PhysioNet team

Tom Pollard, Benjamin Moody, Li-Wei Lehman; Brian Gow, Chen Xie,Dana Moukheiber, Lama Moukheiber; Ken Paik, Leo Celi, Alistair Johnson, Roger Mark

Open Data
@ MIT

PhysioNet : Open Data@MIT

Tom Pollard (@tompollard)

Friday, 28 October 2022

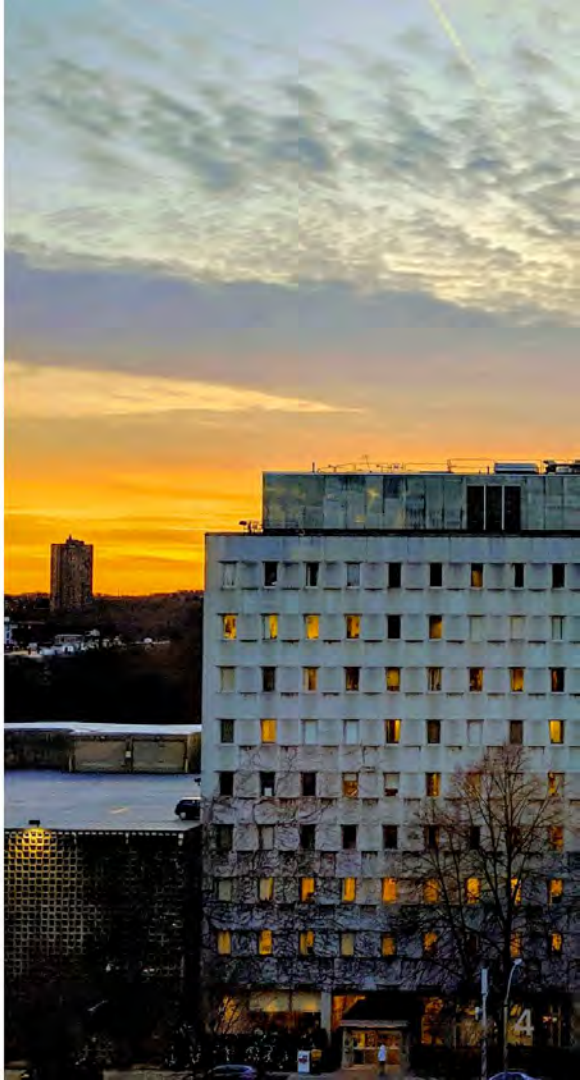
PhysioNet

- Data sharing platform built and maintained by MIT
- Began as outreach arm of a research project
- Rebuilt in 2019 following "FAIR principles"
- >55,000 registered, active users since 2019
- >30TB ecgs, x-rays, echocardiograms...

What are some examples of content shared on PhysioNet?

MIMIC-IV

- Highly-detailed critical care database for >40k patients in the US, comprising:
  - Vital signs, medications, labs..
  - Chest X-rays
  - ECGs, waveforms
  - Free text notes
  - Echocardiograms
- Extensively used across education, research, and industry



MIMIC-IV (version 1.0). PhysioNet. https://doi.org/10.13026/s6n6-xd98.

BRAX, a Brazilian labeled chest X-ray dataset

- 40,967 DICOM images
- De-identified to protect patient privacy
- Annotated by trained radiologists
- Adds geographic diversity in public chest x-ray data

BRAX, a Brazilian labeled chest X-ray dataset (v 1.0.0). PhysioNet. https://doi.org/10.13026/ae9a-f727
(image shows patient with pneumonia)

CheXclusion: Fairness gaps in deep chest X-ray classifiers

Laleh Seyyed-Kalantari [1], Guanxiong Liu, Matthew McDermott, Irene Y Chen, Marzyeh Ghassemi

