Application of Artificial Intelligence to Help Detect Rib Fractures in Radiographs

Gaoxiang Luo*, Andrew Walker*, Le Peng, Christopher Tignanelli, Ju Sun

University of Minnesota
Driven to Discover℠

AI diversity thrives

MIT Technology Review

Health-care institutions are looking an assortment of technologies to achieve better health outcomes, with keen interest in electronic-health-record automation, medical imaging and diagnostics, and patient data and risk analytics.

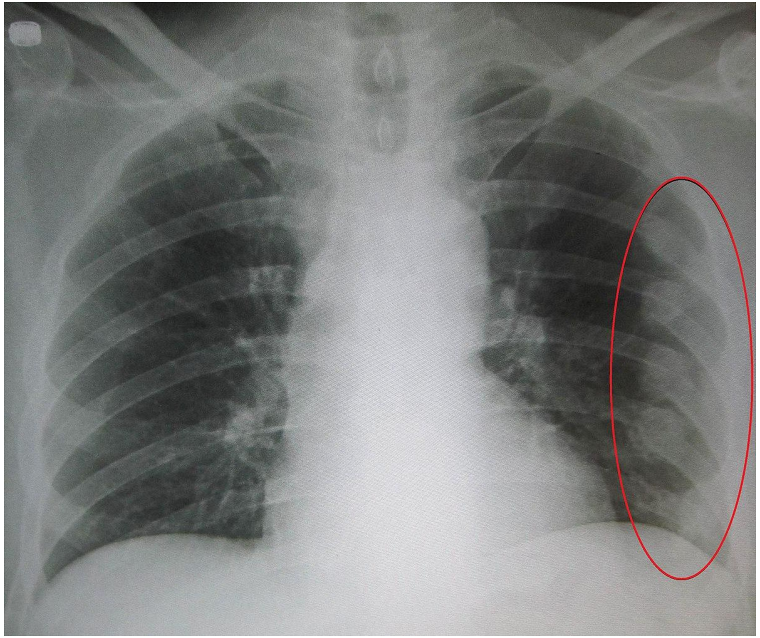
Technology	Adopted	Considering adoption	Total interest
Automation of electronic health records	43%	20%	63%
Medical imaging and diagnostics	41%	23%	64%
Patient data and risk analytics	41%	21%	62%
AI for predictive analytics	40%	23%	63%
AI for patient flow optimization	39%	26%	65%
Virtual nursing assistants	25%	29%	54%
AI-assisted endoscopy	24%	21%	45%
Surgical analytics	23%	23%	46%
Robot-assisted surgery	22%	24%	46%
Analytics for mental health	21%	27%	48%

Source: MIT Technology Review Insights' survey on AI in health care of 908 health-care professionals in the US and UK, fall 2019

40-80%

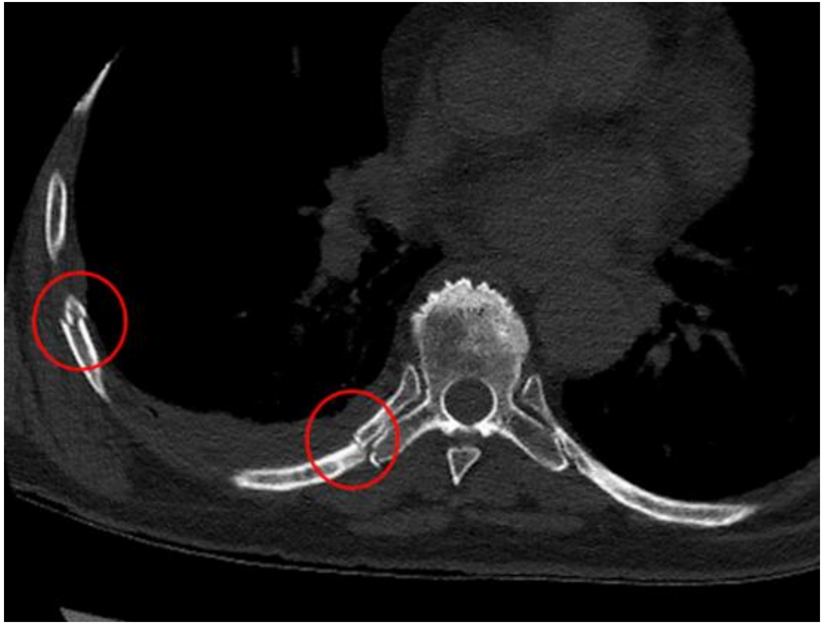
chest trauma patients having **rib fractures** [Lin, 2018]

