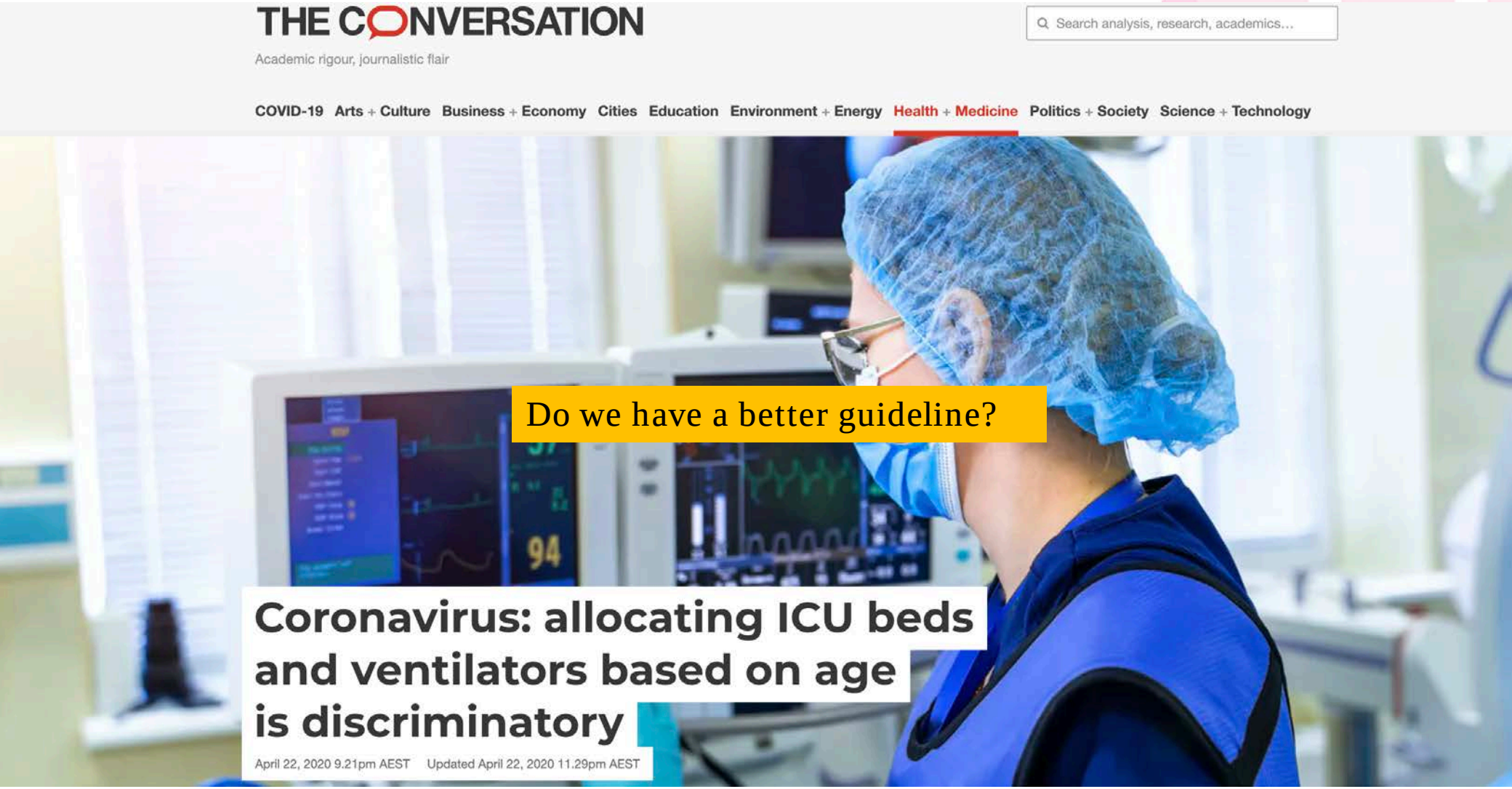
Assignment 3
COVID-19 30-day Mortality Prediction from CXR




