Attention-Based Medical Knowledge Injection in Deep Image Classification Models

Anonymous Authors

I. Appendix

A. Ablation study: σ parameter

The value of the σ parameter determines the threshold used to consider the pixel-level attention assigned by the model as significant, or not significant. This mechanism is fundamental for us to compute the proportion of pixels with high attention over the total number of pixels in the bounding box, to subsequently calculate the custom attention loss. While in the main paper we resort to a single configuration of σ which was empirically shown to be reliable in most cases, here we provide additional experiments to study the sensitivity of our method with varying values of σ.

Since the attention heatmap is normalized to a range from 0 to 1, the candidate values for σ considered in our study are from 0.1 to 0.9, with all values spaced at 0.2 intervals, i.e. $0.1, 0.3, 0.5, 0.7$, and 0.9. We employ PVT, VGG16 and ResNet50 model backbones for all classification tasks on both datasets and use 4 values for $\lambda : \{0.25, 0.5, 0.75,$ and adaptive (*) $\}$. We report detailed results of experiments on testing sets in Table II to Table XXII.

To analyze the distribution of results for F1, Sensitivity, and Specificity across all tables, we consider the number of cases where specific σ values achieve the best performance, leading to the aggregated results in Table I. Aggregating these results further across all model architectures, highlights that the performance is relatively favorable for σ values of 0.1 and 0.9. More specifically, ranking σ values across all experiments based on the number of cases they achieve the best performance (considering F1-Score, Sensitivity, and Specificity), leads to the following result: $\sigma = 0.7 : 50, \sigma = 0.9 : 46,\ \sigma = 0.3 : 43,\ \sigma = 0.5 : 42,\ \sigma = 0.1 : 39$.

When considering F1, the best performance of all models across tasks is achieved with σ values of 0.5 and 0.7. More specifically, ranking σ values across all experiments only based on F1-Score leads to the following result: $\sigma = 0.5 : 21,\ \sigma = 0.7 : 19, \sigma = 0.9 : 17, \sigma = 0.3 : 16, \sigma = 0.1 : 15$.

If we consider Sensitivity, the best performance of all models across tasks is achieved with σ values of 0.1, 0.3 and 0.7. More in detail, ranking σ values across all experiments solely based on Sensitivity leads to the following result: $\sigma = \{0.1, 0.3, 0.7\} : 14,\ \sigma = 0.9 : 13,\ \sigma = 0.5 : 10$. When we consider only Specificity, the performance ranking of the various σ values remains the same as when all metrics are taken into account. Specifically, ranking σ values across all experiments based on Specificity leads to the following outcome: $\sigma = 0.7 : 17,\ \sigma = 0.9 : 14, \sigma = 0.3 : 13, \sigma = 0.5 : 11, \sigma = 0.1 : 10$. In the attempt to identify an ideal value for σ, it is also important to consider that we intend for the model to impose a greater penalty on incorrect predictions (e.g. False Negatives) during training, as referenced by formulas (3), (4), (5), and (7) in the main paper. This implies preferring relatively larger $\mathcal{L}_{attention}$ values, which leads us to prefer higher values of σ. However, it is worth noting that the value of σ should be small enough to avoid overly neglecting the contributions of the light-colored regions in attention heatmaps. Therefore, our conservative choice of $\sigma = 0.3$ for results presented in the main paper is reasonably justified. Overall, we observe that the value of σ has an impact on the final model performance, although no unique trend is observed across models and metrics.

TABLE I
Comparison of σ values: number of cases in which they achieve the best performance across different configurations of λ and for different key metrics (F1, Sensitivity, and Specificity) for all model architectures.

Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
PVT					
F1	5	4	9	7	3
Sensitivity	6	6	5	1	3
Specificity	4	4	5	7	3
VGG16					
F1	8	6	6	6	4
Sensitivity	3	3	3	8	6
Specificity	3	7	2	5	4
ResNet50					
F1	2	6	6	6	10
Sensitivity	5	5	2	5	4
Specificity	3	2	4	5	7

B. Comparison with medical works

In our experiments, we showed that the proposed custom loss has the potential to improve the performance of state-of-the-art model backbones using the conventional cross-entropy loss. However, another relevant aspect is that of comparing model performance with research works devoted specifically to medical image classification.

To this end, one difficulty for a direct fair comparison is presented by our specific experimental setup. For instance, the proposed custom loss led us to the selection of a subset of classes for the multi-class classification setting, based on

TABLE II

RESULTS OF PVT MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

PVT: Effusion vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.660	0.622	0.641	0.660	0.647
	Prec	0.666	0.622	0.643	0.660	0.653
	F1	0.657	0.622	0.640	0.660	0.644
	AUC	0.713	0.702	0.705	0.739	0.702
	PPV	0.634	0.620	0.662	0.667	0.683
	Sensitivity	0.756	0.628	0.577	0.641	0.551
	Specificity	0.564	0.615	0.705	0.679	0.744
	FDR	0.366	0.380	0.338	0.333	0.317
	FOR	0.302	0.377	0.375	0.346	0.376
$\lambda = 0.5$	Acc / Rec	**0.724**	0.667	0.654	0.686	0.667
	Prec	**0.725**	0.668	0.662	0.698	0.667
	F1	**0.724**	0.666	0.649	0.681	0.667
	AUC	0.771	0.732	0.743	**0.774**	0.727
	PPV	0.740	0.686	0.700	**0.746**	0.667
	Sensitivity	0.692	0.615	0.538	0.564	0.667
	Specificity	0.756	0.718	0.769	**0.808**	0.667
	FDR	0.260	0.314	0.300	**0.254**	0.333
	FOR	0.289	0.349	0.375	0.351	0.333
$\lambda = 0.75$	Acc / Rec	0.615	0.673	0.679	0.647	0.647
	Prec	0.617	0.673	0.679	0.653	0.655
	F1	0.614	0.673	0.679	0.644	0.643
	AUC	0.723	0.740	0.740	0.702	0.719
	PPV	0.629	0.671	0.679	0.624	0.621
	Sensitivity	0.564	0.679	0.679	0.744	0.756
	Specificity	0.667	0.667	0.679	0.551	0.538
	FDR	0.371	0.329	0.321	0.376	0.379
	FOR	0.395	0.325	0.321	0.317	0.311
$\lambda = *$	Acc / Rec	0.686	0.712	0.679	0.667	0.673
	Prec	0.711	0.712	0.680	0.672	0.674
	F1	0.676	0.711	0.679	0.664	0.672
	AUC	0.720	0.734	0.760	0.726	0.735
	PPV	0.638	0.720	0.671	0.641	0.659
	Sensitivity	**0.859**	0.692	0.705	0.756	0.718
	Specificity	0.513	0.731	0.654	0.577	0.628
	FDR	0.362	0.280	0.329	0.359	0.341
	FOR	**0.216**	0.296	0.311	0.297	0.310

the available medical knowledge required to setup bounding boxes. This condition does not apply to other studies, where the full set of classes of the NIH dataset is adopted. On the other hand, state-of-the-art papers do not usually publish their code, making replicability in the same experimental conditions unfeasible.