Affiliations + expand

PMID: 33691020

Free article

Abstract

Machine learning systems have received much attention recently for their ability to achieve expert-level performance on clinical tasks, particularly in medical imaging. Here, we examine the extent to which state-of-the-art deep learning classifiers trained to yield diagnostic labels from X-ray images are biased with respect to protected attributes. We train convolution neural networks to predict 14 diagnostic labels in 3 prominent public chest X-ray datasets: MIMIC-CXR, Chest-Xray8, CheXpert, as well as a multi-site aggregation of all those datasets. We evaluate the TPR disparity - the difference in true positive rates (TPR) - among different protected attributes such as patient sex, age, race, and insurance type as a proxy for socioeconomic status. We demonstrate that TPR disparities exist in the state-of-the-art classifiers in all datasets, for all clinical tasks, and all subgroups. A multi-source dataset corresponds to the smallest disparities, suggesting one way to reduce bias. We find that TPR disparities are not significantly correlated with a subgroup's proportional disease burden. As clinical models move from papers to products, we encourage clinical decision makers to carefully audit for algorithmic disparities prior to deployment. Our

Google AI Blog

The latest from Google Research

Multi-task Prediction of Organ Dysfunction in ICUs

Thursday, July 22, 2021

Posted by Subhrajit Roy, Research Scientist and Diana Mincu, Research Software Engineer, Google Research

The intensive care unit (ICU) of a hospital looks after the most medically vulnerable patients, many of whom require organ support, such as mechanical ventilation or dialysis. While always critical, the demand on ICU services during the COVID-19 pandemic has further underscored the importance of data-driven decision-making in healthcare. Furthermore, the ability to accurately predict the clinical outcomes of ICU patients has the potential to guide therapy and may inform decisions about most effective care, including staffing and triage support.

Multitask prediction of organ dysfunction in the intensive care unit using sequential subnetwork routing. Journal of the Medical Informatics Association (JAMIA). https://academic.oup.com/jamia/article/28/9/1986/6307184.

Why do people choose to share on PhysioNet?

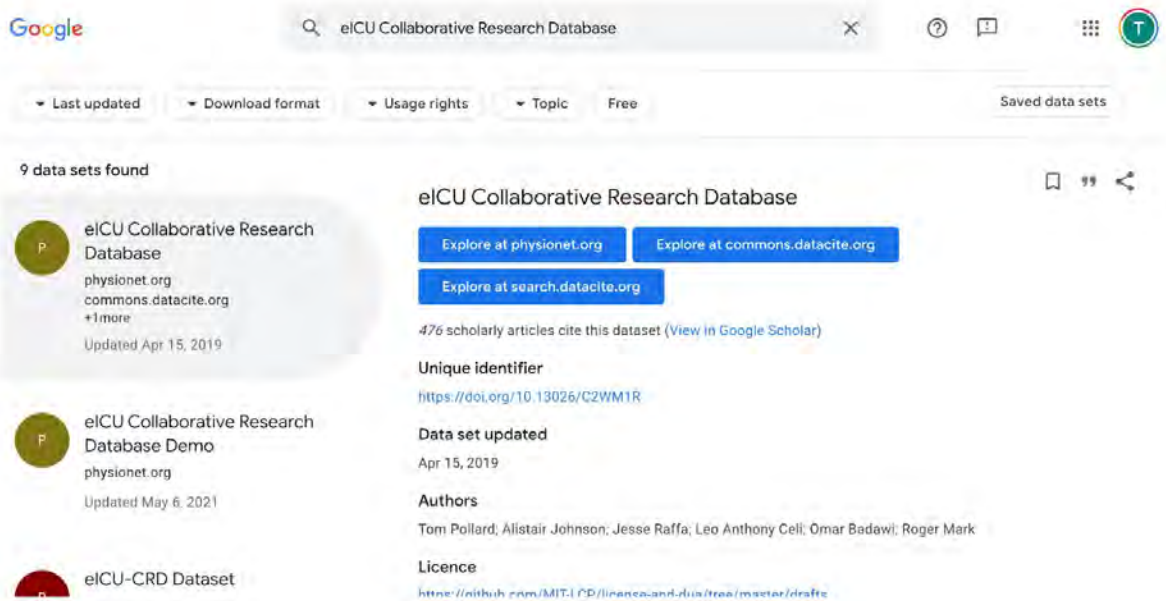
Expert review and curation

- > 20 years experience in safely sharing clinical data
- Assistance with data preparation
- State-of-the-art tools in de-identification

