

## MedCLIP Zero-Shot Disease Classification on IU-CXR Pathology-Rich Sanity Subset (NO filtering; all sanity images)

Number of images: 54

### Setup:

- Start from the pathology-rich IU-CXR sanity subset CSV.
- For each image, MedCLIP PromptClassifier produces zero-shot probabilities for: Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion.
- For each disease, ground-truth labels are derived from Pathology\_Labels\_14 using simple string matching; if ANY matching phrase is present, label=1.

### Evaluation:

- Per-disease metrics: AUROC, AUPRC, Acc, Precision, Recall, F1 (threshold=0.5).
- Micro-averaged metrics flatten all 5 diseases across all images.

Micro-averaged performance (across all diseases & images, threshold 0.5):

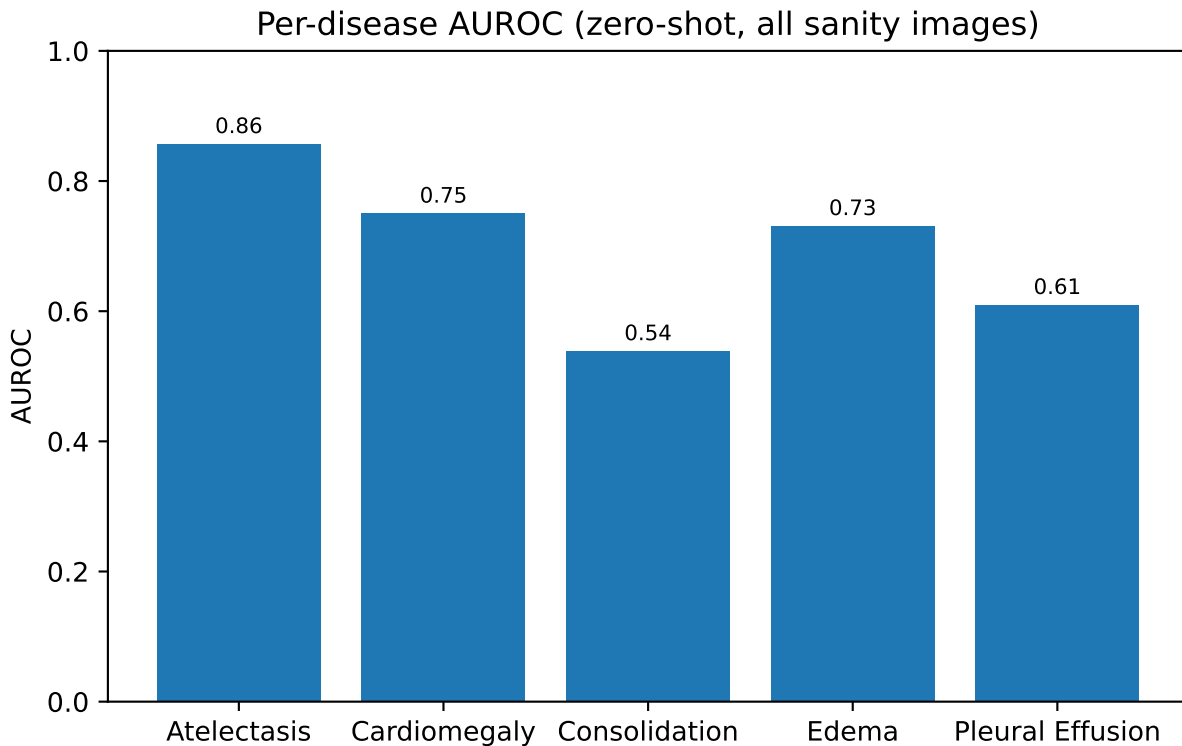
- Micro Accuracy: 0.315
- Micro Precision: 0.315
- Micro Recall: 1.000
- Micro F1: 0.479
- Total disease-label decisions evaluated: 270

Top-K zero-shot disease ranking (only images with  $\geq 1$  positive label):