Rib Fracture



X-ray

Cheap but less reliable (miss >85%)



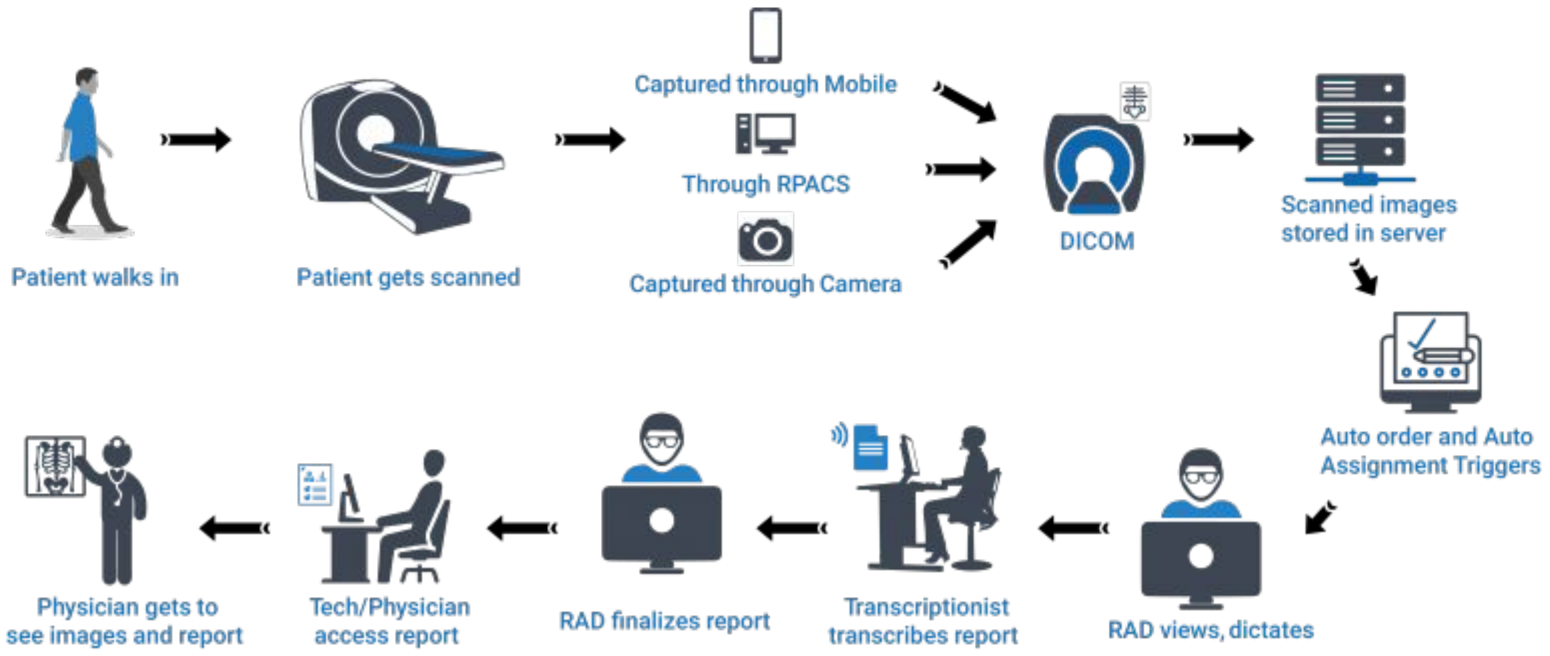
CT

Expensive but more reliable (miss 20.5%)
[Cho, 2012]

