CSE 531A Projects
Why is AI important to healthcare?

Machine learning is leading a potential revolution in healthcare:

• Early disease detection
• Predictive analytics for patient outcomes
• Personalized treatments
• Decision support for healthcare workers

Sources: Bajwa, Junaid et al. “Artificial intelligence in healthcare: transforming the practice of medicine.” Future healthcare journal
Research-Oriented Class on AI for Health

Projects 60%
- Proposal and presentation: 10%
- Demo I: 5%
- Demo II: 5%
- Final report and demo: 40%

Critiques 35%

Participation 5%

- Mini research experience on AIM
- Interdisciplinary: co-taught by researchers from CS and medicine
Project

- Two students per team
  - Need permission for a bigger or smaller team.

- Develop/integrate your own research ideas
- Perform experiments on real-world healthcare dataset
- Write a paper
- Demos
Steps

1. Pick your interested direction
2. Form a team
3. Propose a research topic
4. Analyze and Implement your solution
5. Evaluate your solution
6. Demo 1, 2 and Final Demo
7. Write a technical report
Get Started Early

- Think about topics and ideas
- Talk to TA and me
- Put together a team

- A lot of work (and fun) throughout the semester!
Teaming

- Everyone should be in a **two**-member team
  - unless you receive special approval from TA for a different size

- Use **Piazza** to “Search for Teammates”

- Email TA your team members by **1/25**
  - One email per team

- Join one group on Canvas > People > Groups.

- We will help make sure everyone has a team.
Proposal Presentation

- In class on 02/06

- **4 min** per group
  - 3-min talk + 1-min Q&A
  - 4 slides
  - Rehearse in advance

- Your **elevator pitch!**

- Email TA your slides before class
Written Proposal

- One proposal/team, one page
  - Team members
  - Concise description of project
  - Responsibilities of each member
  - Equipment needed
  - ACM format

- Written proposal due: 02/06
Demo I & II

Demo I

- In class on **03/05** and **03/07**.
- **8 min** per team.
- Must show something **real**.

Demo II

- In class on **04/02** and **04/04**.
- **8 min** per team.
- **Substantial** progress → final demo.
Final Demo

- Preparation & Discussion: in class on **04/23**

- Final demo: in class on **04/25 (1pm - 3:20pm)**.
  - 8 min per team.

- All expected to attend the entire session. It’ll be fun!
Final Report

- Submit by **05/02, 11:59pm**

- Report
  - Should be ACM style (most conference papers in the reading list)
  - 6 pages, double column, 10 pts font
  - Use the LaTeX template on the class web page

- Materials
  - Slides of your final presentation
  - Source code
  - Documentation: README/INSTALL
Suggested Outline

- Abstract
- Introduction
- Related Work
- Methods
- Experiments
- Lessons Learned
- Conclusion and Future Work
Logistics

- TA for projects: Ziqi Xu, Ben Warner
- Email TAs or Professor Lu for appointments to discuss ideas
- All work will be submitted on Canvas
Logistics

- Guidelines and slides are on the class [homepage](http://www.cse.wustl.edu/~lu/cse521s/).

- Communication will be through [Piazza](http://www.cse.wustl.edu/~lu/cse521s/).
  - E.g., search for teammates
Large, freely available database comprising de-identified health-related data
40,000 patients who stayed in critical care units

Static Data
Demographic
Medication history
...

Time-series Data
Vital signs
Laboratory measurements
...

Text Data
Notes from caregivers
Diagnostic Reports
...

MIMIC-III Potential Research Topics

Four example outcomes:

- Mortality (binary classification)
- Real-time detection of decompensation (time series classification)
- Length of stay (regression)
- Phenotype classification (multilabel classification)
MIMIC-III Access

• Create an account on PhysioNet using the following link: https://physionet.org/register/ with institutional email (...@wustl.edu)

• Go to https://physionet.org/content/mimiciii/1.4/ and complete the application process located in the bottom of the page

Files

This is a restricted-access resource. To access the files, you must fulfill all of the following requirements:

• be a credentialed user
• complete required training:
  ○ CITI Data or Specimens Only Research
    You may submit your training here.
• sign the data use agreement for the project
MIMIC-III Access

• **Be a credentialed user:**
• Fill out all required information
• Reference: Chenyang Lu, lu@wustl.edu, Fullgraf Professor, Computer Science and Engineering Department

• Project description:
• Student in CSE531 AI for Health at WUSTL for Spring 2024
• Describe your project ideas, interested research directions
• Don’t be too general
• Each student must apply for individual access, even you are in a group
MIMIC-III Access

- **Complete the required training:**
- PhysioNet requires the CITI Data or Specimens Only Course.
- Follow the instructions here:
- [https://physionet.org/about/citi-course/](https://physionet.org/about/citi-course/)

- Will take a few hours to complete, possibly less if you completed another CITI course.
- Submit the report in your account and wait for approval
Sign the data use agreement for the project:

https://physionet.org/sign-dua/mimiciii/1.4/

Sign Data Use Agreement - MIMIC-III Clinical Database v1.4

PhysioNet Credentialed Health Data Use Agreement 1.5.0

If I am granted access to the database:

1. I will not attempt to identify any individual or institution referenced in PhysioNet restricted data.
2. I will exercise all reasonable and prudent care to avoid disclosure of the identity of any individual or institution referenced in PhysioNet restricted data in any publication or other communication.
3. I will not share access to PhysioNet restricted data with anyone else.
4. I will exercise all reasonable and prudent care to maintain the physical and electronic security of PhysioNet restricted data.
5. If I find information within PhysioNet restricted data that I believe might permit identification of any individual or institution, I will report the location of this information promptly by email to PHI-report@physionet.org, citing the location of the specific information in question.
6. I have requested access to PhysioNet restricted data for the sole purpose of lawful use in scientific research, and I will use my privilege of access, if it is granted, for this purpose and no other.
7. I have completed a training program in human research subject protections and HIPAA regulations, and I am submitting proof of having done so.
8. I will indicate the general purpose for which I intend to use the database in my application.
9. If I openly disseminate my results, I will also contribute the code used to produce those results to a repository that is open to the research community.
10. This agreement may be terminated by either party at any time, but my obligations with respect to PhysioNet data shall continue after termination.

I agree  I do not agree
MIMIC-III Structure

- Consists of data from 40,000 ICU patients at Beth Israel Deaconess in Boston.
  - Full dataset is available once you complete the training.
  - A demo set with 100 patients is also freely available: https://physionet.org/content/mimiciii-demo/1.4/
- Data is organized as a series of CSVs:
  - Rows are connected via identifiers ending in the _ID suffix: SUBJECT_ID, HADM_ID, ICUSTAY_ID, ROW_ID, and ITEM_ID.
  - Tables ending in EVENTS are charted event files.
  - Tables with the prefix D_ are dictionary files that give definitions for identifiers.
  - Patient stays are primarily characterized by ADMISSIONS, PATIENTS, ICUSTAYS, SERVICES, and TRANSFERS.
Recommended Tools

- A list of recommended tools:
  - Traditional ML: scikit-learn, XGBoost, CatBoost
  - Deep Learning: PyTorch, PyTorch Lightning
  - Evaluation: Torch Metrics or Torch Eval
  - Transformer Models: Hugging Face transformers
  - Visualization: Gradio, Seaborn/Matplotlib
  - MIMIC-III tools/models: PyHealth

- This list serves as a starting point, we encourage you to experiment with other tools as you see fit.
- Please do not use OpenAI APIs and external third-party tools, as they violate the PhysioNet DUA (https://physionet.org/news/post/415).
Conferences/Workshops for Ideas

- Conferences:
  - General ML conferences: NeurIPS, ICLR, ICML, AAAI, etc.
  - Data Mining: SIGKDD, CIKM, etc.
  - Health/AI: MLHC, CHIL, ML4H, etc.

- Example Workshops:
  - Workshop on Time Series Representation Learning for Health (ICLR)
  - Learning from Time Series for Health (NeurIPS)
  - Clinical NLP Workshop (NAACL)

- Other:
  - Papers With Code, Hugging Face, Twitter