Finding the correct answer (Ground Truths)

Dr. Judy Gichoya
Associate Professor of Radiology and Informatics
Emory University

HITI LAB

EMORY UNIVERSITY

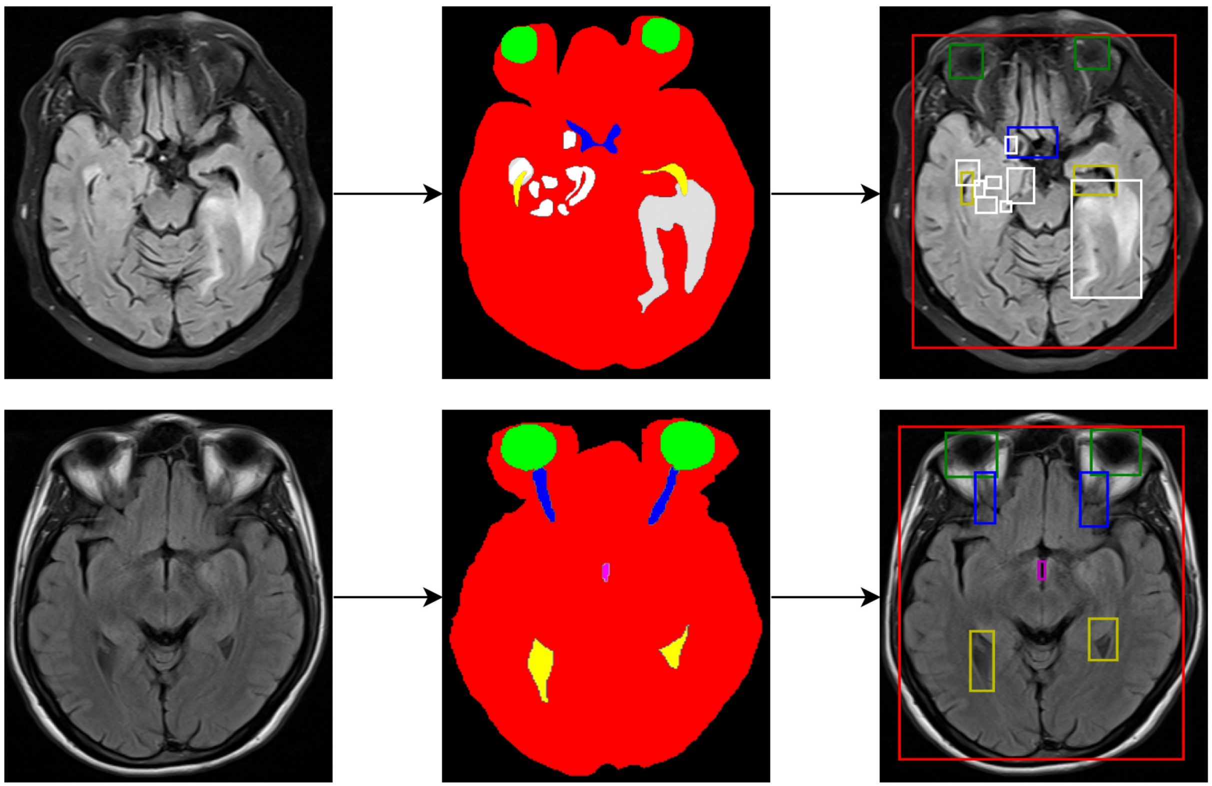
Disclosures

- ACR
  - AI advisory council
- RSNA
  - Associate Editor - Radiology AI Trainee Editorial Board
  - CIRE & RIC Committee member
- SIIM
  - Co-chair – Research Committee
  - Board member
- HL7 and AHLI Board member
  - Association for Health Learning and Inference
- Softbrew LTD
  - Consulting on Global Health /Clinical informatics

- Funding
  - NIBIB MIDRC / COVID -19 Data repository
  - Clairity Consortium
  - NIH AIM AHEAD pilot grant
  - NIH AIM-AHEAD consortium development program
  - RSNA Health disparities grant
  - DeepLook grant for AI validation
  - Harold Amos Faculty Award to study AI bias
  - Lacuna fund for creating diverse medical datasets
  - LUNIT for breast DBT evaluation
  - R01 NHLBI Grant for opportunistic screening for ASCVD using multimodal AI
  - Winship invest cancer disparities pilot grant

Last updated May 29th ,2024

DL tasks for computer vision



(a) (b) (c)

History of Labeling CXRs

ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases

Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald M. Summers

The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. A tremendous number of X-ray imaging studies accompanied by radiological reports are accumulated and stored in many modern hospitals' Picture Archiving and Communication Systems (PACS). On the other side, it is still an open question how this type of hospital-size knowledge database containing invaluable imaging informatics (i.e., loosely labeled) can be used to facilitate the data-hungry deep learning paradigms in building truly large-scale high precision computer-aided diagnosis (CAD) systems.

In this paper, we present a new chest X-ray database, namely "ChestX-ray8", which comprises 108,948 frontal-view X-ray images of 32,717 unique patients with the text-mined eight disease image labels (where each image can have multi-labels), from the associated radiological reports using natural language processing. Importantly, we demonstrate that these commonly occurring thoracic diseases can be detected and even spatially-located via a unified weakly-supervised multi-label image classification and disease localization framework, which is validated using our proposed dataset. Although the initial quantitative results are promising as reported, deep convolutional neural network based "reading chest X-rays" (i.e., recognizing and locating the common disease patterns trained with only image-level labels) remains a strenuous task for fully-automated high precision CAD systems. Data download link: this https URL

History of Labeling CXRs

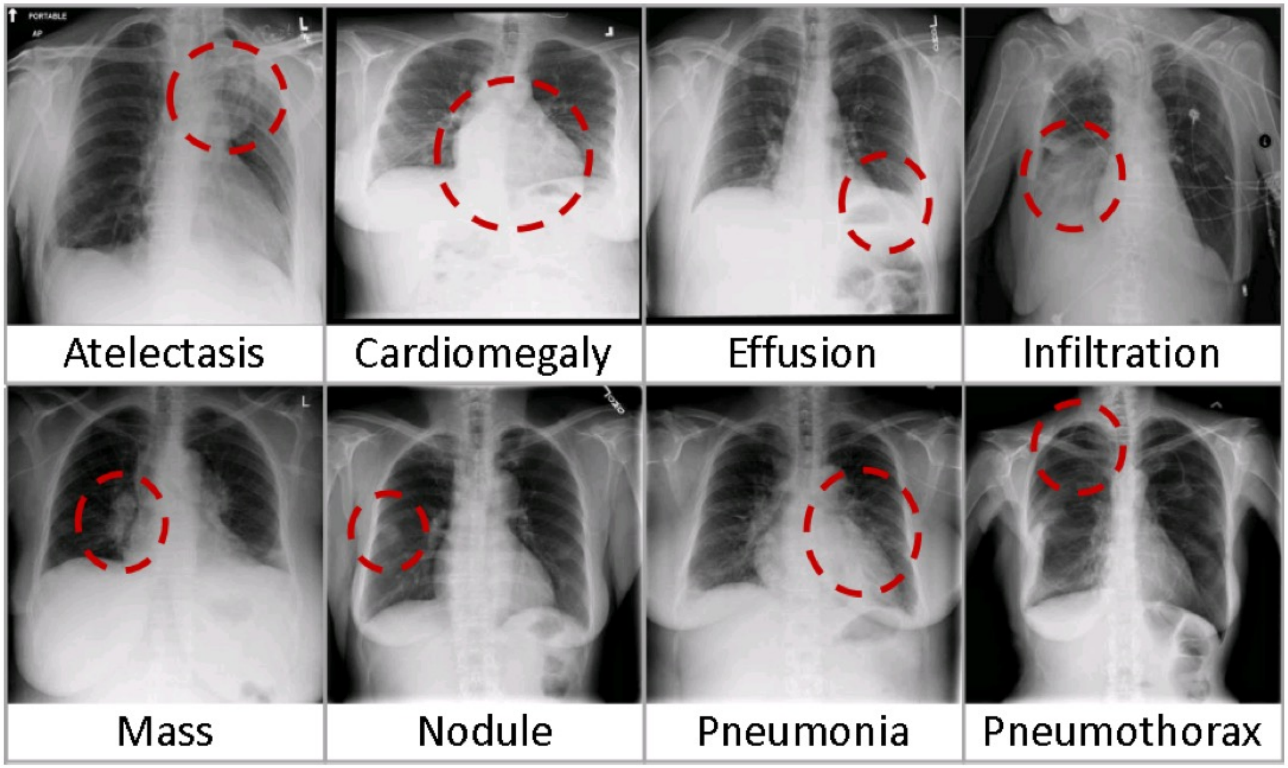
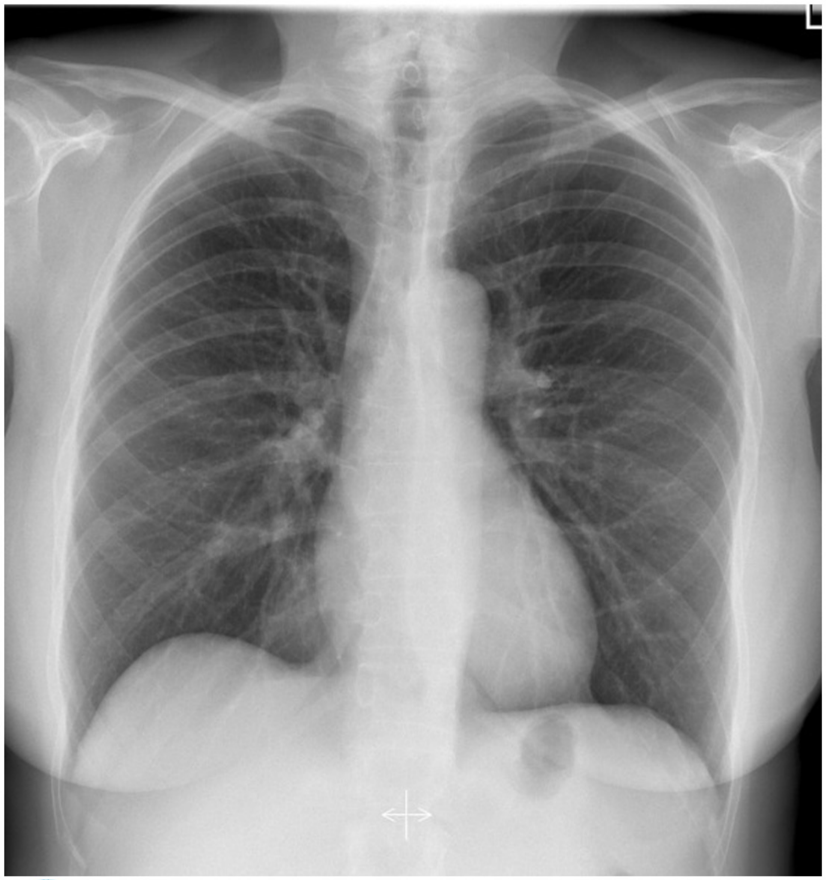


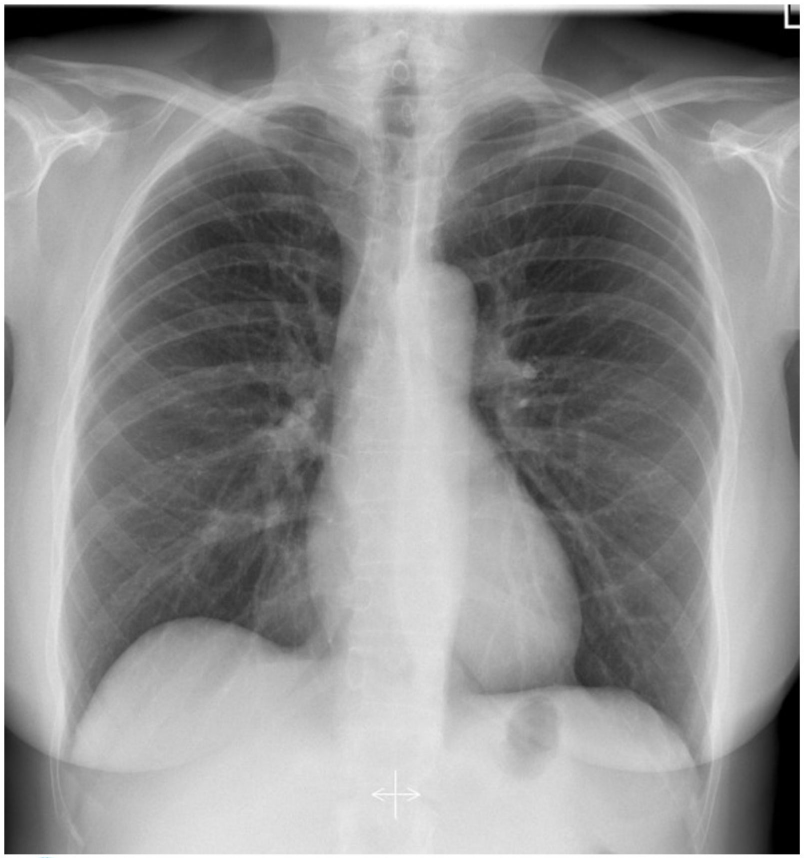
Figure 1. *Eight common thoracic diseases observed in chest X-rays that validate a challenging task of fully-automated diagnosis.*

CXR Report



STUDY:Chest radiograph, AP and lateral views
INDICATION: Pneumothorax?
PRIORS: AP view taken one week ago
FINDINGS: The lungs are well-expanded and no sign of focal consolidation, effusion, or pneumothorax. Cardiac silhouette is within …
IMPRESSION: No acute cardiopulmonary process identified.

CXR report



STUDY:Chest radiograph, AP and lateral views
INDICATION: Pneumothorax?
PRIORS: AP view taken one week ago
FINDINGS: The lungs are well-expanded and no sign of focal consolidation, effusion, or pneumothorax. Cardiac silhouette is within …
IMPRESSION: No acute cardiopulmonary process identified.

