nature biomedical engineering

Supplementary information

https://doi.org/10.1038/s41551-026-01741-4

CLEAR: an auditable foundation model for radiology grounded in clinical concepts

In the format provided by the authors and unedited

Supplementary Information

Table S1: **Patient-level data split summary demonstrating no overlap between training and evaluation.**

Experiment	Split	Dataset	Images	Patients	Institution/Country
Contrastive Pretraining (Stage 1)					
	Train	MIMIC-CXR	377,110	65,379	Beth Israel Deaconess, USA
	Train	CheXpert-Plus	223,228	64,525	Stanford Hospital, USA
	Train	ReXGradient	273,004	109,487	Harvard, USA
Supervised Evaluation (Linear Probing & CBM)					
	Train	MIMIC-CXR	368,960	64,586	Beth Israel Deaconess, USA
Test Data (All Experiments)					
	Test	CheXpert Test	500	500	Stanford Hospital, USA
	Test	VinDr-CXR	3,000	—	Vietnam
	Test	PadChest	7,943	7,272	San Juan Hospital, Spain
	Test	Indiana	7,466	3,851	Indiana University, USA

Table S2: **Determinism validation of Ministral-8B observation extraction** on 500 MIMIC-CXR reports (temperature = 0, greedy decoding). Each report was processed twice with identical inputs.

Metric	Count	Rate
Observation list exact match	491 / 500	98.2%
Total observations (run 1)	1,529	
Total observations (run 2)	1,525	

Table S3: **Representative extraction examples** showing all observations produced by Ministral-8B for each impression. Negated observations are shown in *italics*.

Original Impression	**All Extracted Observations**
"Elevated right hemidiaphragm. Tiny left pleural effusion. No signs of pneumonia."	1. Elevated right hemidiaphragm 2. Tiny left pleural effusion 3. *No signs of pneumonia*
"Mild cardiomegaly, new since the prior study, and enlargement of the azygos compatible with volume overload without frank pulmonary edema."	1. Mild cardiomegaly 2. Enlargement of the azygos 3. *Volume overload without frank pulmonary edema*
"No previous images. The cardiac silhouette is mildly enlarged with left ventricular prominence. However, no vascular congestion, pleural effusion, or acute focal pneumonia."	1. Cardiac silhouette is mildly enlarged 2. Left ventricular prominence 3. *No vascular congestion* 4. *No pleural effusion* 5. *No acute focal pneumonia*
"Normal heart, lungs, hila, mediastinum, and pleural surfaces. No evidence of intrathoracic malignancy or infection."	1. Normal heart 2. Normal lungs 3. Normal hila 4. Normal mediastinum 5. Normal pleural surfaces 6. *No evidence of intrathoracic malignancy* 7. *No evidence of intrathoracic infection*

Table S4: **Radiologist evaluation of extraction fidelity.** Three board-certified radiologists independently rated 50 report-observation pairs (15 negated, 35 non-negated) on faithfulness and negation preservation.

Metric	**All pairs (50)**	**Negated pairs (15)**
Faithful (%)	97.3	100.0
Partially faithful (%)	2.7	0.0
Hallucinated (%)	0.0	0.0
Negation correct (%)	–	93.3
Negation incorrect (%)	–	0.0
Negation N/A (%)	–	6.7
Fleiss' κ (negation)	0.934	

Table S5: **Embedding model selection on CheXpert validation set** (200 images). SFR-Embedding-Mistral achieved the highest macro AUROC and was selected for CLEAR. External test set evaluations were performed only after this selection.

Embedding Model	**Macro AUROC**
SFR-Embedding-Mistral	$\mathbf{0.789 \pm 0.076}$
text-embedding-3-small (OpenAI)	0.692 ± 0.172
Qwen3-Embedding-8B	0.618 ± 0.195
BiomedBERT	0.607 ± 0.195

Table S6: **Prompt variants tested for robustness analysis.** Each variant represents a different way to express the presence or absence of a finding.

Variant	Positive Prompt	Negative Prompt
Standard	"{disease}"	"no {disease}"
Absence	"{disease}"	"absence of {disease}"
Present/Absent	"{disease} present"	"{disease} absent"
Not Present	"{disease}"	"{disease} not present"
Clinical	"findings consistent with {disease}"	"no evidence of {disease}"
Negative	"{disease}"	"negative for {disease}"

Table S7: **Prompt robustness analysis across embedding models and datasets.** Mean AUROC and range (min–max) computed across six prompt variants. Narrower ranges indicate greater robustness to prompt formulation changes.

Model	VinDr-CXR	PadChest	Indiana
Mean AUROC [range]			
SFR-Embedding-Mistral	**0.774** [0.760–0.788]	**0.717** [0.710–0.724]	**0.735** [0.719–0.748]
text-embedding-3-small	0.693 [0.636–0.758]	0.670 [0.645–0.703]	0.672 [0.636–0.718]
Qwen3-Embedding-8B	0.718 [0.681–0.757]	0.672 [0.621–0.717]	0.699 [0.666–0.728]
BiomedBERT	0.549 [0.534–0.571]	0.570 [0.529–0.605]	0.556 [0.525–0.582]

Table S8: **Zero-shot performance on the VinDr-CXR dataset.** AUROC values with 95% confidence intervals. Bold indicates best performance.

