

Localization-Aware Chest X-ray Classification via Segmentation and Gradient-based Attention

Anonymous Authors

I. EVALUATION METRICS

In the classification tasks, instances where the lesion type for object detection is "No Finding" are designated as the negative class. For evaluating the classification results, we utilize the performance metrics provided by scikit-learn, including Accuracy, Precision, Recall, and F1-Score. Since the datasets have been balanced through cleaning to ensure representation across classes, both weighted and macro variants of the metrics are reported. Given the unique characteristics of medical data, we additionally incorporate sensitivity (recall), specificity, positive predictive value (PPV), false discovery rate (FDR), false omission rate (FOR), Youden's index (YI) and Discriminant power (DP) as evaluation metrics, which are formally defined as follows:

$$Sensitivity = \frac{TP}{TP + FN},$$

$$Specificity = \frac{TN}{FP + TN},$$

$$PPV = \frac{TP}{TP + FP},$$

$$FDR = \frac{FP}{FP + TP},$$

$$FOR = \frac{FN}{FN + TN}.$$

$$YI = Sensitivity + Specificity - 1$$

$$DP = \frac{\sqrt{3}}{\pi} \cdot \left(\log \left(\frac{sensitivity}{1 - sensitivity} \right) + \log \left(\frac{specificity}{1 - specificity} \right) \right)$$

TABLE I
RESULTS OF COMPARISON FOR OTTAWA DATASET
(CROSS-ENTROPY/PROPOSED LOSS) OVER THREE MODELS, THE
EXPERIMENTAL SETTING USED FOR ATTENTION LOSS IS THE
BEST-PERFORMING (F1 SCORE) CONFIGURATION FROM ALL SETTINGS OF
EACH MODEL, DP REPRESENTS DISCRIMINANT POWER.

Ottawa: No Finding vs Effusion			
Metric	PVT	VGG16	ResNet50
Sensitivity	0.654/0.705	0.782/0.692	0.744/0.692
Specificity	0.603/0.744	0.564/0.718	0.564/0.718
Youden's index	0.257/0.449	0.410/0.384	0.308/0.410
DP	0.253/0.464	0.367/0.418	0.317/0.418
Ottawa: No Finding vs Pneumothorax			
Metric	PVT	VGG16	ResNet50
Sensitivity	0.517/0.638	0.784/0.784	0.707/0.672
Specificity	0.612/0.647	0.483/0.586	0.397/0.612
Youden's index	0.129/0.285	0.267/0.370	0.104/0.284
DP	0.125/0.281	0.292/0.392	0.111/0.281
Ottawa: No Finding vs Subcutaneous emphysema			
Metric	PVT	VGG16	ResNet50
Sensitivity	0.697/0.780	0.727/0.735	0.621/0.689
Specificity	0.811/0.795	0.773/0.780	0.712/0.712
Youden's index	0.508/0.575	0.500/0.515	0.333/0.401
DP	0.548/0.628	0.528/0.547	0.335/0.407
Ottawa: No finding, Effusion, Pneumothorax			
Metric	PVT	VGG16	ResNet50
Sensitivity	0.457/0.551	0.474/0.543	0.496/0.513
Specificity	0.729/0.776	0.737/0.771	0.748/0.756
Youden's index	0.186/0.327	0.212/0.314	0.244/0.269
DP	0.212/0.368	0.223/0.405	0.263/0.284

TABLE II
RESULTS OF COMPARISON FOR NIH DATASET
(CROSS-ENTROPY/PROPOSED LOSS) OVER THREE MODELS, THE
EXPERIMENTAL SETTING USED FOR ATTENTION LOSS IS THE
BEST-PERFORMING (F1 SCORE) CONFIGURATION FROM ALL SETTINGS OF
EACH MODEL, DP REPRESENTS DISCRIMINANT POWER.