Methods of Labeling

CUI	Concept
Atelectasis	
C0004144	atelectasis
C0264494	discoid atelectasis
C0264496	focal atelectasis
Cardiomegaly	
C0018800	cardiomegaly
Effusion	
C0013687	effusion
C0031039	pericardial effusion
C0032227	pleural effusion disorder
C0747635	bilateral pleural effusion
C0747639	loculated pleural effusion
Pneumonia	
C0032285	pneumonia
C0577702	basal pneumonia
C0578576	left upper zone pneumonia
C0578577	right middle zone pneumonia
C0585104	left lower zone pneumonia
C0585105	right lower zone pneumonia
C0585106	right upper zone pneumonia
C0747651	recurrent aspiration pneumonia
C1960024	lingular pneumonia
Pneumothorax	
C0032326	pneumothorax
C0264557	chronic pneumothorax
C0546333	right pneumothorax
C0546334	left pneumothorax

Table 5. *Sample Target Diseases and their corresponding concept and identifiers (CUIs) in SNOMED-CT.*

- Pathology labeling –
  - Dnorm -> SNOMED
  - MetaMap -> UMLS
- Negation and uncertainity

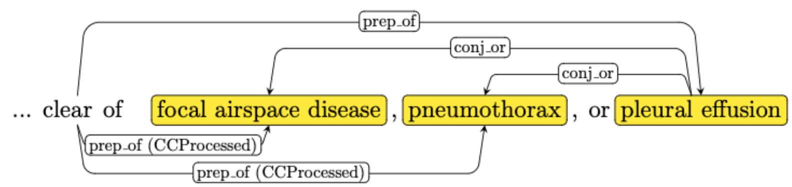
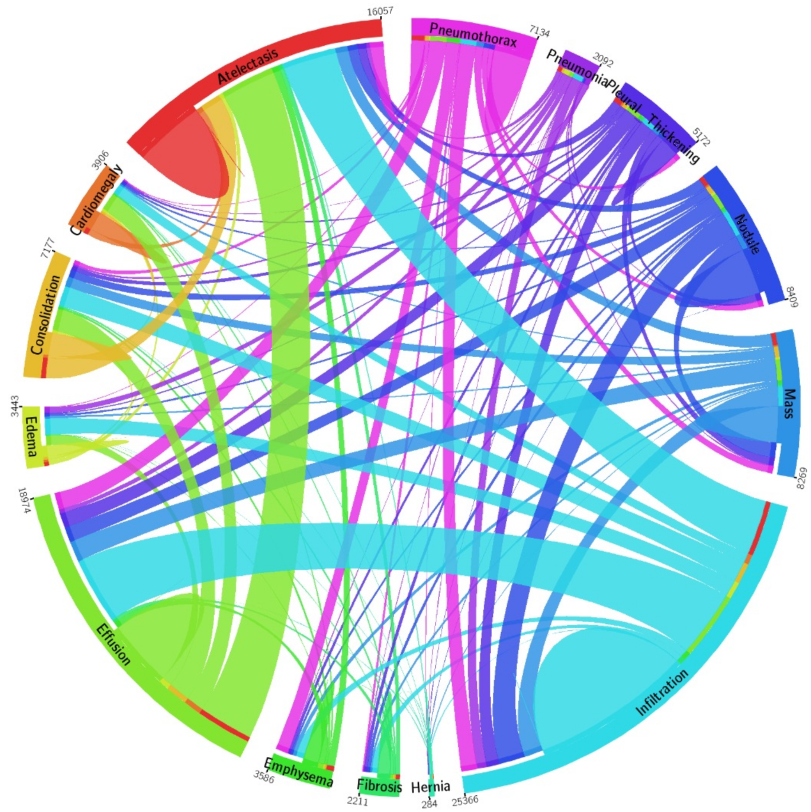


Figure 3. *The dependency graph of text: "clear of focal airspace disease, pneumothorax, or pleural effusion".*

Normal CXR


Atelectasis
16057
Pneumothorax
7134
Pneumonia
2092
Pleural_Thickening
5172
Nodule
8409
Mass
8269
Infiltration
25366
Hernia
284
Fibrosis
2211
Emphysema
3586
Effusion
18974
Edema
3443
Consolidation
7177
Cardiomegaly
3906

QC for Labels

Disease	MetaMap	Our Method
	P / R / F	P / R / F
Atelectasis	0.95 / 0.95 / 0.95	0.99 / 0.85 / 0.91
Cardiomegaly	0.99 / 0.83 / 0.90	1.00 / 0.79 / 0.88
Effusion	0.74 / 0.90 / 0.81	0.93 / 0.82 / 0.87
Infiltration	0.25 / 0.98 / 0.39	0.74 / 0.87 / 0.80
Mass	0.59 / 0.67 / 0.62	0.75 / 0.40 / 0.52
Nodule	0.95 / 0.65 / 0.77	0.96 / 0.62 / 0.75
Normal	0.93 / 0.90 / 0.91	0.87 / 0.99 / 0.93
Pneumonia	0.58 / 0.93 / 0.71	0.66 / 0.93 / 0.77
Pneumothorax	0.32 / 0.82 / 0.46	0.90 / 0.82 / 0.86
Total	0.84 / 0.88 / 0.86	0.90 / 0.91 / 0.90

Table 2. *Evaluation of image labeling results on OpenI dataset. Performance is reported using P, R, F1-score.*

Bounding Box

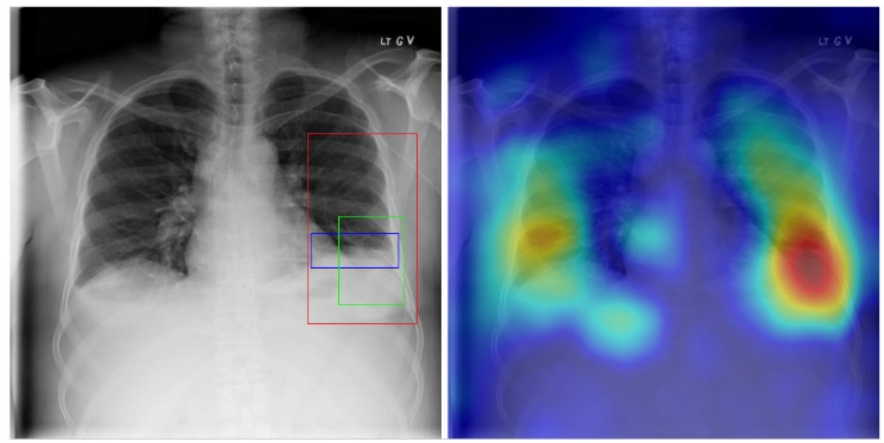
Radiology report	Keyword	Localization Result
findings include: 1. left basilar atelectasis/consolidation. 2. prominent hilum (mediastinal adenopathy). 3. left pic catheter (tip in atriocaval junction). 4. stable, normal appearing cardiomediastinal silhouette. impression: small right pleural effusion otherwise stable abnormal study including left basilar infiltrate/atelectasis, prominent hilum, and position of left pic catheter (tip atriocaval junction).	Effusion; Infiltration; Atelectasis	

Table 8. *A sample of chest x-ray radiology report, mined disease keywords and localization result from the "Atelectasis" Class. Correct bounding box (in green), false positives (in red) and the ground truth (in blue) are plotted over the original image.*

CXR 8 -> CXR 14

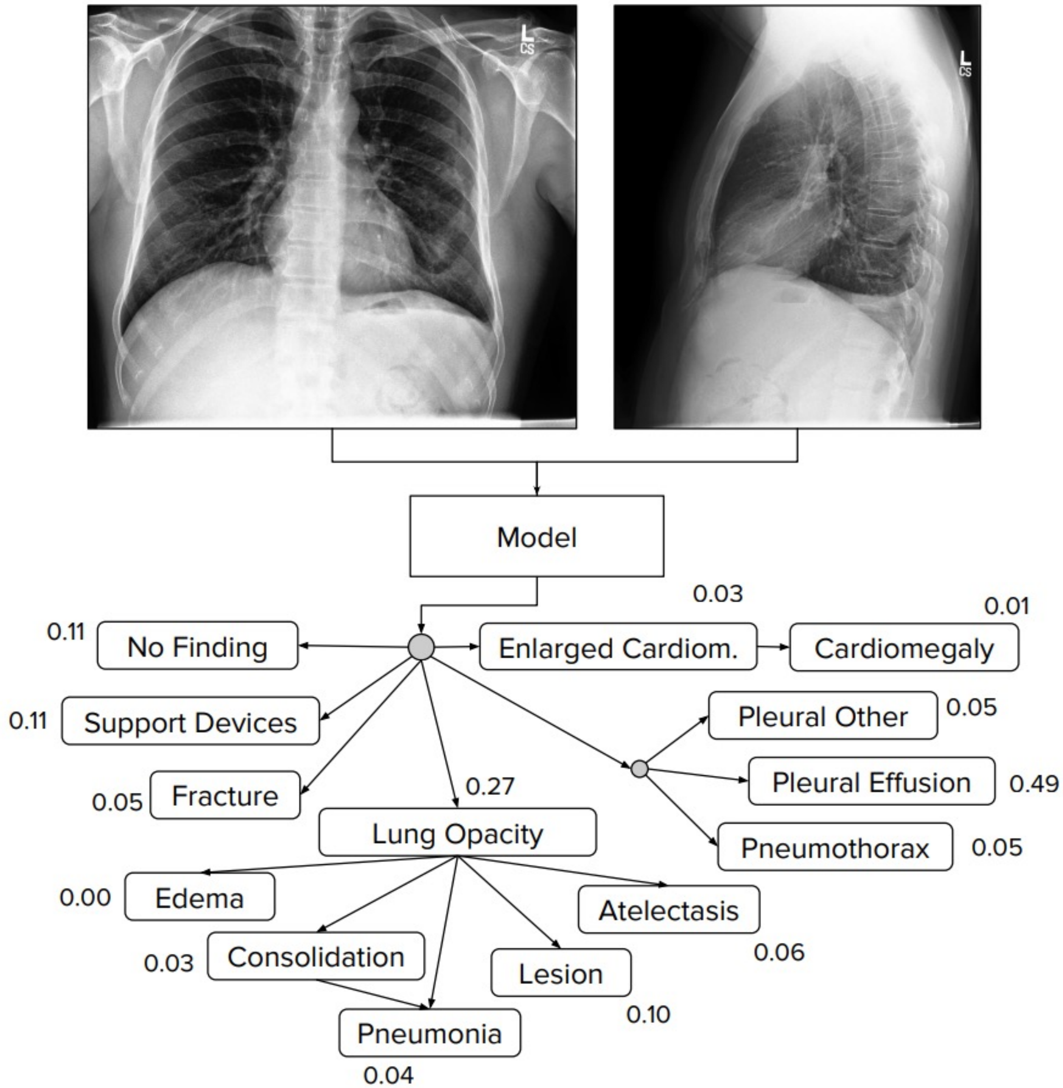
Item #	X-ray8	Ov.	X-ray14	Ov.
Report	108,948	-	112,120	-
Atelectasis	5,789	3,286	11,535	7,323
Cardiomegaly	1,010	475	2,772	1,678
Effusion	6,331	4,017	13,307	9,348
Infiltration	10,317	4,698	19,871	10,319
Mass	6,046	3,432	5,746	2,138
Nodule	1,971	1,041	6,323	3,617
Pneumonia	1,062	703	1,353	1,046
Pneumothorax	2,793	1,403	5,298	3,099
Consolidation	-	-	4,667	3,353
Edema	-	-	2,303	1,669
Emphysema	-	-	2,516	1,621
Fibrosis	-	-	1,686	959
PT	-	-	3,385	2,258
Hernia	-	-	227	117
No findings	84,312	0	60,412	0

Table 16. *Total number (#) and # of Overlap (Ov.) of the corpus in ChestX-ray8 and ChestX-ray14 datasets.PT: Pleural Thickening*

Chexpert

CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison

Jeremy Irvin,[1,*] Pranav Rajpurkar,[1,*] Michael Ko,[1] Yifan Yu,[1]
Silviana Ciurea-Ilcus,[1] Chris Chute,[1] Henrik Marklund,[1] Behzad Haghgoo,[1]
Robyn Ball,[2] Katie Shpanskaya,[3] Jayne Seekins,[3] David A. Mong,[3]
Safwan S. Halabi,[3] Jesse K. Sandberg,[3] Ricky Jones,[3] David B. Larson,[3]
Curtis P. Langlotz,[3] Bhavik N. Patel,[3] Matthew P. Lungren,[3,†] Andrew Y. Ng[1,†]
[1]Department of Computer Science, Stanford University
[2]Department of Medicine, Stanford University
[3]Department of Radiology, Stanford University
[*]Equal contribution
[†]Equal contribution
{jirvin16, pranavsr}@cs.stanford.edu



L
CS
L
CS
Model
0.11
No Finding
0.03
Enlarged Cardiom.
0.01
Cardiomegaly
0.11
Support Devices
Pleural Other
0.05
Pleural Effusion
0.49
0.05
Fracture
0.27
Lung Opacity
Pneumothorax
0.05
0.00
Edema
Atelectasis
0.06
0.03
Consolidation
Lesion
0.10
Pneumonia
0.04

Pathology	Positive (%)	Uncertain (%)	Negative (%)
No Finding	16627 (8.86)	0 (0.0)	171014 (91.14)
Enlarged Cardiom.	9020 (4.81)	10148 (5.41)	168473 (89.78)
Cardiomegaly	23002 (12.26)	6597 (3.52)	158042 (84.23)
Lung Lesion	6856 (3.65)	1071 (0.57)	179714 (95.78)
Lung Opacity	92669 (49.39)	4341 (2.31)	90631 (48.3)
Edema	48905 (26.06)	11571 (6.17)	127165 (67.77)
Consolidation	12730 (6.78)	23976 (12.78)	150935 (80.44)
Pneumonia	4576 (2.44)	15658 (8.34)	167407 (89.22)
Atelectasis	29333 (15.63)	29377 (15.66)	128931 (68.71)
Pneumothorax	17313 (9.23)	2663 (1.42)	167665 (89.35)
Pleural Effusion	75696 (40.34)	9419 (5.02)	102526 (54.64)
Pleural Other	2441 (1.3)	1771 (0.94)	183429 (97.76)
Fracture	7270 (3.87)	484 (0.26)	179887 (95.87)
Support Devices	105831 (56.4)	898 (0.48)	80912 (43.12)

1. *unremarkable* cardiomediastinal silhouette

2. diffuse reticular pattern, which can be seen with an atypical infection **or** chronic fibrotic change. *no* focal consolidation.