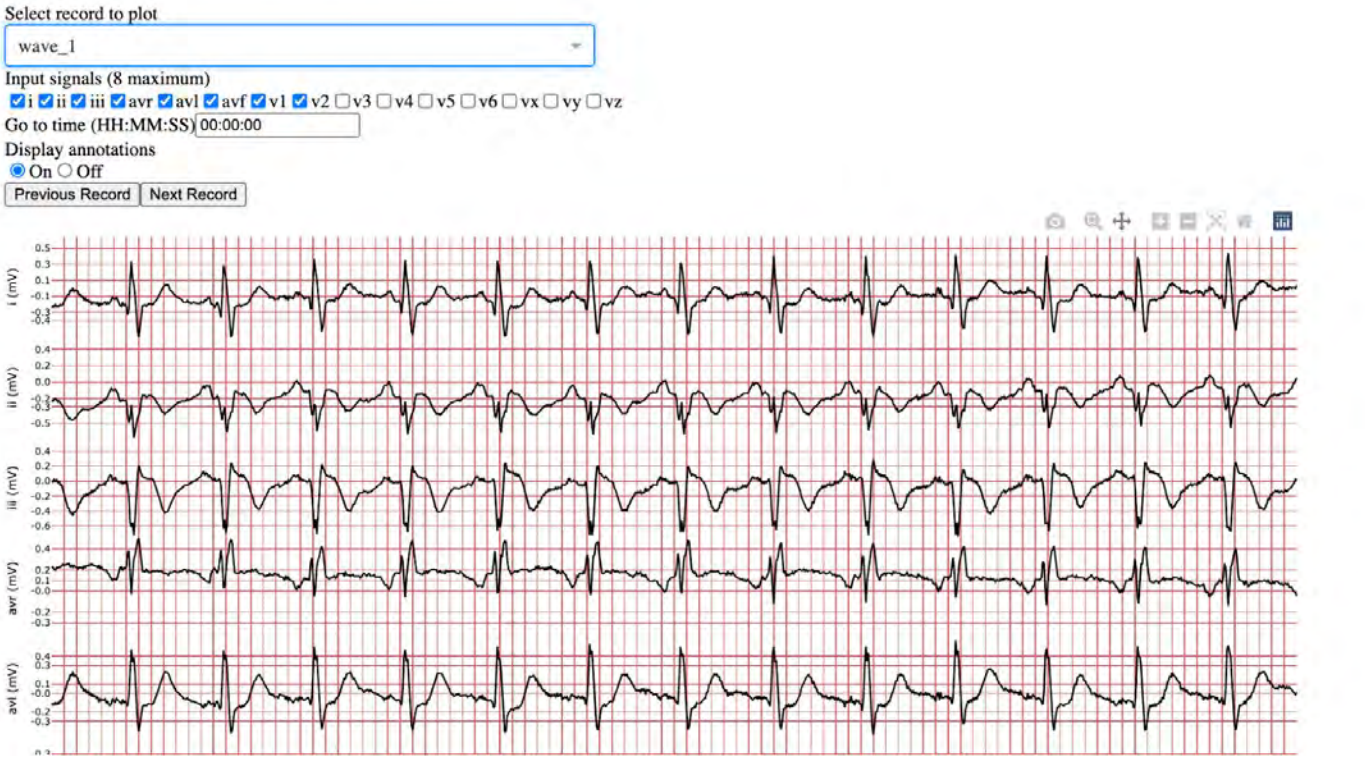
Fine-grained access control

- **Open data**
- **Restricted**:
  - Data Use Agreement.
- **Credentialed**:
  - Data Use Agreement
  - Training in human research.
  - Identity check.
- **Contributor-managed**:
  - Approval of the contributor.

