MedCLIP Zero-Shot Disease Classification
IU-CXR sanity subset (pathology-focused)

Tag / Image variant: raw
Total images in CSV: 29

Ground truth source:
  - chexpert_label (single-label per image)

Evaluation setup:
- For each image, MedCLIP PromptClassifier (CheXpert head) produces
  probabilities for 5 diseases:
    Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion.
- This script converts those scores into disease-present vs absent labels
  using a 0.5 decision threshold and computes per-disease metrics.

Micro-averaged performance across all 5 diseases (threshold 0.5):
  • Micro Accuracy : 0.138
  • Micro Precision: 0.138
  • Micro Recall   : 1.000
  • Micro F1   : 0.242
  • Total label decisions evaluated: 145

Top-K disease ranking (only images with ≥1 positive label):
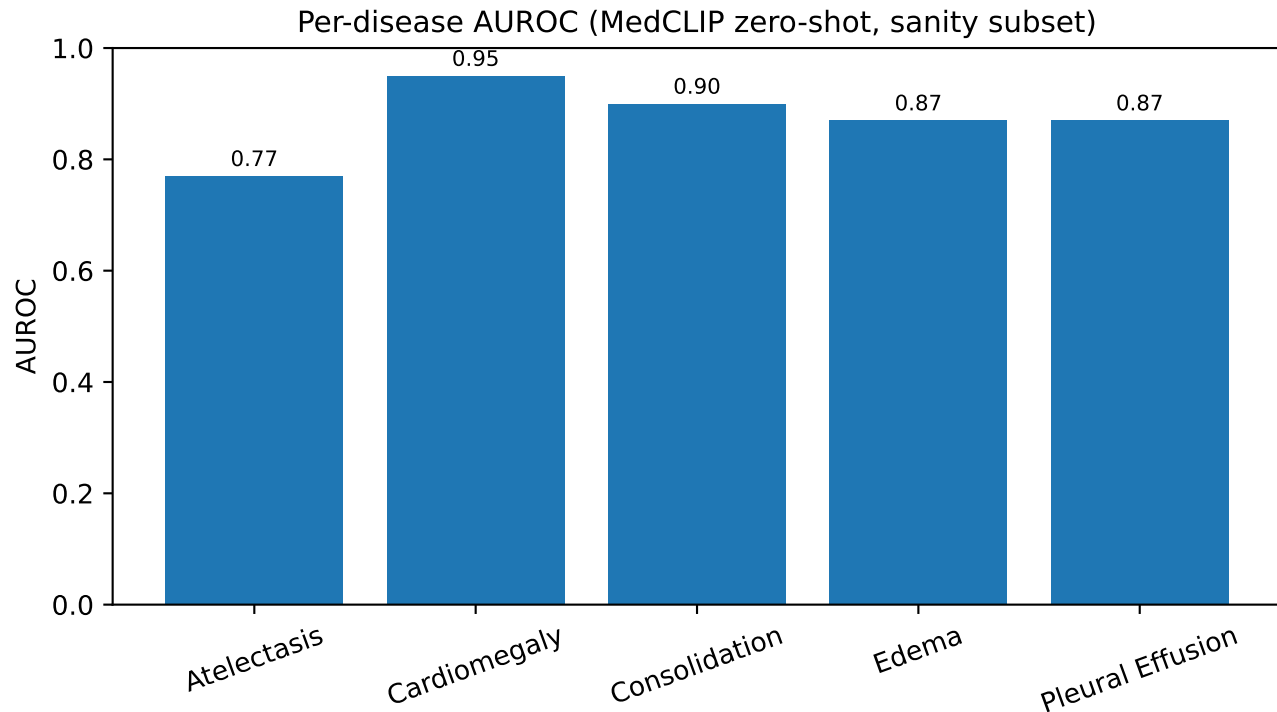 • Top-1 accuracy: 0.600 (N=20)
  • Top-3 accuracy: 0.950 (N=20)
    (Top-K is counted correct if any true disease appears in the top-K scores.)
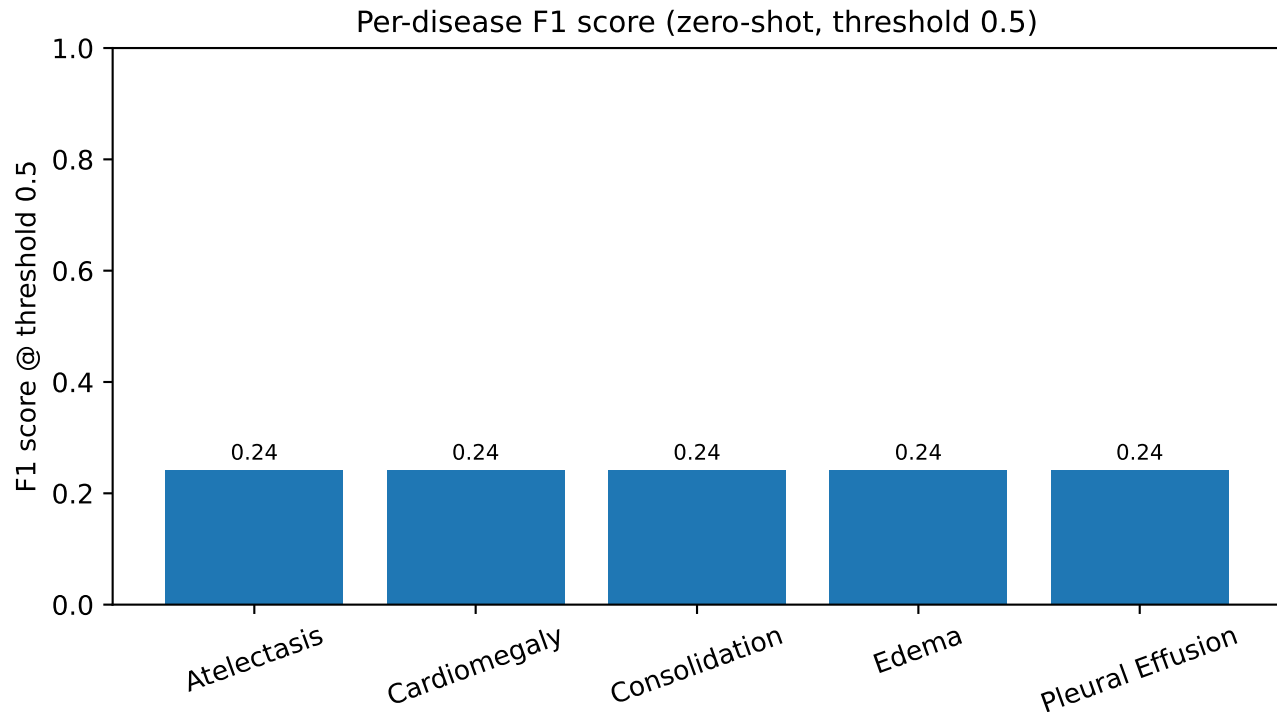
Any-hit metric (at least one disease present & predicted):
  We reduce each image to a single binary decision:
    GT ≥1 of the 5 diseases present vs none,
    Pred ≥1 of the 5 diseases positive vs none.
  • Any-hit Accuracy : 0.690
  • Any-hit Precision: 0.690
  • Any-hit Recall   : 1.000
  • Any-hit F1   : 0.816

Per-disease AUROC (MedCLIP zero-shot, sanity subset)

