

## ✓ CMU 10-742 (Fall 2024) - Machine Learning in Healthcare

### Assignment 0: Training and Data Access

Out: Tues Aug 27 2024

Due: Thurs Sep 5 2024

This assignment counts for 3 points out of the 35 total points allocated to the course problem sets.

In the subsequent assignments for this class, you'll be using real-world healthcare datasets that contain information about actual patients and their medical history. Although these datasets will be de-identified, even de-identified datasets contain information regarding the health journeys of real people, and as such we must treat the data with care and respect.

One dataset you'll be using is MIMIC, a publicly available set of data on patient visits to critical care units of the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts. The MIMIC datasets are widely used in healthcare AI research. If you continue pursuing healthcare AI after this course ends, you will be glad to have access to MIMIC.

MIMIC comes in two versions:

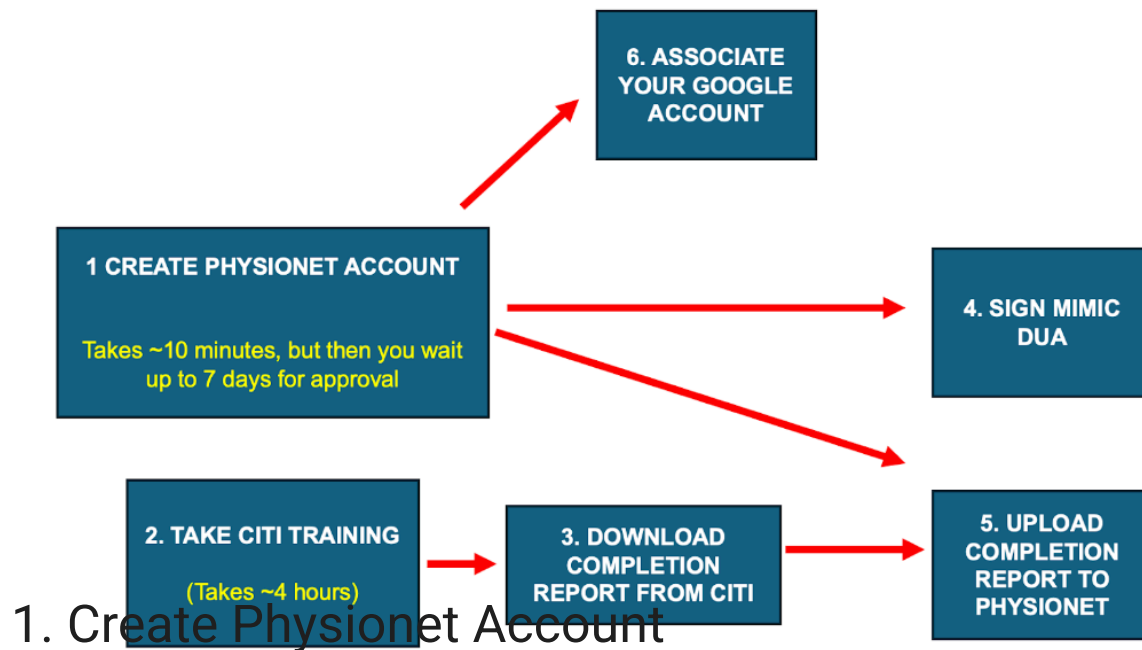
- MIMIC-III contains data on over 50K admissions to BIDMC between 2001 and 2012. [Here](#) is a 2016 Nature paper describing MIMIC-III.
- MIMIC-IV contains data on encounters for over 300K encounters at BIDMC from 2008-2019. [Here](#) is a 2023 Nature paper describing MIMIC-IV.

To gain access to MIMIC requires several steps. You must complete them all. There are a few dependencies between the steps. Here's simple state diagram to show you how the steps fit together. The blue boxes are activities you need to perform (detailed below), and the red arrows are dependencies. For example, you cannot perform step 5 until you have completed steps 1 and 3.

Each step is detailed further below in this workbook.

WARNING: Step 2 (CITI Training) takes a long time -- be sure to budget around five hours of your time. Also note: Step 1 (Create Physionet Account) takes very little of your time, but it requires you wait up to several days before Physionet reviews (and hopefully approves) your application.

**All of which to say: get started early!**



Register for a physionet account: <https://physionet.org/register/>. When prompted for an email, use your CMU email. During the application, you will be prompted for the following data; below is the responses you should provide.

If you already have a physionet account but don't have credential access, go under your profile name > settings > credentialing and fill in the below information to get access.

| Category                  | Details  |
|---------------------------|--|
| Reference Category        | Supervisor   |
| Reference Name            | Adam Berger  |
| Phone Number/Contact Info | 412-268-1593   |
| Email                     | <a href="mailto:adam@andrew.cmu.edu">adam@andrew.cmu.edu</a>   |
| Job Title                 | Professor  |
| Research Area             | I am a student in Carnegie Mellon University course 10-742 "Machine Learning for Healthcare", Fall 20<br>Several of the course assignments require use of Physionet datasets, including MIMIC. |

## 2. Take CITI Training

You can complete this stage while waiting for Physionet to approve your application.

The training is at this URL: <https://about.citiprogram.org/>

Login in: Find the Login' button on the top right, then select 'Log In Through My Organization' using your Andrew ID. Select "No CME Credit" when prompted. You will then encounter eight questions; say no to all of them except Q2: Human Subjects Research. For this question, select "HSR Basics for Biomedical Research."

(If you were not prompted with "No CME Credit" immediately, go under My Courses > fill out the information about your role at CMU > Add a Course and you will encounter the eight questions suggested above.)

That will bring up a training course. To make sure you're enrolled in the correct training, look for the title "Biomedical Research - Basic/Refresher." There are 17 modules for this, starting with "Belmont Report and Its Principles (ID 1127)."

It will take you several hours to complete and pass this course. The "supplemental modules" are optional.

### 3. Download Completion Report from CITI

When you complete this training course, download your Completion Report (not the "Completion Certificate") to your computer. You will need it in step 5.

### 4. Sign DUA

During this class we will use either/both of MIMIC III and MIMIC IV datasets.

Go to the bottom of these pages and click on "sign the data use agreement" for both of these datasets.

MIMIC 3: <https://physionet.org/content/mimiciii/1.4/#files>

MIMIC 4: <https://physionet.org/content/mimiciv/2.2/#files>

### 5. Upload CITI Completion Report to Physionet

Once you have completed the CITI training and you have received approval from Physionet, visit <https://physionet.org/settings/training/> and upload the Completion Report you downloaded in step 3.

### 6. Associate your Google account with Physionet

Visit <https://physionet.org/settings/cloud/>. In the Google (email) field, enter your andrew CMU id.