Finding	N	CLEAR	CheXzero	BiomedCLIP	OpenAICLIP
Pneumonia	246	**0.937 (0.922, 0.950)**	0.889 (0.865, 0.912)	0.675 (0.636, 0.713)	0.545 (0.507, 0.583)
Pleural effusion	111	0.904 (0.882, 0.923)	**0.959 (0.934, 0.979)**	0.871 (0.833, 0.908)	0.529 (0.469, 0.587)
Lung tumor	80	**0.884 (0.847, 0.918)**	0.810 (0.757, 0.860)	0.784 (0.731, 0.833)	0.541 (0.480, 0.598)
Pneumothorax	18	**0.879 (0.837, 0.918)**	0.828 (0.744, 0.898)	0.662 (0.487, 0.822)	0.450 (0.342, 0.571)
Mediastinal shift	20	**0.878 (0.825, 0.923)**	0.846 (0.750, 0.919)	0.828 (0.720, 0.918)	0.485 (0.354, 0.606)
Consolidation	96	0.877 (0.855, 0.896)	**0.905 (0.869, 0.939)**	0.615 (0.552, 0.679)	0.422 (0.365, 0.478)
ILD	221	0.816 (0.789, 0.840)	**0.838 (0.811, 0.866)**	0.711 (0.675, 0.749)	0.603 (0.565, 0.640)
Lung Opacity	84	0.809 (0.768, 0.845)	**0.812 (0.768, 0.851)**	0.679 (0.619, 0.736)	0.558 (0.488, 0.625)
Nodule/Mass	176	**0.793 (0.754, 0.829)**	0.662 (0.622, 0.700)	0.536 (0.490, 0.583)	0.565 (0.521, 0.607)
Atelectasis	86	**0.785 (0.735, 0.834)**	0.745 (0.696, 0.789)	0.709 (0.649, 0.769)	0.554 (0.496, 0.615)
Cardiomegaly	309	0.782 (0.753, 0.810)	**0.936 (0.924, 0.948)**	0.769 (0.740, 0.796)	0.591 (0.557, 0.623)
Rib fracture	11	**0.776 (0.659, 0.873)**	0.579 (0.394, 0.767)	0.511 (0.327, 0.683)	0.491 (0.325, 0.643)
Pulmonary fibrosis	217	**0.768 (0.735, 0.800)**	0.668 (0.631, 0.706)	0.657 (0.617, 0.696)	0.604 (0.568, 0.643)
Other lesion	94	**0.763 (0.720, 0.808)**	0.745 (0.695, 0.793)	0.560 (0.499, 0.624)	0.571 (0.517, 0.625)
Infiltration	58	0.725 (0.652, 0.791)	**0.843 (0.780, 0.894)**	0.451 (0.373, 0.534)	0.519 (0.437, 0.600)
Pleural thickening	169	**0.722 (0.677, 0.765)**	0.688 (0.648, 0.728)	0.635 (0.588, 0.681)	0.608 (0.564, 0.651)
Aortic enlargement	220	0.709 (0.673, 0.745)	**0.758 (0.726, 0.788)**	0.726 (0.690, 0.761)	0.543 (0.506, 0.579)
No finding	2051	**0.693 (0.672, 0.714)**	0.139 (0.125, 0.154)	0.369 (0.345, 0.390)	0.506 (0.484, 0.528)
Tuberculosis	164	0.666 (0.623, 0.710)	**0.731 (0.690, 0.770)**	0.647 (0.600, 0.692)	0.502 (0.460, 0.546)
Other disease	657	**0.636 (0.616, 0.656)**	0.604 (0.579, 0.627)	0.455 (0.430, 0.481)	0.489 (0.464, 0.514)
Calcification	194	0.617 (0.573, 0.658)	**0.760 (0.728, 0.791)**	0.604 (0.564, 0.644)	0.382 (0.336, 0.424)

Table S9: **Zero-shot performance on the PadChest dataset.** AUROC values with 95% confidence intervals. Bold indicates best performance.