Motivation 1: AI-assisted Diagosis

- Mortality rate of 20% for patients over 65 [Peek, 2020]
- Pose the risks of getting a collapsed lung or internal bleeding [Tignanelli, 2020]
- Poorest risk-adjusted outcomes [Macheel, 2020]

Motivation 2: Efficient Healthcare System



Problem Definition

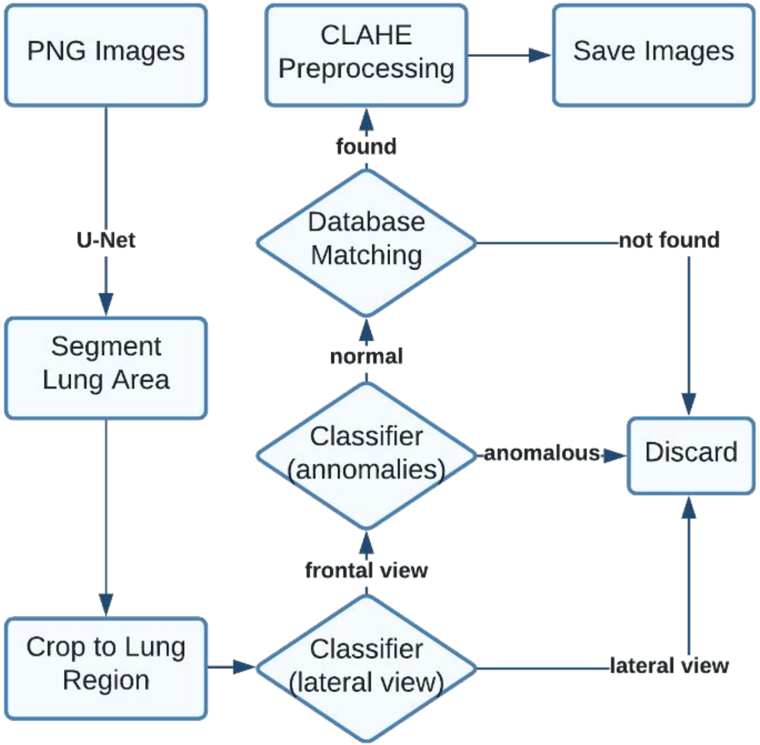
- Input
  - High-Resolution Chest X-rays and Electronic Health Record (EHR)
  - Positive/Negative Labels
- Output
  - A predictive model for the new/incoming chest x-rays and EHR
- Objective
  - High AUPRC (the area under the precision-recall curve) and High AUROC

The AUPRC is a useful performance metric for imbalanced data in a problem setting where you care a lot more finding the positive examples.

- Constrains
  - Fractures are usually in a small and elongated shape, which is more observable in CT scans rather than 2D images such as x-rays.
  - Computational constraints.

Data Preparation

This module ensures the quality of the input data by removing data that is incorrectly formatted. It detects the lung region on the x-ray with a U-Net [Ronneberger, 2015] and crops to that area before feeding into the following submodules. From here, a Resnet-18 model is used to discard lateral view images, a conditional-GAN [Mirza, 2014] is used to discard other anomalous images, and the label is double-checked against the database.



CLAHE Transform [Saix, 2020]

Original

CLAHE

Reason to use: 4% of AUPRC increase on rib fracture classification in CheXpert dataset.