3. *no* pleural effusion or pneumothorax

4. mild degenerative changes in the lumbar spine and old right rib fractures.

Observation	**Labeler Output**
No Finding	
Enlarged Cardiom.	0
Cardiomegaly	
Lung Opacity	1
Lung Lesion	
Edema	
Consolidation	0
Pneumonia	u
Atelectasis	
Pneumothorax	0
Pleural Effusion	0
Pleural Other	
Fracture	1
Support Devices	

Method

- Mention Extraction - Impression
- Mention classification - Negative ,Uncertain or positive
- Mention aggregation

Category	Mention F1 NIH	Mention F1 Ours	Negation F1 NIH	Negation F1 Ours	Uncertain F1 NIH	Uncertain F1 Ours
Atelectasis	0.976	**0.998**	0.526	**0.833**	0.661	**0.936**
Cardiomegaly	0.647	**0.973**	0.000	**0.909**	0.211	**0.727**
Consolidation	0.996	**0.999**	0.879	**0.981**	0.438	**0.924**
Edema	0.978	**0.993**	0.873	**0.962**	0.535	**0.796**
Pleural Effusion	0.985	**0.996**	0.951	**0.971**	0.553	**0.707**
Pneumonia	0.660	**0.992**	0.703	**0.750**	0.250	**0.817**
Pneumothorax	0.993	**1.000**	0.971	**0.977**	0.167	**0.762**
Enlarged Cardiom.	N/A	0.935	N/A	0.959	N/A	0.854
Lung Lesion	N/A	0.896	N/A	0.900	N/A	0.857
Lung Opacity	N/A	0.966	N/A	0.914	N/A	0.286
Pleural Other	N/A	0.850	N/A	1.000	N/A	0.769
Fracture	N/A	0.975	N/A	0.807	N/A	0.800
Support Devices	N/A	0.933	N/A	0.720	N/A	N/A
No Finding	N/A	0.769	N/A	N/A	N/A	N/A
Macro-average	N/A	0.948	N/A	0.899	N/A	0.770
Micro-average	N/A	0.969	N/A	0.952	N/A	0.848

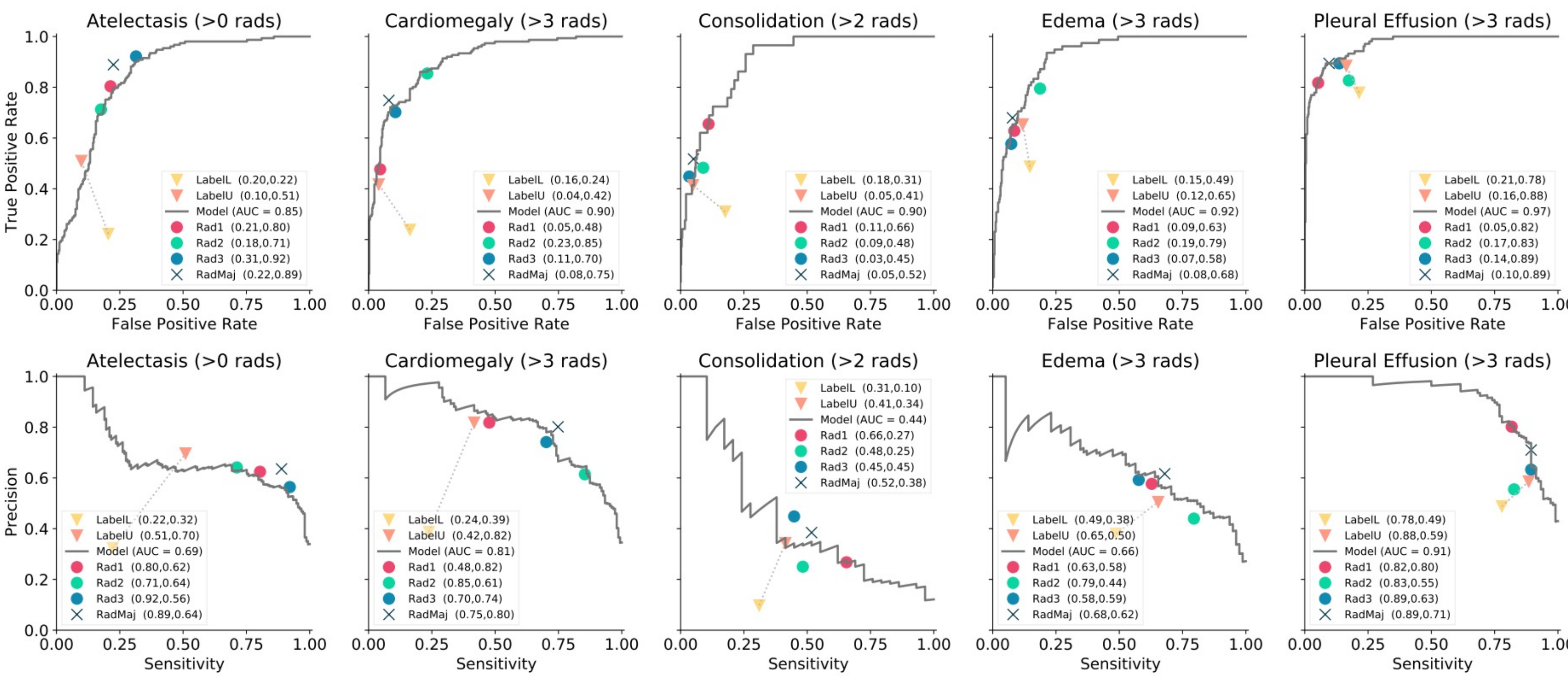
Label Uncertainty

	Atelectasis	Cardiomegaly	Consolidation	Edema	Pleural Effusion
U-Ignore	0.818 (0.759,0.877)	*0.828 (0.769,0.888)*	0.938 (0.905,0.970)	0.934 (0.893,0.975)	0.928 (0.894,0.962)
U-Zeros	*0.811 (0.751,0.872)*	0.840 (0.783,0.897)	0.932 (0.898,0.966)	0.929 (0.888,0.970)	0.931 (0.897,0.965)
U-Ones	**0.858 (0.806,0.910)**	0.832 (0.773,0.890)	0.899 (0.854,0.944)	0.941 (0.903,0.980)	0.934 (0.901,0.967)
U-SelfTrained	0.833 (0.776,0.890)	0.831 (0.770,0.891)	0.939 (0.908,0.971)	0.935 (0.896,0.974)	0.932 (0.899,0.966)
U-MultiClass	0.821 (0.763,0.879)	**0.854 (0.800,0.909)**	0.937 (0.905,0.969)	0.928 (0.887,0.968)	0.936 (0.904,0.967)

Cardiomegaly

- Most cases are borderline – “minimal cardiac enlargement”
- Difficult to categorize when the heart is mentioned “unchanged appearance of the heart” or “stable cardiac contours”

Label QC


Atelectasis (>0 rads)
True Positive Rate
False Positive Rate
LabelL (0.20,0.22)
LabelU (0.10,0.51)
Model (AUC = 0.85)
Rad1 (0.21,0.80)
Rad2 (0.18,0.71)
Rad3 (0.31,0.92)
RadMaj (0.22,0.89)
Cardiomegaly (>3 rads)
False Positive Rate
LabelL (0.16,0.24)
LabelU (0.04,0.42)
Model (AUC = 0.90)
Rad1 (0.05,0.48)
Rad2 (0.23,0.85)
Rad3 (0.11,0.70)
RadMaj (0.08,0.75)
Consolidation (>2 rads)
False Positive Rate
LabelL (0.18,0.31)
LabelU (0.05,0.41)
Model (AUC = 0.90)
Rad1 (0.11,0.66)
Rad2 (0.09,0.48)
Rad3 (0.03,0.45)
RadMaj (0.05,0.52)
Edema (>3 rads)
False Positive Rate
LabelL (0.15,0.49)
LabelU (0.12,0.65)
Model (AUC = 0.92)
Rad1 (0.09,0.63)
Rad2 (0.19,0.79)
Rad3 (0.07,0.58)
RadMaj (0.08,0.68)
Pleural Effusion (>3 rads)
False Positive Rate
LabelL (0.21,0.78)
LabelU (0.16,0.88)
Model (AUC = 0.97)
Rad1 (0.05,0.82)
Rad2 (0.17,0.83)
Rad3 (0.14,0.89)
RadMaj (0.10,0.89)
Atelectasis (>0 rads)
Precision
Sensitivity
LabelL (0.22,0.32)
LabelU (0.51,0.70)
Model (AUC = 0.69)
Rad1 (0.80,0.62)
Rad2 (0.71,0.64)
Rad3 (0.92,0.56)
RadMaj (0.89,0.64)
Cardiomegaly (>3 rads)
Sensitivity
LabelL (0.24,0.39)
LabelU (0.42,0.82)
Model (AUC = 0.81)
Rad1 (0.48,0.82)
Rad2 (0.85,0.61)
Rad3 (0.70,0.74)
RadMaj (0.75,0.80)
Consolidation (>2 rads)
Sensitivity
LabelL (0.31,0.10)
LabelU (0.41,0.34)
Model (AUC = 0.44)
Rad1 (0.66,0.27)
Rad2 (0.48,0.25)
Rad3 (0.45,0.45)
RadMaj (0.52,0.38)
Edema (>3 rads)
Sensitivity
LabelL (0.49,0.38)
LabelU (0.65,0.50)
Model (AUC = 0.66)
Rad1 (0.63,0.58)
Rad2 (0.79,0.44)
Rad3 (0.58,0.59)
RadMaj (0.68,0.62)
Pleural Effusion (>3 rads)
Sensitivity
LabelL (0.78,0.49)
LabelU (0.88,0.59)
Model (AUC = 0.91)
Rad1 (0.82,0.80)
Rad2 (0.83,0.55)
Rad3 (0.89,0.63)
RadMaj (0.89,0.71)

Differences between NIH and ChexPert

- Do not use automatic mention extractors like MetaMap or Dnorm
- Incorporated several rules to capture variation in negation and uncertainity
- Split uncertainity classification into pre negation and post negation

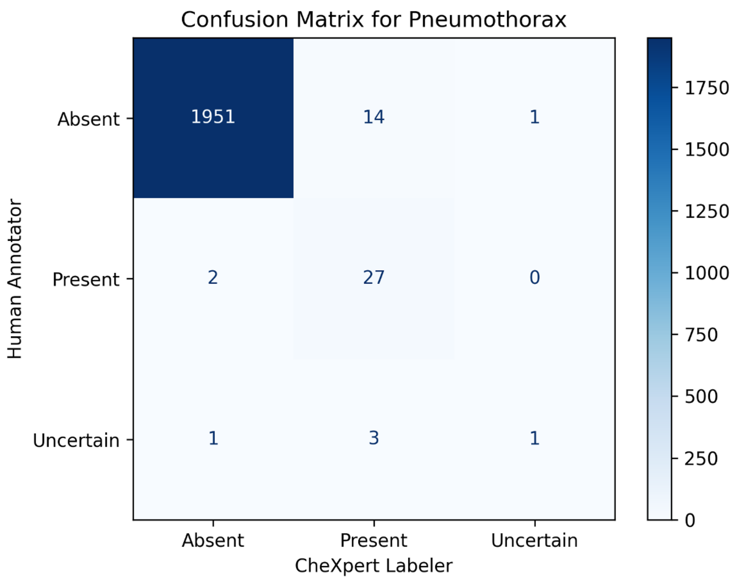
- "cannot exclude pneumothorax." conveys uncertainty in the presence of pneumothorax. Without the pre-negation stage, the 'pneumothorax' match is classified as negative due to the 'exclude XXX' rule. However, by applying the 'cannot exclude' rule in the pre-negation stage, this observation can be correctly classified as uncertain.

LLM Based labeling

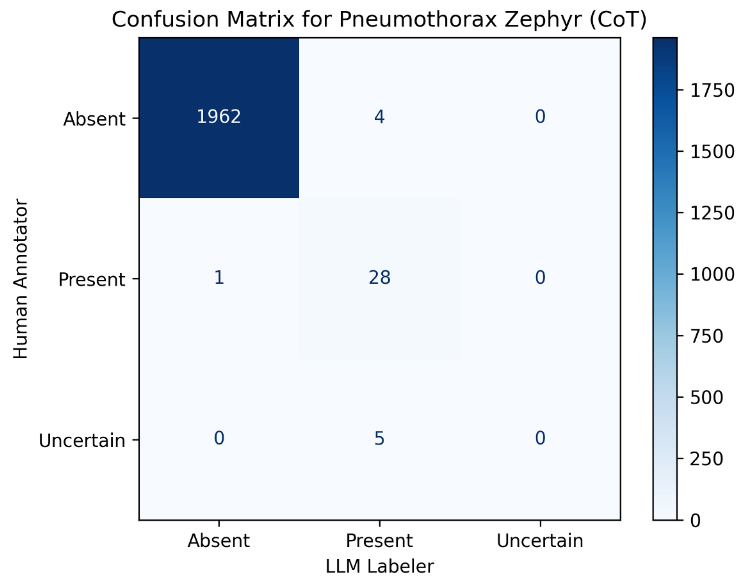
- Zephyr-β
  - 7 billion parameters
  - Base: Mistral-7 v0.01
  - DPO fine tuned
- Chain of thought prompting
- MIMIC-CXR: 220,000 CXRs with their accompanying reports
  - 2000 were manually annotated
  - Another 500 were used for prompt tuning
- CANDID-PTX: 19,000 CXRs with expert pixel-level pneumothorax segmentation

Results

CheXpert labeler



LLM Labeler (Zephyr 7b)



Extraction performance

Labeler	Accuracy	Sensitivity	Specificity	F1 Score
CheXpert Labeler	99.1%	93.1%	99.2%	0.771
LLM Labeler (Zephyr 7b)	99.7%	96.5%	99.7%	0.981

