

Cethraian-X Supplementary Methods and Results

Mohammad Badhan

Independent Researcher, Dhaka, Bangladesh

Correspondence: muhammedbadhon@gmail.com

ORCID: <https://orcid.org/0009-0009-7411-9569>

This supplement accompanies Cethraian-X: A Leakage-Clean, Multi-Seed Benchmark of Chest X-Ray Classification Under Weak Labels. It reports descriptive secondary details from the same locked evaluation. It does not introduce new training, inference, or test tuning.

S1. Reproducibility snapshot

The primary leakage-clean manifest SHA-256 was

353107b77f0ff25742d6647bfe7b3b5904af094cb19e6eea778f815c05e0eabf. The retained cohort contained 112,060 images from 30,799 patients: 77,984 training, 8,480 validation, and 25,596 test images. All three finalized seeds used the same manifest, architecture, preprocessing, objective, and evaluation workflow; only the random seed differed. Each selected checkpoint received one intended locked test inference. The selected epochs were 7, 5, and 7. The initial seed-1337 transition failed before epoch 1, before checkpoint creation, and before test access; its recovery reused the same seed, configuration, and manifest.

The software environment was Python 3.11.9; PyTorch 2.6.0+cu126; CUDA 12.6; Albumentations 2.0.8; OpenCV 5.0.0.93; NumPy 2.4.6; SciPy 1.17.1; pandas 3.0.3; scikit-learn 1.9.0; Matplotlib 3.11.0; PyYAML 6.0.3; psutil 7.2.2; and an NVIDIA GeForce RTX 4080 SUPER. Random generators and cuDNN deterministic settings were controlled, but global deterministic-algorithm enforcement was not enabled.

S2. Per-seed performance and calibration

Seed	Selected epoch	Validation ROC-AUC	Test ROC-AUC	Test AP	Raw ECE	Scaled ECE
42	7	0.83396	0.80951	0.28042	0.24222	0.22687
1337	5	0.83205	0.81005	0.28187	0.24934	0.22561
2026	7	0.83137	0.80973	0.27984	0.23119	0.21634

Patient-clustered 95% bootstrap ROC-AUC intervals (1,000 resamples) were 0.80302–0.81495 for seed 42, 0.80303–0.81617 for seed 1337, and 0.80368–0.81522 for seed 2026. These are per-seed intervals, not pooled cross-seed intervals.

S3. Complete pathology-level discrimination

Pathology	Positives	Prevalence	ROC-AUC, mean $\pm$ SD	AP, mean $\pm$ SD
Atelectasis	3,279	0.12811	0.76756 $\pm$ 0.00249	0.33787 $\pm$ 0.00477
Cardiomegaly	1,069	0.04176	0.88291 $\pm$ 0.00359	0.32300 $\pm$ 0.00201
Consolidation	1,815	0.07091	0.74324 $\pm$ 0.00108	0.16040 $\pm$ 0.00573
Edema	925	0.03614	0.84498 $\pm$ 0.00529	0.16688 $\pm$ 0.00610
Effusion	4,658	0.18198	0.82420 $\pm$ 0.00326	0.50434 $\pm$ 0.00652
Emphysema	1,093	0.04270	0.91289 $\pm$ 0.00289	0.42498 $\pm$ 0.00800
Fibrosis	435	0.01699	0.82541 $\pm$ 0.00315	0.10202 $\pm$ 0.00532
Hernia	86	0.00336	0.91302 $\pm$ 0.00926	0.37202 $\pm$ 0.01584
Infiltration	6,112	0.23879	0.70344 $\pm$ 0.00295	0.39722 $\pm$ 0.00253
Mass	1,748	0.06829	0.81932 $\pm$ 0.00551	0.31211 $\pm$ 0.00289
Nodule	1,623	0.06341	0.74764 $\pm$ 0.00519	0.21471 $\pm$ 0.01085
Pleural Thickening	1,143	0.04466	0.77835 $\pm$ 0.00164	0.14905 $\pm$ 0.00308
Pneumonia	555	0.02168	0.71211 $\pm$ 0.00828	0.05258 $\pm$ 0.00258
Pneumothorax	2,665	0.10412	0.86157 $\pm$ 0.00841	0.41276 $\pm$ 0.00894

S4. Positive weights and temperature parameters

Positive weights were calculated from the retained training split as (N - positives) / positives, capped at 20, and applied in binary cross-entropy with logits.