Center Crop

before

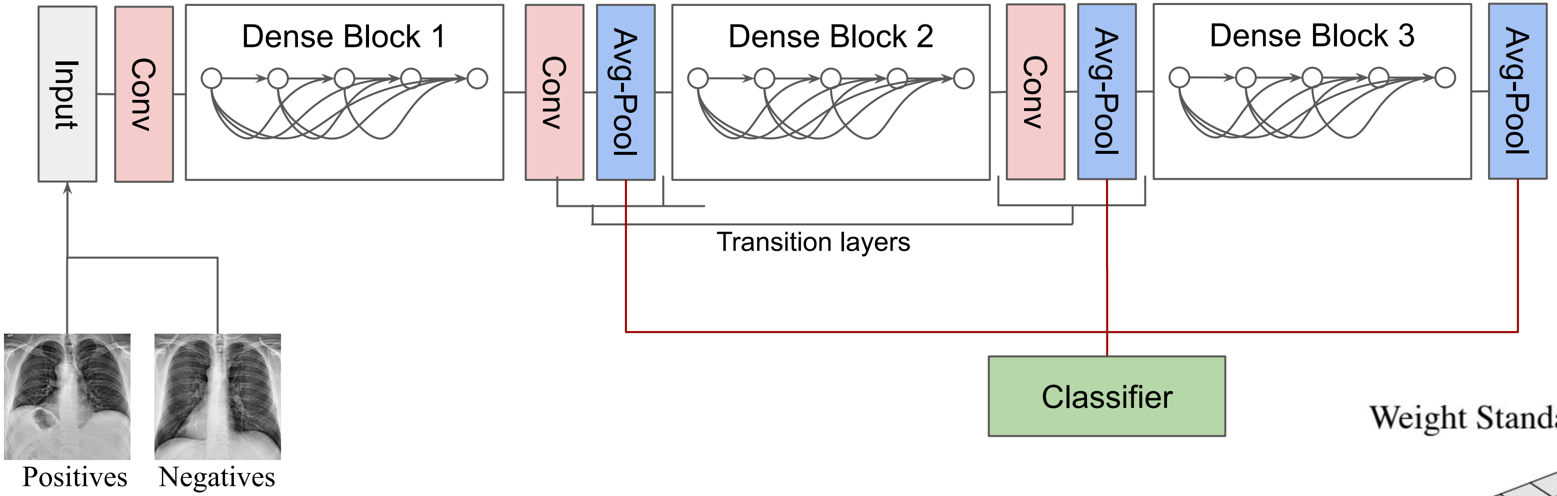
after

Dataset

	Presence of Rib Fracture	N	Age Median (IQR: 25th-75th)	Male %
Train	Positive	1,990	63 (51-77)	57.3%
	Negative	11,097	57 (40-72)	47.0%
Validation	Positive	272	62 (49-77)	60.7%
	Negative	1,569	58 (41-72)	45.6%
Hold-out Test	Positive	551	62 (50-79)	55.0%
	Negative	3,187	57 (40-72)	48.5%

Imbalance Ratio = 1:6 (reflecting the real-world distribution in clinical setting)

Image Model



Weight Standardization

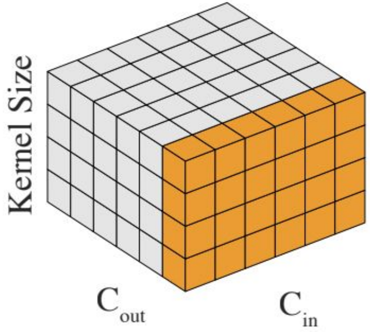
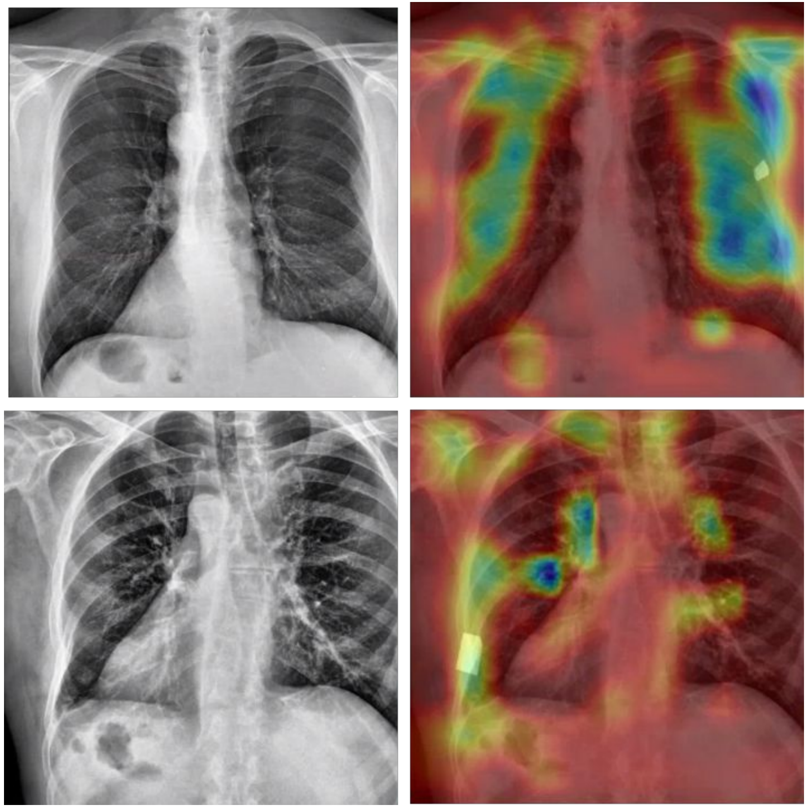