AUROC (Area Under ROC Curve) measures how well the model ranks positive cases above negatives for each disease.
Values closer to 1.0 indicate better ranking performance, regardless of any fixed threshold.
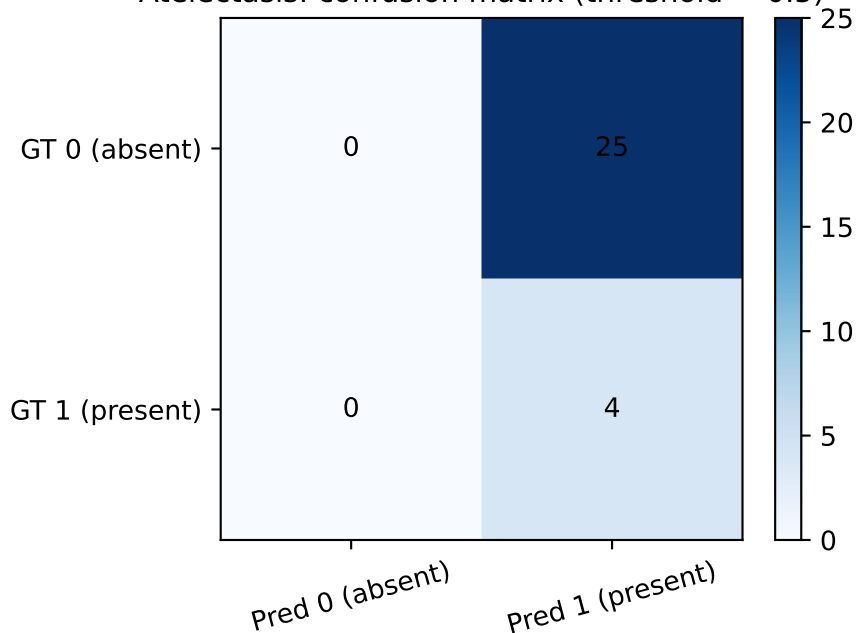
Per-disease F1 score (zero-shot, threshold 0.5)

F1 score combines precision and recall for each disease at the 0.5 probability threshold.
It is especially useful when classes are imbalanced (few positives vs many negatives).

Per-disease zero-shot metrics (all images in this subset)

Disease	Positives	Negatives	AUROC	AUPRC	Acc@0.5	Prec@0.5	Rec@0.5	F1@0.5	TP	FP	TN	FN
Atelectasis	4	25	0.770	0.353	0.138	0.138	1.000	0.242	4	25	0	0
Cardiomegaly	4	25	0.950	0.812	0.138	0.138	1.000	0.242	4	25	0	0
Consolidation	4	25	0.900	0.600	0.138	0.138	1.000	0.242	4	25	0	0
Edema	4	25	0.870	0.412	0.138	0.138	1.000	0.242	4	25	0	0
Pleural Effusion	4	25	0.870	0.415	0.138	0.138	1.000	0.242	4	25	0	0

Each row summarizes MedCLIP's zero-shot behavior for one disease. We report ranking metrics (AUROC, AUPRC)
and threshold-based metrics (Accuracy, Precision, Recall, F1 at 0.5).

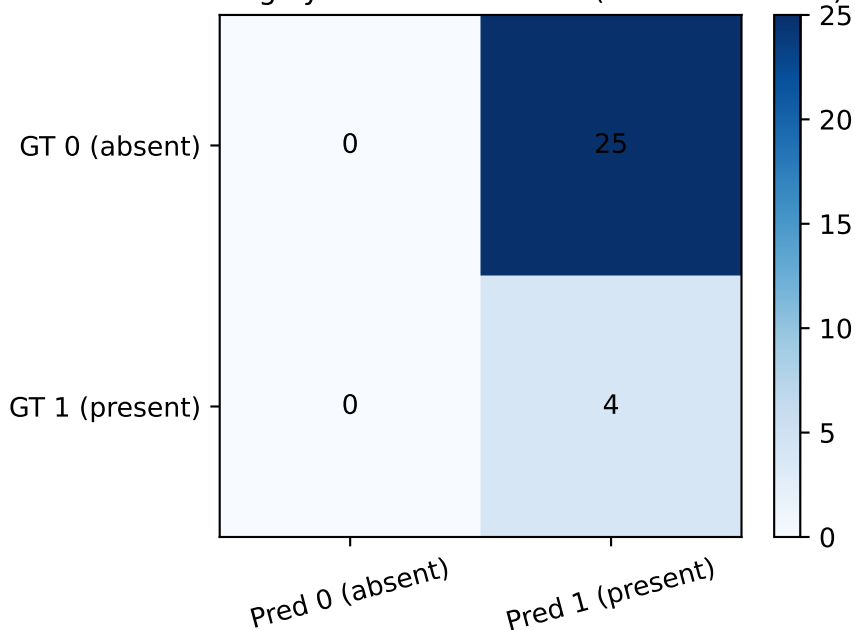Atelectasis: confusion matrix (threshold = 0.5)