Finding	N	CLEAR	CheXzero	BiomedCLIP	OpenAICLIP
pleural effusion	376	**0.966 (0.958, 0.972)**	0.965 (0.955, 0.974)	0.928 (0.915, 0.942)	0.565 (0.538, 0.592)
pulmonary mass	42	**0.941 (0.907, 0.968)**	0.883 (0.820, 0.934)	0.888 (0.826, 0.943)	0.567 (0.484, 0.646)
consolidation	63	**0.916 (0.893, 0.936)**	0.875 (0.823, 0.921)	0.757 (0.698, 0.815)	0.530 (0.452, 0.606)
heart insufficiency	95	0.915 (0.896, 0.931)	**0.924 (0.906, 0.939)**	0.789 (0.748, 0.833)	0.349 (0.296, 0.403)
hilar congestion	117	0.845 (0.820, 0.869)	**0.888 (0.867, 0.909)**	0.725 (0.682, 0.769)	0.576 (0.524, 0.629)
hypoexpansion	27	**0.842 (0.792, 0.885)**	0.764 (0.670, 0.852)	0.634 (0.530, 0.742)	0.450 (0.346, 0.556)
interstitial pattern	368	**0.841 (0.822, 0.860)**	0.819 (0.795, 0.839)	0.745 (0.719, 0.772)	0.424 (0.394, 0.456)
laminar atelectasis	235	**0.840 (0.814, 0.865)**	0.685 (0.650, 0.716)	0.500 (0.462, 0.539)	0.515 (0.479, 0.549)
pneumonia	336	**0.835 (0.811, 0.856)**	0.816 (0.792, 0.840)	0.645 (0.614, 0.674)	0.498 (0.465, 0.530)
cardiomegaly	739	0.825 (0.813, 0.835)	**0.896 (0.887, 0.905)**	0.773 (0.758, 0.788)	0.499 (0.475, 0.521)
atelectasis	145	**0.820 (0.782, 0.853)**	0.786 (0.748, 0.820)	0.561 (0.508, 0.621)	0.434 (0.387, 0.481)
lobar atelectasis	35	0.818 (0.760, 0.874)	0.721 (0.634, 0.803)	**0.826 (0.743, 0.899)**	0.534 (0.429, 0.638)
infiltrates	253	**0.795 (0.767, 0.821)**	0.750 (0.716, 0.780)	0.569 (0.528, 0.610)	0.483 (0.446, 0.519)
alveolar pattern	271	0.791 (0.769, 0.813)	**0.907 (0.886, 0.926)**	0.428 (0.388, 0.467)	0.448 (0.412, 0.485)
ground glass pattern	31	**0.779 (0.705, 0.847)**	0.582 (0.461, 0.698)	0.625 (0.522, 0.728)	0.553 (0.455, 0.647)
rib fracture	32	**0.777 (0.729, 0.820)**	0.488 (0.378, 0.596)	0.628 (0.513, 0.742)	0.354 (0.252, 0.469)
bronchiectasis	122	**0.762 (0.726, 0.801)**	0.744 (0.697, 0.786)	0.658 (0.609, 0.702)	0.510 (0.459, 0.560)
pleural thickening	47	0.762 (0.700, 0.821)	**0.793 (0.732, 0.851)**	0.657 (0.577, 0.736)	0.508 (0.426, 0.587)
costophrenic angle blunting	345	0.759 (0.734, 0.784)	**0.787 (0.764, 0.809)**	0.483 (0.450, 0.515)	0.546 (0.515, 0.577)
tuberculosis sequelae	42	**0.758 (0.691, 0.817)**	0.616 (0.528, 0.702)	0.593 (0.505, 0.678)	0.523 (0.436, 0.613)
hypoexpansion basal	36	0.736 (0.676, 0.794)	**0.753 (0.691, 0.808)**	0.524 (0.425, 0.620)	0.484 (0.384, 0.576)
hiatal hernia	115	0.734 (0.704, 0.764)	**0.759 (0.724, 0.792)**	0.489 (0.440, 0.540)	0.573 (0.523, 0.622)
hemidiaphragm elevation	149	**0.726 (0.694, 0.755)**	0.631 (0.579, 0.679)	0.660 (0.614, 0.701)	0.528 (0.484, 0.572)
aortic button enlargement	49	**0.726 (0.666, 0.785)**	0.716 (0.647, 0.782)	0.672 (0.599, 0.745)	0.407 (0.326, 0.492)
descendent aortic elongation	35	0.717 (0.666, 0.766)	**0.759 (0.701, 0.811)**	0.554 (0.457, 0.658)	0.508 (0.398, 0.625)
aortic atheromatosis	322	**0.699 (0.676, 0.722)**	0.567 (0.536, 0.598)	0.557 (0.524, 0.589)	0.598 (0.568, 0.629)
supra aortic elongation	28	0.694 (0.630, 0.754)	**0.704 (0.630, 0.776)**	0.564 (0.437, 0.691)	0.431 (0.320, 0.539)
chronic changes	749	**0.685 (0.668, 0.703)**	0.658 (0.640, 0.677)	0.529 (0.509, 0.550)	0.561 (0.540, 0.583)
aortic elongation	623	0.685 (0.668, 0.700)	**0.692 (0.673, 0.709)**	0.575 (0.552, 0.599)	0.547 (0.522, 0.572)
nodule	149	**0.665 (0.628, 0.705)**	0.546 (0.499, 0.597)	0.521 (0.472, 0.571)	0.542 (0.493, 0.589)
osteopenia	44	**0.663 (0.583, 0.729)**	0.657 (0.569, 0.737)	0.639 (0.569, 0.706)	0.564 (0.474, 0.655)
fibrotic band	250	**0.654 (0.622, 0.685)**	0.575 (0.539, 0.609)	0.518 (0.482, 0.553)	0.547 (0.512, 0.583)
hyperinflated lung	36	0.650 (0.559, 0.737)	**0.768 (0.683, 0.843)**	0.728 (0.641, 0.808)	0.545 (0.460, 0.625)
calcified granuloma	141	**0.647 (0.602, 0.689)**	0.513 (0.465, 0.562)	0.452 (0.407, 0.498)	0.533 (0.485, 0.583)
hilar enlargement	91	**0.646 (0.594, 0.695)**	0.622 (0.565, 0.682)	0.548 (0.486, 0.605)	0.482 (0.422, 0.539)
vertebral anterior compression	116	**0.639 (0.586, 0.690)**	0.633 (0.583, 0.685)	0.537 (0.484, 0.590)	0.581 (0.528, 0.632)
kyphosis	221	0.630 (0.598, 0.663)	**0.653 (0.619, 0.686)**	0.527 (0.487, 0.566)	0.428 (0.390, 0.464)
copd signs	905	0.615 (0.599, 0.633)	**0.752 (0.735, 0.768)**	0.548 (0.528, 0.566)	0.584 (0.564, 0.603)
calcified densities	65	**0.612 (0.552, 0.672)**	0.559 (0.500, 0.614)	0.426 (0.350, 0.498)	0.435 (0.357, 0.511)
granuloma	35	**0.611 (0.530, 0.691)**	0.516 (0.436, 0.595)	0.475 (0.389, 0.566)	0.565 (0.463, 0.665)
mediastinic lipomatosis	28	0.610 (0.541, 0.671)	**0.654 (0.560, 0.745)**	0.499 (0.383, 0.602)	0.604 (0.503, 0.702)

Table S10: **Zero-shot performance on the PadChest dataset (continued).** AUROC values with 95% confidence intervals. Bold indicates best performance.