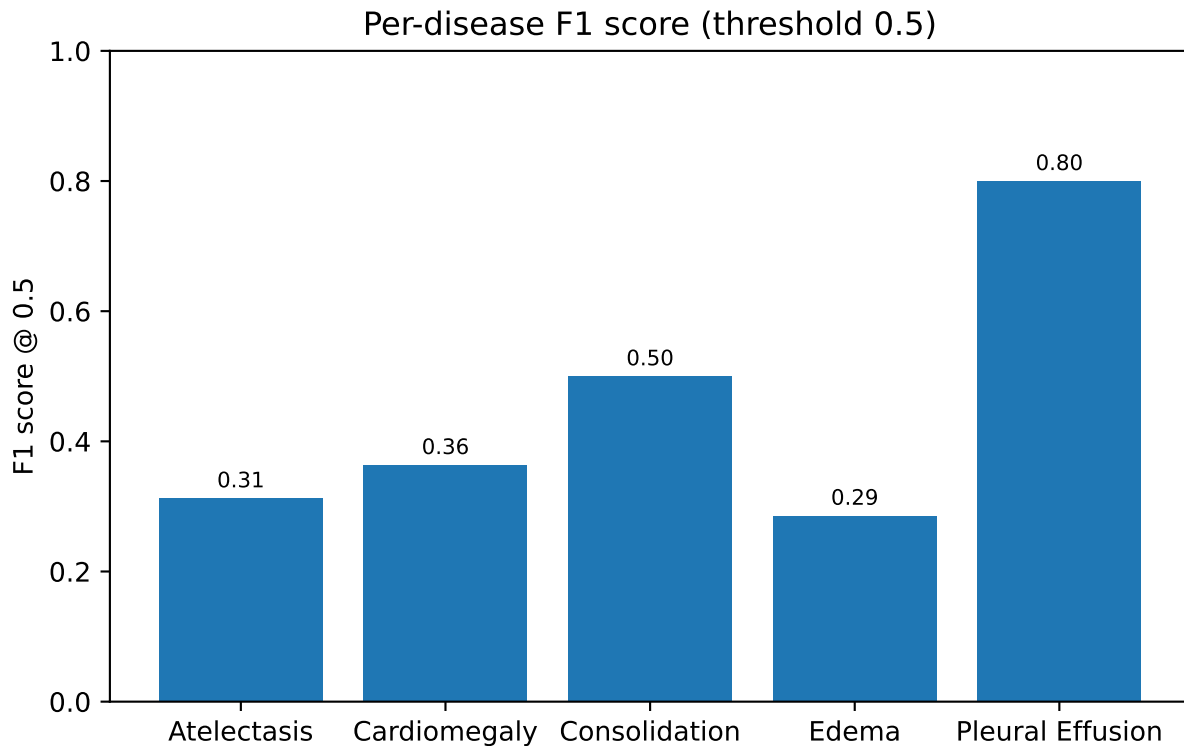
- Top-1 accuracy: 0.391 (N=46)
- Top-3 accuracy: 0.696 (N=46)

Global 'any disease present & predicted' metric (threshold 0.5):

- We mark an image as correctly classified if at least one of the 5 diseases is present in the ground truth AND predicted as positive.
- Any-hit Accuracy: 0.852
- Any-hit Precision: 0.852
- Any-hit Recall: 1.000
- Any-hit F1: 0.920



AUROC measures how well the model ranks positive vs negative cases per disease. High AUROC indicates good separation even without probability calibration.



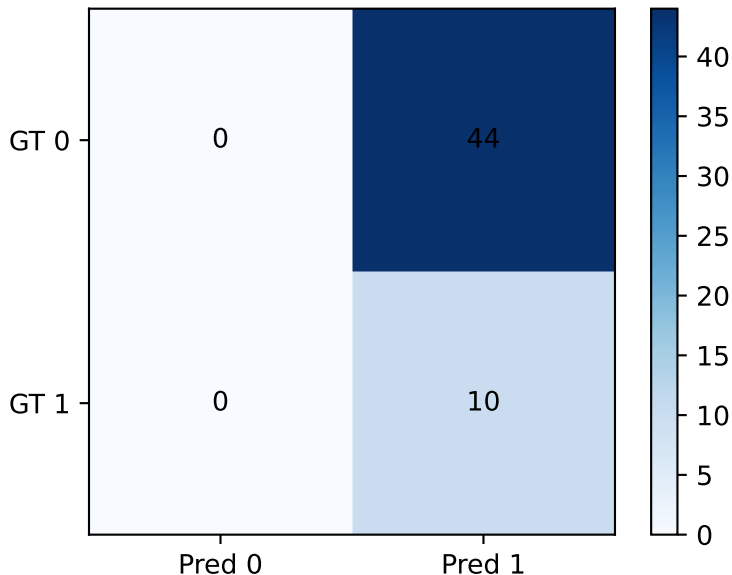
F1@0.5 combines precision and recall for each disease.  
It reflects how well zero-shot scores translate into present/absent decisions  
with a simple 0.5 decision threshold.