THE CONVERSATION
Academic rigour, journalistic flair
Search analysis, research, academics...
COVID-19 Arts + Culture Business + Economy Cities Education Environment + Energy Health + Medicine Politics + Society Science + Technology
Do we have a better guideline?
Coronavirus: allocating ICU beds and ventilators based on age is discriminatory
April 22, 2020 9.21pm AEST Updated April 22, 2020 11.29pm AEST

Stanford ML Group

CheXpert

A Large Chest X-Ray Dataset And Competition

What is CheXpert?

CheXpert is a large dataset of chest X-rays and competition for automated chest x-ray interpretation, which features uncertainty labels and radiologist-labeled reference standard evaluation sets.

READ THE PAPER (IRVIN & RAJPURKAR ET AL.)

Why CheXpert?

Chest radiography is the most common imaging examination globally, critical for screening, diagnosis, and management of many life threatening diseases.

Leaderboard

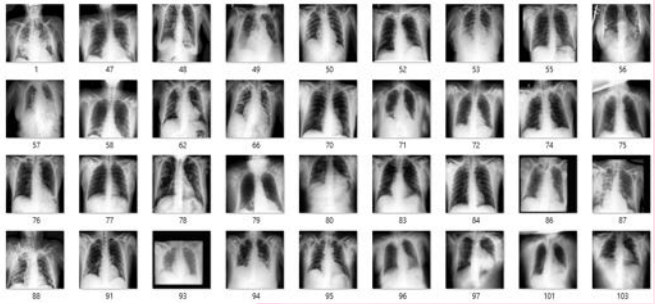
Will your model perform as well as radiologists in detecting different pathologies in chest X-rays?

Rank	Date	Model	AUC	Num Rads Below Curve
1	Aug 31, 2020	DeepAUC-v1 *ensemble*	0.930	2.8
2	Sep 01, 2019	Hierarchical-Learning-V1 *(ensemble) Vingroup Big*	0.930	2.6

Goal

- To build a machine learning model to predict the 30-day mortality of patients
- That is, whether a patient will die in the hospital within 30 days
- Use chest X-ray (CXR) images for model training

About Training Dataset

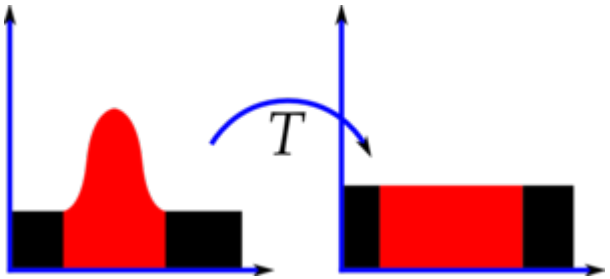


Please keep the data confidential!

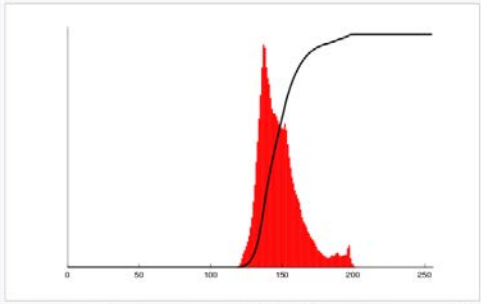
- There are **1,393** patients in the training dataset and each with a CXR JPG image and its file name is an unique patient ID.
- **1,393** hospital outcome for all patients will be provided and used as the labels.
- The size of CXR has been processed to **320 x 320.**
- Each image has been processed with histogram equalization.