Nevertheless, we made our best effort to analyze recent works and ascertain that the performance scores achieved with our proposed method are in line (or outperform) those reported by other studies. For instance, [1] report a mean AUC on the official NIH dataset split of 0.745. The work in [2] reports a mean AUC of 0.755. More recently, influential works reported a mean AUC of 0.807 [3], 0.817 [4], and 0.819 [5]. Very recent research reported a mean AUC of 0.821 [6] and 0.827 [7]. In our results, it is possible to observe that we can achieve mean values of AUC corresponding to 0.839 ($\lambda = 0.75$, PVT model) as shown in Table VIII, and 0.826 ($\lambda = 0.5$, VGG16 model) as shown in Table XV. Besides the aforementioned challenges exacerbating the difficulty of an exact comparison, we confidently observe that our results are competitive with results obtained thus far by state-of-the-art works on this dataset.

TABLE III

RESULTS OF PVT MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

PVT: Pneumothorax vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.612	0.509	0.612	0.573	0.578
	Prec	0.612	0.509	0.613	0.573	0.580
	F1	0.612	0.508	0.611	0.573	0.575
	AUC	0.656	0.556	0.664	0.627	0.610
	PPV	0.618	0.509	0.625	0.571	0.592
	Sensitivity	0.586	0.491	0.560	0.586	0.500
	Specificity	0.638	0.526	0.664	0.560	0.655
	FDR	0.382	0.491	0.375	0.429	0.408
	FOR	0.393	0.492	0.398	0.425	0.433
$\lambda = 0.5$	Acc / Rec	0.552	0.586	**0.634**	0.591	0.569
	Prec	0.555	0.590	0.634	0.607	0.569
	F1	0.544	0.582	**0.633**	0.575	0.568
	AUC	0.590	0.628	0.653	0.612	0.599
	PPV	0.541	0.609	0.645	0.648	0.574
	Sensitivity	**0.681**	0.483	0.595	0.397	0.534
	Specificity	0.422	0.690	0.672	**0.784**	0.603
	FDR	0.459	0.391	0.355	0.352	0.426
	FOR	0.430	0.429	0.376	0.435	0.435
$\lambda = 0.75$	Acc / Rec	0.556	0.573	0.582	0.616	0.608
	Prec	0.556	0.575	0.582	0.617	0.609
	F1	0.555	0.570	0.582	0.616	0.607
	AUC	0.595	0.628	0.605	0.662	0.652
	PPV	0.561	0.588	0.587	0.626	0.598
	Sensitivity	0.517	0.491	0.552	0.578	0.655
	Specificity	0.595	0.655	0.612	0.655	0.560
	FDR	0.439	0.412	0.413	0.374	0.402
	FOR	0.448	0.437	0.423	0.392	0.381
$\lambda = *$	Acc / Rec	0.621	0.578	0.634	0.603	0.625
	Prec	0.622	0.578	**0.638**	0.609	0.625
	F1	0.620	0.576	0.631	0.598	0.625
	AUC	0.663	0.621	**0.712**	0.646	0.666
	PPV	0.611	0.587	**0.663**	0.633	0.628
	Sensitivity	0.664	0.526	0.543	0.491	0.612
	Specificity	0.578	0.629	0.724	0.716	0.638
	FDR	0.389	0.413	**0.337**	0.367	0.372
	FOR	**0.368**	0.430	0.387	0.415	0.378

TABLE IV

RESULTS OF PVT MODEL WITH DIFFERENT σ VALUES (MULTI-CLASS CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

PVT: Effusion, Pneumothorax and No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.470	0.517	0.462	0.449	0.466
	Prec	0.471	0.514	0.480	0.447	0.462
	F1	0.469	0.511	0.464	0.442	0.461
	AUC	0.661	0.685	0.654	0.667	0.657
$\lambda = 0.5$	Acc / Rec	**0.521**	0.491	0.470	0.457	0.504
	Prec	**0.523**	0.494	0.466	0.463	0.502
	F1	**0.522**	0.492	0.463	0.459	0.490
	AUC	0.695	0.656	0.659	0.657	0.643
$\lambda = 0.75$	Acc / Rec	0.470	0.479	**0.521**	0.500	0.509
	Prec	0.466	0.480	0.522	0.499	0.509
	F1	0.467	0.478	0.518	0.490	0.505
	AUC	0.663	0.649	0.680	0.678	**0.708**
$\lambda = *$	Acc / Rec	0.474	0.449	0.483	0.466	0.470
	Prec	0.477	0.447	0.483	0.459	0.468
	F1	0.473	0.447	0.476	0.458	0.467
	AUC	0.647	0.639	0.663	0.672	0.677