Per-disease zero-shot metrics (all sanity images)

Disease	Positives	Negatives	AUROC	AUPRC	Acc@0.5	Prec@0.5	Rec@0.5	F1@0.5	TP	FP	TN	FN
Atelectasis	10	44	0.857	0.740	0.185	0.185	1.000	0.312	10	44	0	0
Cardiomegaly	12	42	0.750	0.592	0.222	0.222	1.000	0.364	12	42	0	0
Consolidation	18	36	0.539	0.425	0.333	0.333	1.000	0.500	18	36	0	0
Edema	9	45	0.731	0.466	0.167	0.167	1.000	0.286	9	45	0	0
Pleural Effusion	36	18	0.610	0.778	0.667	0.667	1.000	0.800	36	18	0	0

Metrics computed disease-wise using 0.5 as the decision threshold.  
This lets you compare MedCLIP behavior across the 5 CheXpert diseases.

Atelectasis: Confusion Matrix (threshold=0.5)

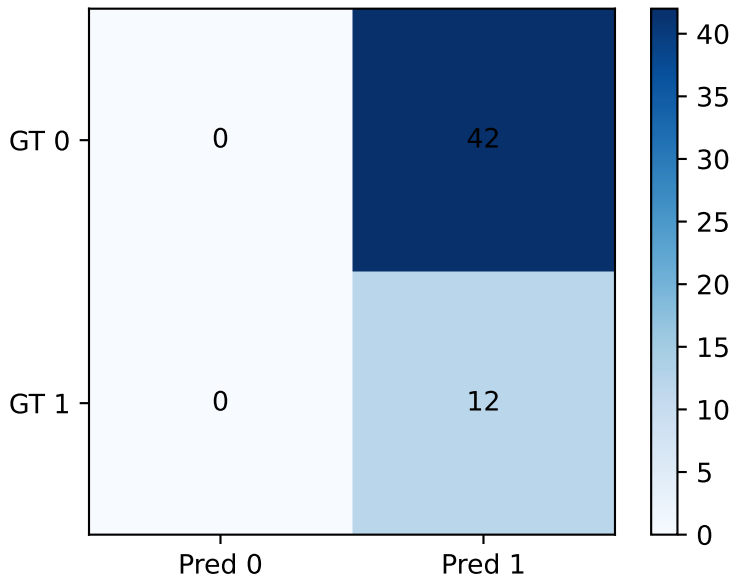


For Atelectasis, positives=10, negatives=44.

- TN=0: images without Atelectasis correctly predicted as negative.
- FP=44: images without Atelectasis incorrectly flagged as present.
- FN=0: images with Atelectasis missed by the model.
- TP=10: images with Atelectasis correctly detected.

Here, the model predicts this disease for nearly every case (no true negatives), so recall is high but spec

Cardiomegaly: Confusion Matrix (threshold=0.5)