Finding	N	CLEAR	CheXzero	BiomedCLIP	OpenAICLIP
sclerotic bone lesion	30	0.598 (0.487, 0.696)	0.415 (0.313, 0.510)	0.556 (0.449, 0.650)	**0.658 (0.574, 0.734)**
bullas	33	**0.597 (0.501, 0.688)**	0.459 (0.369, 0.548)	0.506 (0.433, 0.587)	0.465 (0.365, 0.566)
vascular hilar enlargement	276	0.588 (0.556, 0.619)	**0.625 (0.594, 0.657)**	0.546 (0.513, 0.580)	0.542 (0.507, 0.578)
pseudonodule	145	0.587 (0.544, 0.634)	0.468 (0.425, 0.507)	0.551 (0.502, 0.596)	**0.610 (0.565, 0.653)**
diaphragmatic eventration	76	**0.586 (0.536, 0.632)**	0.584 (0.526, 0.639)	0.423 (0.360, 0.487)	0.545 (0.479, 0.608)
nipple shadow	76	0.576 (0.510, 0.642)	0.559 (0.500, 0.618)	0.495 (0.433, 0.554)	**0.585 (0.521, 0.646)**
flattened diaphragm	36	**0.568 (0.486, 0.649)**	0.514 (0.435, 0.594)	0.527 (0.443, 0.605)	0.433 (0.333, 0.533)
scoliosis	544	0.566 (0.543, 0.589)	**0.602 (0.576, 0.627)**	0.554 (0.528, 0.578)	0.500 (0.474, 0.523)
volume loss	156	0.563 (0.523, 0.602)	**0.618 (0.576, 0.661)**	0.547 (0.496, 0.598)	0.577 (0.531, 0.620)
apical pleural thickening	292	0.558 (0.524, 0.589)	**0.573 (0.541, 0.604)**	0.537 (0.502, 0.572)	0.551 (0.518, 0.580)
goiter	54	0.541 (0.473, 0.603)	0.570 (0.505, 0.638)	**0.616 (0.553, 0.673)**	0.539 (0.462, 0.616)
air trapping	354	0.514 (0.484, 0.545)	0.470 (0.442, 0.498)	**0.581 (0.549, 0.611)**	0.564 (0.534, 0.593)
gynecomastia	62	0.512 (0.452, 0.568)	0.494 (0.428, 0.556)	**0.675 (0.608, 0.737)**	0.641 (0.573, 0.709)
azygos lobe	36	**0.475 (0.377, 0.568)**	0.414 (0.327, 0.502)	0.467 (0.386, 0.545)	0.469 (0.381, 0.553)

Table S11: **Zero-shot performance on the Indiana dataset.** AUROC values with 95% confidence intervals. Bold indicates best performance.