PATIENT ID	hospital_outcome
1382	0
1619	1
1710	0
83	0
534	0
1105	0
113	0
1140	0
1460	0
1960	0
989	0
748	0
1564	1
831	0
2324	0
1441	0
2195	0
1653	0
1050	0
693	1
2224	0
2218	0

Histogram equalization

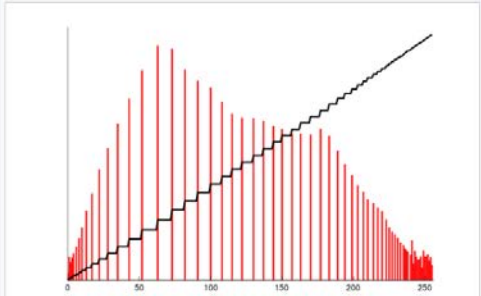


Before Histogram Equalization

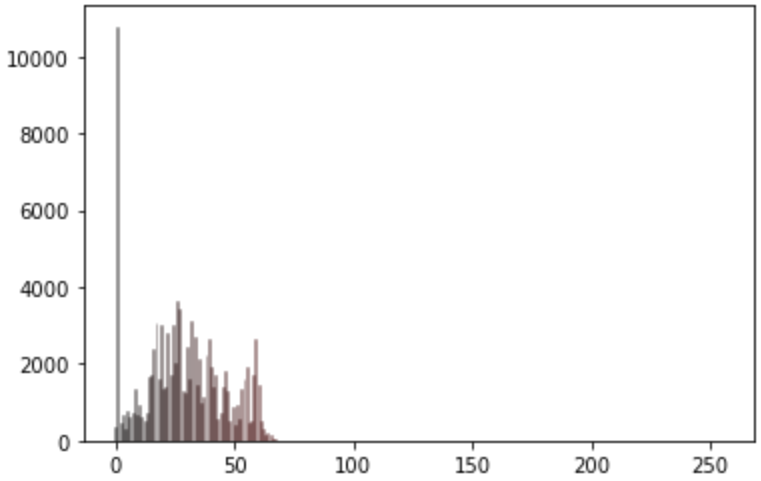


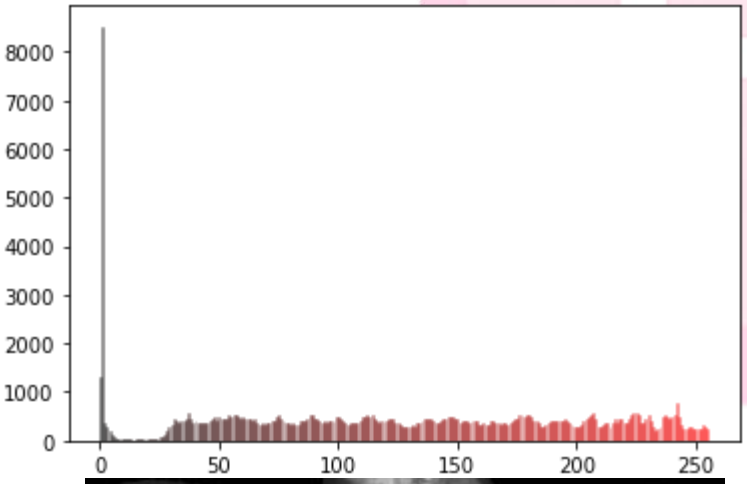
Corresponding histogram (red) and cumulative histogram (black)

After Histogram Equalization



Corresponding histogram (red) and cumulative histogram (black)