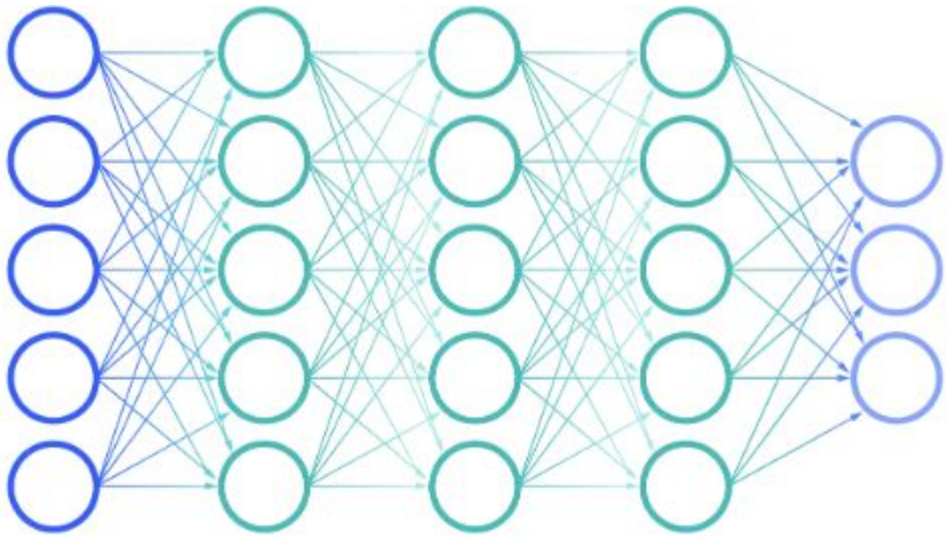
Image Model Results

Visualize Learned Features



The heatmaps generated from the last activation layer of DenseNet121 using GradCAM++ [Chattopadhay, 2018]. The blue areas are the regions of high responses (i.e., the model pays attention in decision making). The white area is where the rib fracture is located, annotated by a radiologist.