Finding	N	CLEAR	CheXzero	BiomedCLIP	OpenAICLIP
pleural effusion	286	0.935 (0.920, 0.948)	**0.937 (0.921, 0.951)**	0.775 (0.748, 0.802)	0.567 (0.530, 0.605)
consolidation	53	**0.926 (0.902, 0.948)**	0.874 (0.818, 0.926)	0.715 (0.645, 0.785)	0.485 (0.408, 0.568)
pulmonary edema	87	0.887 (0.852, 0.915)	**0.920 (0.885, 0.948)**	0.725 (0.676, 0.772)	0.457 (0.392, 0.518)
pneumonia	77	**0.882 (0.845, 0.915)**	0.823 (0.774, 0.877)	0.553 (0.491, 0.613)	0.532 (0.469, 0.594)
pulmonary fibrosis	34	**0.881 (0.817, 0.933)**	0.770 (0.688, 0.849)	0.655 (0.559, 0.738)	0.669 (0.568, 0.765)
pulmonary congestion	135	0.872 (0.844, 0.897)	**0.892 (0.866, 0.917)**	0.664 (0.619, 0.707)	0.525 (0.476, 0.577)
airspace disease	224	**0.845 (0.821, 0.868)**	0.729 (0.695, 0.763)	0.601 (0.567, 0.636)	0.503 (0.462, 0.544)
cardiomegaly	655	0.840 (0.827, 0.853)	**0.902 (0.892, 0.913)**	0.662 (0.642, 0.681)	0.485 (0.460, 0.509)
infiltrate	116	**0.798 (0.759, 0.831)**	0.733 (0.678, 0.786)	0.652 (0.599, 0.707)	0.504 (0.443, 0.562)
mass	32	**0.794 (0.732, 0.851)**	0.697 (0.613, 0.775)	0.657 (0.582, 0.728)	0.499 (0.406, 0.591)
pulmonary atelectasis	616	**0.776 (0.758, 0.794)**	0.756 (0.736, 0.774)	0.509 (0.483, 0.533)	0.512 (0.488, 0.536)
pulmonary emphysema	74	0.750 (0.680, 0.811)	**0.833 (0.790, 0.871)**	0.584 (0.523, 0.648)	0.559 (0.484, 0.628)
pneumothorax	44	0.735 (0.658, 0.806)	**0.776 (0.706, 0.841)**	0.524 (0.435, 0.607)	0.379 (0.300, 0.465)
cicatrix	355	**0.729 (0.704, 0.754)**	0.628 (0.600, 0.655)	0.533 (0.501, 0.564)	0.550 (0.520, 0.580)
opacity	840	0.717 (0.699, 0.734)	**0.752 (0.735, 0.770)**	0.517 (0.497, 0.537)	0.540 (0.521, 0.561)
normal	2695	**0.681 (0.669, 0.693)**	0.632 (0.620, 0.644)	0.530 (0.517, 0.543)	0.483 (0.469, 0.496)
atherosclerosis	233	0.661 (0.630, 0.692)	**0.665 (0.636, 0.696)**	0.546 (0.507, 0.582)	0.585 (0.548, 0.623)
arthritis	65	**0.652 (0.584, 0.718)**	0.352 (0.292, 0.413)	0.490 (0.422, 0.556)	0.500 (0.433, 0.569)
emphysema	116	0.647 (0.598, 0.697)	**0.869 (0.846, 0.891)**	0.482 (0.426, 0.538)	0.534 (0.476, 0.589)
deformity	210	**0.641 (0.606, 0.678)**	0.576 (0.535, 0.614)	0.546 (0.505, 0.585)	0.465 (0.426, 0.506)
kyphosis	58	0.624 (0.559, 0.689)	**0.704 (0.647, 0.761)**	0.504 (0.417, 0.585)	0.355 (0.271, 0.432)
calcified granuloma	510	**0.605 (0.580, 0.629)**	0.478 (0.452, 0.502)	0.495 (0.470, 0.522)	0.512 (0.486, 0.537)
calcinosis	558	**0.586 (0.562, 0.608)**	0.429 (0.406, 0.453)	0.506 (0.480, 0.532)	0.511 (0.485, 0.535)
nodule	211	**0.586 (0.546, 0.625)**	0.518 (0.480, 0.559)	0.510 (0.474, 0.546)	0.561 (0.522, 0.599)
granuloma	103	**0.577 (0.523, 0.626)**	0.411 (0.355, 0.466)	0.521 (0.466, 0.579)	0.547 (0.493, 0.602)
spondylosis	126	0.566 (0.514, 0.619)	0.439 (0.393, 0.485)	0.470 (0.420, 0.520)	**0.577 (0.532, 0.625)**
osteophyte	137	0.545 (0.502, 0.589)	**0.602 (0.556, 0.647)**	0.477 (0.422, 0.526)	0.542 (0.494, 0.591)
granulomatous disease	213	0.518 (0.482, 0.551)	**0.555 (0.518, 0.595)**	0.471 (0.437, 0.505)	0.498 (0.459, 0.535)
scoliosis	177	0.505 (0.458, 0.549)	**0.600 (0.557, 0.642)**	0.493 (0.450, 0.535)	0.526 (0.485, 0.568)
lucency	51	0.466 (0.391, 0.544)	0.332 (0.264, 0.408)	**0.528 (0.457, 0.602)**	0.487 (0.404, 0.578)

Table S12: **Zero-shot calibration metrics for CLIP-based models.** ECE = Expected Calibration Error, Brier = Brier Score. Zero-shot evaluation is deterministic (no training), so no std is reported. Note: predictions cluster around 0.5 due to softmax over normalized similarities.

Dataset	CLEAR		CheXzero	
	ECE	Brier Score	ECE	Brier Score
CheXpert	**0.299**	**0.248**	0.300	0.249
VinDr-CXR	**0.429**	**0.242**	0.433	0.247
PadChest	0.487	0.248	**0.486**	**0.247**
Indiana	**0.478**	**0.245**	0.481	0.248

Table S13: **Supervised performance on the CheXpert dataset.** AUROC values with 95% confidence intervals. Bold indicates best performance.

Finding	CLEAR	CheXzero	BiomedCLIP	OpenAICLIP
Atelectasis	**0.876 (0.876, 0.876)**	0.844 (0.844, 0.845)	0.735 (0.733, 0.737)	0.726 (0.725, 0.727)
Cardiomegaly	0.867 (0.867, 0.868)	**0.881 (0.881, 0.882)**	0.724 (0.724, 0.725)	0.689 (0.688, 0.691)
Consolidation	**0.905 (0.905, 0.905)**	0.897 (0.896, 0.897)	0.797 (0.795, 0.799)	0.739 (0.738, 0.739)
Edema	0.884 (0.884, 0.884)	**0.906 (0.905, 0.906)**	0.804 (0.803, 0.804)	0.771 (0.770, 0.773)
Enlarged Cardiomediastinum	**0.727 (0.727, 0.728)**	0.727 (0.726, 0.729)	0.694 (0.691, 0.699)	0.705 (0.704, 0.706)
Fracture	**0.720 (0.719, 0.720)**	0.673 (0.661, 0.682)	0.411 (0.400, 0.421)	0.289 (0.283, 0.298)
Lung Lesion	**0.886 (0.885, 0.886)**	0.847 (0.844, 0.850)	0.647 (0.640, 0.651)	0.717 (0.711, 0.724)
Lung Opacity	**0.933 (0.933, 0.933)**	0.895 (0.895, 0.896)	0.777 (0.777, 0.778)	0.794 (0.792, 0.796)
Pleural Effusion	**0.945 (0.945, 0.945)**	0.935 (0.935, 0.936)	0.809 (0.807, 0.811)	0.762 (0.761, 0.765)
Pleural Other	0.908 (0.908, 0.908)	**0.947 (0.945, 0.949)**	0.915 (0.913, 0.917)	0.652 (0.633, 0.683)
Pneumonia	**0.839 (0.838, 0.841)**	0.752 (0.750, 0.756)	0.606 (0.603, 0.614)	0.705 (0.700, 0.709)
Pneumothorax	**0.894 (0.894, 0.894)**	0.679 (0.678, 0.681)	0.624 (0.621, 0.627)	0.487 (0.485, 0.489)
Support Devices	**0.928 (0.928, 0.928)**	0.852 (0.851, 0.852)	0.796 (0.796, 0.796)	0.753 (0.751, 0.756)

Table S14: **Calibration metrics for linear probing on CheXpert test set** (mean $\pm$ std over 20 seeds).