Examples

Report	CheXpert	LLM	Ground Truth
Skin fold projecting over the left hemithorax should not be mistaken for pneumothorax. Left lower lobe is still densely consolidated. Less severe consolidation of the right lung base is unchanged. Both are likely atelectasis. Heart is top-normal size exaggerated by low lung volumes. Widening of the upper mediastinum is due to combination of adenopathy and mediastinal fat.	**Pneumothorax Present**	**Pneumothorax Absent**	**Pneumothorax Absent**

Examples

Report	CheXpert	LLM	Ground Truth
Left pigtail pleural catheter remains in place in the lower left hemithorax. Large left pleural effusion with small hydropneumothorax components appears slightly larger than on the prior study, and is associated with substantial atelectasis of the adjacent left mid and lower lung regions. On the right, a new patchy opacity has developed at the right lung base, and may reflect atelectasis, aspiration, and less likely a developing pneumonia. Known mediastinal and left hilar lymphadenopathy have been more fully characterized on recent CT of ___.	**Pneumothorax Present**	**Pneumothorax Absent**	**Pneumothorax Present**

Beyond the 14 labels

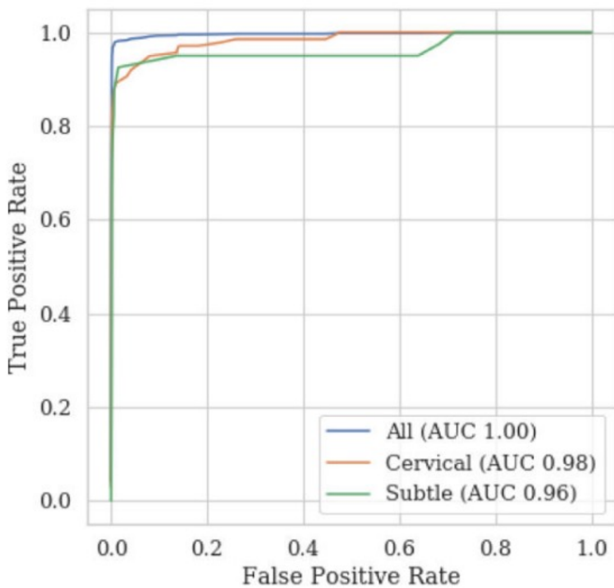
Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging

Luke Oakden-Rayner,* Jared Dunnmon,* Gustavo Carneiro, and Christopher Ré

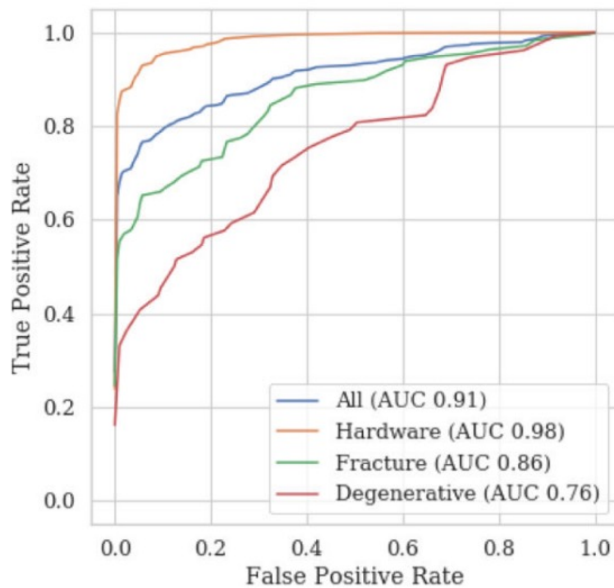
Methods

- Schema completion – provide complete set of subclasses that should be labeled
- Error auditing - the auditor examines model outputs for unexpected regularities, for example a difference in the distribution of a recognizable subclass in the correct and incorrect model prediction groups
- Algorithmic measurement - the algorithm developer designs a method to search for subclasses automatically. (clustering)

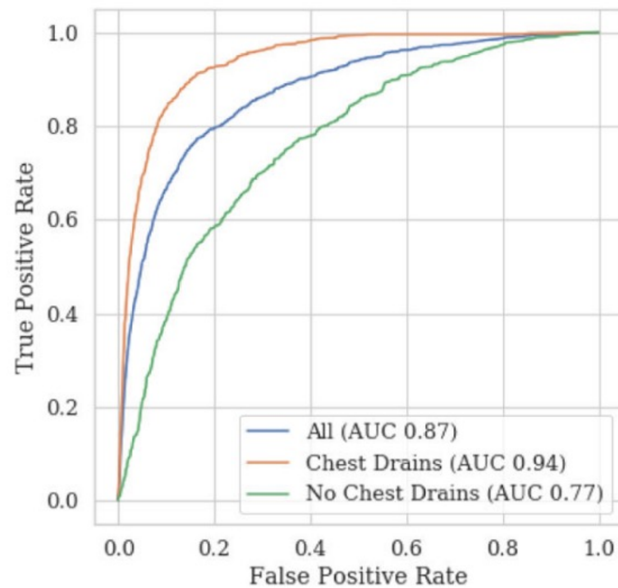
Results



(a) Adelaide Hip Abnormal

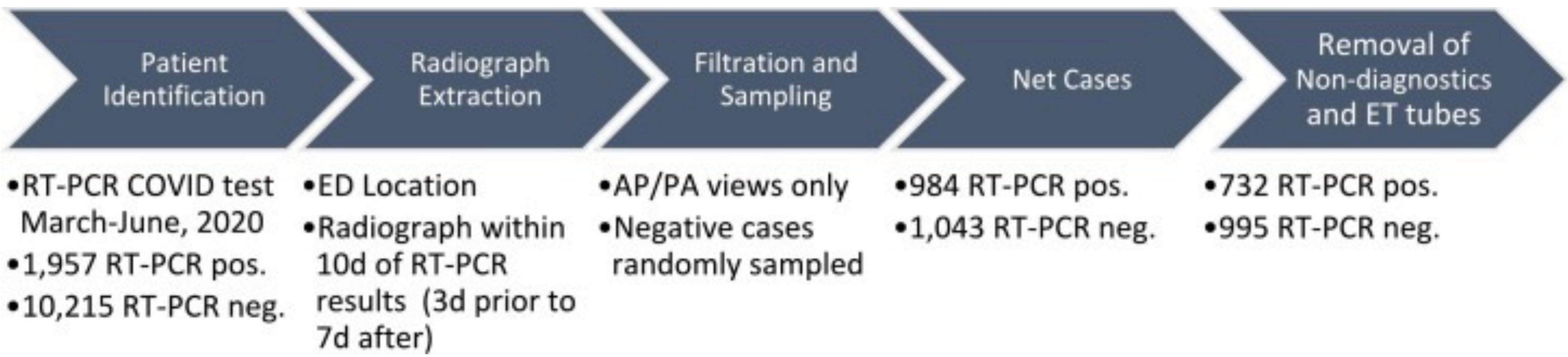


(b) MURA Abnormal



(c) CXR14 Pneumothorax

Multireader evaluation of radiologist performance for COVID-19 detection on emergency department chest radiographs



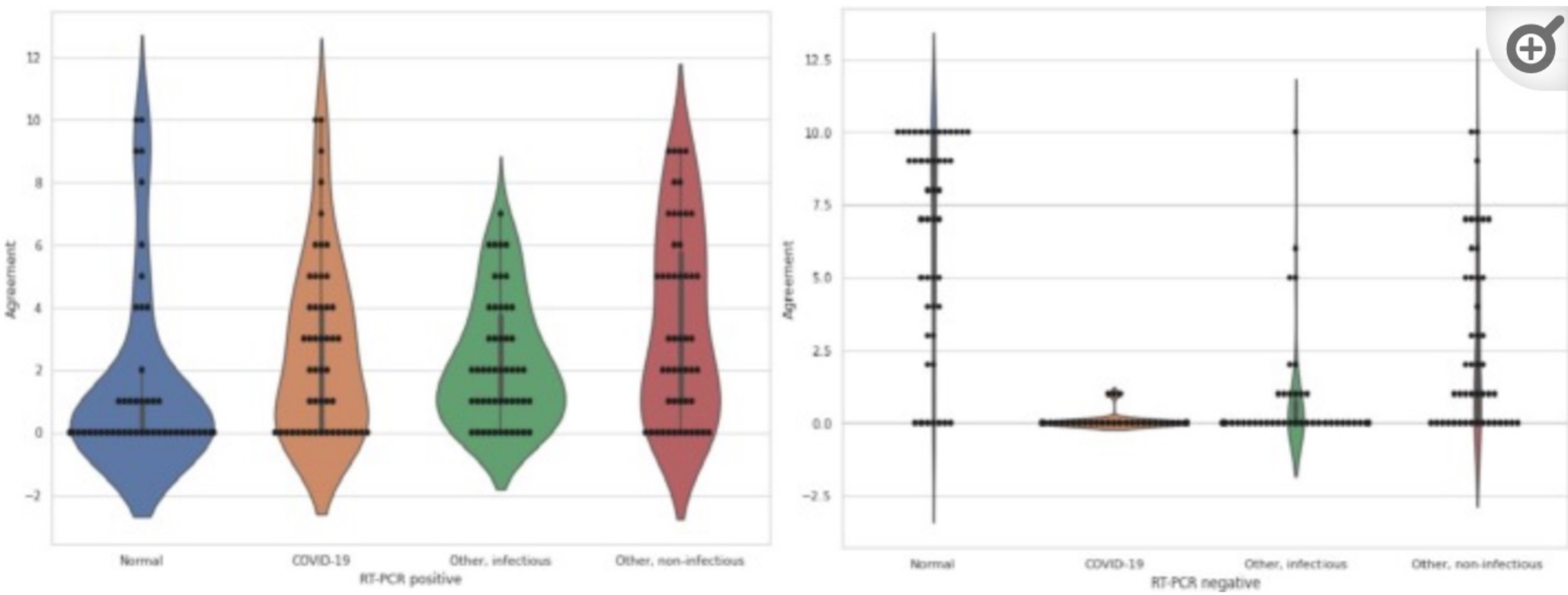
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8585957/

Multireader evaluation of radiologist performance for COVID-19 detection on emergency department chest radiographs

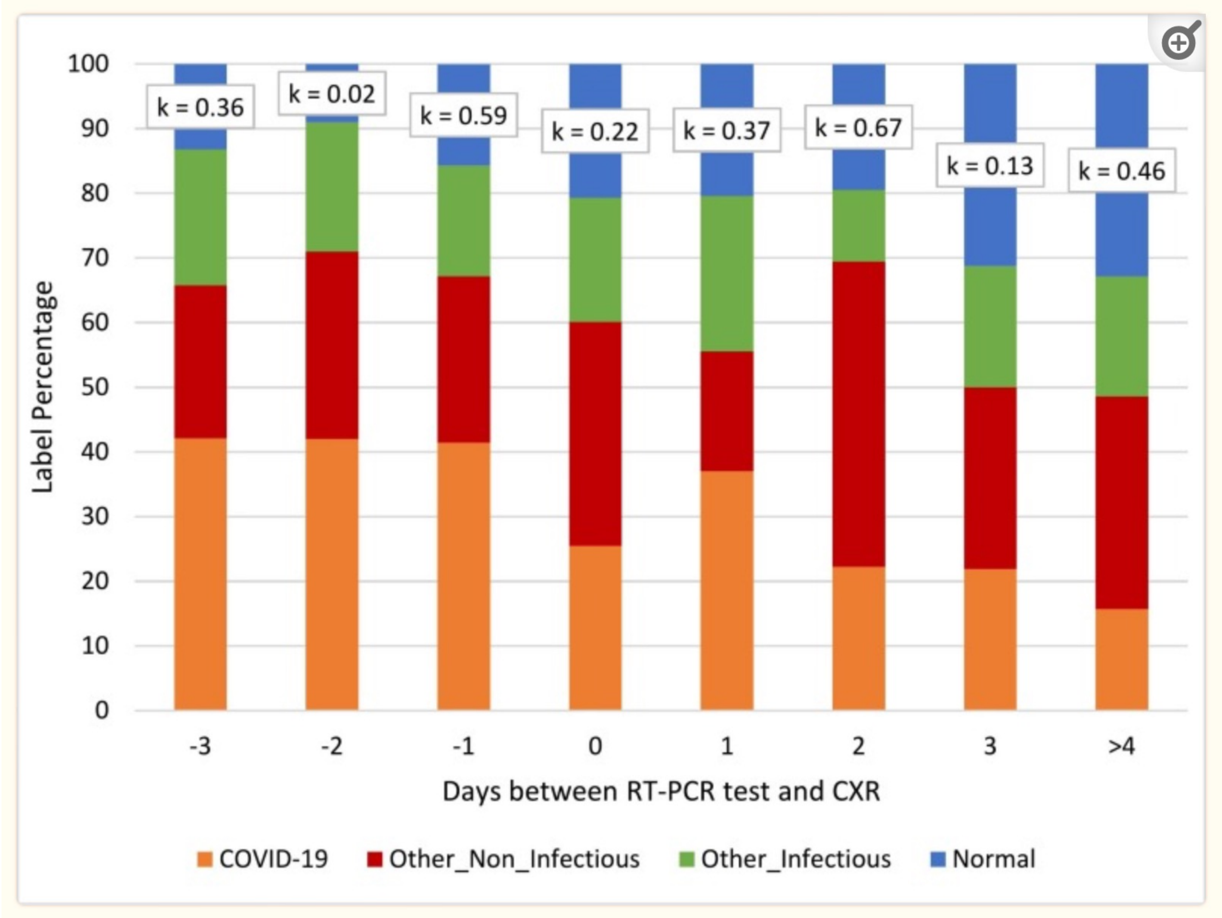
Summary of readers.