NIH: No Finding vs Effusion			
Metric	PVT	VGG16	ResNet50
Sensitivity	0.752/0.810	0.861/0.823	0.746/0.828
Specificity	0.882/0.899	0.807/0.853	0.872/0.824
Youden's index	0.634/0.709	0.668/0.676	0.618/0.652
DP	0.747/0.871	0.779/ 0.750	0.789/0.746

NIH: No Finding vs Pneumothorax			
Metric	PVT	VGG16	ResNet50
Sensitivity	0.745/0.815	0.795/0.795	0.715/0.806
Specificity	0.797/0.793	0.736/0.788	0.738/0.770
Youden's index	0.542/0.608	0.531/0.583	0.453/0.576
DP	0.584/0.677	0.570/0.639	0.468/0.630

NIH: No finding, Effusion, Pneumothorax			
Metric	PVT	VGG16	ResNet50
Sensitivity	0.651/0.689	0.674/0.683	0.631/0.632
Specificity	0.826/0.845	0.837/0.842	0.815/0.816
Youden's index	0.477/0.534	0.511/0.525	0.446/0.449
DP	0.525/0.603	0.569/0.590	0.498/0.492

TABLE III

RESULTS WITH DIFFERENT λ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). $\lambda = 1$: CROSS-ENTROPY LOSS, $\lambda = *$: ADAPTIVE CUSTOM LOSS, BOLD: BEST RESULTS WITHIN THE SAME MODEL, UNDERLINE: BEST RESULTS OVER ALL MODELS, ITALICS: BEST RESULTS OVER ALL MODELS PRIOR TO “ATTENTION LOSS” BEING USED.

Effusion vs No Finding						
Model	Metric	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = *$
PVT	Acc / Rec	0.628	0.660	<u>0.692</u>	0.654	0.628
	Prec	0.629	0.662	<u>0.693</u>	0.659	0.630
	F1	0.628	0.659	<u>0.692</u>	0.651	0.627
	AUC	<i>0.743</i>	0.723	<u>0.752</u>	0.678	0.683
	PPV	0.622	0.681	<u>0.703</u>	0.630	0.614
	Sensitivity	0.654	0.603	0.718	<u>0.744</u>	0.692
	Specificity	<i>0.603</i>	<u>0.718</u>	0.718	0.564	0.564
	FDR	0.378	0.319	<u>0.297</u>	0.370	0.386
	FOR	0.365	0.356	0.317	<u>0.313</u>	0.353
VGG16	Acc / Rec	<i>0.673</i>	0.667	0.673	0.654	<u>0.692</u>
	Prec	<i>0.682</i>	0.667	0.674	0.662	<u>0.692</u>
	F1	<i>0.669</i>	0.666	0.672	0.649	<u>0.692</u>
	AUC	0.726	0.735	0.745	0.745	<u>0.760</u>
	PPV	<i>0.642</i>	0.676	0.659	0.625	<u>0.697</u>
	Sensitivity	<u>0.782</u>	0.641	0.718	0.769	0.679
	Specificity	0.564	0.692	0.628	0.538	<u>0.705</u>
	FDR	<i>0.358</i>	0.324	0.341	0.375	<u>0.303</u>
	FOR	<u>0.279</u>	0.341	0.310	0.300	0.313
ResNet50	Acc / Rec	0.654	0.635	0.628	<u>0.673</u>	0.667
	Prec	0.659	0.638	0.630	<u>0.673</u>	0.669
	F1	0.651	0.632	0.627	<u>0.673</u>	0.665
	AUC	0.736	0.672	<u>0.739</u>	0.737	0.711
	PPV	0.630	0.615	0.616	<u>0.667</u>	0.648
	Sensitivity	<u>0.744</u>	0.718	0.679	0.692	0.731
	Specificity	0.564	0.551	0.577	<u>0.654</u>	0.603
	FDR	0.370	0.385	0.384	<u>0.333</u>	0.352
	FOR	0.313	0.338	0.357	0.320	<u>0.309</u>

REFERENCES

TABLE IV

RESULTS WITH DIFFERENT λ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). $\lambda = 1$: CROSS-ENTROPY LOSS, $\lambda = *$: ADAPTIVE CUSTOM LOSS, BOLD: BEST RESULTS WITHIN THE SAME MODEL, UNDERLINE: BEST RESULTS OVER ALL MODELS, ITALICS: BEST RESULTS OVER ALL MODELS PRIOR TO “ATTENTION LOSS” BEING USED.