TABLE V

RESULTS OF PVT MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

PVT: Effusion vs No Finding						
λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.826	**0.839**	0.832	0.820	0.831
	Prec	0.826	**0.839**	0.833	0.821	0.834
	F1	0.826	**0.839**	0.832	0.820	0.831
	AUC	0.894	0.903	0.890	0.878	0.904
	PPV	0.832	0.846	0.847	0.841	0.865
	Sensitivity	0.815	0.828	0.812	0.789	0.785
	Specificity	0.836	0.850	0.853	0.851	0.877
	FDR	0.168	0.154	0.153	0.159	0.135
	FOR	0.181	0.168	0.181	0.199	0.197
$\lambda = 0.5$	Acc / Rec	0.830	0.831	0.824	0.838	0.803
	Prec	0.835	0.831	0.824	**0.839**	0.806
	F1	0.829	0.831	0.824	0.838	0.803
	AUC	**0.911**	0.895	0.895	0.903	0.881
	PPV	0.878	0.840	0.816	0.855	0.831
	Sensitivity	0.766	0.817	0.837	0.814	0.761
	Specificity	0.894	0.845	0.812	0.862	0.846
	FDR	0.122	0.160	0.184	0.145	0.169
	FOR	0.207	0.178	0.167	0.177	0.220
$\lambda = 0.75$	Acc / Rec	0.821	0.834	0.822	0.834	0.838
	Prec	0.830	0.836	0.823	0.834	0.838
	F1	0.820	0.834	0.822	0.834	0.838
	AUC	0.897	0.908	0.882	0.902	0.896
	PPV	**0.884**	0.863	0.847	0.838	0.836
	Sensitivity	0.740	0.795	0.785	0.828	**0.839**
	Specificity	**0.903**	0.874	0.858	0.839	0.836
	FDR	**0.116**	0.137	0.153	0.162	0.164
	FOR	0.224	0.190	0.200	0.170	**0.161**
$\lambda = *$	Acc / Rec	0.816	0.825	0.826	0.815	0.826
	Prec	0.816	0.826	0.833	0.817	0.829
	F1	0.816	0.825	0.825	0.814	0.826
	AUC	0.879	0.887	0.901	0.879	0.907
	PPV	0.828	0.841	**0.884**	0.845	0.859
	Sensitivity	0.798	0.802	0.750	0.771	0.780
	Specificity	0.834	0.848	0.901	0.858	0.872
	FDR	0.172	0.159	**0.116**	0.155	0.141
	FOR	0.195	0.190	0.217	0.210	0.201

TABLE VI

RESULTS OF PVT MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

PVT: Pneumothorax vs No Finding						
λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.777	0.761	0.788	0.743	0.772
	Prec	0.781	0.761	0.793	0.767	0.776
	F1	0.776	0.761	0.787	0.737	0.771
	AUC	0.854	0.837	0.870	0.857	0.849
	PPV	0.816	0.752	0.756	**0.847**	0.742
	Sensitivity	0.715	0.779	**0.852**	0.592	0.834
	Specificity	0.838	0.743	0.724	**0.893**	0.711
	FDR	0.184	0.248	0.244	**0.153**	0.258
	FOR	0.254	0.229	**0.170**	0.313	0.190
$\lambda = 0.5$	Acc / Rec	0.777	0.780	0.788	0.772	0.772
	Prec	0.777	0.786	0.789	0.776	0.774
	F1	0.777	0.779	0.788	0.772	0.772
	AUC	0.845	0.865	0.868	0.841	0.851
	PPV	0.765	0.746	0.776	0.806	0.794
	Sensitivity	0.800	0.850	0.811	0.718	0.736
	Specificity	0.754	0.711	0.765	0.827	0.809
	FDR	0.235	0.254	0.224	0.194	0.206
	FOR	0.210	0.175	0.198	0.255	0.246
$\lambda = 0.75$	Acc / Rec	0.780	0.771	0.754	0.764	0.769
	Prec	0.780	0.771	0.761	0.764	0.769
	F1	0.780	0.771	0.752	0.764	0.769
	AUC	0.855	0.851	0.843	0.857	0.845
	PPV	0.783	0.779	0.719	0.772	0.777
	Sensitivity	0.774	0.756	0.834	0.749	0.754
	Specificity	0.786	0.786	0.674	0.779	0.784
	FDR	0.217	0.221	0.281	0.228	0.223
	FOR	0.223	0.237	0.198	0.243	0.239
$\lambda = *$	Acc / Rec	0.772	0.769	0.780	**0.794**	0.774
	Prec	0.774	0.769	0.782	**0.794**	0.775
	F1	0.772	0.769	0.780	**0.794**	0.774
	AUC	0.861	0.841	0.870	**0.871**	0.851
	PPV	0.794	0.757	0.806	0.797	0.767
	Sensitivity	0.736	0.793	0.738	0.788	0.788
	Specificity	0.809	0.745	0.822	0.800	0.761
	FDR	0.206	0.243	0.194	0.203	0.233
	FOR	0.246	0.218	0.242	0.209	0.218

TABLE VII

Results of PVT model with different σ values (binary classification - NIH dataset). Values in bold indicate the best results within the same model for this task.

PVT: Cardiomegaly vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
	Acc / Rec	0.722	0.757	0.709	0.702	0.709
	Prec	0.725	0.762	0.714	0.715	0.710
	F1	0.722	0.756	0.707	0.697	0.708
	AUC	0.783	0.829	0.785	0.810	0.794
$\lambda = 0.25$	PPV	0.749	**0.798**	0.681	0.768	0.726
	Sensitivity	0.670	0.688	0.784	0.578	0.670
	Specificity	0.775	0.826	0.633	0.826	0.748
	FDR	0.251	**0.202**	0.319	0.232	0.274
	FOR	0.299	0.274	0.254	0.338	0.306
	Acc / Rec	0.716	0.729	0.720	**0.771**	0.734
	Prec	0.732	0.735	0.726	**0.772**	0.734
	F1	0.710	0.728	0.718	**0.770**	0.734
	AUC	0.823	0.772	0.800	**0.832**	0.811
$\lambda = 0.5$	PPV	0.794	0.698	0.761	0.795	0.726
	Sensitivity	0.583	**0.807**	0.642	0.729	0.752
	Specificity	**0.849**	0.651	0.798	0.812	0.716
	FDR	0.206	0.302	0.239	0.205	0.274
	FOR	0.330	**0.228**	0.310	0.250	0.257
	Acc / Rec	0.718	0.709	0.700	0.736	0.704
	Prec	0.718	0.709	0.707	0.743	0.704
	F1	0.718	0.708	0.697	0.734	0.704
	AUC	0.792	0.789	0.797	0.822	0.783
$\lambda = 0.75$	PPV	0.729	0.697	0.746	0.785	0.709
	Sensitivity	0.693	0.739	0.606	0.651	0.693
	Specificity	0.743	0.679	0.794	0.821	0.716
	FDR	0.271	0.303	0.254	0.215	0.291
	FOR	0.293	0.278	0.332	0.298	0.300
	Acc / Rec	0.681	0.704	0.755	0.743	0.739
	Prec	0.684	0.707	0.755	0.747	0.744
	F1	0.680	0.703	0.754	0.742	0.737
	AUC	0.756	0.782	0.814	0.818	0.814
$\lambda = *$	PPV	0.707	0.731	0.744	0.779	0.783
	Sensitivity	0.619	0.647	0.775	0.679	0.661
	Specificity	0.743	0.761	0.734	0.807	0.817
	FDR	0.293	0.269	0.256	0.221	0.217
	FOR	0.339	0.317	0.234	0.285	0.294

TABLE VIII

Results of PVT model with different σ values (multi-class classification - NIH dataset). Values in bold indicate the best results within the same model for this task.