Model	ECE	Brier Score
CLEAR	$\mathbf{0.072 \pm 0.001}$	0.119 ± 0.000
CheXzero	0.076 ± 0.000	$\mathbf{0.118 \pm 0.000}$
BiomedCLIP	$\mathbf{0.072 \pm 0.001}$	0.135 ± 0.000
OpenAI-CLIP	0.086 ± 0.000	0.142 ± 0.000

Table S15: **Additional selected concepts for building the CBM.**

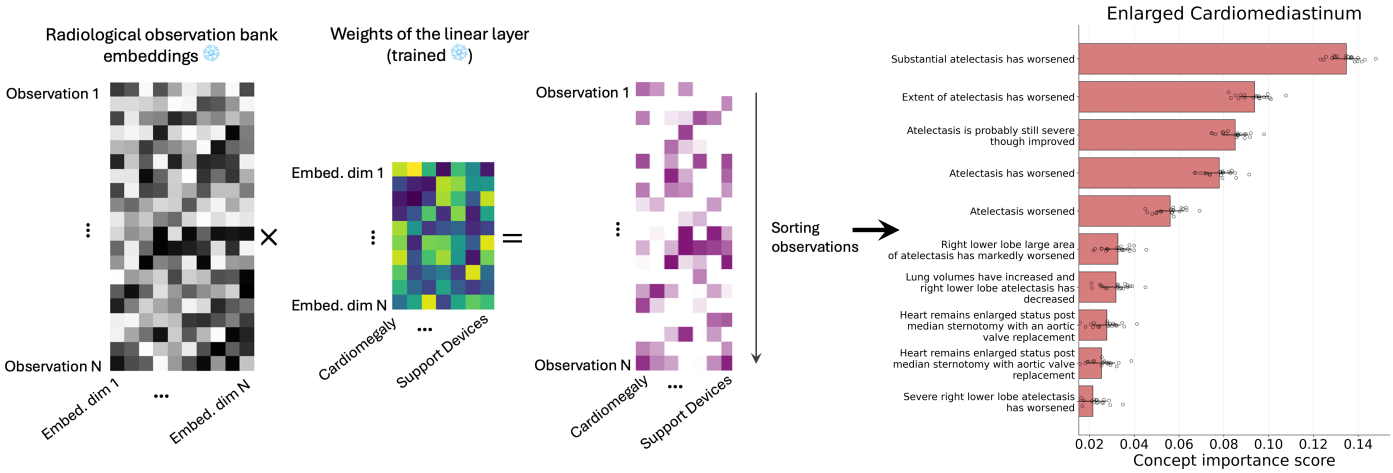
• Large bibasilar atelectasis	• Heart size continues to be severely enlarged
• Heart size is chronically and severely enlarged	• Heart size remains moderately to severely enlarged
• Heart is moderately to severely enlarged though improved	• Heart chronically enlarged
• Two, now confluent left lower lobe and a consolidative right upper lobe opacity are all worsening	• Worsening consolidation in the left lower lobe now involving the left upper lobe
• Consolidation in the right midlung and left lower lung has worsened	• Severe bilateral perihilar pulmonary consolidation has worsened on the left, improved on the right
• Left lower lobe consolidation more severe than right	• Large consolidation in the left lower lobe has worsened
• Interval worsening of bilateral interstitial pulmonary edema and vascular congestion	• Pulmonary vascular congestion has worsened to the point of early pulmonary edema
• Mild-to-moderate pulmonary edema bilaterally has worsened	• Severe pulmonary edema worsened
• Widespread pulmonary edema has worsened	• Bilateral atelectasis has slightly improved on the right and substantially worsened on the left
• Right middle and lower lobe atelectasis worsened	• Left lower lung atelectasis has worsened
• Right lower lobe atelectasis has worsened substantially, perhaps lobar collapse	• Bilateral lower lobe atelectasis has improved substantially
• Old lower left lateral rib fractures	• Chronic right lateral rib fracture with adjacent atelectasis
• Compression deformity of the lower thoracic spine at least a year old	• Severe widespread opacities have worsened in the lower lungs and left upper lobe but improved in right upper lobe
• Severe worsening of bilateral upper lobe predominant airspace opacities, right greater than left	• Bilateral extensive asymmetric right greater than left lung opacities have worsened
• Bilateral opacities on the right and consolidation on the left have increased in extent and severity	• Bilateral lower lung consolidations have worsened
• Small bilateral pleural effusions have worsened	• Bilateral pleural effusion has worsened
• Bilateral pleural effusions improved	• Most of the left upper lobe and the right lung apex chronically scarred
• Extensive fibrotic changes in the upper lobes bilaterally	• Severe right lung scarring and concurrent emphysema
• Chronic interstitial changes in the bilateral lungs	• Combination of emphysematous changes and scarring in bilateral lungs
• Worsening multifocal pulmonary opacities, particularly within the right upper lobe and left lower lobe	• Right lower lobe and left upper lobe opacities have worsened
• Right lower lobe consolidation which has worsened	• Right apical pneumothorax reduced with chest tube in place
• No right apical pneumothorax status post removal of right chest tube	• Left apical pneumothorax resolved
• Improving small right apical pneumothorax	• Transvenous right atrial biventricular pacer leads
• Single-lead pacemaker placed with tip in left ventricle	

Table S16: **Calibration metrics for concept bottleneck models** (mean ± std over 20 seeds). ECE = Expected Calibration Error, Brier = Brier Score.