Pneumothorax vs No Finding						
Model	Metric	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = *$
PVT	Acc / Rec	0.565	0.578	0.595	0.595	<u>0.642</u>
	Prec	0.565	0.578	0.595	0.596	<u>0.642</u>
	F1	0.564	0.576	0.594	0.594	<u>0.642</u>
	AUC	0.626	0.632	<u>0.654</u>	0.623	0.653
	PPV	0.571	0.587	0.602	0.587	<u>0.643</u>
	Sensitivity	0.517	0.526	0.560	<u>0.638</u>	<u>0.638</u>
	Specificity	<i>0.612</i>	0.629	0.629	0.552	<u>0.647</u>
	FDR	0.429	0.413	0.398	0.413	<u>0.357</u>
	FOR	0.441	0.430	0.411	0.396	<u>0.359</u>
VGG16	Acc / Rec	<i>0.634</i>	0.642	<u>0.681</u>	0.647	0.629
	Prec	<i>0.647</i>	0.643	<u>0.681</u>	0.648	0.629
	F1	<i>0.625</i>	0.642	<u>0.681</u>	0.646	0.629
	AUC	<i>0.707</i>	0.681	0.714	<u>0.740</u>	0.680
	PPV	<i>0.603</i>	0.636	<u>0.684</u>	0.663	0.627
	Sensitivity	<u>0.784</u>	0.664	0.672	0.595	0.638
	Specificity	0.483	0.621	0.690	<u>0.698</u>	0.621
	FDR	<i>0.397</i>	0.364	<u>0.316</u>	0.337	0.373
	FOR	<u>0.309</u>	0.351	0.322	0.367	0.368
ResNet50	Acc / Rec	0.552	0.547	0.504	<u>0.608</u>	0.556
	Prec	0.557	0.549	0.505	<u>0.609</u>	0.558
	F1	0.541	0.544	0.496	<u>0.607</u>	0.553
	AUC	0.591	0.569	0.504	<u>0.627</u>	0.583
	PPV	0.539	0.540	0.506	<u>0.619</u>	0.567
	Sensitivity	<u>0.707</u>	0.638	0.379	0.560	0.474
	Specificity	0.397	0.457	0.629	<u>0.655</u>	0.638
	FDR	0.461	0.460	0.494	<u>0.381</u>	0.433
	FOR	0.425	0.442	0.497	<u>0.402</u>	0.452

TABLE V

RESULTS WITH DIFFERENT λ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). $\lambda = 1$: CROSS-ENTROPY LOSS, $\lambda = *$: ADAPTIVE CUSTOM LOSS, BOLD: BEST RESULTS WITHIN THE SAME MODEL, UNDERLINE: BEST RESULTS OVER ALL MODELS, ITALICS: BEST RESULTS OVER ALL MODELS PRIOR TO “ATTENTION LOSS” BEING USED.

Subcutaneous emphysema vs No Finding						
Model	Metric	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = *$
PVT	Acc / Rec	<i>0.754</i>	0.735	0.727	<u>0.758</u>	0.739
	Prec	<i>0.757</i>	0.740	0.727	<u>0.759</u>	0.740
	F1	<i>0.753</i>	0.733	0.727	<u>0.757</u>	0.738
	AUC	<i>0.828</i>	0.810	0.811	<u>0.831</u>	0.799
	PPV	<u>0.786</u>	0.777	0.721	0.779	0.756
	Sensitivity	0.697	0.659	<u>0.742</u>	0.720	0.705
	Specificity	<u>0.811</u>	<u>0.811</u>	0.712	0.795	0.773
	FDR	<u>0.214</u>	0.223	0.279	0.221	0.244
	FOR	0.272	0.296	0.266	<u>0.261</u>	0.277
VGG16	Acc / Rec	0.750	0.742	0.742	0.731	<u>0.758</u>
	Prec	0.751	0.742	0.751	0.731	<u>0.776</u>
	F1	0.750	0.742	0.740	0.731	<u>0.753</u>
	AUC	0.790	0.811	<u>0.829</u>	0.813	0.811
	PPV	0.762	0.739	0.796	0.729	<u>0.847</u>
	Sensitivity	<i>0.727</i>	<u>0.750</u>	0.652	0.735	0.629
	Specificity	0.773	0.735	0.833	0.727	<u>0.886</u>
	FDR	0.238	0.261	0.204	0.271	<u>0.153</u>
	FOR	<i>0.261</i>	<u>0.254</u>	0.295	0.267	0.295
ResNet50	Acc / Rec	0.667	<u>0.678</u>	0.621	0.652	0.652
	Prec	0.668	<u>0.683</u>	0.621	0.657	0.652
	F1	0.666	<u>0.676</u>	0.621	0.649	0.651
	AUC	0.707	0.712	0.667	<u>0.715</u>	0.696
	PPV	0.683	<u>0.712</u>	0.619	0.685	0.664
	Sensitivity	0.621	0.598	<u>0.629</u>	0.561	0.614
	Specificity	0.712	<u>0.758</u>	0.614	0.742	0.689
	FDR	0.317	<u>0.288</u>	0.381	0.315	0.336
	FOR	0.347	<u>0.346</u>	0.377	0.372	0.359