PVT: Effusion, Pneumothorax and No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
	Acc / Rec	**0.767**	0.655	0.683	0.672	0.674
$\lambda = 0.25$	Prec	0.680	0.656	0.686	0.678	0.682
	F1	0.676	0.655	0.683	0.669	0.673
	AUC	0.845	0.815	0.840	0.841	0.838
	Acc / Rec	0.671	0.667	0.674	0.651	0.678
$\lambda = 0.5$	Prec	0.675	0.667	0.674	0.651	0.680
	F1	0.670	0.667	0.674	0.651	0.679
	AUC	0.837	0.835	0.840	0.822	0.835
	Acc / Rec	0.666	0.685	0.675	0.686	0.664
$\lambda = 0.75$	Prec	0.675	**0.693**	0.680	0.688	0.665
	F1	0.664	0.684	0.675	0.685	0.664
	AUC	0.830	**0.846**	0.844	0.843	0.832
	Acc / Rec	0.694	0.686	0.671	0.664	0.681
$\lambda = *$	Prec	**0.693**	0.692	0.676	0.669	0.682
	F1	**0.693**	0.685	0.672	0.665	0.679
	AUC	0.839	**0.846**	0.825	0.829	0.845

TABLE IX

Results of VGG16 model with different σ values (binary classification - Ottawa dataset). Values in bold indicate the best results within the same model for this task.

VGG16 : Effusion vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
	Acc / Rec	0.635	0.647	0.635	**0.718**	0.667
	Prec	0.635	0.650	0.641	**0.719**	0.667
	F1	0.634	0.646	0.630	**0.718**	0.667
	AUC	0.707	0.727	0.685	**0.756**	0.736
$\lambda = 0.25$	PPV	0.630	0.672	0.672	**0.730**	0.671
	Sensitivity	0.654	0.577	0.526	0.692	0.654
	Specificity	0.615	0.718	0.744	0.744	0.679
	FDR	0.370	0.328	0.328	**0.270**	0.329
	FOR	0.360	0.371	0.389	0.293	0.338
	Acc / Rec	0.590	0.679	0.583	0.647	0.686
	Prec	0.594	0.681	0.586	0.647	0.686
	F1	0.585	0.679	0.580	0.647	0.686
	AUC	0.673	**0.756**	0.708	0.737	0.750
$\lambda = 0.5$	PPV	0.613	0.694	0.571	0.649	0.693
	Sensitivity	0.487	0.641	0.667	0.641	0.667
	Specificity	0.692	0.718	0.500	0.654	0.705
	FDR	0.387	0.306	0.429	0.351	0.307
	FOR	0.426	0.333	0.400	0.354	0.321
	Acc / Rec	0.622	0.673	0.673	0.654	0.660
	Prec	0.624	0.680	0.678	0.656	0.660
	F1	0.620	0.670	0.671	0.652	0.660
	AUC	0.706	0.744	0.741	0.722	0.726
$\lambda = 0.75$	PPV	0.642	0.714	0.708	0.676	0.658
	Sensitivity	0.551	0.577	0.590	0.590	0.667
	Specificity	0.692	**0.769**	0.756	0.718	0.654
	FDR	0.358	0.286	0.292	0.324	0.342
	FOR	0.393	0.355	0.352	0.364	0.338
	Acc / Rec	0.660	0.673	0.679	0.647	0.654
	Prec	0.662	0.673	0.687	0.668	0.655
	F1	0.660	0.673	0.676	0.637	0.653
	AUC	0.740	0.725	0.716	0.731	0.728
$\lambda = *$	PPV	0.676	0.675	0.649	0.610	0.667
	Sensitivity	0.615	0.667	0.782	**0.821**	0.615
	Specificity	0.705	0.679	0.577	0.474	0.692
	FDR	0.324	0.325	0.351	0.390	0.333
	FOR	0.353	0.329	**0.274**	0.275	0.357

TABLE X

RESULTS OF VGG16 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

VGG16: Pneumothorax vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	**0.668**	0.642	0.651	0.616	0.616
	Prec	0.668	0.643	0.651	0.617	0.617
	F1	**0.668**	0.642	0.651	0.616	0.616
	AUC	0.698	0.718	**0.728**	0.687	0.719
	PPV	0.664	0.654	0.647	0.629	0.612
	Sensitivity	0.681	0.603	0.664	0.569	0.638
	Specificity	0.655	0.681	0.638	0.664	0.595
	FDR	0.336	0.346	0.353	0.371	0.388
	FOR	0.327	0.368	0.345	0.394	0.378
$\lambda = 0.5$	Acc / Rec	**0.668**	0.569	0.634	0.629	0.651
	Prec	**0.672**	0.579	0.634	0.631	0.651
	F1	0.666	0.554	0.633	0.628	0.651
	AUC	0.698	0.694	0.679	0.704	0.679
	PPV	0.647	0.551	0.642	0.615	0.645
	Sensitivity	0.741	0.750	0.603	0.690	0.672
	Specificity	0.595	0.388	0.664	0.569	0.629
	FDR	0.353	0.449	0.358	0.385	0.355
	FOR	0.303	0.392	0.374	0.353	0.342
$\lambda = 0.75$	Acc / Rec	0.664	0.659	0.642	0.634	0.608
	Prec	0.664	0.662	0.660	0.634	0.629
	F1	0.664	0.658	0.632	0.633	0.591
	AUC	0.713	0.710	0.694	0.686	0.687
	PPV	0.661	0.683	0.606	0.642	0.577
	Sensitivity	0.672	0.595	**0.810**	0.603	**0.810**
	Specificity	0.655	0.724	0.474	0.664	0.405
	FDR	0.339	0.317	0.394	0.358	0.423
	FOR	0.333	0.359	**0.286**	0.374	0.319
$\lambda = *$	Acc / Rec	0.638	0.651	0.642	0.599	0.651
	Prec	0.639	0.651	0.643	0.607	0.656
	F1	0.637	0.651	0.642	0.592	0.648
	AUC	0.708	0.681	0.685	0.664	0.694
	PPV	0.654	0.661	0.634	0.578	**0.684**
	Sensitivity	0.586	0.621	0.672	0.733	0.560
	Specificity	0.690	0.681	0.612	0.466	**0.741**
	FDR	0.346	0.339	0.366	0.422	**0.316**
	FOR	0.375	0.358	0.349	0.365	0.372

