
MIMIC Code Repository DOI 10.5281/zenodo.6818823

The MIMIC Code Repository is intended to be a central hub for sharing, refining, and reusing code used for analysis of the MIMIC critical care database. To find out more about MIMIC, please see: <https://mimic.mit.edu>. Source code for the website is in the [mimic-website](#) GitHub repository.

You can read more about the code repository in the following open access paper: [The MIMIC Code Repository: enabling reproducibility in critical care research](#).

Cloud access to datasets

The various MIMIC databases are available on Google Cloud Platform (GCP) and Amazon Web Services (AWS). To access the data on the cloud, simply add the relevant cloud identifier to your PhysioNet profile. Then request access to the dataset for the particular cloud platform via the PhysioNet project page. Further instructions are available on the MIMIC website.

Navigating this repository

This repository contains code for five databases on PhysioNet:

- MIMIC-III - critical care data for patients admitted to ICUs at the BIDMC between 2001 - 2012
- MIMIC-IV - hospital and critical care data for patients admitted to the ED or ICU between 2008 - 2019
- MIMIC-IV-ED - emergency department data for individuals attending the ED between 2011 - 2019
- MIMIC-IV Waveforms (TBD) - this dataset has yet to be published.
- MIMIC-CXR - chest x-ray imaging and deidentified free-text radiology reports for patients admitted to the ED from 2012 - 2016

The repository contains one top-level folder containing community developed code for each datasets:

- `mimic-iii` - build scripts for MIMIC-III, derived concepts which are available on the [physionet-data.mimiciii_derived](#) dataset on BigQuery, and tutorials.
- `mimic-iv` - build scripts for MIMIC-IV, derived concepts which are available on the [physionet-data.mimic_derived](#) dataset on BigQuery, and tutorials.
- `mimic-iv-cxr` - code for loading and analyzing both dicom (`mimic-iv-cxr/dcm`) and text (`mimic-iv-cxr/txt`) data. In order to clearly indicate that MIMIC-CXR can be linked with MIMIC-IV, we have named this folder `mimic-iv-cxr`, and any references to MIMIC-CXR / MIMIC-IV-CXR are interchangeable.

-
- mimic-iv-ed - build scripts for MIMIC-IV-ED.
 - mimic-iv-waveforms - TBD

Each subfolder has a README with further detail regarding its content.

Launch MIMIC-III in AWS

MIMIC-III is available on AWS (and MIMIC-IV will be available in the future). Use the below Launch Stack button to deploy access to the MIMIC-III dataset into your AWS account. This will give you real-time access to the MIMIC-III data in your AWS account without having to download a copy of the MIMIC-III dataset. It will also deploy a Jupyter Notebook with access to the content of this GitHub repository in your AWS account. Prior to launching this, please login to the MIMIC PhysioNet website, input your AWS account number, and request access to the MIMIC-III Clinical Database on AWS.

To start this deployment, click the Launch Stack button. On the first screen, the template link has already been specified, so just click next. On the second screen, provide a Stack name (letters and numbers) and click next, on the third screen, just click next. On the forth screen, at the bottom, there is a box that says **I acknowledge that AWS CloudFormation might create IAM resources..** Check that box, and then click **Create**. Once the Stack has complete deploying, look at the **Outputs** tab of the AWS CloudFormation console for links to your Jupyter Notebooks instance.



Other useful tools

- Bloatectomy (paper) - A python based package for removing duplicate text in clinical notes
- Medication categories - Python script for extracting medications from free-text notes
- MIMIC Extract (paper) - A python based package for transforming MIMIC-III data into a machine learning friendly format
- FIDDLE (paper) - A python based package for a Flexible Data-Driven pipeLine (FIDDLE), transforming structured EHR data into a machine learning friendly format

Acknowledgement

If you use code or concepts available in this repository, we would be grateful if you would:

- cite the dataset(s) you use as described in the PhysioNet project page: MIMIC-III, MIMIC-IV, MIMIC-IV-ED , and/or MIMIC-CXR

-
- cite the Zenodo repository directly as it contains a static copy of the code. Be sure to select the release of MIMIC Code you used from the menu on the right side of the page on Zenodo: <https://zenodo.org/record/6818823>
 - cite the MIMIC code repository paper: The MIMIC Code Repository: enabling reproducibility in critical care research

```
1 @article{johnson2018mimic,  
2   title={The MIMIC Code Repository: enabling reproducibility in  
3     critical care research},  
4   author={Johnson, Alistair E W and Stone, David J and Celi, Leo A and  
5     Pollard, Tom J},  
6   journal={Journal of the American Medical Informatics Association},  
7   volume={25},  
8   number={1},  
9   pages={32--39},  
10  year={2018},  
11  publisher={Oxford University Press}  
12 }
```

Contributing

Our team has worked hard to create and share the MIMIC datasets. We encourage you to share the code that you use for data processing and analysis. Sharing code helps to make studies reproducible and promotes collaborative research. To contribute, please:

- Fork the repository using the following link: <https://github.com/MIT-LCP/mimic-code/fork>. For a background on GitHub forks, see: <https://help.github.com/articles/fork-a-repo/>
- Commit your changes to the forked repository.
- Submit a pull request to the MIMIC code repository, using the method described at: <https://help.github.com/articles/using-pull-requests/>

We encourage users to share concepts they have extracted by writing code which generates a materialized view. These materialized views can then be used by researchers around the world to speed up data extraction. For example, ventilation durations can be acquired by creating the ventdurations view in `concepts/durations/ventilation_durations.sql`.

License

By committing your code to the MIMIC Code Repository you agree to release the code under the MIT License attached to the repository.

Coding style

Please refer to the style guide for guidelines on formatting your code for the repository.