For Atelectasis, positives = 4, negatives = 25.
Interpretation:
 • TN (top-left): correctly predicted as no disease.
 • FP (top-right): predicted disease but GT is negative.
 • FN (bottom-left): missed true disease cases.
 • TP (bottom-right): correctly detected disease.

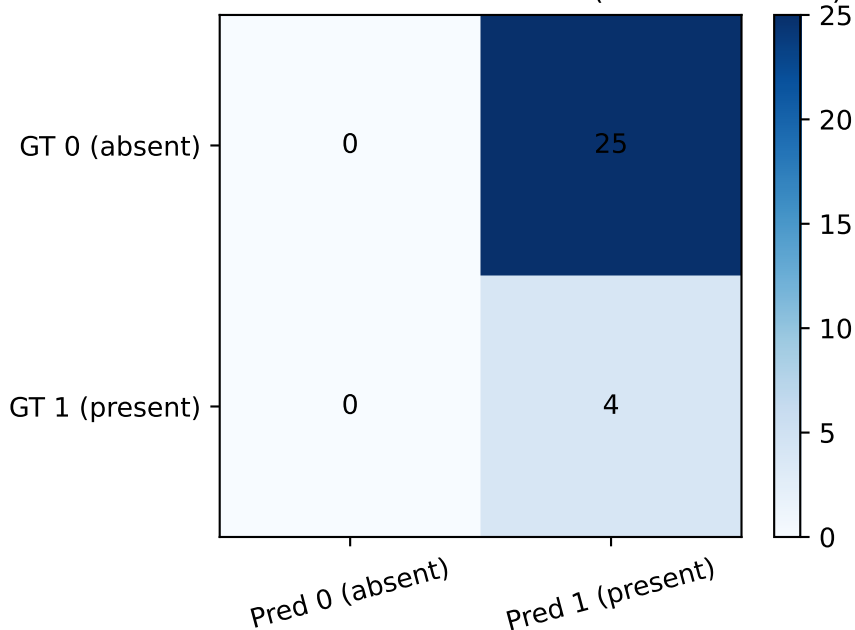Cardiomegaly: confusion matrix (threshold = 0.5)

For Cardiomegaly, positives = 4, negatives = 25.
Interpretation:
  • TN (top-left): correctly predicted as no disease.
  • FP (top-right): predicted disease but GT is negative.
  • FN (bottom-left): missed true disease cases.
  • TP (bottom-right): correctly detected disease.