TABLE XI

RESULTS OF VGG16 MODEL WITH DIFFERENT σ VALUES (MULTI-CLASS CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

VGG16: Effusion, Pneumothorax and No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.487	0.500	0.504	0.474	0.491
	Prec	0.498	0.501	0.505	0.473	0.483
	F1	0.488	0.500	0.489	0.472	0.483
	AUC	0.685	0.684	0.700	0.681	0.686
$\lambda = 0.5$	Acc / Rec	0.521	0.483	0.521	0.496	0.466
	Prec	0.522	0.490	0.520	0.497	0.453
	F1	0.521	0.478	0.521	0.496	0.444
	AUC	0.719	0.707	0.703	0.688	0.696
$\lambda = 0.75$	Acc / Rec	0.487	0.491	0.496	**0.551**	0.474
	Prec	0.482	0.503	0.476	**0.547**	0.464
	F1	0.481	0.484	0.451	**0.546**	0.449
	AUC	0.687	0.661	0.691	**0.735**	0.696
$\lambda = *$	Acc / Rec	0.530	0.496	0.517	0.517	0.453
	Prec	0.532	0.492	0.513	0.519	0.446
	F1	0.529	0.491	0.510	0.518	0.436
	AUC	0.710	0.706	0.718	0.719	0.679

TABLE XII

RESULTS OF VGG16 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

VGG16: Effusion vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.820	0.820	0.829	**0.832**	0.825
	Prec	0.822	0.822	0.831	**0.835**	0.826
	F1	0.820	0.820	0.829	**0.832**	0.825
	AUC	0.897	0.898	0.895	**0.904**	0.897
	PPV	0.840	0.848	0.857	0.861	0.849
	Sensitivity	0.791	0.780	0.790	0.793	0.790
	Specificity	0.850	0.860	0.869	0.872	0.860
	FDR	0.160	0.152	0.143	0.139	0.151
	FOR	0.197	0.204	0.195	0.192	0.196
$\lambda = 0.5$	Acc / Rec	0.806	0.819	0.810	0.813	0.817
	Prec	0.807	0.822	0.814	0.814	0.821
	F1	0.806	0.818	0.809	0.813	0.816
	AUC	0.878	0.890	0.899	0.891	0.895
	PPV	0.826	0.857	0.850	0.830	0.860
	Sensitivity	0.775	0.765	0.752	0.786	0.756
	Specificity	0.837	0.872	0.867	0.839	0.877
	FDR	0.174	0.143	0.150	0.170	0.140
	FOR	0.212	0.212	0.222	0.203	0.218
$\lambda = 0.75$	Acc / Rec	0.812	0.819	0.817	0.823	0.819
	Prec	0.812	0.820	0.818	0.823	0.820
	F1	0.812	0.818	0.817	0.823	0.818
	AUC	0.892	0.889	0.890	0.898	0.892
	PPV	0.823	0.837	0.835	0.812	0.838
	Sensitivity	0.794	0.791	0.791	**0.841**	0.790
	Specificity	0.829	0.846	0.843	0.805	0.847
	FDR	0.177	0.163	0.165	0.188	0.162
	FOR	0.199	0.198	0.198	**0.165**	0.199
$\lambda = *$	Acc / Rec	0.826	0.827	0.814	0.809	0.818
	Prec	0.826	0.832	0.819	0.810	0.819
	F1	0.826	0.827	0.814	0.809	0.818
	AUC	0.894	0.903	0.894	0.887	0.893
	PPV	0.833	**0.871**	0.857	0.829	0.840
	Sensitivity	0.814	0.769	0.755	0.779	0.785
	Specificity	0.837	**0.886**	0.874	0.839	0.851
	FDR	0.167	**0.129**	0.143	0.171	0.160
	FOR	0.182	0.207	0.219	0.209	0.202

TABLE XIII

RESULTS OF VGG16 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

VGG16: Pneumothorax vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.788	0.778	**0.789**	0.779	0.788
	Prec	**0.791**	0.779	0.790	0.779	0.788
	F1	0.788	0.778	**0.789**	0.779	0.788
	AUC	**0.865**	0.837	0.863	0.853	0.849
	PPV	0.817	0.793	0.802	0.770	0.784
	Sensitivity	0.743	0.752	0.768	0.795	0.795
	Specificity	0.834	0.804	0.811	0.763	0.781
	FDR	0.183	0.207	0.198	0.230	0.216
	FOR	0.236	0.236	0.223	0.212	0.208
$\lambda = 0.5$	Acc / Rec	**0.789**	0.787	0.768	0.780	0.767
	Prec	0.789	0.787	0.768	0.780	0.775
	F1	**0.789**	0.787	0.768	0.780	0.765
	AUC	0.853	0.863	0.837	0.852	0.854
	PPV	0.794	0.779	0.773	0.780	**0.823**
	Sensitivity	0.781	0.802	0.759	0.781	0.679
	Specificity	0.797	0.772	0.777	0.779	**0.854**
	FDR	0.206	0.221	0.227	0.220	**0.177**
	FOR	0.215	0.204	0.237	0.219	0.273
$\lambda = 0.75$	Acc / Rec	0.777	0.778	0.759	0.762	0.771
	Prec	0.778	0.778	0.761	0.762	0.773
	F1	0.777	0.778	0.758	0.762	0.771
	AUC	0.849	0.855	0.840	0.841	0.846
	PPV	0.797	0.789	0.735	0.769	0.749
	Sensitivity	0.743	0.759	0.809	0.749	0.815
	Specificity	0.811	0.797	0.708	0.774	0.727
	FDR	0.203	0.211	0.265	0.231	0.251
	FOR	0.241	0.232	0.213	0.244	0.203
$\lambda = *$	Acc / Rec	0.774	0.785	0.759	0.771	0.764
	Prec	0.781	0.785	0.759	0.773	0.764
	F1	0.773	0.785	0.758	0.771	0.764
	AUC	0.843	0.847	0.831	0.840	0.835
	PPV	0.738	0.791	0.752	0.793	0.767
	Sensitivity	**0.852**	0.774	0.772	0.733	0.759
	Specificity	0.697	0.795	0.745	0.809	0.770
	FDR	0.262	0.209	0.248	0.207	0.233
	FOR	**0.175**	0.221	0.234	0.248	0.239

