeon Lee; 310-43 LT; lee.10029@osu.edu; https://yeon-lab.github.io/
- Mr. Changchang Yin; 310-53 LT; yin.731@osu.edu; https://yinchangchang.github.io/
- Mr. Thai-Hoang Pham; 310-47 LT; pham.375@osu.edu; https://pth1993.github.io/

Level and credits: G, 3
Office Hours: Tue 10:00 AM – 11:00 AM; Thu 2:00 PM – 3:00 PM (In-person and/or via Zoom)
Zoom Link: https://osu.zoom.us/j/94052044513?pwd=XDeAmKUgAyDfievmUgZqaPbl1YCba3.1

Course Description:
Electronic health records (EHRs) provide massive, longitudinal data on thousands or even millions of patients, including demographics, laboratory results, diagnosis codes, prescriptions, and physician notes. As EHR databases are becoming more standardized and integrated across multiple hospital systems, they are gaining increasing attention from the informatics community as a resource to be mined, for example, to assess the quality of patient care, develop early prediction models for disease, and define disease phenotypes.

   The course introduces trainees to core data mining/machine learning algorithms, and their applications on predictive analytics on EHRs. Examples of major topics to be discussed in the class include 1) EHR data and data preprocessing; 2) classification algorithms and their applications to prognosis, risk stratification, and disease management; 3) clustering algorithms and their applications to disease progression modeling and patient subtyping; 4) Frequent pattern mining algorithms and their applications to EHR data visualization.

Learning Objectives:
Upon successful completion of the course, students will be able to:
- Recognize the benefits of using AI for EHR predictive analytics
- Gain knowledge of the most popular data mining/machine learning algorithms
- Select the best data mining/machine learning algorithms for an EHR prediction task
- Discuss current challenges and opportunities of AI in predictive analytics on EHRs

Prerequisites: Programming for Biomedical Informatics (BMI 5780), or permission of instructor

Textbooks:

- (Required) Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to data Mining (1st or 2nd ed), 2006
- (Required) Eric Topol, Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, 2019
- (Optional) Jiawei Han, Micheline Kamber, and Jian Pei, Data Mining: Concepts and Techniques (3rd ed), 2011
- (Optional) Charu C. Aggarwal, Data Mining: The Textbook, Springer, 2015.
- (Optional) Mohammed J. Zaki and Wagner Meira Jr., Data Mining and Analysis: Fundamental Concepts and Algorithms, 2014.

Software:
- (Required) Python
- (Required) NumPy, SciPy, Scikit-learn
- (Optional) PyTorch, Keras, MXNet

Grading Plan:
- Assignment 1: 25%
  - EHR data exploration and data preprocessing
- Assignment 2: 25%
  - Classification algorithms and mortality risk prediction
- Assignment 3: 25%
  - Clustering algorithms and patient subphenotyping
- Assignment 4: 25%
  - EHR data visualization

Health and safety requirements

All students, faculty and staff are required to comply with and stay up to date on all university safety and health guidance (https://safeandhealthy.osu.edu), which includes wearing a face mask in any indoor space and maintaining a safe physical distance at all times. Non-compliance will be warned first and disciplinary actions will be taken for repeated offenses.

Carmen access

You will need to use BuckeyePass multi-factor authentication to access your courses in Carmen. To ensure that you are able to connect to Carmen at all times, it is recommended that you take the following steps:
- Register multiple devices in case something happens to your primary device. Visit the BuckeyePass - Adding a Device help article for step-by-step instructions.
- Request passcodes to keep as a backup authentication option. When you see the Duo login screen on your computer, click **Enter a Passcode** and then click the **Text me new codes** button that appears. This will text you ten passcodes good for 365 days that can each be used once.
- Download the Duo Mobile application to all of your registered devices for the ability to generate one-time codes in the event that you lose cell, data, or Wi-Fi service.

- If none of these options will meet the needs of your situation, you can contact the IT Service Desk at 614-688-4357 (HELP) and IT support staff will work out a solution with you.

Course Technology Resources

For help with your password, university email, Carmen, or any other technology issues, questions, or requests, contact the Ohio State IT Service Desk. Standard support hours are available at ocio.osu.edu/help/hours, and support for urgent issues is available 24/7.

- **Self-Service and Chat support:** ocio.osu.edu/help
- **Phone:** 614-688-4357(HELP)
- **Email:** servicedesk@osu.edu
- **TDD:** 614-688-8743

Baseline technical skills for online courses

- Basic computer and web-browsing skills
- Navigating Carmen: for questions about specific functionality, see the Canvas Student Guide.

Required Technology skills specific to this course

- CarmenZoom virtrual meetings

Required equipment

- Computer: current Mac (OS X) or PC (Windows 7+) with high-speed internet connection
- Webcam: built-in or external webcam, fully installed and tested
- Microphone: built-in laptop or tablet mic or external microphone
- Other: a mobile device (smartphone or tablet) or landline to use for BuckeyePass authentication

Required software

- Microsoft Office 365: All Ohio State students are now eligible for free Microsoft Office 365 ProPlus through Microsoft's Student Advantage program. Full instructions for downloading and installation can be found at go.osu.edu/office365help.

Copyright disclaimer

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Ohio State Academic Integrity Policy

Academic integrity is essential to maintaining an environment that fosters excellence in teaching, research, and other educational and scholarly activities. Thus, The Ohio State University and the Committee on Academic Misconduct (COAM) expect that all students have read and understand the university's *Code of Student Conduct*, and that all students will complete all academic and scholarly assignments with fairness and honesty. Students must recognize that failure to follow the rules and guidelines established in the university's *Code of Student Conduct* and this syllabus may constitute "Academic Misconduct."