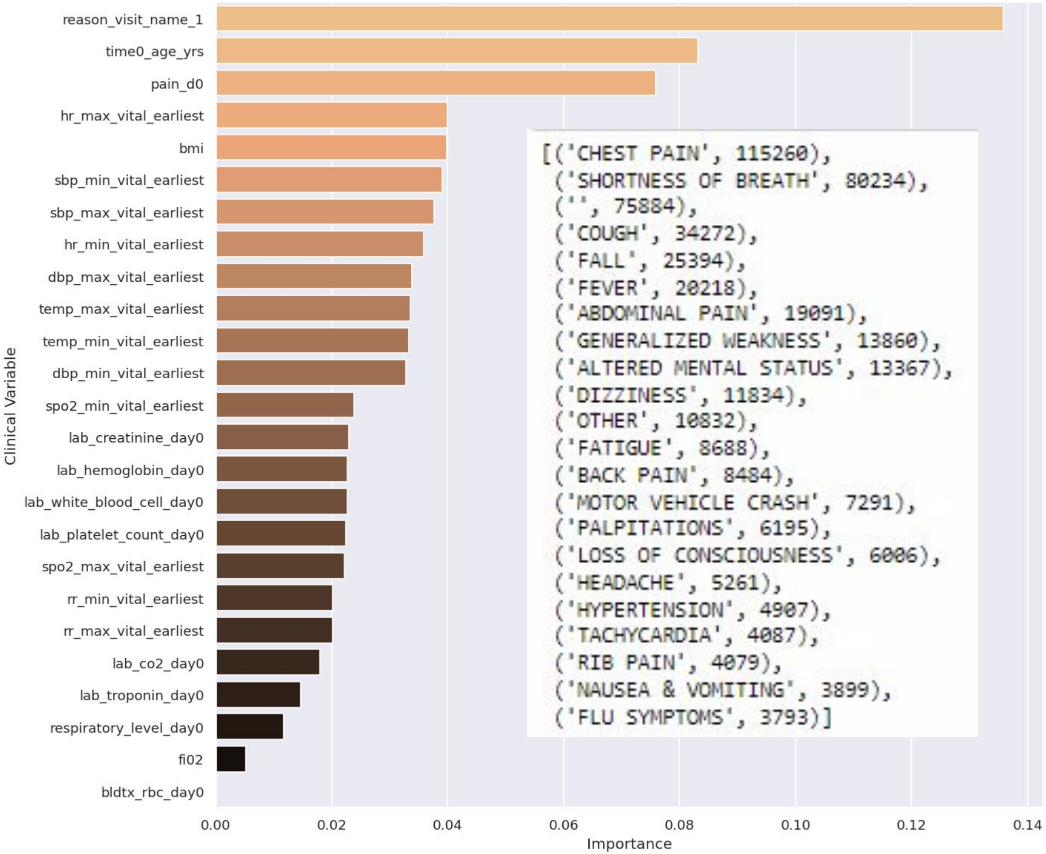
Tabular Model



Clinical Variables
bmi
hr_min_vital*
lab_co2
rr_min_vital*
lab_creatinine
sbp_min_vital*
lab_hemoglobin
dbp_min_vital*
lab_platelet_count
temp_min_vital*
lab_troponin
spo2_min_vital*
lab_white_blood_cell
hr_max_vital*
bldtx_rbc
rr_max_vital*
reason for visit[1]
sbp_max_vital*
respiratory_level
dbp_max_vital*
pain
temp_max_vital*
age
spo2_max_vital*
fio2[2]
reason_visit_name_1