Pathology	Applied positive weight
Atelectasis	9.47046
Cardiomegaly	20
Consolidation	20
Edema	20
Effusion	8.95710
Emphysema	20
Fibrosis	20
Hernia	20
Infiltration	5.28396
Mass	20
Nodule	17.50154
Pleural Thickening	20
Pneumonia	20
Pneumothorax	20

One temperature per pathology and seed was fitted on validation logits only, with bounds 0.05–20. Temperatures are reported for provenance; they are not evidence of reliable calibrated probabilities.

Pathology	Seed 42	Seed 1337	Seed 2026
Atelectasis	1.51269	1.59490	0.86280
Cardiomegaly	0.84849	0.73148	0.89488
Consolidation	0.92052	0.91142	0.84373
Edema	0.83687	1.04401	0.84190
Effusion	1.25867	1.22388	1.11561
Emphysema	0.85002	0.83419	0.76333
Fibrosis	0.67626	0.47888	0.54627
Hernia	0.61637	0.78394	0.65780
Infiltration	0.99995	1.43088	0.97065
Mass	1.13650	0.77509	0.90660
Nodule	0.79088	0.74060	1.25029
Pleural Thickening	0.70655	0.55305	0.71247
Pneumonia	0.40713	0.43121	0.42189
Pneumothorax	1.10992	0.76165	0.93309

S5. Validation-locked thresholds and raw test operating characteristics

Thresholds were selected independently for each label and seed using Youden's J statistic on validation predictions. The raw test sensitivity and specificity below are descriptive consequences of those locked thresholds, not clinically validated operating points.

Seed 42

Pathology	Validation-locked threshold	Raw test sensitivity	Raw test specificity
Atelectasis	0.52627	0.85300	0.50943
Cardiomegaly	0.17003	0.89616	0.66237
Consolidation	0.32032	0.89917	0.44292
Edema	0.10352	0.93189	0.57103
Effusion	0.43099	0.92293	0.50143
Emphysema	0.25432	0.86276	0.80088
Fibrosis	0.24268	0.74023	0.75359
Hernia	0.01592	0.84884	0.79667
Infiltration	0.46331	0.86273	0.35506
Mass	0.43300	0.83638	0.61158
Nodule	0.42608	0.69439	0.66216
Pleural Thickening	0.21878	0.87752	0.48505
Pneumonia	0.13112	0.80901	0.49084
Pneumothorax	0.37242	0.91295	0.64720

Seed 1337

Pathology	Validation-locked threshold	Raw test sensitivity	Raw test specificity
Atelectasis	0.60347	0.71455	0.67536
Cardiomegaly	0.21510	0.81572	0.78497
Consolidation	0.39790	0.92727	0.36996
Edema	0.34490	0.92108	0.60018
Effusion	0.39665	0.92722	0.48324
Emphysema	0.37968	0.84172	0.84712
Fibrosis	0.28549	0.62299	0.85211
Hernia	0.08153	0.69767	0.97589
Infiltration	0.50271	0.87598	0.34680
Mass	0.42503	0.72998	0.73553
Nodule	0.40295	0.77264	0.58040
Pleural Thickening	0.27254	0.81715	0.57895
Pneumonia	0.16053	0.85405	0.43073
Pneumothorax	0.34314	0.82214	0.74336