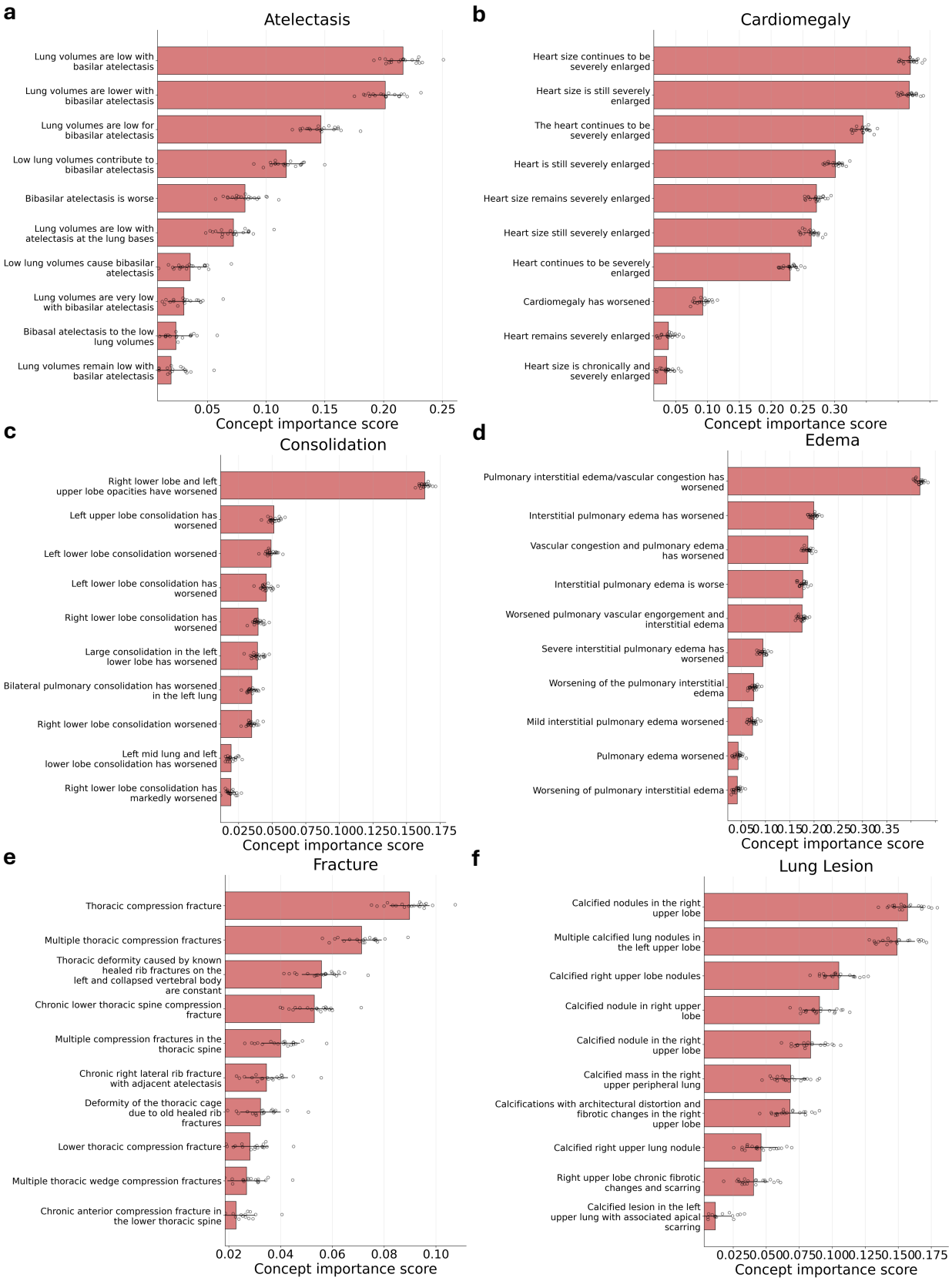
Dataset	**Standard CBM**		**CLEAR CBM**	
	ECE	**Brier Score**	**ECE**	**Brier Score**
CheXpert	0.043 ± 0.002	$\mathbf{0.099 \pm 0.001}$	$\mathbf{0.040 \pm 0.001}$	0.100 ± 0.000
VinDr-CXR	0.020 ± 0.000	$\mathbf{0.026 \pm 0.000}$	$\mathbf{0.017 \pm 0.001}$	0.029 ± 0.000
PadChest	0.052 ± 0.002	0.030 ± 0.000	$\mathbf{0.025 \pm 0.001}$	$\mathbf{0.025 \pm 0.000}$
Indiana	0.025 ± 0.002	0.032 ± 0.000	$\mathbf{0.007 \pm 0.000}$	0.032 ± 0.000


a
Mediastinal shift
Cardiomegaly
Pleural effusion
b
Aortic aneurysm
Pulmonary mass
Consolidation
c
Tracheostomy tube
Pacemaker
CVC