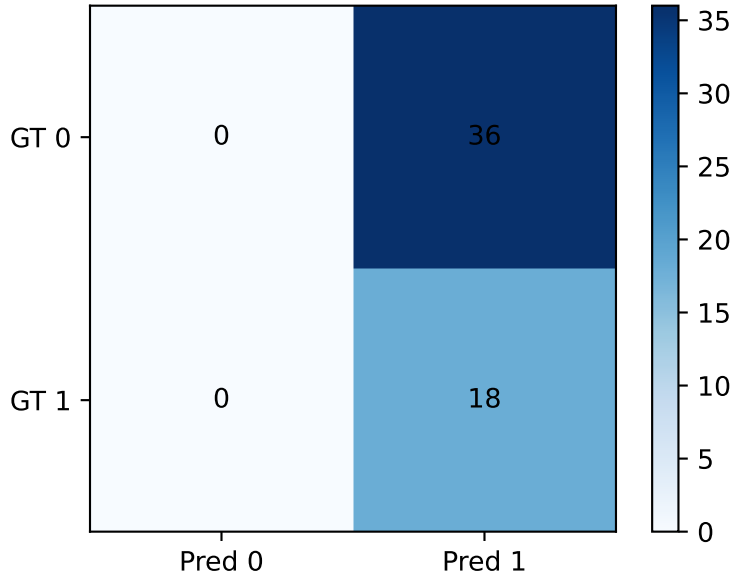


For Cardiomegaly, positives=12, negatives=42.

- TN=0: images without Cardiomegaly correctly predicted as negative.
- FP=42: images without Cardiomegaly incorrectly flagged as present.
- FN=0: images with Cardiomegaly missed by the model.
- TP=12: images with Cardiomegaly correctly detected.

Here, the model predicts this disease for nearly every case (no true negatives), so recall is high but spec

## Consolidation: Confusion Matrix (threshold=0.5)

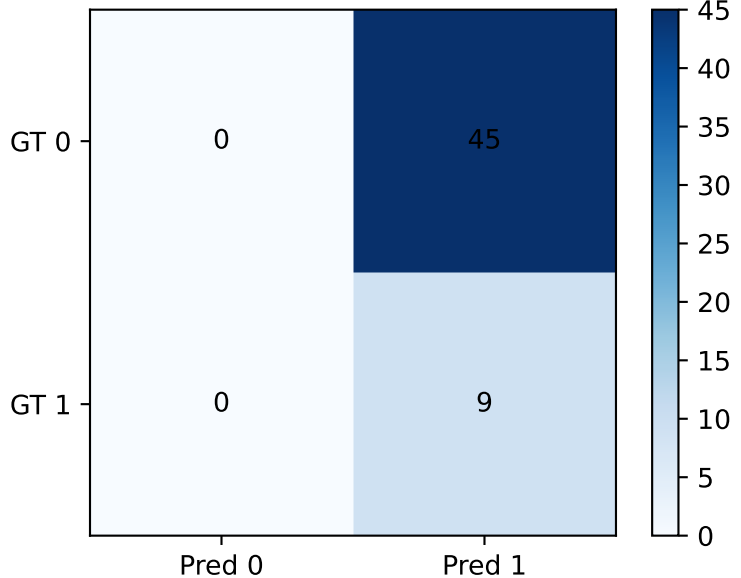


For Consolidation, positives=18, negatives=36.

- TN=0: images without Consolidation correctly predicted as negative.
- FP=36: images without Consolidation incorrectly flagged as present.
- FN=0: images with Consolidation missed by the model.
- TP=18: images with Consolidation correctly detected.

Here, the model predicts this disease for nearly every case (no true negatives), so recall is high but spec

Edema: Confusion Matrix (threshold=0.5)



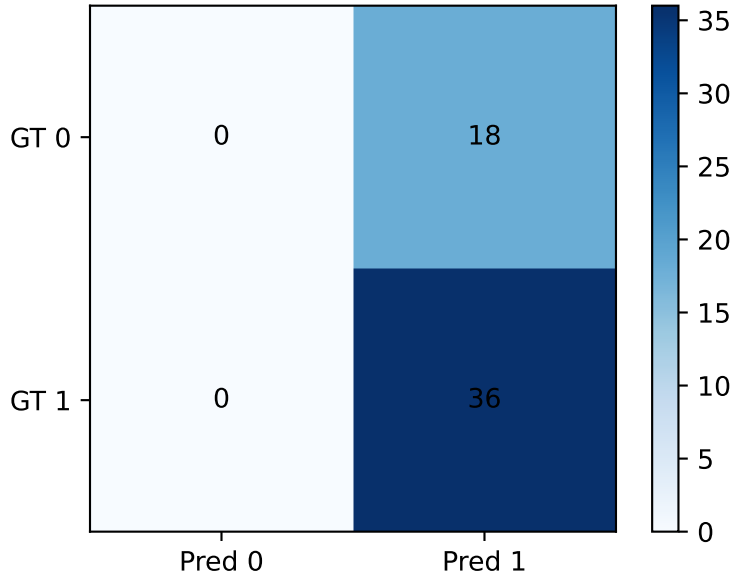
For Edema, positives=9, negatives=45.