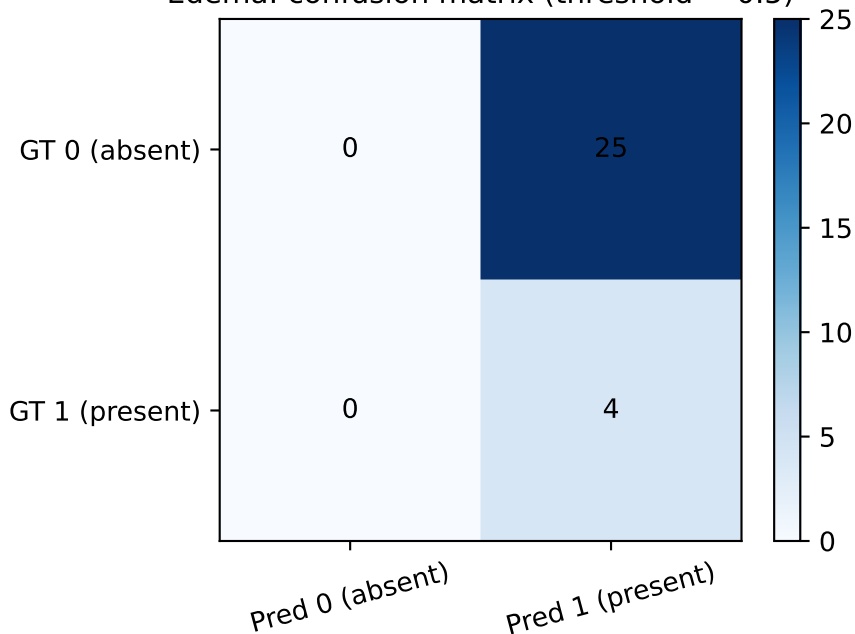Consolidation: confusion matrix (threshold = 0.5)

For Consolidation, positives = 4, negatives = 25.
Interpretation:
- TN (top-left): correctly predicted as no disease.
- FP (top-right): predicted disease but GT is negative.
- FN (bottom-left): missed true disease cases.
- TP (bottom-right): correctly detected disease.

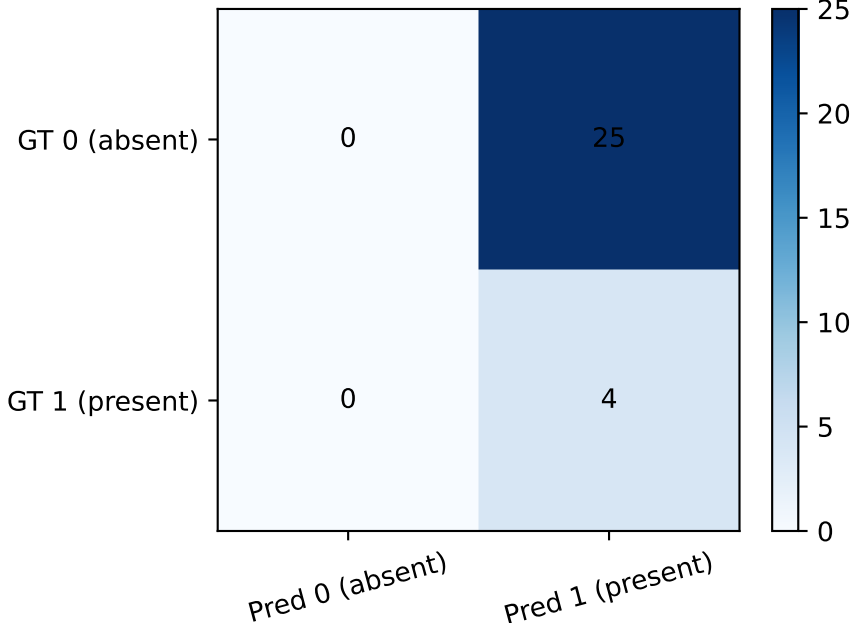Edema: confusion matrix (threshold = 0.5)

For Edema, positives = 4, negatives = 25.
Interpretation:
  • TN (top-left): correctly predicted as no disease.
  • FP (top-right): predicted disease but GT is negative.
  • FN (bottom-left): missed true disease cases.
  • TP (bottom-right): correctly detected disease.