10000
8000
6000
4000
2000
0
0
50
100
150
200
250

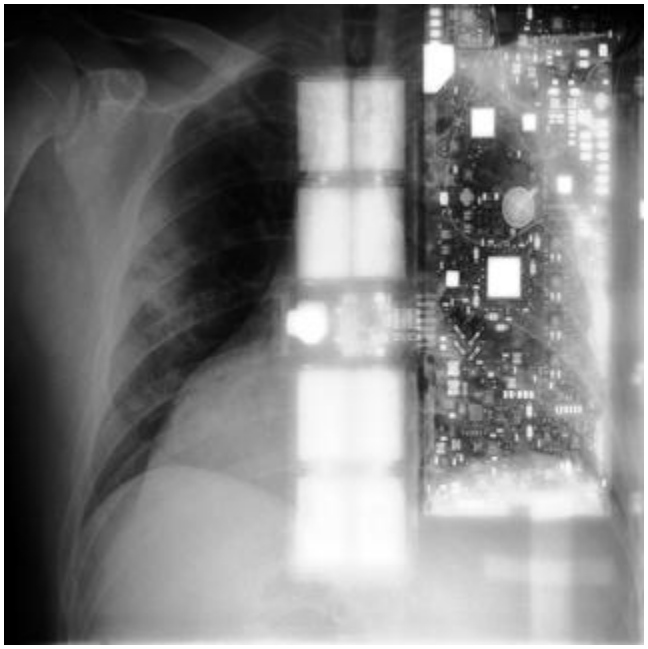

8000
7000
6000
5000
4000
3000
2000
1000
0
0
50
100
150
200
250

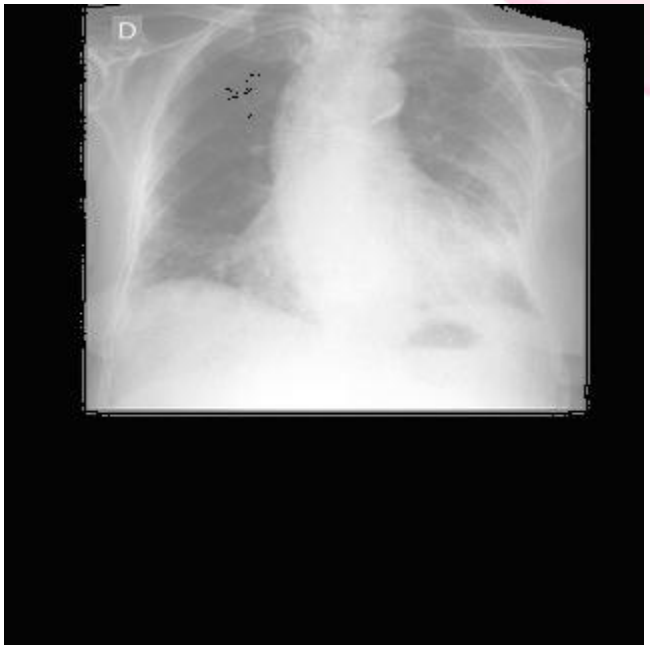
Data description

Alive

Expired

Something you may find





Assignment 3

- Do it individually!
- Announce date: 2020/12/03
- Deadline for prediction, report, and model: 2020/12/31 23:59

  No score if the model is not provided.
- Hand in your 3 files separately in the following format
  - Prediction: student ID.csv
  - Report: student ID_HW3_Report.pdf
  - Model: student ID_HW3_Model

Grading Policy

Item	Score
Report	30%
Model performance	70%

Grading Policy of Model (70%)

- We have **457** patients in the **testing** dataset to evaluate your model
- We will use F1-score to measure your performance

Report (30%)

- ☐ Description of the model and features (or any extra preprocessing steps) you used
- ☐ Description of how to use the model file
- ☐ Summarizing the works you have done

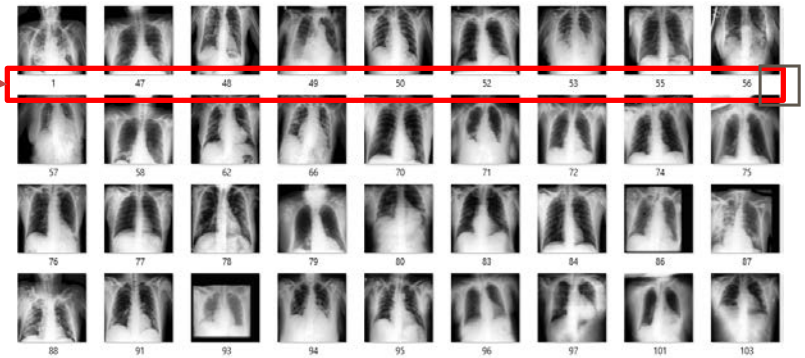
Bonus (+30%)

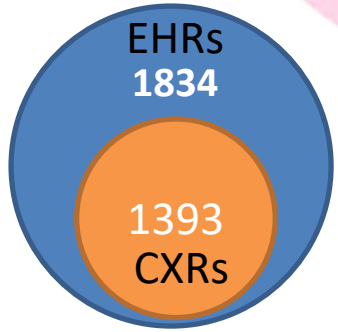
Combine CXRs and EHRs for better prediction!