Reader	Institution	Fellowship	Years in practice	Sensitivity	Specificity	PPV	NPV
1	Institution 1	Body	7	0.273	0.983	0.942	0.577
2	Institution 2	Emergency	11	0.247	0.991	0.968	0.559
3	Institution 1	Breast	3	0.074	0.996	0.947	0.507
4	Institution 1	Cardiothoracic	5	0.335	0.984	0.951	0.610
5	Institution 3	Breast	3	0.387	0.974	0.941	0.598
6	Institution 1	Body	10	0.374	0.984	0.955	0.635
7	Institution 2	None	13	0.416	0.972	0.932	0.643
8	Institution 4	Cardiothoracic	1	0.260	0.980	0.922	0.597
9	Institution 2	None	11	0.542	0.915	0.854	0.685
10	Institution 1	Neuro	3	0.118	0.996	0.967	0.519

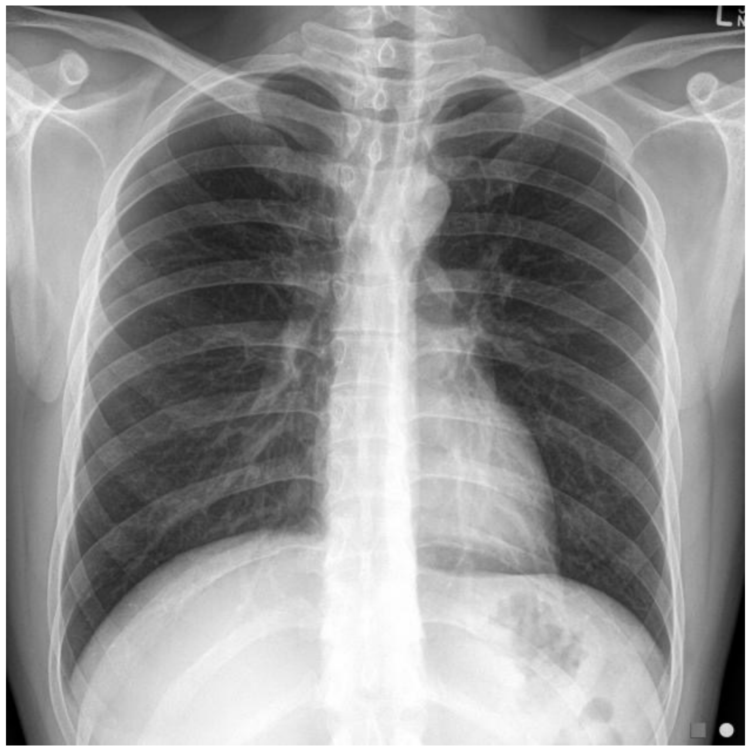
Multireader evaluation of radiologist performance for COVID-19 detection on emergency department chest radiographs



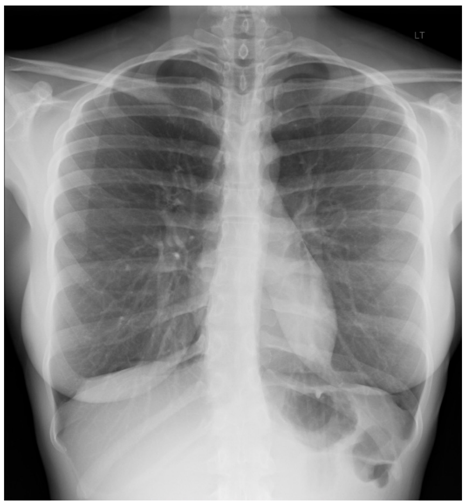
Reader agreement



PCR +ve Normal CXR



RECAP



ELSEVIER

The Lancet Digital Health

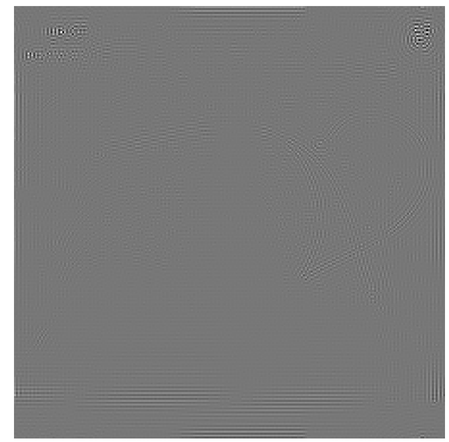
Available online 11 May 2022

In Press, Corrected Proof

Articles

AI recognition of patient race in medical imaging: a modelling study

Judy Wawira Gichoya MD [a], Imon Banerjee PhD [c], Ananth Reddy Bhimireddy MS [a], John L Burns MS [d], Leo Anthony Celi MD [e, g], Li-Ching Chen BS [h], Ramon Correa BS [c], Natalie Dullerud MS [i], Marzyeh Ghassemi PhD [e, f], Shih-Cheng Huang [j], Po-Chih Kuo PhD [h], Matthew P Lungren MD [j], Lyle J Palmer PhD [k, l], Brandon J Price MD [m], Saptarshi Purkayastha PhD [d], Ayis T Pyrros MD [n], Lauren Oakden-Rayner MD [k], Chima Okechukwu MS [o] ... Haoran Zhang MS [i]



1) Performance of deep learning models to detect race from medical images across modalities and external datasets
2) Assessment of possible anatomic and phenotype confounders such as body habitus and disease distribution
3) Investigation into underlying mechanisms by which AI models can recognize race.

ORIGINAL ARTICLE DATA SCIENCE | VOLUME 19, ISSUE 1, P184-191, JANUARY 01, 2022

Detecting Racial/Ethnic Health Disparities Using Deep Learning From Frontal Chest Radiography

Ayis Pyrros, MD • Jorge Mario Rodríguez-Fernández, MD • Stephen M
Judy Wawira Gichoya, MD • Jeanne M. Horowitz, MD • Brian Fornelli, MS • Na
Yury Velichko, PhD • Oluwasanmi Koyejo, PhD • William Galanter, MD, PhD •

Chest radiography as a biomarker of ageing: artificial intelligence-based, multi-institutional model development and validation in Japan

Yasuhito Mitsuyama, Toshimasa Matsumoto, Hiroyuki Tatekawa, Shannon L Walston, Tatsuo Kimura, Akira Yamamoto, Toshio Watanabe, Yukio Miki, Daiju Ueda

> JACC Cardiovasc Imaging. 2021 Nov;14(11):2226-2236. doi: 10.1016/j.jcmg.2021.01.008. Epub 2021 Mar 17.

Deep Learning to Estimate Biological Age From Chest Radiographs

Vineet K Raghu [1], Jakob Weiss [2], Udo Hoffmann [3], Hugo J W L Aerts [4], Michael T Lu [3]

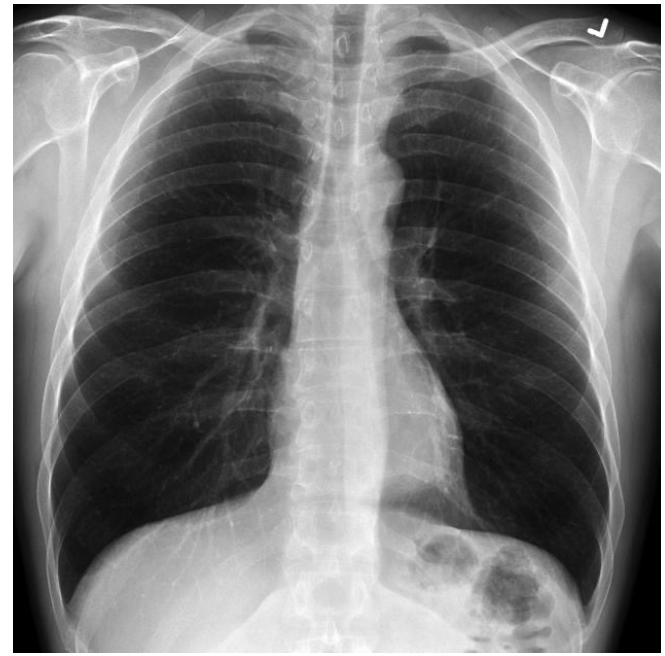
articles > article

Article | Open Access | Published: 18 May 2022

Prediction of future healthcare expenses of patients from chest radiographs using deep learning: a pilot study

Jae Ho Sohn, Yixin Chen, Dmytro Lituiev, Jaewon Yang, Karen Ordovas, Dexter Hadley, Thienkhai H. Vu, Benjamin L. Franc & Youngho Seo

Judy is “Black”, F, 60 yrs (CXR age = 78 yrs), SDI 45, ICD codes – COPD, CHF, 15,000 USD



Opportunistic screening

nature communications

Article https://doi.org/10.1038/s41467-023-39631-x

Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs

Received: 14 December 2022

Accepted: 19 June 2023

Published online: 07 July 2023

Check for updates

Ayis Pyrros[1,2,20], Stephen M. Borstelmann[3,20], Ramana Mantravadi[4], Zachary Zaiman[5], Kaesha Thomas[5], Brandon Price[6], Eugene Greenstein[7], Nasir Siddiqui[1], Melinda Willis[1], Ihar Shulhan[8], John Hines-Shah[1], Jeanne M. Horowitz[9], Paul Nikolaidis[9], Matthew P. Lungren[10,11,12], Jorge Mario Rodríguez-Fernández[13], Judy Wawira Gichoya[5], Sanmi Koyejo[14], Adam E Flanders[15], Nishith Khandwala[16], Amit Gupta[17], John W. Garrett[18], Joseph Paul Cohen[11], Brian T. Layden[19], Perry J. Pickhardt[18] & William Galanter[19]

Image Based Precision Medicine using Opportunistic Screening

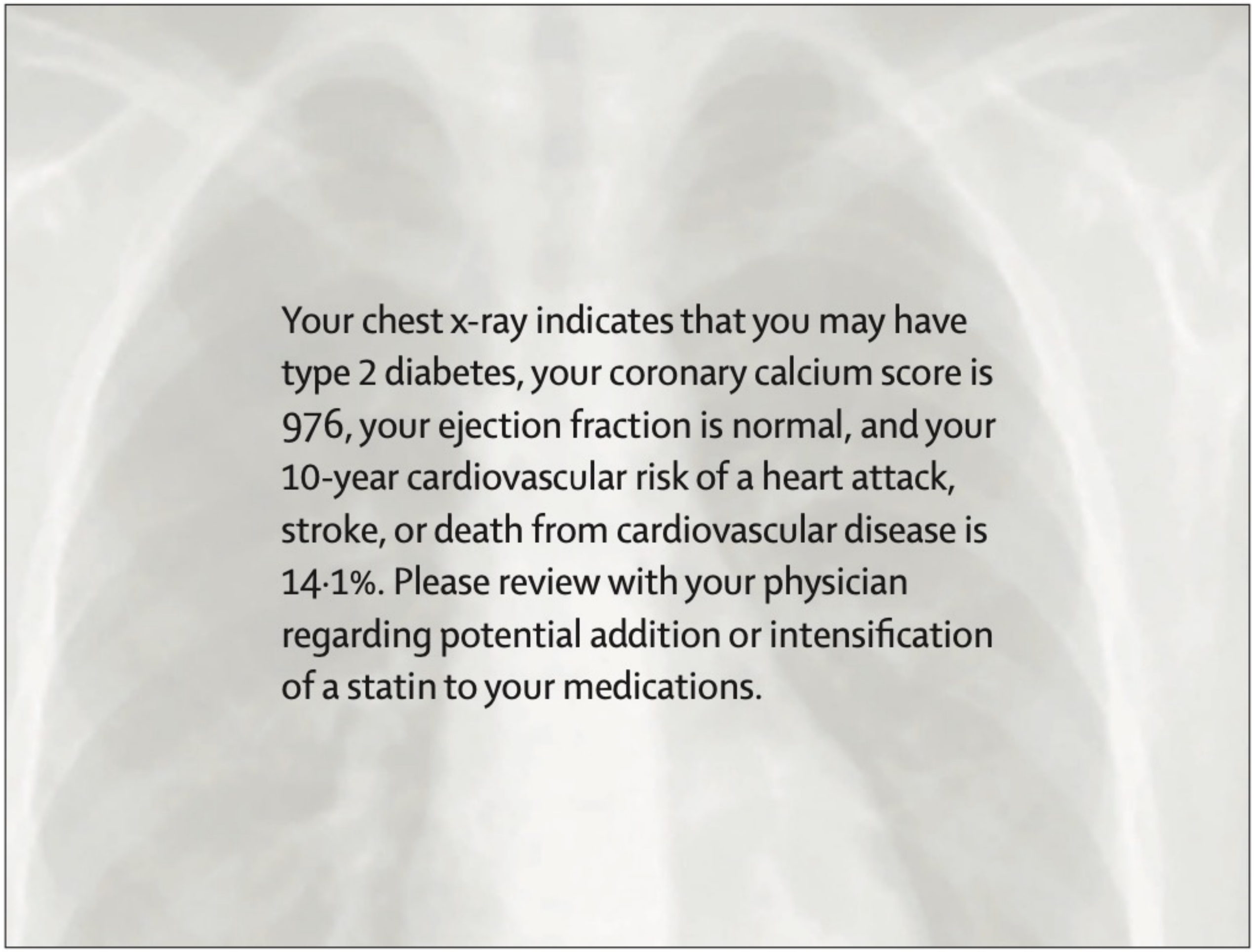
Imaging	Opportunistic Detection	Citation
Chest X-ray	Cardiovascular risk 10-year MACE*	Weiss J, Annals Int Med, 2024
Chest X-ray	Type 2 Diabetes	Pyrros A, Nature Comms, 2023
Chest X-ray	Ejection fraction	Ueda D, Lancet Digital Health, 2023
Chest X-ray	Coronary calcium score, CV risk	Kamel P, Radiol Cardiothoraic Imaging, 2021
Chest CT	Pancreatic cancer	Cao K, Nature Medicine, 2023
Chest CT	Coronary artery disease risk	Eng D, Nature Digital Medicine, 2021
Lung CT	Cardiovascular Risk	Chao H, Nature Comms, 2021
Mammography	Cardiovascular Risk	Iribarren C, Circulation: CV Imaging, 2022
Abdominal CT	Type 2 Diabetes	Tallam A, Radiology, 2022
Abdominal CT	Cardiovascular Risk	Magudia K, Am J Roentgenology, 2022

*major adverse cardiac events, better than traditional risk score

@erictopol

https://erictopol.substack.com/p/opportunistic-ai-for-medical-scans?utm_campaign=post&utm_medium=web&triedRedirect=true

Topol, Eric J. "AI-enabled opportunistic medical scan interpretation." *The Lancet* 403.10439 (2024): 1842.