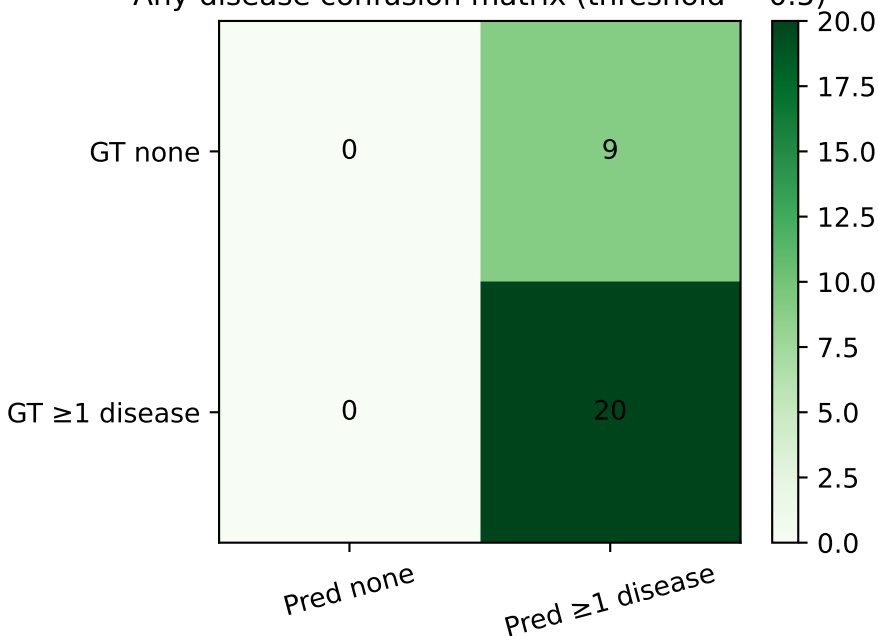Pleural Effusion: confusion matrix (threshold = 0.5)

For Pleural Effusion, positives = 4, negatives = 25.
Interpretation:
- TN (top-left): correctly predicted as no disease.
- FP (top-right): predicted disease but GT is negative.
- FN (bottom-left): missed true disease cases.
- TP (bottom-right): correctly detected disease.

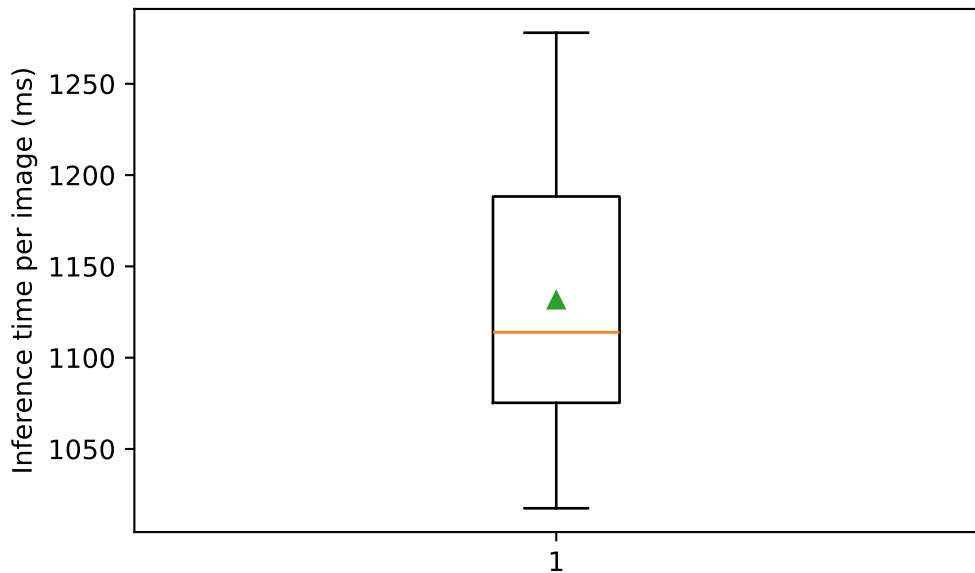Any-disease confusion matrix (threshold = 0.5)

We compress the multi-label task into a single 'any disease' decision:
  • GT ≥1 disease vs GT none
  • Pred ≥1 disease vs Pred none (prob ≥ 0.5).

TN=0, FP=9, FN=0, TP=20
Accuracy=0.690, Precision=0.690, Recall=1.000, F1=0.816.