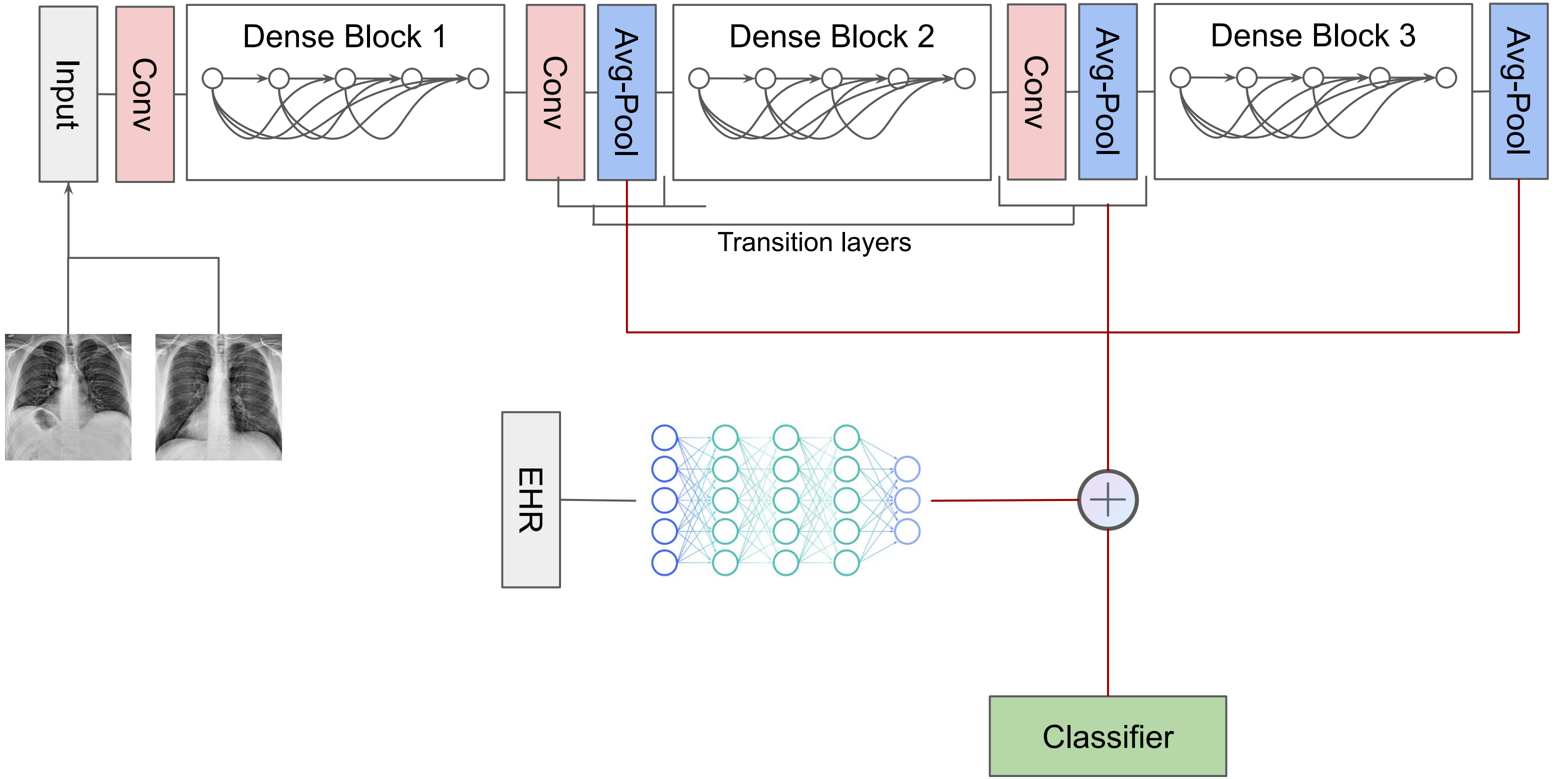
Tabular Model Results

AUROC	0.9099
AUPRC	0.6613




```
[('CHEST PAIN', 115260),
 ('SHORTNESS OF BREATH', 80234),
 ('', 75884),
 ('COUGH', 34272),
 ('FALL', 25394),
 ('FEVER', 20218),
 ('ABDOMINAL PAIN', 19091),
 ('GENERALIZED WEAKNESS', 13860),
 ('ALTERED MENTAL STATUS', 13367),
 ('DIZZINESS', 11834),
 ('OTHER', 10832),
 ('FATIGUE', 8688),
 ('BACK PAIN', 8484),
 ('MOTOR VEHICLE CRASH', 7291),
 ('PALPITATIONS', 6195),
 ('LOSS OF CONSCIOUSNESS', 6006),
 ('HEADACHE', 5261),
 ('HYPERTENSION', 4907),
 ('TACHYCARDIA', 4087),
 ('RIB PAIN', 4079),
 ('NAUSEA & VOMITING', 3899),
 ('FLU SYMPTOMS', 3793)]
```

Multimodal Model



Multimodal Model Results

	AUROC	AUPRC	PPV	Sens	Spec	NPV	F1
Image-only	0.77~0.80	0.42~0.46	0.28	0.80	0.64	0.95	0.41
EHR-only	0.90~0.92	0.64~0.68	0.48	0.83	0.84	0.97	0.61
Multimodal	0.90~0.93	0.67~0.71	0.49	0.84	0.85	0.97	0.62

	Internal Validation	
Ratio	AUROC	AUPRC
1	0.8866~0.9232	0.8785~0.9165
2	0.9033~0.9343	0.8421~0.8818
5	0.9001~0.9271	0.6963~0.7381

after running the Youden Index to decide the optimal threshold for binary prediction