TABLE XIV

RESULTS OF VGG16 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

VGG16: Cardiomegaly vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.736	0.702	0.750	0.722	0.752
	Prec	0.745	0.720	0.756	0.723	0.753
	F1	0.734	0.696	0.749	0.722	0.752
	AUC	0.835	0.837	0.829	0.829	0.837
	PPV	0.791	0.782	0.795	0.712	0.767
	Sensitivity	0.642	0.560	0.674	0.748	0.725
	Specificity	0.830	**0.844**	0.826	0.697	0.780
	FDR	0.209	0.218	0.205	0.288	0.233
	FOR	0.301	0.343	0.283	0.266	0.261
$\lambda = 0.5$	Acc / Rec	0.750	0.729	0.755	0.752	0.741
	Prec	0.754	0.730	0.756	0.760	0.744
	F1	0.749	0.729	0.754	0.750	0.740
	AUC	0.834	0.826	0.840	0.842	0.834
	PPV	0.785	0.719	0.776	**0.806**	0.772
	Sensitivity	0.688	0.752	0.716	0.665	0.683
	Specificity	0.812	0.706	0.794	0.839	0.798
	FDR	0.215	0.281	0.224	**0.194**	0.228
	FOR	0.278	0.260	0.264	0.285	0.284
$\lambda = 0.75$	Acc / Rec	0.745	0.741	0.745	0.757	0.734
	Prec	0.749	0.743	0.746	0.762	0.735
	F1	0.745	0.740	0.745	0.756	0.734
	AUC	0.833	0.828	0.838	**0.846**	0.814
	PPV	0.777	0.766	0.761	0.798	0.722
	Sensitivity	0.688	0.693	0.716	0.688	**0.761**
	Specificity	0.803	0.789	0.775	0.826	0.706
	FDR	0.223	0.234	0.239	0.202	0.278
	FOR	0.280	0.280	0.268	0.274	**0.252**
$\lambda = *$	Acc / Rec	0.741	0.732	**0.764**	0.741	0.734
	Prec	0.743	0.737	**0.765**	0.743	0.737
	F1	0.740	0.730	**0.763**	0.740	0.733
	AUC	0.837	0.832	0.844	0.834	0.827
	PPV	0.764	0.773	0.786	0.766	0.766
	Sensitivity	0.697	0.656	0.725	0.693	0.674
	Specificity	0.784	0.807	0.803	0.789	0.794
	FDR	0.236	0.227	0.214	0.234	0.234
	FOR	0.278	0.299	0.255	0.280	0.291

TABLE XV

RESULTS OF VGG16 MODEL WITH DIFFERENT σ VALUES (MULTI-CLASS CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

VGG16: Effusion, Pneumothorax and No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	**0.674**	0.646	0.648	0.669	0.651
	Prec	0.674	0.646	0.648	0.672	0.652
	F1	**0.673**	0.646	0.648	0.669	0.651
	AUC	**0.834**	0.821	0.815	0.827	0.817
$\lambda = 0.5$	Acc / Rec	0.672	0.648	0.646	0.667	0.642
	Prec	**0.679**	0.652	0.658	0.670	0.647
	F1	0.668	0.649	0.646	0.668	0.643
	AUC	0.833	0.825	0.827	0.825	0.821
$\lambda = 0.75$	Acc / Rec	0.652	0.651	0.657	0.635	0.667
	Prec	0.654	0.651	0.659	0.645	0.667
	F1	0.651	0.651	0.652	0.635	0.667
	AUC	0.823	0.812	0.824	0.822	0.831
$\lambda = *$	Acc / Rec	0.645	0.651	0.657	0.652	0.663
	Prec	0.648	0.656	0.659	0.653	0.663
	F1	0.646	0.653	0.651	0.653	0.663
	AUC	0.817	0.817	0.826	0.811	0.828

TABLE XVI

RESULTS OF RESNET50 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

ResNet50: Effusion vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.660	0.628	0.667	0.673	0.622
	Prec	0.660	0.628	0.672	**0.687**	0.622
	F1	0.660	0.628	0.664	0.667	0.622
	AUC	0.719	0.702	**0.748**	**0.748**	0.693
	PPV	0.658	0.628	0.641	0.636	0.617
	Sensitivity	0.667	0.628	0.756	**0.808**	0.641
	Specificity	0.654	0.628	0.577	0.538	0.603
	FDR	0.342	0.372	0.359	0.364	0.383
	FOR	0.338	0.372	0.297	**0.263**	0.373
$\lambda = 0.5$	Acc / Rec	0.641	0.647	**0.679**	0.667	0.615
	Prec	0.641	0.655	0.679	0.669	0.624
	F1	0.641	0.643	**0.679**	0.665	0.609
	AUC	0.731	0.722	0.728	0.731	0.694
	PPV	0.645	0.621	**0.679**	0.648	0.592
	Sensitivity	0.628	0.756	0.679	0.731	0.744
	Specificity	0.654	0.538	**0.679**	0.603	0.487
	FDR	0.355	0.379	**0.321**	0.352	0.408
	FOR	0.363	0.311	0.321	0.309	0.345
$\lambda = 0.75$	Acc / Rec	0.660	0.647	0.667	0.641	0.667
	Prec	0.670	0.650	0.668	0.641	0.669
	F1	0.655	0.646	0.666	0.641	0.665
	AUC	0.724	0.709	0.724	0.716	0.725
	PPV	0.629	0.629	0.655	0.637	0.648
	Sensitivity	0.782	0.718	0.705	0.654	0.731
	Specificity	0.538	0.577	0.628	0.628	0.603
	FDR	0.371	0.371	0.345	0.363	0.352
	FOR	0.288	0.328	0.319	0.355	0.309
$\lambda = *$	Acc / Rec	0.654	0.660	0.615	0.673	0.660
	Prec	0.654	0.660	0.618	0.674	0.660
	F1	0.654	0.660	0.613	0.673	0.660
	AUC	0.731	0.702	0.707	0.709	0.700
	PPV	0.654	0.662	0.600	0.663	0.658
	Sensitivity	0.654	0.654	0.692	0.705	0.667
	Specificity	0.654	0.667	0.538	0.641	0.654
	FDR	0.346	0.338	0.400	0.337	0.342
	FOR	0.346	0.342	0.364	0.315	0.338