MedCLIP zero-shot disease head – per-image latency

Boxplot summarizing median, interquartile range, mean and outliers for per-image inference time. This is useful to compare runtime across different image variants.

filename	chexpert_	Atelectasi	Atelectasi	Cardiome	Cardiome	Consolida	Consolida	Edema_gt	Edema_p	Pleural Eff	Pleural Eff usion_p
3638_IM-1804-10	Cardiome	0	0.547	1	0.583	0	0.549	0	0.572	0	0.538
2716_IM-1181-20	Consolida	0	0.609	0	0.622	1	0.591	0	0.688	0	0.745
2422_IM-0965-20	Consolida	0	0.577	0	0.570	1	0.592	0	0.568	0	0.593
3249_IM-1539-20	Edema	0	0.607	0	0.603	0	0.574	1	0.616	0	0.780
1439_IM-0282-10	Atelectasi	1	0.563	0	0.563	0	0.544	0	0.546	0	0.575
1439_IM-0282-10	Atelectasi	1	0.584	0	0.562	0	0.576	0	0.539	0	0.616
3249_IM-1539-10	Edema	0	0.615	0	0.637	0	0.562	1	0.627	0	0.762
91_IM-2415-2001	Pneumoth	0	0.555	0	0.548	0	0.541	0	0.542	0	0.545
3396_IM-1640-30	Pleural Eff	0	0.612	0	0.544	0	0.623	0	0.543	1	0.760
2502_IM-1027-10	Pleural Eff	0	0.561	0	0.565	0	0.615	0	0.601	1	0.708
2716_IM-1181-10	Consolida	0	0.572	0	0.564	1	0.583	0	0.681	0	0.783
2422_IM-0965-10	Consolida	0	0.574	0	0.552	1	0.627	0	0.558	0	0.555
3069_IM-1432-10	Pneumoni	0	0.554	0	0.562	0	0.620	0	0.531	0	0.534
353_IM-1726-100	Edema	0	0.533	0	0.561	0	0.558	1	0.680	0	0.677
1262_IM-0178-00	Pneumoth	0	0.535	0	0.536	0	0.538	0	0.553	0	0.536

filename	chexpert_	Atelectasi	Atelectasi	Cardiome	Cardiome	Consolida	Consolida	Edema_gt	Edema_p	Pleural Eff	Pleural Effusion_p
1262_IM-0178-10	Pneumoth	0	0.567	0	0.548	0	0.544	0	0.546	0	0.550
353_IM-1726-200	Edema	0	0.563	0	0.642	0	0.569	1	0.668	0	0.680
3348_IM-1605-20	Cardiome	0	0.558	1	0.749	0	0.551	0	0.627	0	0.530
541_IM-2147-100	Fracture	0	0.556	0	0.552	0	0.548	0	0.563	0	0.554
3348_IM-1605-10	Cardiome	0	0.542	1	0.755	0	0.544	0	0.581	0	0.527
3396_IM-1640-10	Pleural Eff	0	0.658	0	0.554	0	0.589	0	0.528	1	0.764
3069_IM-1432-20	Pneumoni	0	0.553	0	0.542	0	0.547	0	0.556	0	0.543
3638_IM-1804-20	Cardiome	0	0.548	1	0.641	0	0.551	0	0.573	0	0.544
3233_IM-1530-20	Atelectasi	1	0.592	0	0.561	0	0.559	0	0.546	0	0.554
1430_IM-0277-20	Pneumoni	0	0.581	0	0.560	0	0.550	0	0.576	0	0.619
2352_IM-0918-10	Fracture	0	0.545	0	0.545	0	0.541	0	0.534	0	0.656
2502_IM-1027-10	Pleural Eff	0	0.553	0	0.562	0	0.551	0	0.684	1	0.710
3233_IM-1530-10	Atelectasi	1	0.644	0	0.551	0	0.556	0	0.555	0	0.548
1262_IM-0178-20	Pneumoth	0	0.565	0	0.564	0	0.543	0	0.558	0	0.554