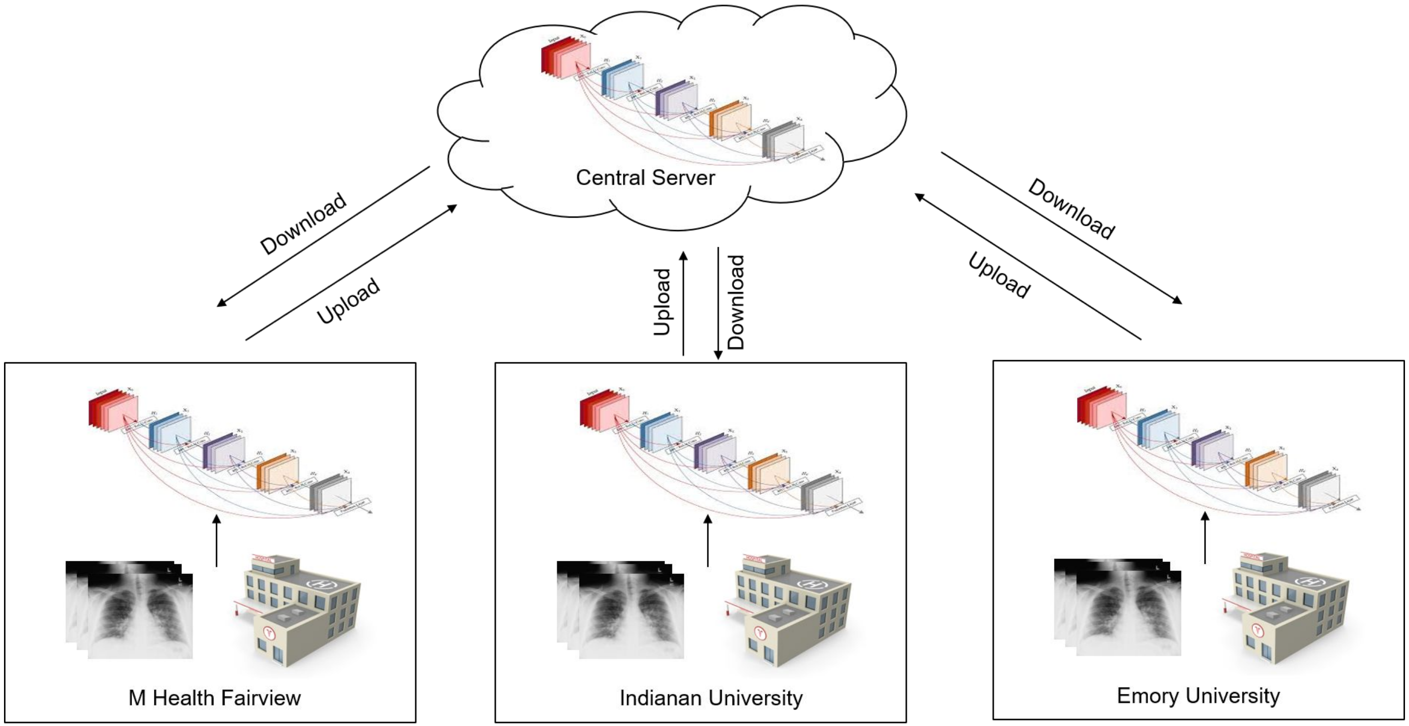
Subgroup Analysis

	N	AUROC	AUPRC
Gender			
Male	1889	0.9157~0.9303	0.7297~0.7796
Female	1849	0.8773~0.9139	0.5749~0.6324
Race			
White	3060	0.9040~0.9323	0.7008~0.7462
Black	375	0.7941~0.9146	0.1986~0.3172
Asian	85	0.8727~0.9468	0.6302~0.9053
Others	45	0.7617~0.9938	0.3540~0.6948

Delong Test

Test on	AUC-ROC		P-value
	Image-only Model	Multimodal model	
Internal	0.7845	0.9146	p<0.05
	EHR-only Model	Multimodal model	
Internal	0.9099	0.9146	0.65

Federated Learning



Acknowledgement

We thank Cisco Research for funding this project, and we thank M Health Fairview for providing the dataset and the GPU server. We thank Le Peng, Hengyue Liang, and Prof. Ju Sun (advisor) from GLOVEX at the University of Minnesota for providing technical assistance and mentoring.