- TN=0: images without Edema correctly predicted as negative.
- FP=45: images without Edema incorrectly flagged as present.
- FN=0: images with Edema missed by the model.
- TP=9: images with Edema correctly detected.

Here, the model predicts this disease for nearly every case (no true negatives), so recall is high but spec



Pleural Effusion: Confusion Matrix (threshold=0.5)

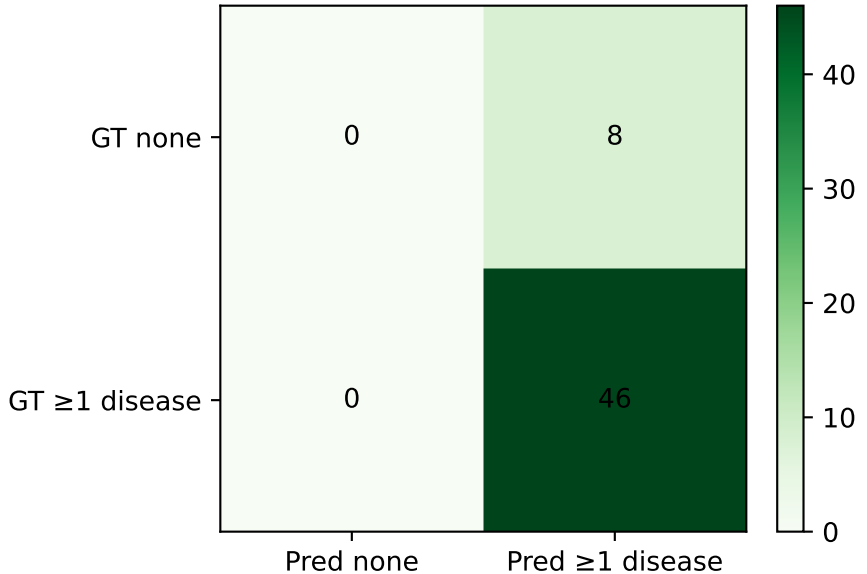


For Pleural Effusion, positives=36, negatives=18.

- TN=0: images without Pleural Effusion correctly predicted as negative.
- FP=18: images without Pleural Effusion incorrectly flagged as present.
- FN=0: images with Pleural Effusion missed by the model.
- TP=36: images with Pleural Effusion correctly detected.

Here, the model predicts this disease for nearly every case (no true negatives), so recall is high but spec

Any-disease confusion matrix (threshold=0.5)



We compress the multi-label task into a single decision per image:

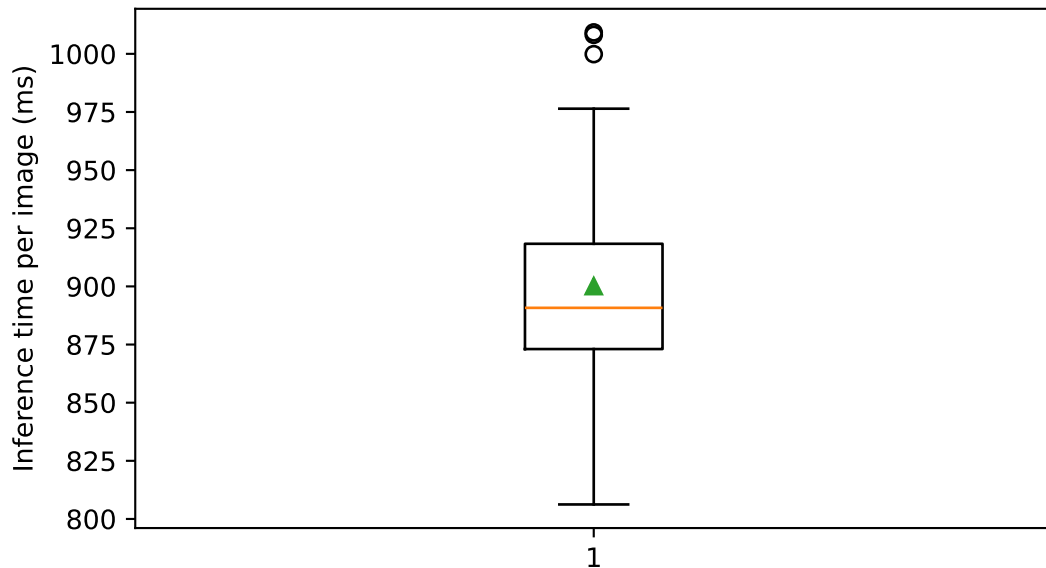
- GT ≥1 disease vs GT none (over the 5 CheXpert labels).
- Pred ≥1 disease vs Pred none, using prob ≥ 0.5.

TN=0, FP=8, FN=0, TP=46

Accuracy: 0.852, Precision: 0.852, Recall: 1.000, F1: 0.920.

This view answers: did the model detect at least one true disease on each image?

## MedCLIP zero-shot disease head - CPU inference time



Boxplot summarizing median, IQR, outliers and mean of per-image inference time. This helps compare disease-head latency with view-classification experiments.

Per-image labels & scores (rows 1-15 of 54)

UID	Image_ID	Pathology_Labels_14	Atelectasis	Atelectasis	Cardiomeg	Cardiomeg	Consolidat	Consolidat	Edema_gt	Edema_p	Pleural Eff	Pleural Effusion_
174	174_IM-0488-1001.dcm.png	pneumothorax,support_devices	0	0.549	0	0.567	0	0.544	0	0.553	0	0.547
174	174_IM-0488-2001.dcm.png	pneumothorax,support_devices	0	0.607	0	0.563	0	0.557	0	0.556	0	0.572
206	206_IM-0697-6001.dcm.png	cardiomegaly,lung_lesion	0	0.568	1	0.579	0	0.572	0	0.628	0	0.533
289	289_IM-1292-0001-0001.dcm.png	edema,consolidation,pleural_effusion,support_devices	0	0.551	0	0.568	1	0.556	1	0.548	1	0.564
289	289_IM-1292-0001-0002.dcm.png	edema,consolidation,pleural_effusion,support_devices	0	0.543	0	0.560	1	0.541	1	0.549	1	0.541
295	295_IM-1348-1001.dcm.png	consolidation,pneumothorax,pleural_effusion,fracture	0	0.578	0	0.566	1	0.544	0	0.554	1	0.544
295	295_IM-1348-2001.dcm.png	consolidation,pneumothorax,pleural_effusion,fracture	0	0.608	0	0.558	1	0.560	0	0.581	1	0.564
332	332_IM-1588-1001.dcm.png	lung_opacity,pneumonia,pleural_effusion	0	0.621	0	0.561	0	0.557	0	0.551	1	0.548
332	332_IM-1588-2001.dcm.png	lung_opacity,pneumonia,pleural_effusion	0	0.589	0	0.563	0	0.552	0	0.570	1	0.555
342	342_IM-1655-0001-0001.dcm.png	cardiomegaly	0	0.606	1	0.565	0	0.554	0	0.552	0	0.547
342	342_IM-1655-0001-0002.dcm.png	cardiomegaly	0	0.577	1	0.598	0	0.563	0	0.583	0	0.538
403	403_IM-2052-2001.dcm.png	lung_opacity,fracture	0	0.576	0	0.557	0	0.555	0	0.581	0	0.544
403	403_IM-2052-1001.dcm.png	lung_opacity,fracture	0	0.559	0	0.553	0	0.538	0	0.563	0	0.565
552	552_IM-2155-1001.dcm.png	consolidation,pleural_effusion	0	0.553	0	0.561	1	0.543	0	0.548	1	0.544
552	552_IM-2155-2001.dcm.png	consolidation,pleural_effusion	0	0.602	0	0.562	1	0.556	0	0.556	1	0.561