The Ohio State University's *Code of Student Conduct* (Section 3335-23-04) defines academic misconduct as: "Any activity that tends to compromise the academic integrity of the university or subvert the educational

process." Examples of academic misconduct include (but are not limited to) plagiarism, collusion (unauthorized collaboration), copying the work of another student, and possession of unauthorized materials during an examination. Ignorance of the university's *Code of Student Conduct* is never considered an excuse for academic misconduct, so I recommend that you review the *Code of Student Conduct* and, specifically, the sections dealing with academic misconduct.

If I suspect that a student has committed academic misconduct in this course, I am obligated by university rules to report my suspicions to the Committee on Academic Misconduct. If COAM determines that you have violated the university's *Code of Student Conduct* (i.e., committed academic misconduct), the sanctions for the misconduct could include a failing grade in this course and suspension or dismissal from the university.

If you have any questions about the above policy or what constitutes academic misconduct in this course, please contact me.

Other sources of information on academic misconduct (integrity) to which you can refer include:

- The Committee on Academic Misconduct web pages (COAM Home)
- *Ten Suggestions for Preserving Academic Integrity (Ten Suggestions)*
- *Eight Cardinal Rules of Academic Integrity* (www.northwestern.edu/uacc/8cards.htm)

Statement on Title IX

All students and employees at Ohio State have the right to work and learn in an environment free from harassment and discrimination based on sex or gender, and the university can arrange interim measures, provide support resources, and explain investigation options, including referral to confidential resources.

If you or someone you know has been harassed or discriminated against based on your sex or gender, including sexual harassment, sexual assault, relationship violence, stalking, or sexual exploitation, you may find information about your rights and options at titleix.osu.edu or by contacting the Ohio State Title IX Coordinator at titleix@osu.edu. Title IX is part of the Office of Institutional Equity (OIE) at Ohio State, which responds to all bias-motivated incidents of harassment and discrimination, such as race, religion, national origin and disability. For more information on OIE, visit equity.osu.edu or email equity@osu.edu.

Your Mental Health

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you find yourself feeling isolated, anxious or overwhelmed, please know that there are resources to help: ccs.osu.edu. You can reach an on-call counselor when CCS is closed at (614) 292-5766 and 24 hour emergency help is also available through the 24/7 National Prevention Hotline at 1-(800)-273-TALK or at suicidepreventionlifeline.org. The Ohio State Wellness app is also a great resource available at go.osu.edu/wellnessapp.

Week	Date	Topic	Assignment
1	Aug 21	Course Introduction	
1	Aug 23	Introduction: AI on EHRs + Data Access Tutorial	A 1.1 Out
2	Aug 26	Introduction: MIMIC-IV Data (Part I)	
2	Aug 28	Introduction: MIMIC-IV Data (Part II)	
2	Aug 30	Data Exploration (Lecture + Demo)	A 1.2 Out
3	Sep 2	Labor Day - no classes	
3	Sep 4	Data Preprocessing (Lecture + Demo)	
3	Sep 6	Introduction to Classification	
4	Sep 9	Decision Tree (Part I)	
4	Sep 11	Decision Tree (Part II)	A 1.1 Due; A 1.3 Out
4	Sep 13	Evaluation Metrics	
5	Sep 16	Evaluation Metrics and Model Evaluation	A 1.2 Due
5	Sep 18	Logistic Regression	
5	Sep 20	Nearest-Neighbor Classifiers	
6	Sep 23	Naïve Bayes Classifier	
6	Sep 25	Ensemble Classifier	
6	Sep 27	Demo: Building a Classification Model + Assignment	A 2 Out
7	Sep 30	Classification Tasks on EHRs (Part I) + Demo	A 1.3 Due
7	Oct 2	Classification Tasks on EHRs (Part II) + Demo	
7	Oct 4	Classification Tasks on EHRs (Part III)	
8	Oct 7	Introduction to Clustering	
8	Oct 9	K-means	
8	Oct 11	Autumn Break - no classes	
9	Oct 14	Strengths and Weaknesses of K-means	A 2 Due
9	Oct 16	Hierarchical Clustering	
9	Oct 18	Strengths and Weaknesses of Hierarchical Clustering	
10	Oct 21	DBSCAN Clustering	

10	Oct 23	Cluster Evaluation	A 3 Out
10	Oct 25	Clustering Using Mixture Models	
11	Oct 28	Demo: Building a Clustering Model + Assignment	
11	Oct 30	Clustering Tasks on EHRs (Part I) + Demo	
11	Nov 1	Clustering Tasks on EHRs (Part I) + Demo	
12	Nov 4	Clustering Tasks on EHRs (Part II)	
12	Nov 6	Introduction to Data Visualization	A 3 Due
12	Nov 8	Basic Data Visualization: Line Plot, Bar Plot, Area Plot, Histogram	
13	Nov 11	Veterans Day observed - no classes	
13	Nov 13	Specialized Data Visualization: Pie Chart, Box Plot, Scatter Plot	
13	Nov 15	Advanced Data Visualization: Waffle Chart, Word Cloud, Regression Plot	
14	Nov 18	Creating Dashboard with Plotly	
14	Nov 20	Interactive Dashboard with Dash	
14	Nov 22	Data Visualization for EHR data	A 4 Out
15	Nov 25	Advanced Topic 1: Treatment Effect Estimation on EHR	
15	Nov 27	Thanksgiving Break - no classes	
15	Nov 29	Thanksgiving Break - no classes	
16	Dec 2	Advanced Topic 2: Human-Centered Decision Support	
16	Dec 4	Course Summary	A 4 Due
16	Dec 6	Final examinations - no classes	
17	Dec 9	Final examinations - no classes	
17	Dec 11	Final examinations - no classes	
17	Dec 13	Last day of the semester	Post All Grades