TABLE XVII

RESULTS OF RESNET50 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

ResNet50: Pneumothorax vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.595	0.547	**0.651**	0.612	0.509
	Prec	0.595	0.549	**0.655**	0.612	0.511
	F1	0.595	0.544	**0.649**	0.612	0.478
	AUC	0.618	0.579	0.666	0.637	0.538
	PPV	0.589	0.540	0.630	0.612	0.517
	Sensitivity	0.629	0.638	**0.733**	0.612	0.267
	Specificity	0.560	0.457	0.569	0.612	**0.750**
	FDR	0.411	0.460	0.370	0.388	0.483
	FOR	0.398	0.442	**0.320**	0.388	0.494
$\lambda = 0.5$	Acc / Rec	0.517	0.599	0.573	0.560	0.491
	Prec	0.518	0.599	0.574	0.562	0.491
	F1	0.515	0.599	0.573	0.557	0.490
	AUC	0.556	0.646	0.599	0.583	0.492
	PPV	0.520	0.598	0.578	0.551	0.490
	Sensitivity	0.448	0.603	0.543	0.647	0.431
	Specificity	0.586	0.595	0.603	0.474	0.552
	FDR	0.480	0.402	0.422	0.449	0.510
	FOR	0.485	0.400	0.431	0.427	0.508
$\lambda = 0.75$	Acc / Rec	0.599	0.569	0.616	0.582	0.556
	Prec	0.601	0.572	0.616	0.582	0.558
	F1	0.597	0.565	0.616	0.582	0.552
	AUC	0.648	0.622	0.625	0.607	0.617
	PPV	0.586	0.558	0.615	0.586	0.570
	Sensitivity	0.672	0.664	0.621	0.560	0.457
	Specificity	0.526	0.474	0.612	0.603	0.655
	FDR	0.414	0.442	0.385	0.414	0.430
	FOR	0.384	0.415	0.383	0.421	0.453
$\lambda = *$	Acc / Rec	0.547	0.569	0.582	0.608	0.603
	Prec	0.548	0.569	0.582	0.608	0.609
	F1	0.546	0.569	0.582	0.608	0.598
	AUC	0.596	0.600	0.598	0.614	**0.683**
	PPV	0.543	0.573	0.580	0.609	**0.633**
	Sensitivity	0.603	0.543	0.595	0.603	0.491
	Specificity	0.491	0.595	0.569	0.612	0.716
	FDR	0.457	0.427	0.420	0.391	**0.367**
	FOR	0.447	0.434	0.416	0.393	0.415

TABLE XVIII

RESULTS OF RESNET50 MODEL WITH DIFFERENT σ VALUES (MULTI-CLASS CLASSIFICATION - OTTAWA DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

ResNet50: Effusion, Pneumothorax and No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.453	0.474	0.474	0.491	0.440
	Prec	0.438	0.456	0.473	0.487	0.436
	F1	0.436	0.440	0.470	0.488	0.437
	AUC	0.609	0.673	0.675	0.655	0.658
$\lambda = 0.5$	Acc / Rec	0.453	0.432	0.427	0.479	0.483
	Prec	0.448	0.428	0.426	0.463	0.348
	F1	0.450	0.426	0.425	0.465	0.479
	AUC	0.666	0.610	0.625	0.660	0.678
$\lambda = 0.75$	Acc / Rec	0.453	0.462	0.479	0.483	0.491
	Prec	0.451	0.468	0.469	0.477	0.485
	F1	0.452	0.458	0.472	0.478	0.480
	AUC	0.673	0.679	0.661	0.650	0.665
$\lambda = *$	Acc / Rec	0.517	0.462	0.432	0.457	**0.538**
	Prec	0.522	0.455	0.438	0.449	**0.537**
	F1	0.513	0.453	0.434	0.451	**0.536**
	AUC	0.645	0.634	0.631	0.655	**0.702**

TABLE XIX

RESULTS OF RESNET50 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

ResNet50: Effusion vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.790	0.799	0.784	0.813	**0.824**
	Prec	0.792	0.799	0.787	0.815	**0.824**
	F1	0.789	0.799	0.783	0.813	**0.824**
	AUC	0.859	0.875	0.852	0.878	0.884
	PPV	0.819	0.793	0.818	**0.840**	0.822
	Sensitivity	0.743	0.809	0.729	0.772	0.828
	Specificity	0.836	0.789	0.838	**0.853**	0.820
	FDR	0.181	0.207	0.182	**0.160**	0.178
	FOR	0.235	0.195	0.244	0.211	**0.173**
$\lambda = 0.5$	Acc / Rec	0.796	0.822	0.798	0.790	0.791
	Prec	0.796	0.822	0.798	0.793	0.792
	F1	0.796	0.822	0.798	0.790	0.791
	AUC	0.863	0.879	0.863	0.870	0.862
	PPV	0.810	0.829	0.812	0.823	0.809
	Sensitivity	0.774	0.813	0.775	0.740	0.761
	Specificity	0.818	0.832	0.820	0.841	0.820
	FDR	0.190	0.171	0.188	0.177	0.191
	FOR	0.217	0.184	0.215	0.237	0.226
$\lambda = 0.75$	Acc / Rec	0.786	0.821	0.807	0.796	0.796
	Prec	0.786	0.822	0.807	0.796	0.797
	F1	0.786	0.821	0.807	0.796	0.796
	AUC	0.859	**0.887**	0.873	0.858	0.861
	PPV	0.794	0.834	0.806	0.799	0.807
	Sensitivity	0.771	0.802	0.809	0.790	0.780
	Specificity	0.800	0.841	0.805	0.802	0.813
	FDR	0.206	0.166	0.194	0.201	0.193
	FOR	0.222	0.190	0.192	0.208	0.213
$\lambda = *$	Acc / Rec	0.803	0.806	0.777	0.794	0.802
	Prec	0.803	0.807	0.777	0.794	0.802
	F1	0.803	0.806	0.777	0.794	0.801
	AUC	0.860	0.866	0.843	0.865	0.870
	PPV	0.808	0.790	0.785	0.788	0.811
	Sensitivity	0.796	**0.833**	0.762	0.804	0.786
	Specificity	0.810	0.779	0.791	0.784	0.817
	FDR	0.192	0.210	0.215	0.212	0.189
	FOR	0.201	0.176	0.231	0.200	0.207

TABLE XX

RESULTS OF RESNET50 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK.