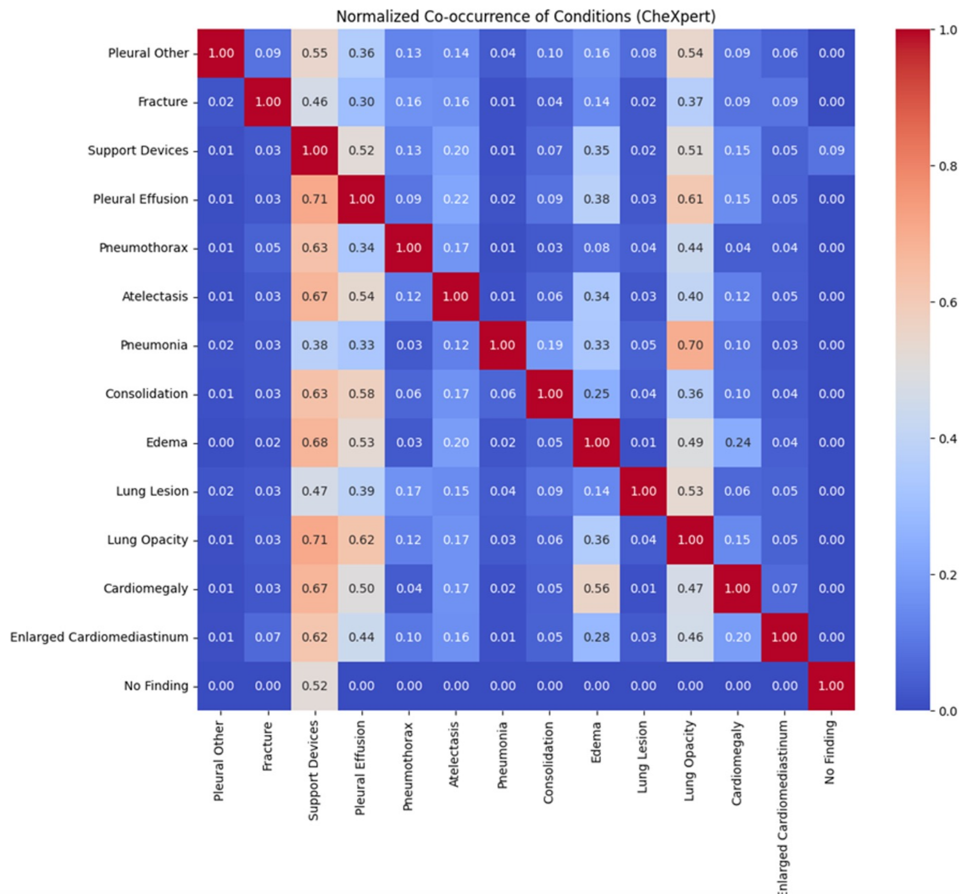


Topol EJ. AI-enabled opportunistic medical scan interpretation. The Lancet. 2024 May 11;403(10439):1842.

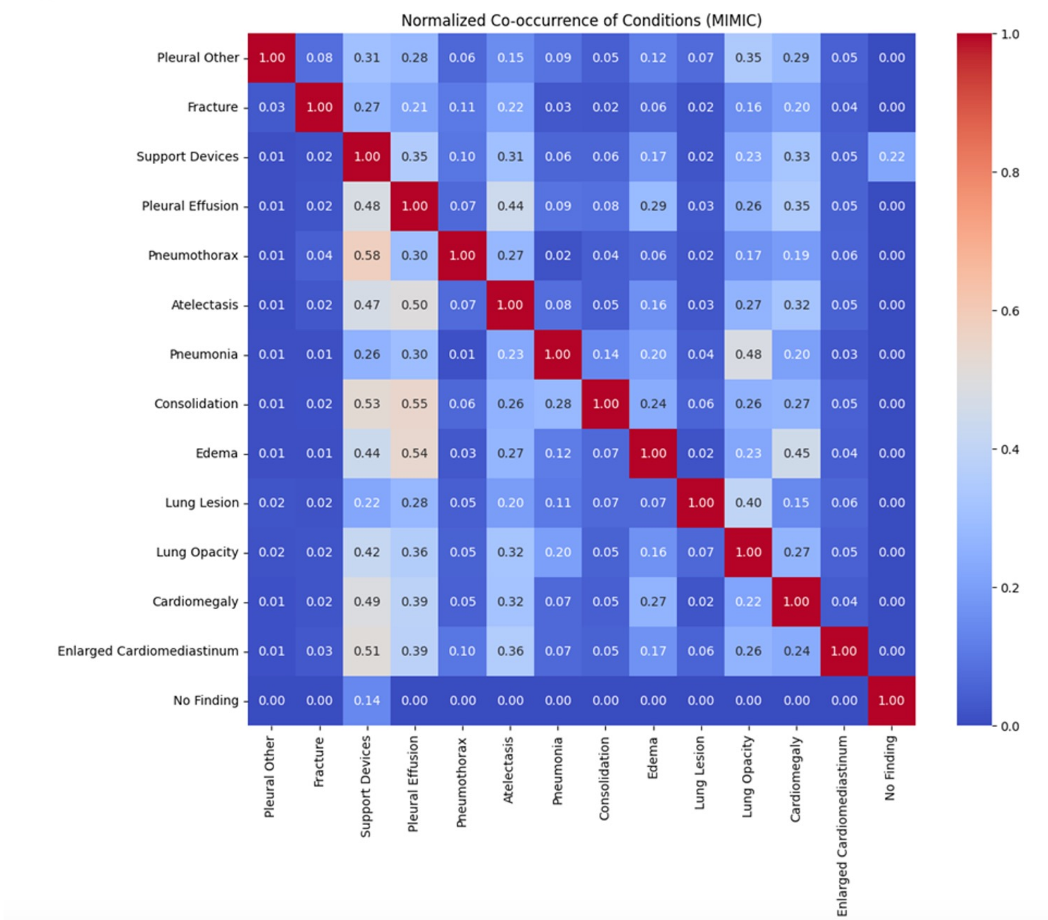
Label quality

Source Labeling	Ground Truth Labeling	Fracture	Pneumothorax
CheXpert	Human	0.790	0.947
CheXpert	CXP	0.707	0.904
CheXpert	Zephyr LLM CoT	0.773	0.935
CheXpert	Zephyr LLM NoCoT	0.771	0.935
CheXpert	LLama LLM CoT	0.748	0.913
CheXpert	LLama LLM NoCoT	0.760	0.935
CheXpert	Phi LLM CoT	0.785	0.924
CheXpert	Phi LLM NoCoT	0.787	0.935

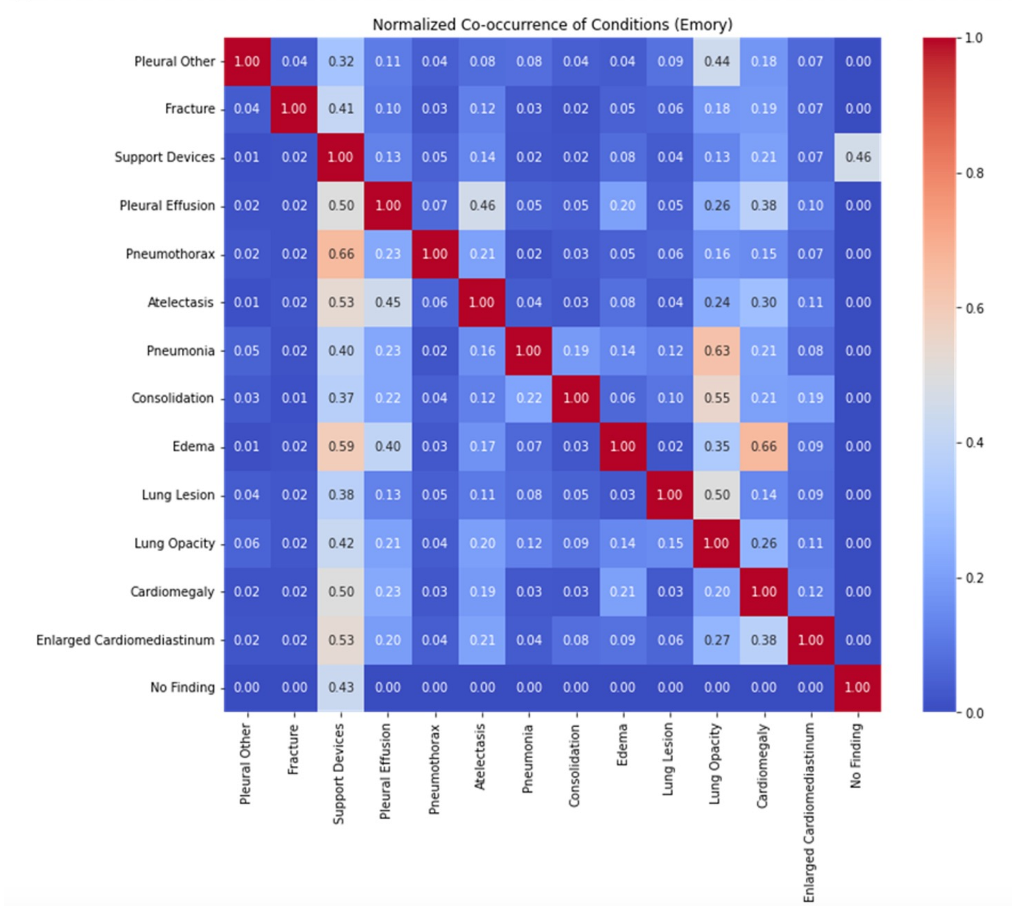
Label correlation - CheXpert



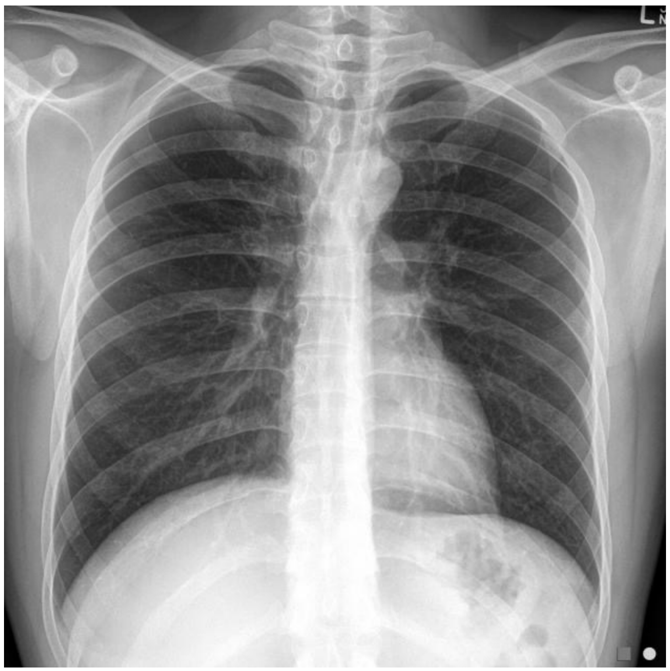
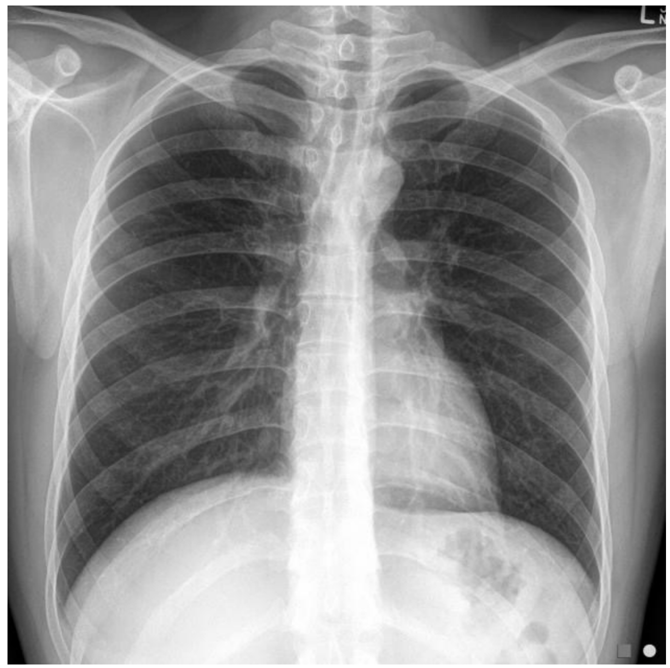
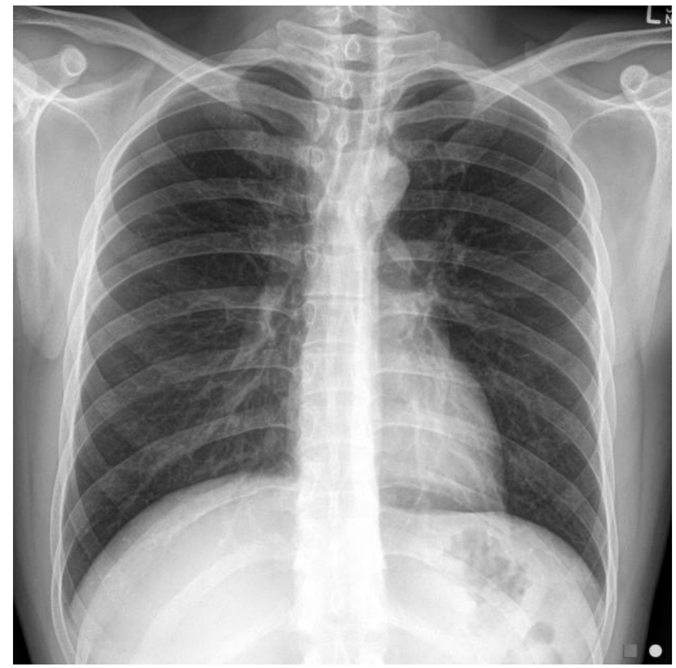
Label correlation - MIMIC



Label correlation - Emory

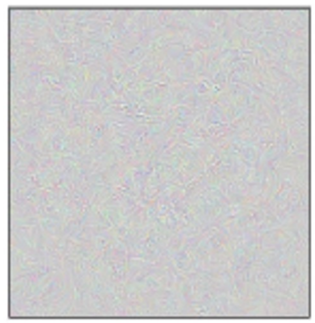


Ensemble models for labeling



Shortcut learning in deep neural networks

Robert Geirhos[1,2,4] ✉, Jörn-Henrik Jacobsen[3,4], Claudio Michaelis[1,2,4], Richard Zemel[3,5], Wieland Brendel[1,5], Matthias Bethge[1,5] and Felix A. Wichmann[1,5]



Article: Super Bowl 50

Paragraph: "Peython Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had a jersey number 37 in Champ Bowl XXXIV."