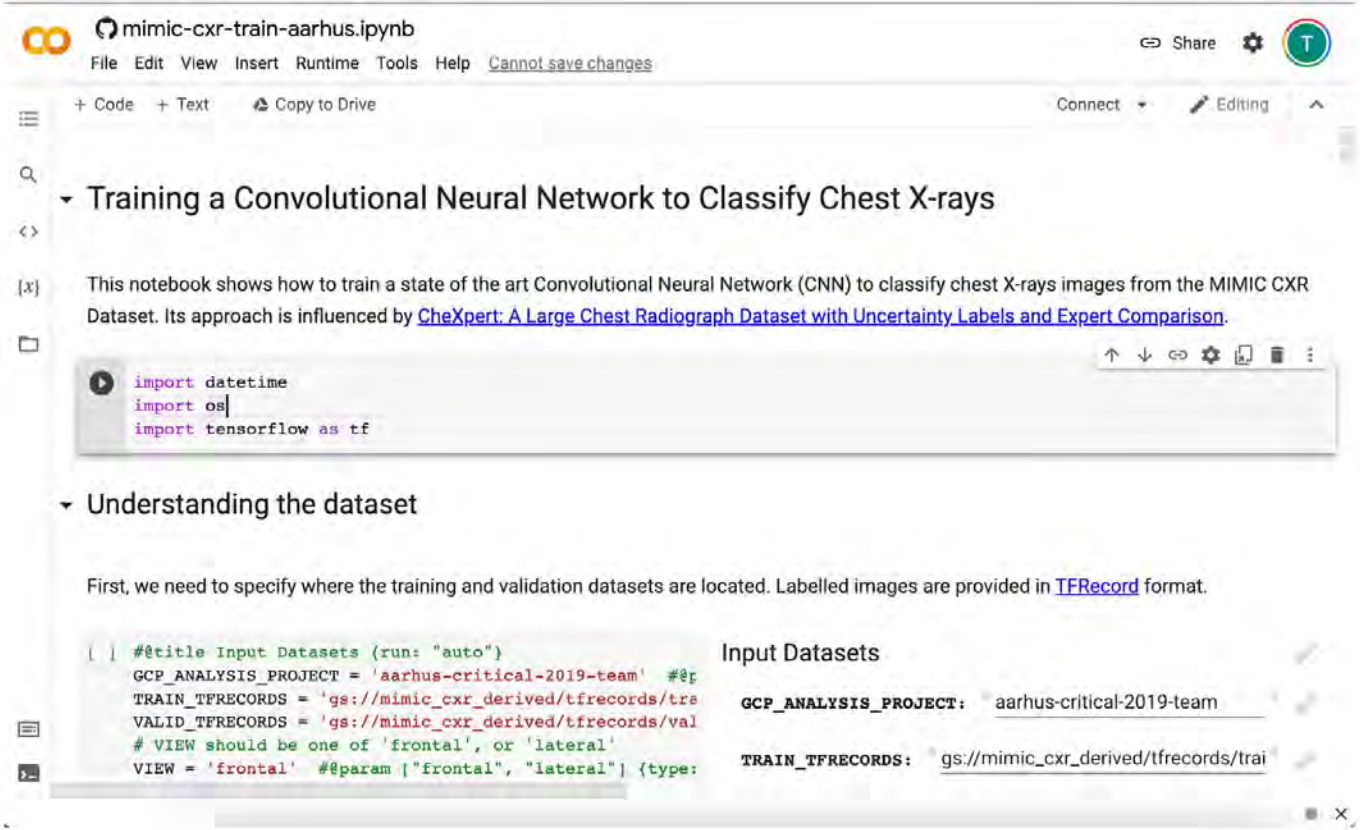
Enhanced discovery



Integrated viewers

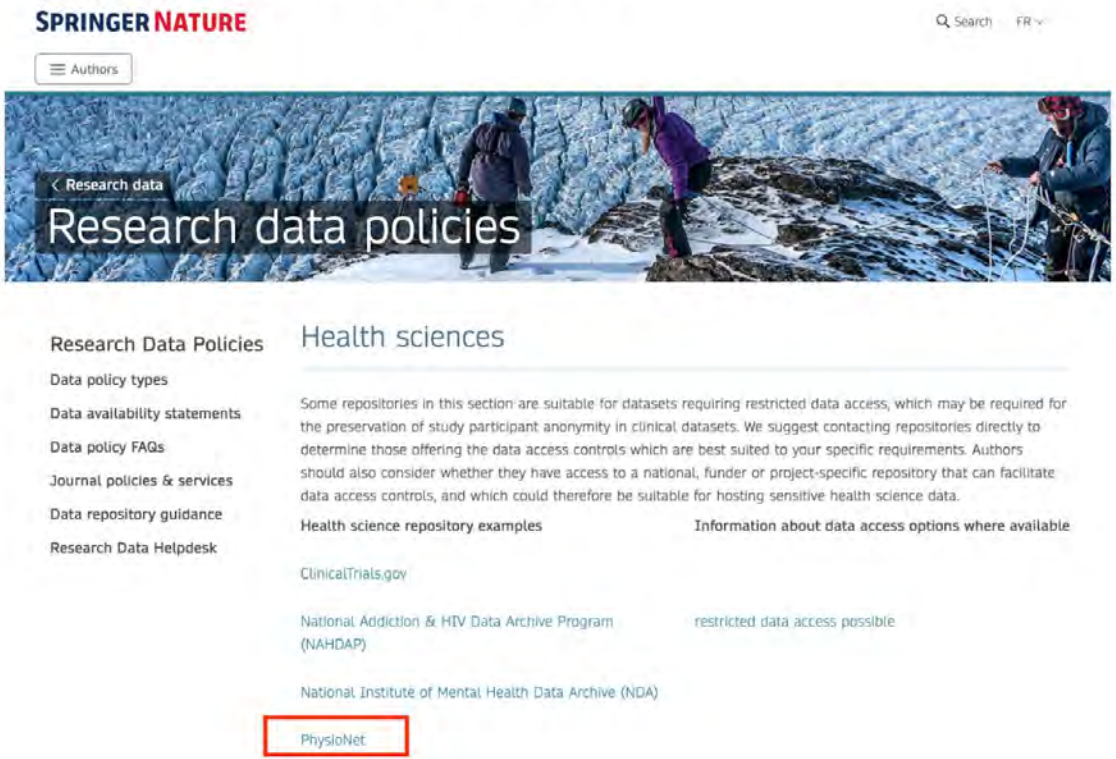


Cloud integration