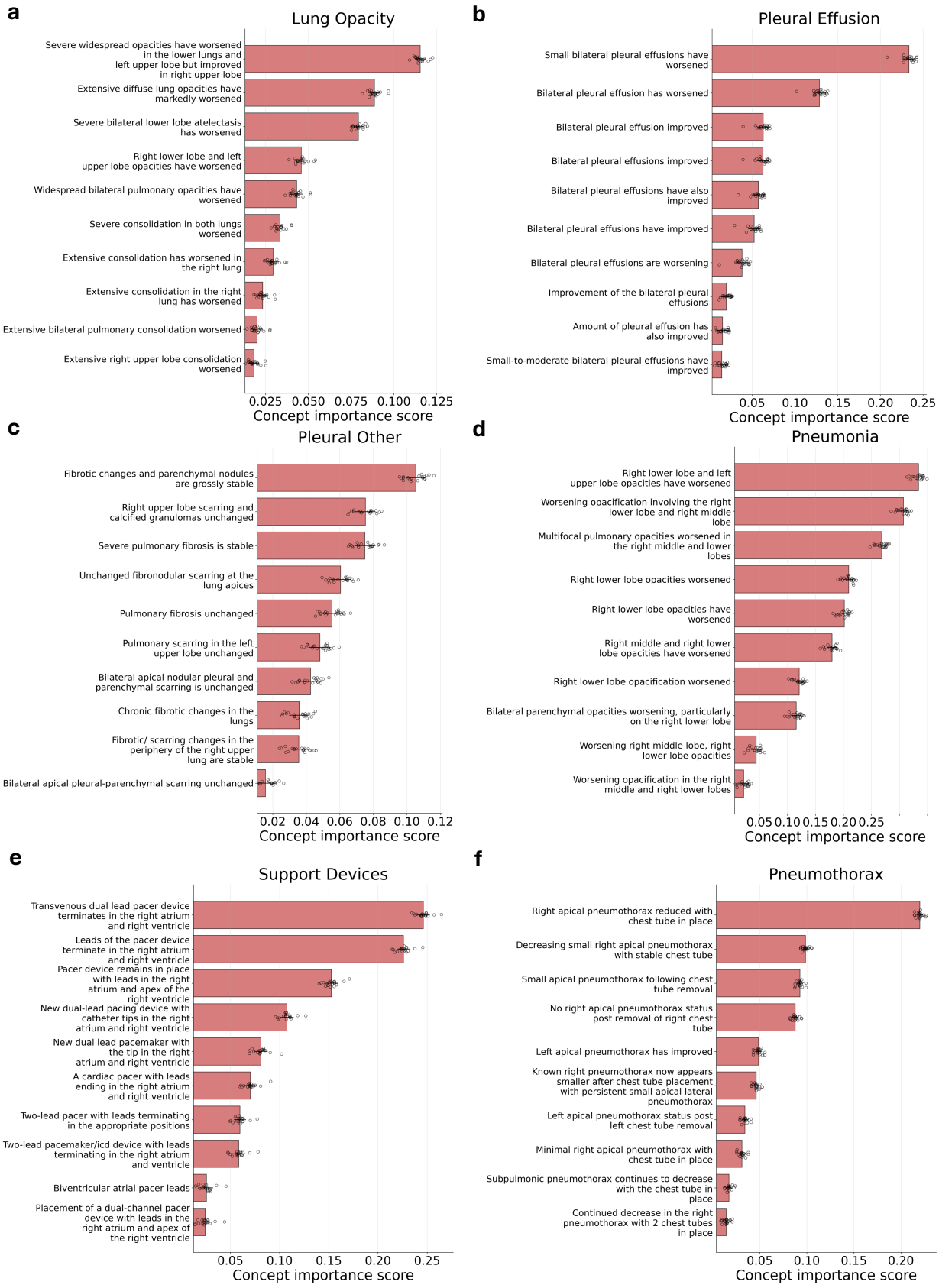
Supplementary Fig. 1: **CXRs with high concept presence scores calculated using CLEAR.** The concept presence score represents the degree to which a radiological observation is present in an image. Each row displays the six highest-scoring CXRs retrieved from the MIMIC-CXR training pool (377,110 images from 65,379 patients) for the indicated concept; no statistics are derived. **(a)** Anatomical findings: mediastinal shift, cardiomegaly, and pleural effusion. **(b)** Pathological findings: aortic aneurysm, pulmonary mass, and consolidation. **(c)** Medical devices: tracheostomy tube, pacemaker, and central venous catheter (CVC). Images demonstrate CLEAR's ability to retrieve clinically relevant examples with visible radiological features corresponding to each concept.



Supplementary Fig. 2: **Concept importance analysis pipeline for supervised linear probing.** The left panel shows the computational workflow: radiological observation bank embeddings are multiplied by trained linear layer weights to produce observation-specific importance scores for each diagnostic label. The resulting importance matrix is sorted to identify the most influential concepts. The right panel demonstrates the output for classification of EC.



Supplementary Fig. 3: **Concept importance output for atelectasis, cardiomegaly, consolidation, edema, fracture, and lung lesion classifications.** Bars show the mean concept-importance score across $n = 20$ independent linear probes trained on MIMIC-CXR with different random seeds; error bars denote ± 1 standard deviation.



Supplementary Fig. 4: **Concept importance output for lung opacity, pleural effusion, pleural other, pneumonia, support devices, and pneumothorax classifications.** Bars show the mean concept-importance score across $n = 20$ independent linear probes trained on MIMIC-CXR with different random seeds; error bars denote ± 1 standard deviation.