Per-image labels & scores (rows 16–30 of 54)

UID	Image_ID	Pathology_Labels_14	Atelectasis	Atelectasis	Cardiomeg	Cardiomeg	Consolidat	Consolidat	Edema_gt	Edema_p	Pleural Eff	Pleural Effusion
558	558_IM-2158-0001-0001.dcm.png	consolidation,pneumothorax	0	0.542	0	0.568	1	0.541	0	0.548	0	0.543
558	558_IM-2158-0001-0002.dcm.png	consolidation,pneumothorax	0	0.545	0	0.567	1	0.544	0	0.547	0	0.538
621	621_IM-2203-1001.dcm.png	pneumonia,pneumothorax	0	0.549	0	0.565	0	0.545	0	0.552	0	0.547
621	621_IM-2203-2001.dcm.png	pneumonia,pneumothorax	0	0.569	0	0.568	0	0.547	0	0.553	0	0.545
729	729_IM-2288-1001.dcm.png	lung_opacity,lung_lesion,pneumothorax,pleural effusion	0	0.562	0	0.664	0	0.547	0	0.610	1	0.545
729	729_IM-2288-2001.dcm.png	lung_opacity,lung_lesion,pneumothorax,pleural effusion	0	0.552	0	0.679	0	0.550	0	0.600	1	0.535
1007	1007_IM-0008-1001.dcm.png	lung_opacity,pleural_effusion	0	0.539	0	0.564	0	0.544	0	0.552	1	0.542
1007	1007_IM-0008-3001.dcm.png	lung_opacity,pleural_effusion	0	0.543	0	0.563	0	0.542	0	0.554	1	0.546
1007	1007_IM-0008-2001.dcm.png	lung_opacity,pleural_effusion	0	0.549	0	0.574	0	0.547	0	0.554	1	0.548
1053	1053_IM-0040-1001.dcm.png	cardiomegaly,consolidation,atelectasis,pneumo thorax	1	0.744	1	0.563	1	0.568	0	0.564	0	0.614
1053	1053_IM-0040-3003.dcm.png	cardiomegaly,consolidation,atelectasis,pneumo thorax	1	0.628	1	0.555	1	0.567	0	0.584	0	0.716
1059	1059_IM-0041-1001.dcm.png	lung_opacity,pneumonia,pleural_effusion	0	0.607	0	0.585	0	0.557	0	0.574	1	0.562
1375	1375_IM-0241-1001.dcm.png	cardiomegaly,edema,pneumonia	0	0.551	1	0.719	0	0.554	1	0.674	0	0.556
1375	1375_IM-0241-3001.dcm.png	cardiomegaly,edema,pneumonia	0	0.550	1	0.721	0	0.554	1	0.675	0	0.548
1375	1375_IM-0241-4001.dcm.png	cardiomegaly,edema,pneumonia	0	0.551	1	0.696	0	0.548	1	0.712	0	0.572

Per-image labels & scores (rows 31-45 of 54)