Question: "What is the name of the quarterback who was 38 in Super Bowl XXXIII?"

Original prediction: John Elway

Prediction under adversary: Jeff Dean

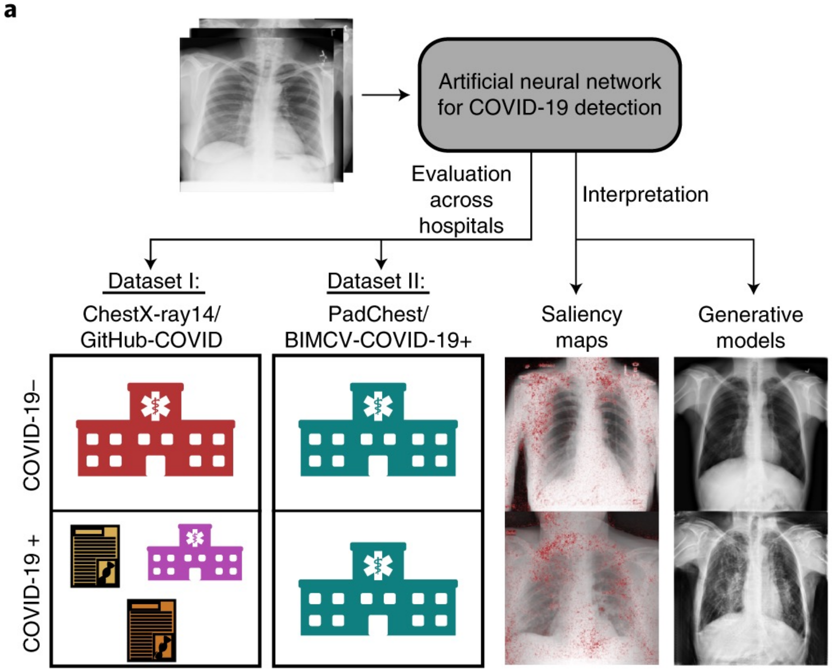
	Shane 2018		Zech 2018	Jia 2017
Task for DNN	Caption image	Recognize object	Recognize pneumonia	Answer question
Problem	Describes green hillside as grazing sheep	Hallucinates teapot if certain patterns are present	Fails on scans from new hospitals	Changes answer if irrelevant information is added
Shortcut	Uses background to recognize primary object	Uses features unrecognizable to humans	Looks at hospital token, not lung	Only looks at last sentence and ignores context

AI for radiographic COVID-19 detection selects shortcuts over signal

Alex J. DeGrave, Joseph D. Janizek & Su-In Lee

Nature Machine Intelligence 3, 610–619 (2021) | Cite this article

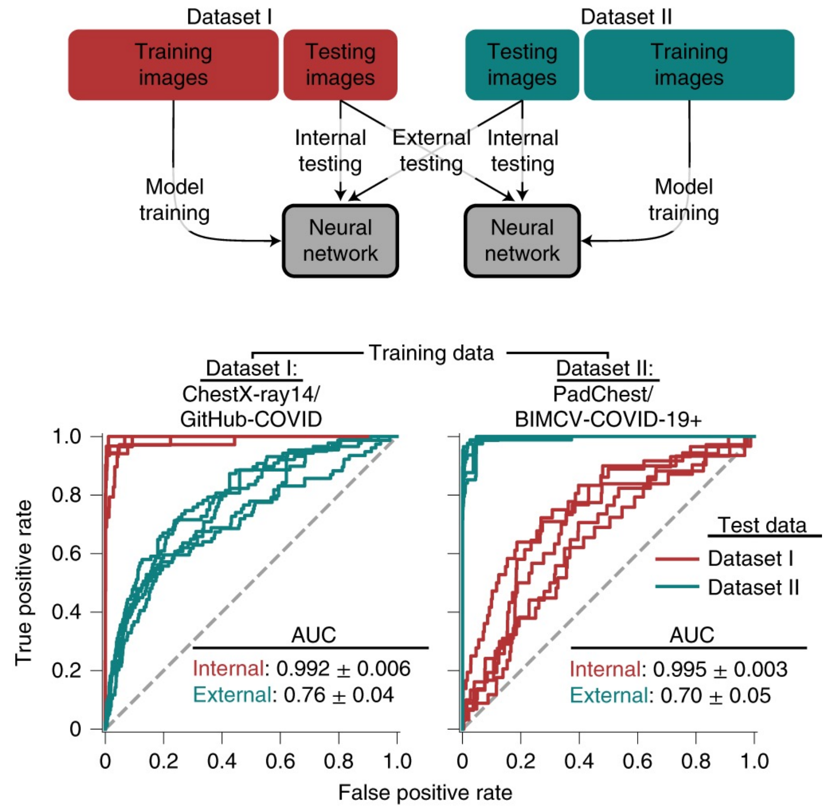
a



b

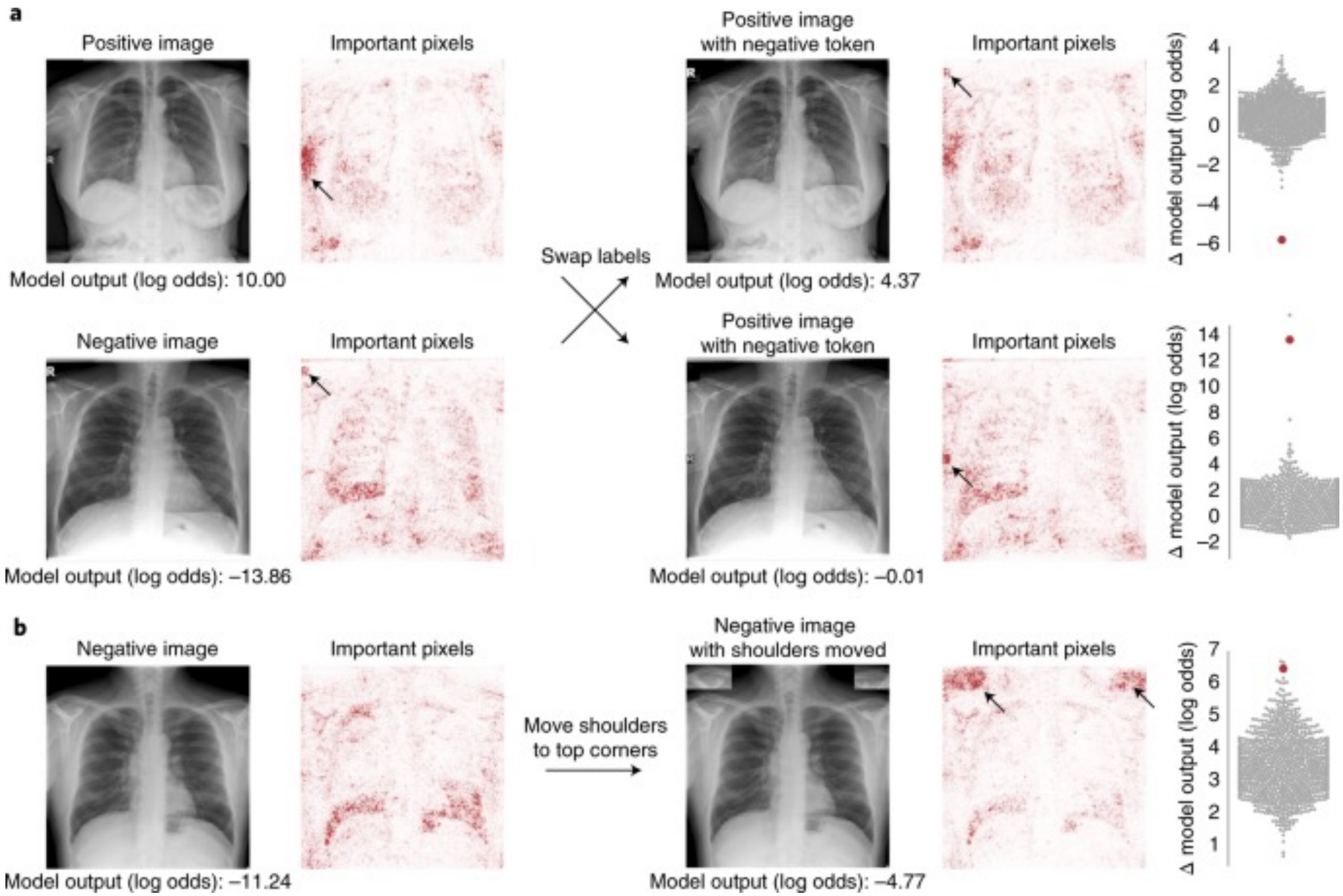
	Dataset I			Dataset II		
	Combined	Chest-X-ray14	GitHub-COVID	Combined	PadChest	BIMCV-COVID-19+
No. radiographs	112,528	112,120	408	97,866	96,270	1,596
No. patients	31,067	30,805	262	64,954	63,939	1,105
% COVID-19+	0.2	0	76.5	1.6	0	100
% AP images	39.9	40	26	5.6	4.7	58.1