Seed 2026

Pathology	Validation-locked threshold	Raw test sensitivity	Raw test specificity
Atelectasis	0.35987	0.83227	0.54084
Cardiomegaly	0.18600	0.87278	0.72528
Consolidation	0.25791	0.95537	0.31815
Edema	0.10748	0.93946	0.51644
Effusion	0.35599	0.93602	0.47220
Emphysema	0.14208	0.86734	0.80076
Fibrosis	0.13988	0.83908	0.63841
Hernia	0.20224	0.59302	0.99577
Infiltration	0.49587	0.83606	0.42096
Mass	0.47982	0.76201	0.71796
Nodule	0.55089	0.71288	0.63855
Pleural Thickening	0.18875	0.86964	0.50337
Pneumonia	0.10989	0.81622	0.45673
Pneumothorax	0.33324	0.87280	0.66944

S6. Acquisition-view associations

For each label, the table gives PA-minus-AP differences in test performance across the three seeds. Values are descriptive and unadjusted; they do not isolate view position from clinical setting, portable acquisition, severity, devices, positioning, or workflow.

Pathology	PA - AP ROC-AUC, mean $\pm$ SD	PA - AP AP, mean $\pm$ SD
Atelectasis	0.03161 $\pm$ 0.00801	0.02565 $\pm$ 0.01218
Cardiomegaly	0.09255 $\pm$ 0.00868	0.15316 $\pm$ 0.00429
Consolidation	0.06037 $\pm$ 0.00338	-0.06656 $\pm$ 0.01053
Edema	0.07775 $\pm$ 0.01389	-0.00435 $\pm$ 0.03463
Effusion	0.06159 $\pm$ 0.00376	0.05301 $\pm$ 0.00612
Emphysema	-0.02702 $\pm$ 0.00542	0.00291 $\pm$ 0.01637
Fibrosis	-0.01543 $\pm$ 0.01708	0.06021 $\pm$ 0.00624
Hernia	0.07551 $\pm$ 0.03551	0.33343 $\pm$ 0.08730
Infiltration	-0.04829 $\pm$ 0.01024	-0.17346 $\pm$ 0.01029
Mass	0.06074 $\pm$ 0.00492	0.16734 $\pm$ 0.01710
Nodule	0.01501 $\pm$ 0.01074	0.10028 $\pm$ 0.01977
Pleural Thickening	-0.00890 $\pm$ 0.00346	0.07802 $\pm$ 0.00810
Pneumonia	0.03789 $\pm$ 0.01308	-0.01280 $\pm$ 0.00150
Pneumothorax	0.01297 $\pm$ 0.00992	0.12370 $\pm$ 0.00761

S7. Sensitivity comparison and saliency scope

The original patient-disjoint analysis is retained solely as a pre-remediation sensitivity comparator. Its aggregate macro ROC-AUC was 0.81030 and average precision 0.28275, versus 0.80976 and 0.28071 for the primary

leakage-clean analysis. The changes (-0.00054 and -0.00204) have no paired statistical-significance analysis. Nodule had the largest ROC-AUC deterioration (-0.00906); Cardiomegaly had the largest average-precision deterioration (-0.01069); and Hernia had the largest ROC-AUC improvement (+0.00866) but only 86 test positives.

No leakage-clean Grad-CAM++ inference was performed. Earlier Grad-CAM++ software checks concerned the original patient-disjoint analysis and were limited by one seed, a restricted bounding-box subset, a pointing-game metric rather than segmentation, no clinician review, and high cross-class heatmap correlation in a small audit. They are not localization or clinical-reasoning results for the primary analysis.

S8. Provenance and availability

The released materials include code, the leakage-clean manifest, derived analysis artifacts, documentation, and reproducibility instructions at <https://github.com/muhammedbadhon-oss/Cethraian-X>. NIH ChestX-ray14 images are not redistributed and remain subject to their original access terms. Model checkpoints and prediction arrays are not included in the public repository. The immutable experiment records retain the frozen manifest hash and seed-level provenance. An archival copy of this release is available at <https://doi.org/10.5281/zenodo.21583726>.