TABLE VI

RESULTS WITH DIFFERENT λ VALUES (MULTI-CLASS CLASSIFICATION - OTTAWA DATASET). $\lambda = 1$: CROSS-ENTROPY LOSS, $\lambda = *$: ADAPTIVE CUSTOM LOSS, BOLD: BEST RESULTS WITHIN THE SAME MODEL, UNDERLINE: BEST RESULTS OVER ALL MODELS, ITALICS: BEST RESULTS OVER ALL MODELS PRIOR TO “ATTENTION LOSS” BEING USED.

Effusion, Pneumothorax and No Finding						
Model	Metric	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = *$
PVT	Acc / Rec	0.457	0.491	0.453	0.474	0.496
	Prec	0.465	0.488	0.462	0.472	0.491
	F1	0.443	0.488	0.453	0.470	0.488
	AUC	0.643	0.663	0.644	0.650	0.677
VGG16	Acc / Rec	0.474	0.512	0.512	0.496	0.474
	Prec	0.466	0.518	0.516	0.486	0.474
	F1	0.459	0.517	0.507	0.482	0.445
	AUC	<i>0.697</i>	0.709	0.722	0.688	0.667
ResNet50	Acc / Rec	0.496	0.419	0.466	0.487	0.453
	Prec	0.493	0.407	0.448	0.486	0.437
	F1	0.488	0.408	0.450	0.484	0.441
	AUC	0.683	0.610	0.675	0.679	0.674

TABLE VII

RESULTS WITH DIFFERENT λ VALUES (BINARY CLASSIFICATION - NIH DATASET). $\lambda = 1$: CROSS-ENTROPY LOSS, $\lambda = *$: ADAPTIVE CUSTOM LOSS, BOLD: BEST RESULTS WITHIN THE SAME MODEL, UNDERLINE: BEST RESULTS OVER ALL MODELS, ITALICS: BEST RESULTS OVER ALL MODELS PRIOR TO “ATTENTION LOSS” BEING USED.

Effusion vs No Finding						
Model	Metric	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = *$
PVT	Acc / Rec	0.817	0.822	0.828	0.821	0.855
	Prec	0.823	0.824	0.828	0.821	0.857
	F1	0.817	0.821	0.828	0.821	0.854
	AUC	0.884	0.896	0.896	0.894	0.909
	PPV	0.865	0.851	0.838	0.829	0.889
	Sensitivity	0.752	0.780	0.813	0.809	0.810
	Specificity	0.882	0.863	0.843	0.833	0.899
	FDR	0.135	0.149	0.162	0.171	0.111
	FOR	0.219	0.203	0.182	0.186	0.174
VGG16	Acc / Rec	0.834	0.821	0.821	0.820	0.825
	Prec	0.835	0.822	0.822	0.824	0.826
	F1	0.834	0.821	0.821	0.820	0.825
	AUC	0.907	0.900	0.896	0.900	0.891
	PPV	0.817	0.843	0.840	0.858	0.846
	Sensitivity	0.861	0.789	0.793	0.769	0.794
	Specificity	0.807	0.853	0.850	0.872	0.856
	FDR	0.183	0.157	0.160	0.142	0.154
	FOR	0.147	0.198	0.196	0.210	0.194
ResNet50	Acc / Rec	0.809	0.808	0.784	0.815	0.809
	Prec	0.814	0.809	0.788	0.815	0.810
	F1	0.808	0.808	0.784	0.815	0.809
	AUC	0.870	0.863	0.860	0.874	0.872
	PPV	0.854	0.803	0.819	0.805	0.821
	Sensitivity	0.746	0.818	0.731	0.831	0.790
	Specificity	0.872	0.799	0.838	0.799	0.828
	FDR	0.146	0.197	0.181	0.195	0.179
	FOR	0.226	0.186	0.243	0.175	0.202