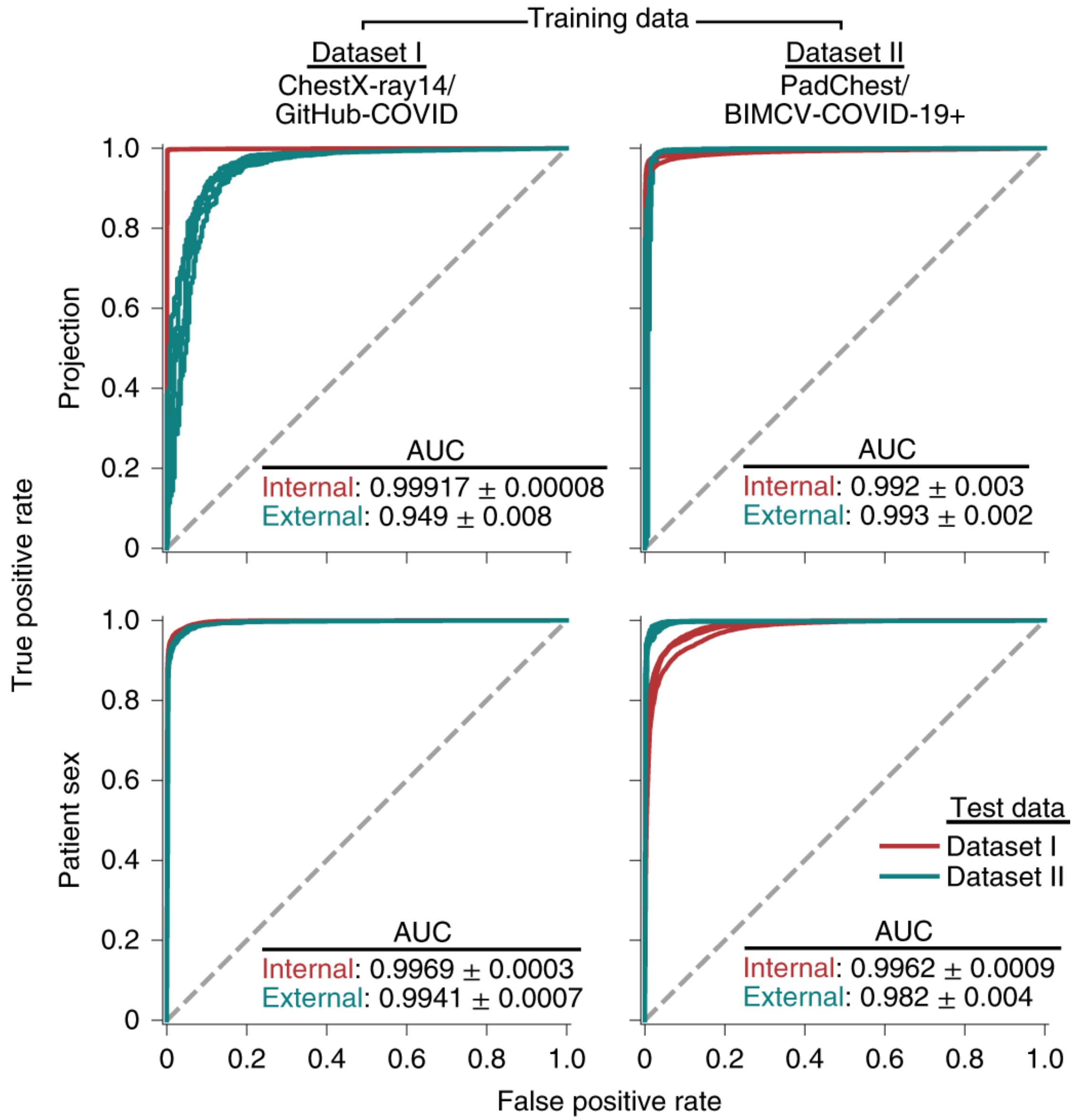
c

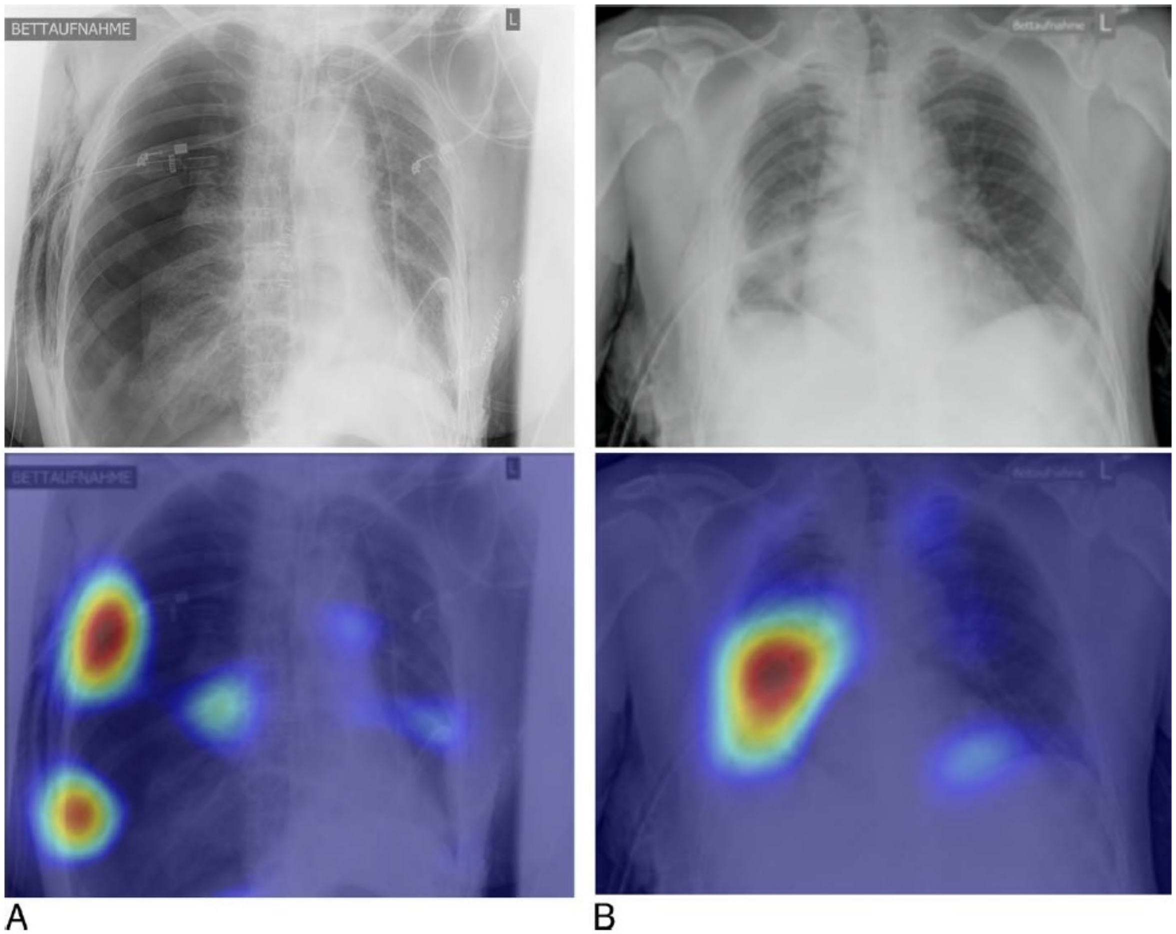


https://www.nature.com/articles/s42256-021-00338-7

Experimental confirmation of insights from saliency maps and CycleGANs via radiograph modification.




Training data
Dataset I
ChestX-ray14/
GitHub-COVID
Dataset II
PadChest/
BIMCV-COVID-19+
True positive rate
Projection
Patient sex
1.0
0.8
0.6
0.4
0.2
0
AUC
Internal: 0.99917 ± 0.00008
External: 0.949 ± 0.008
AUC
Internal: 0.992 ± 0.003
External: 0.993 ± 0.002
AUC
Internal: 0.9969 ± 0.0003
External: 0.9941 ± 0.0007
Test data
Dataset I
Dataset II
AUC
Internal: 0.9962 ± 0.0009
External: 0.982 ± 0.004
0
0.2
0.4
0.6
0.8
1.0
False positive rate



ORIGINAL ARTICLES

Impact of Confounding Thoracic Tubes and Pleural Dehiscence Extent on Artificial Intelligence Pneumothorax Detection in Chest Radiographs

Rueckel, Johannes MD[*]; Trappmann, Lena[*]; Schachtner, Balthasar PhD[*,†]; Wesp, Philipp[*]; Hoppe, Boj Friedrich MD[*]; Fink, Nicola MD[*]; Ricke, Jens MD[*]; Dinkel, Julien MD[*,†]; Ingrisch, Michael PhD[*]; Sabel, Bastian Oliver MD[*]

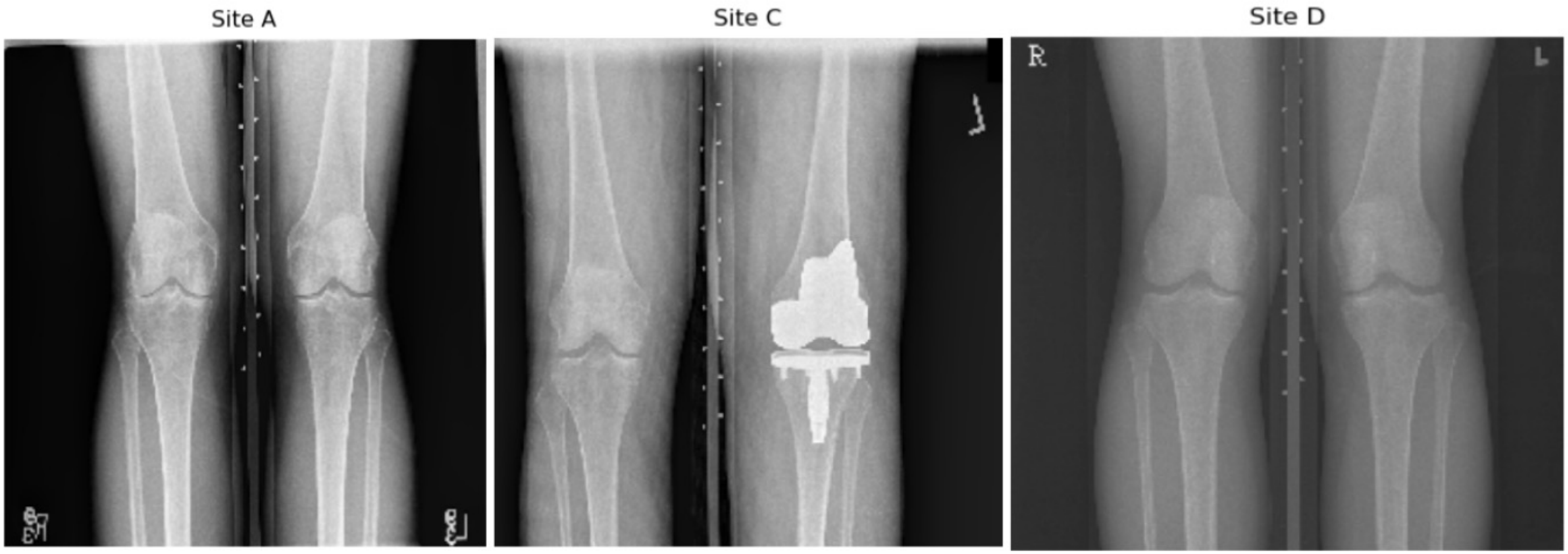
Author Information

Investigative Radiology 55(12):p 792-798, December 2020. | ***DOI:*** 10.1097/RLI.0000000000000707

SDC

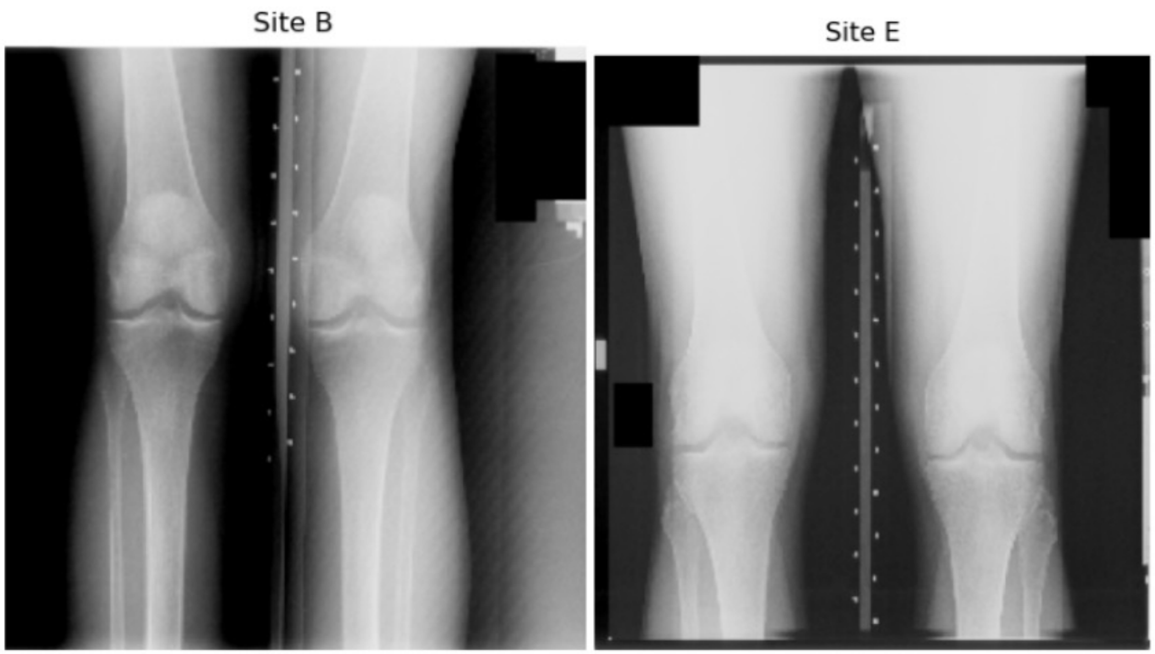
Metrics

Algorithmic Shortcutting in Medical Image Analysis



Frances Koback et al"Algorithmic shortcutting in medical image analysis", Proc. SPIE 12927, Medical Imaging 2024: Computer-Aided Diagnosis, 129272W (3 April 2024)

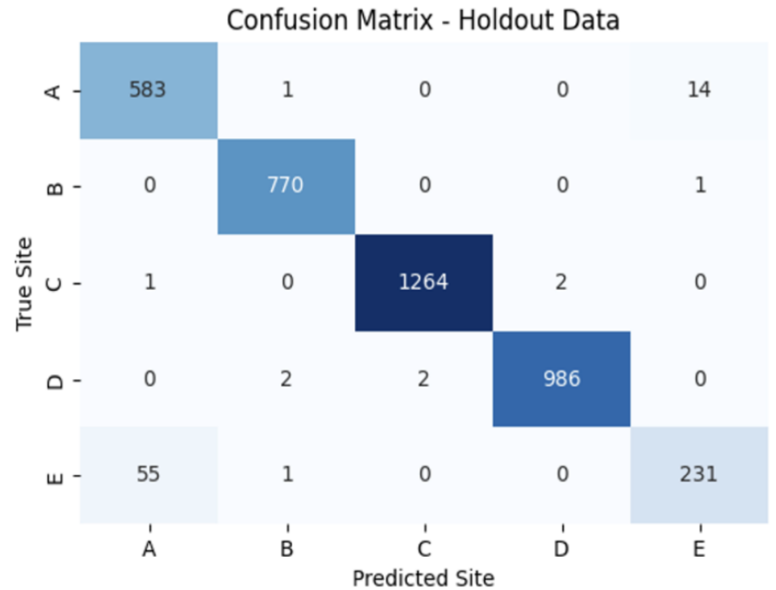
Algorithmic Shortcutting in Medical Image Analysis



Frances Koback et al"Algorithmic shortcutting in medical image analysis", Proc. SPIE 12927, Medical Imaging 2024: Computer-Aided Diagnosis, 129272W (3 April 2024)

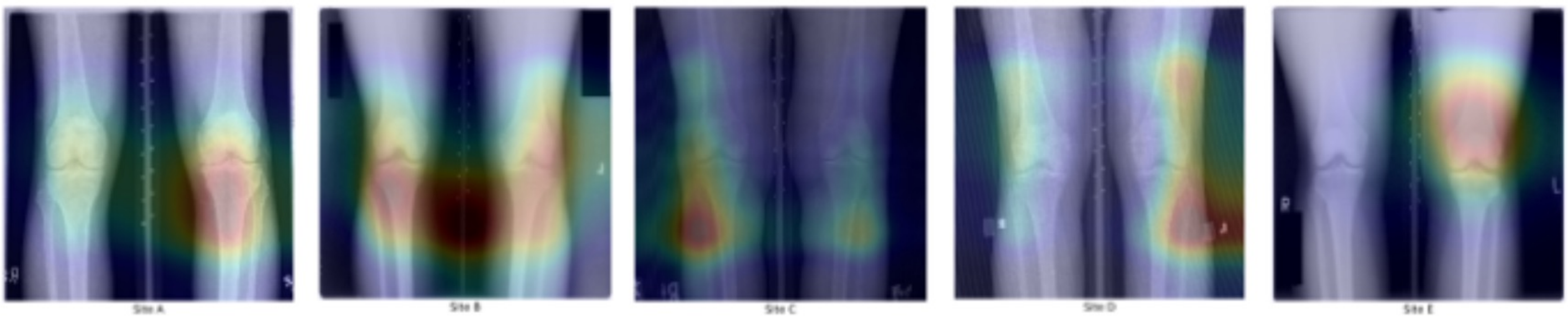
Algorithmic Shortcutting in Medical Image Analysis

Site	Precision	Recall	F1-Score	Support
A	0.912	0.975	0.943	598
B	0.995	0.999	0.997	771
C	0.998	0.998	0.998	1,267
D	0.998	0.996	0.997	990
E	0.939	0.805	0.867	287
Macro Avg	0.969	0.954	0.960	3,913
Weighted Avg	0.980	0.980	0.979	3,913



Frances Koback et al"Algorithmic shortcutting in medical image analysis", Proc. SPIE 12927, Medical Imaging 2024: Computer-Aided Diagnosis, 129272W (3 April 2024)

Algorithmic Shortcutting in Medical Image Analysis



Frances Koback et al"Algorithmic shortcutting in medical image analysis", Proc. SPIE 12927, Medical Imaging 2024: Computer-Aided Diagnosis, 129272W (3 April 2024)

Summary

- Different DL tasks
- 8 labels -> 14 labels
- NLP /Chexpert / LLM based labeler
- Hidden stratification
- Reader study
- Lab based labeling
- Labeling for opportunistic screening
- Label quality
- Label correlation
- Ensemble model labeling
- Shortcut learning

Demo

- https://github.com/BardiaKh/RadPrompter/

- judywawira@emory.edu
- @judywawira

Healthcare Innovation and Translational Informatics
(HITI) Lab at Emory