Recommended repository for leading journals

- Springer Nature, eLife, PLOS...



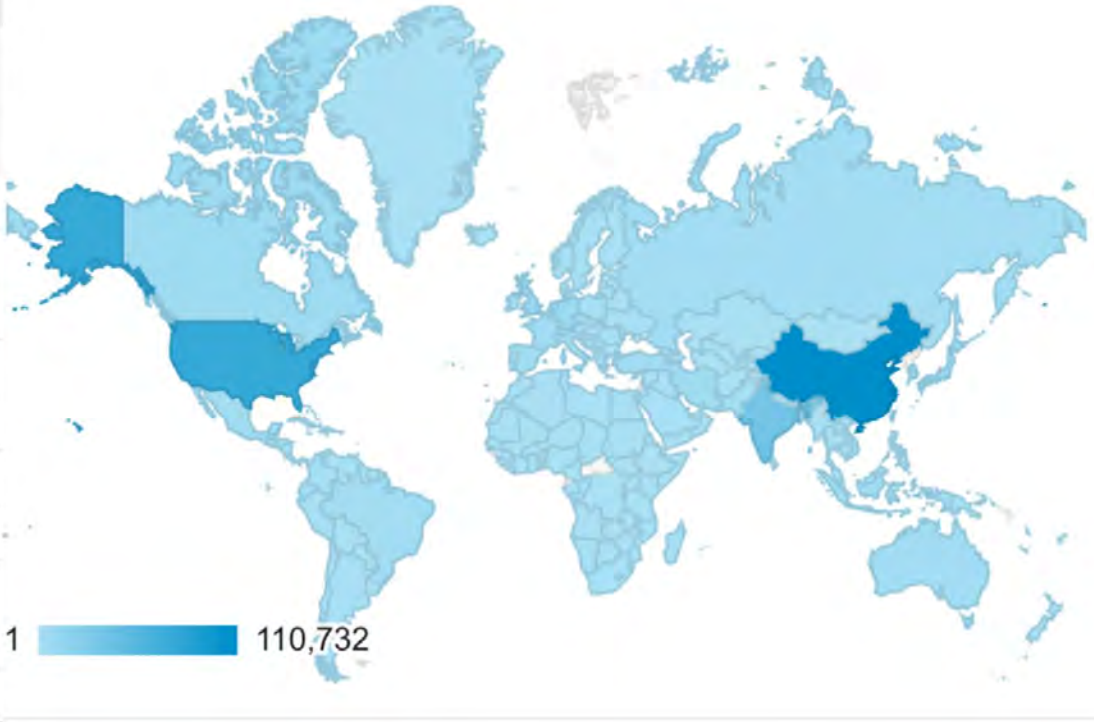
Who is using PhysioNet?