UID	Image_ID	Pathology_Labels_14	Atelectasis	Atelectasis	Cardiomeg	Cardiomeg	Consolidat	Consolidat	Edema_gt	Edema_p	Pleural Eff	Pleural Effusion
1437	1437_IM-0281-3001.dcm.png	lung_opacity,pneumothorax,pleural_effusion	0	0.545	0	0.563	0	0.542	0	0.552	1	0.555
1437	1437_IM-0281-2001.dcm.png	lung_opacity,pneumothorax,pleural_effusion	0	0.546	0	0.568	0	0.547	0	0.552	1	0.561
1577	1577_IM-0375-1001.dcm.png	lung_opacity,pleural_effusion	0	0.579	0	0.561	0	0.546	0	0.547	1	0.537
1577	1577_IM-0375-2001.dcm.png	lung_opacity,pleural_effusion	0	0.546	0	0.564	0	0.541	0	0.540	1	0.538
1581	1581_IM-0378-1001.dcm.png	lung_lesion,consolidation,pneumothorax,pleural_effusion	0	0.557	0	0.564	1	0.545	0	0.541	1	0.542
1581	1581_IM-0378-2001.dcm.png	lung_lesion,consolidation,pneumothorax,pleural_effusion	0	0.538	0	0.561	1	0.542	0	0.548	1	0.540
1678	1678_IM-0447-0001-0001.dcm.png	lung_opacity,lung_lesion,consolidation,pleural_effusion,support_devices	0	0.608	0	0.541	1	0.585	0	0.537	1	0.755
1678	1678_IM-0447-0001-0002.dcm.png	lung_opacity,lung_lesion,consolidation,pleural_effusion,support_devices	0	0.649	0	0.604	1	0.550	0	0.547	1	0.726
1815	1815_IM-0527-1001.dcm.png	pleural_effusion	0	0.561	0	0.560	0	0.551	0	0.540	1	0.593
1815	1815_IM-0527-2001.dcm.png	pleural_effusion	0	0.561	0	0.564	0	0.567	0	0.541	1	0.551
1867	1867_IM-0560-4004.dcm.png	cardiomegaly,lung_opacity,edema,atelectasis,pleural_effusion,fracture	1	0.588	1	0.641	0	0.562	1	0.569	1	0.619
1867	1867_IM-0560-1001.dcm.png	cardiomegaly,lung_opacity,edema,atelectasis,pleural_effusion,fracture	1	0.613	1	0.579	0	0.564	1	0.570	1	0.577
2077	2077_IM-0710-1001.dcm.png	consolidation,atelectasis,pneumothorax,pleural_effusion,support_devices	1	0.652	0	0.542	1	0.550	0	0.535	1	0.549
2077	2077_IM-0710-2001.dcm.png	consolidation,atelectasis,pneumothorax,pleural_effusion,support_devices	1	0.579	0	0.561	1	0.563	0	0.590	1	0.565
2095	2095_IM-0725-4001.dcm.png	atelectasis,pneumothorax,pleural_effusion,fracture	1	0.632	0	0.557	0	0.548	0	0.562	1	0.567

Per-image labels & scores (rows 46–54 of 54)

UID	Image_ID	Pathology_Labels_14	Atelectasis	Atelectasis	Cardiomeg	Cardiomeg	Consolidat	Consolidat	Edema_gt	Edema_p	Pleural Eff	Pleural Effusion
2095	2095_IM-0725-2001.dcm.png	atelectasis,pneumothorax,pleural_effusion,fracture	1	0.543	0	0.572	0	0.549	0	0.704	1	0.686
2396	2396_IM-0945-1001.dcm.png	lung_opacity,consolidation,pneumothorax,pleural_effusion	0	0.570	0	0.554	1	0.596	0	0.558	1	0.588
2396	2396_IM-0945-2001.dcm.png	lung_opacity,consolidation,pneumothorax,pleural_effusion	0	0.584	0	0.564	1	0.601	0	0.556	1	0.614
2752	2752_IM-1202-0001-0001.dcm.png	cardiomegaly,lung_opacity,edema,atelectasis,pleural_effusion	1	0.642	1	0.597	0	0.616	1	0.589	1	0.779
2752	2752_IM-1202-0001-0002.dcm.png	cardiomegaly,lung_opacity,edema,atelectasis,pleural_effusion	1	0.721	1	0.561	0	0.588	1	0.559	1	0.759
3074	3074_IM-1435-1001.dcm.png	lung_opacity,pneumonia	0	0.572	0	0.566	0	0.577	0	0.557	0	0.549
3074	3074_IM-1435-2001.dcm.png	lung_opacity,pneumonia	0	0.560	0	0.644	0	0.550	0	0.589	0	0.552
3636	3636_IM-1803-1001.dcm.png	lung_opacity,pneumothorax,pleural_effusion	0	0.511	0	0.539	0	0.541	0	0.543	1	0.675
3636	3636_IM-1803-2001.dcm.png	lung_opacity,pneumothorax,pleural_effusion	0	0.545	0	0.566	0	0.547	0	0.548	1	0.571