TABLE VIII

RESULTS WITH DIFFERENT λ VALUES (BINARY CLASSIFICATION - NIH DATASET). $\lambda = 1$: CROSS-ENTROPY LOSS, $\lambda = *$: ADAPTIVE CUSTOM LOSS, BOLD: BEST RESULTS WITHIN THE SAME MODEL, UNDERLINE: BEST RESULTS OVER ALL MODELS, ITALICS: BEST RESULTS OVER ALL MODELS PRIOR TO “ATTENTION LOSS” BEING USED.

Pneumothorax vs No Finding						
Model	Metric	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = *$
PVT	Acc / Rec	0.771	0.793	0.784	0.759	0.790
	Prec	0.772	0.793	0.784	0.760	0.791
	F1	0.771	0.793	0.783	0.758	0.790
	AUC	0.842	0.872	0.858	0.841	0.868
	PPV	0.786	0.804	0.769	0.778	0.796
	Sensitivity	0.745	0.774	0.811	0.724	0.781
	Specificity	0.797	0.811	0.756	0.793	0.800
	FDR	0.214	0.196	0.231	0.222	0.204
	FOR	0.242	0.218	0.200	0.258	0.215
VGG16	Acc / Rec	0.765	0.763	0.768	0.776	0.778
	Prec	0.766	0.764	0.768	0.776	0.778
	F1	0.765	0.763	0.768	0.776	0.778
	AUC	0.841	0.828	0.848	0.848	0.860
	PPV	0.751	0.753	0.771	0.768	0.769
	Sensitivity	0.795	0.784	0.761	0.790	0.795
	Specificity	0.736	0.743	0.774	0.761	0.761
	FDR	0.249	0.247	0.229	0.232	0.231
	FOR	0.218	0.226	0.236	0.216	0.212
ResNet50	Acc / Rec	0.727	0.749	0.731	0.738	0.743
	Prec	0.727	0.751	0.733	0.739	0.746
	F1	0.727	0.749	0.731	0.738	0.742
	AUC	0.817	0.821	0.807	0.818	0.804
	PPV	0.732	0.733	0.713	0.723	0.719
	Sensitivity	0.715	0.784	0.774	0.772	0.797
	Specificity	0.738	0.715	0.688	0.704	0.688
	FDR	0.268	0.267	0.287	0.277	0.281
	FOR	0.278	0.232	0.247	0.244	0.228

TABLE IX

RESULTS WITH DIFFERENT λ VALUES (MULTI-CLASS CLASSIFICATION - NIH DATASET). $\lambda = 1$: CROSS-ENTROPY LOSS, $\lambda = *$: ADAPTIVE CUSTOM LOSS, BOLD: BEST RESULTS WITHIN THE SAME MODEL, UNDERLINE: BEST RESULTS OVER ALL MODELS, ITALICS: BEST RESULTS OVER ALL MODELS PRIOR TO “ATTENTION LOSS” BEING USED.

Effusion, Pneumothorax and No Finding						
Model	Metric	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = *$
PVT	Acc / Rec	0.651	0.651	0.670	0.669	0.677
	Prec	0.651	0.651	0.672	0.669	0.678
	F1	0.652	0.651	0.671	0.669	0.677
	AUC	0.812	0.819	0.843	0.828	0.839
VGG16	Acc / Rec	0.674	0.653	0.640	0.660	0.641
	Prec	0.674	0.657	0.641	0.660	0.641
	F1	0.674	0.651	0.640	0.660	0.641
	AUC	0.831	0.822	0.820	0.822	0.817
ResNet50	Acc / Rec	0.631	0.610	0.601	0.620	0.632
	Prec	0.631	0.609	0.598	0.622	0.632
	F1	0.626	0.609	0.598	0.609	0.630
	AUC	0.800	0.781	0.781	0.793	0.794