Diverse, active community

- Regular workshops, challenges, and datathons based around PhysioNet datasets

Photo: 49th Annual Conference of the Society of Critical Care Medicine (SCCM 2020)

Country	Users
	394,415 % of Total: 100.00% (394,415)
1. China	110,732 (26.92%)
2. United States	75,734 (18.41%)
3. India	37,302 (9.07%)
4. South Korea	10,671 (2.59%)
5. United Kingdom	10,324 (2.51%)
6. Germany	9,536 (2.32%)
7. Canada	8,561 (2.08%)
8. Japan	8,208 (2.00%)
9. Taiwan	7,640 (1.86%)
10. Iran	7,359 (1.79%)



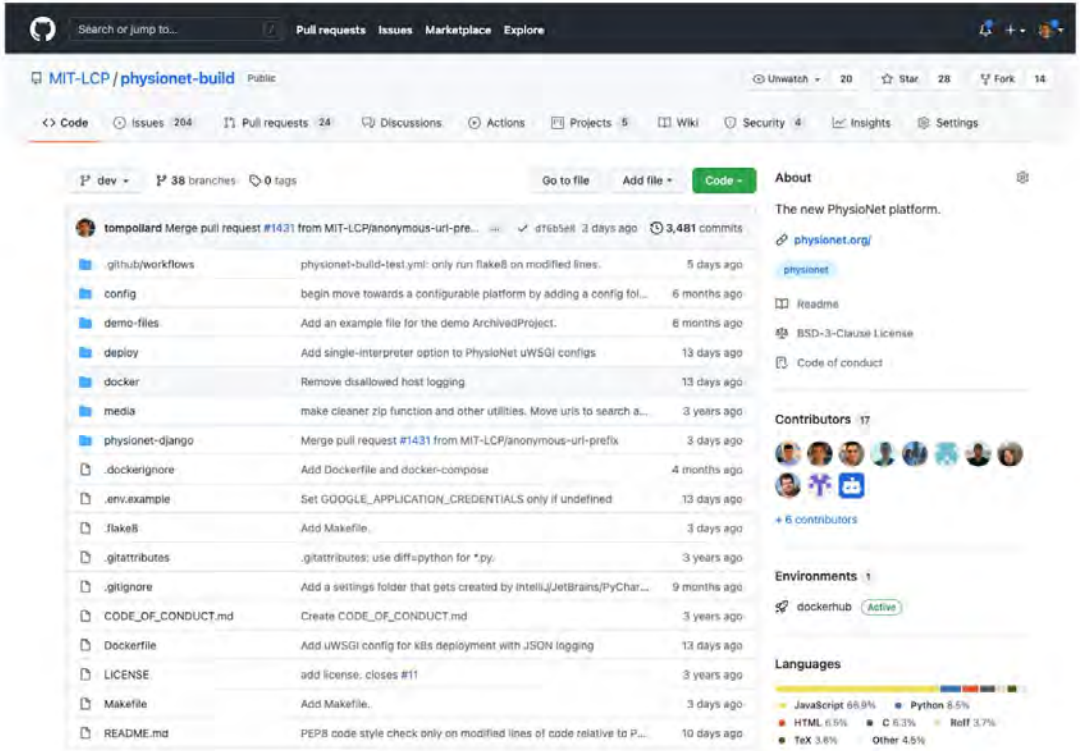
PhysioNet in 2021

In the year 2021:

- 18,553 people created an account on PhysioNet
- 7,399 people were granted "credentialed" access
- ~7 TB of new data
- ~6,610 academic citations

What's next?

Open source code



https://github.com/MIT-LCP/physionet-build

Infrastructure Overview

Creating a network of repositories (pilots in Toronto and Uganda)

https://tcairem.utoronto.ca/

Thanks!

Benjamin Moody
Dr Alistair Johnson
Dr Leo Celi
Brian Gow
Lucas Bulgarelli
Dr Li-Wei Lehman
Dana Moukheiber
Lama Moukheiber
Dr Ken Paik
Lucas Mccullum
Chen Xie
Felipe Torres
Dr Jesse Raffa
Prof Roger Mark