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	PATIENT ID	age	sex	admission_datetime	ed_diagnosis	vitals_temp_ed_first	vitals_sbp_	vitals_dbp_	vitals_hr_e	vitals_spo2	pmhx_diab	pmhx_hld	pmhx_htn	pmhx_ihd
2	1478	72	MALE	2020/3/18 14:45	sx_breathing_difficulty	37.2	111	63	94	60	0	1	1	1
3	94	50	FEMALE	2020/3/19 19:34	sx_breathing_difficulty	37.2			93	94	0	0	0	0
4	2511	78	FEMALE	2020/4/29 16:04	sx_breathing_difficulty	36.4	146	96	85	96	0	1	0	0
5	1931	50	MALE	2020/3/24 11:47	sx_breathing_difficulty	37.7			124	94	0	0	0	0
6	2070	38	MALE	2020/3/19 12:31	sx_breathing_difficulty						0	0	0	0
7	1231	95	MALE	2020/4/4 00:51	sx_breathing_difficulty		160	11	76	97				
8	2054	83	FEMALE	2020/3/16 19:07	sx_others	38.5	138	65	80	73	0	0	0	0
9	2451	79	MALE	2020/4/27 15:15	sx_breathing_difficulty		143	86	111	89	0	0	0	0
10	952	55	MALE	2020/3/20 16:36	sx_flu						0	0	0	0
11	2270	76	MALE	2020/4/10 16:46	sx_others	35.6	160	95	55	99	1	0	1	0
12	778	49	MALE	2020/3/27 20:26	sx_fever						0	0	0	0
13	1965	79	FEMALE	2020/3/26 17:35	sx_breathing_difficulty						1	1	0	0
14	879	81	MALE	2020/3/10 10:45	sx_others	36	105	50	80		0	0	0	0
15	722	70	MALE	2020/3/23 10:27	sx_breathing_difficulty	36			83	95	0	0	0	0

Patient ID

+



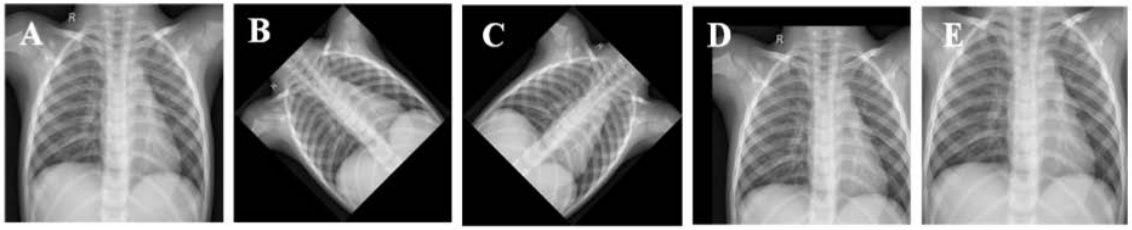


Hand in your 2 extra files

- ☐ Prediction: Bonus_student ID.csv
- ☐ Model: Bonus_ student ID_HW3_Model

Possible solution

- [ ] **CNN-based model**
- [ ] **Data augmentation**



- [ ] **Transfer learning (or using pretrained model)**
- [ ] **Some useful datasets**
  - [ ] https://github.com/ieee8023/covid-chestxray-dataset
  - [ ] https://stanfordmlgroup.github.io/competitions/chexpert/
  - [ ] https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community

Let's do it!