ResNet50: Pneumothorax vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.751	0.738	0.741	0.731	0.759
	Prec	0.751	0.739	0.742	0.733	**0.763**
	F1	0.751	0.738	0.741	0.731	0.758
	AUC	0.826	0.827	0.817	0.815	0.817
	PPV	0.743	0.725	0.732	0.711	0.729
	Sensitivity	0.765	0.768	0.761	0.779	**0.822**
	Specificity	0.736	0.708	0.722	0.683	0.695
	FDR	0.257	0.275	0.268	0.289	0.271
	FOR	0.242	0.247	0.249	0.244	**0.204**
$\lambda = 0.5$	Acc / Rec	0.733	0.720	0.749	0.756	0.739
	Prec	0.734	0.721	0.750	0.757	0.740
	F1	0.733	0.720	0.749	0.756	0.739
	AUC	0.803	0.784	0.825	0.831	0.809
	PPV	0.748	0.708	**0.758**	0.747	0.755
	Sensitivity	0.704	0.749	0.733	0.774	0.708
	Specificity	0.763	0.690	0.765	0.738	**0.770**
	FDR	0.252	0.292	**0.242**	0.253	0.245
	FOR	0.280	0.266	0.258	0.234	0.275
$\lambda = 0.75$	Acc / Rec	**0.760**	0.747	0.731	0.745	0.731
	Prec	0.760	0.747	0.732	0.745	0.731
	F1	**0.760**	0.747	0.731	0.745	0.731
	AUC	0.832	0.828	0.813	0.810	0.802
	PPV	0.750	0.740	0.719	0.736	0.725
	Sensitivity	0.779	0.763	0.759	0.763	0.745
	Specificity	0.740	0.731	0.704	0.727	0.718
	FDR	0.250	0.260	0.281	0.264	0.275
	FOR	0.230	0.245	0.255	0.246	0.262
$\lambda = *$	Acc / Rec	0.756	0.736	0.745	0.720	**0.760**
	Prec	0.758	0.736	0.747	0.720	0.760
	F1	0.756	0.736	0.744	0.720	**0.760**
	AUC	0.814	0.816	0.821	0.799	**0.842**
	PPV	0.735	0.733	0.723	0.722	0.749
	Sensitivity	0.802	0.743	0.793	0.715	0.781
	Specificity	0.711	0.729	0.697	0.724	0.738
	FDR	0.265	0.267	0.277	0.278	0.251
	FOR	0.218	0.261	0.229	0.282	0.229

TABLE XXI

RESULTS OF RESNET50 MODEL WITH DIFFERENT σ VALUES (BINARY CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK

ResNet50: Cardiomegaly vs No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.706	**0.725**	0.700	0.663	**0.725**
	Prec	0.707	**0.725**	0.700	0.668	**0.725**
	F1	0.706	**0.725**	0.700	0.660	**0.725**
	AUC	0.779	**0.805**	0.773	0.750	0.804
	PPV	0.721	0.725	0.704	0.698	0.733
	Sensitivity	0.674	**0.725**	0.688	0.573	0.706
	Specificity	0.739	0.725	0.711	0.752	0.743
	FDR	0.279	0.275	0.296	0.302	0.267
	FOR	0.306	**0.275**	0.305	0.362	0.283
$\lambda = 0.5$	Acc / Rec	0.667	0.704	0.683	0.697	0.693
	Prec	0.676	0.705	0.687	0.703	0.702
	F1	0.663	0.704	0.682	0.695	0.689
	AUC	0.764	0.768	0.744	0.781	0.777
	PPV	0.716	0.715	0.711	0.736	**0.744**
	Sensitivity	0.555	0.679	0.619	0.615	0.587
	Specificity	0.780	0.729	0.748	0.780	**0.798**
	FDR	0.284	0.285	0.289	0.264	**0.256**
	FOR	0.363	0.306	0.337	0.331	0.341
$\lambda = 0.75$	Acc / Rec	0.695	0.667	0.693	0.658	0.693
	Prec	0.702	0.674	0.698	0.671	0.695
	F1	0.692	0.664	0.691	0.652	0.692
	AUC	0.776	0.745	0.767	0.754	0.767
	PPV	0.740	0.706	0.731	0.717	0.719
	Sensitivity	0.601	0.573	0.610	0.523	0.633
	Specificity	0.789	0.761	0.775	0.794	0.752
	FDR	0.260	0.294	0.269	0.283	0.281
	FOR	0.336	0.359	0.335	0.375	0.328
$\lambda = *$	Acc / Rec	0.709	0.681	0.709	0.704	0.716
	Prec	0.710	0.684	0.712	0.705	0.716
	F1	0.708	0.680	0.707	0.704	0.715
	AUC	0.782	0.770	0.781	0.776	0.773
	PPV	0.726	0.705	0.741	0.715	0.726
	Sensitivity	0.670	0.624	0.642	0.679	0.693
	Specificity	0.748	0.739	0.775	0.729	0.739
	FDR	0.274	0.295	0.259	0.285	0.274
	FOR	0.306	0.337	0.316	0.306	0.294

TABLE XXII

RESULTS OF RESNET50 MODEL WITH DIFFERENT σ VALUES (MULTI-CLASS CLASSIFICATION - NIH DATASET). VALUES IN BOLD INDICATE THE BEST RESULTS WITHIN THE SAME MODEL FOR THIS TASK

ResNet50: Effusion, Pneumothorax and No Finding

λ	Metric	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
$\lambda = 0.25$	Acc / Rec	0.630	0.628	0.610	0.619	**0.641**
	Prec	0.630	0.627	0.612	0.617	**0.641**
	F1	0.630	0.627	0.611	0.616	**0.638**
	AUC	0.796	0.797	0.781	0.786	**0.804**
$\lambda = 0.5$	Acc / Rec	0.626	0.616	0.633	0.614	0.600
	Prec	0.624	0.615	0.631	0.613	0.599
	F1	0.624	0.615	0.630	0.614	0.597
	AUC	0.797	0.788	0.803	0.795	0.776
$\lambda = 0.75$	Acc / Rec	0.618	0.603	0.620	0.631	0.617
	Prec	0.618	0.603	0.619	0.631	0.617
	F1	0.618	0.603	0.618	0.631	0.615
	AUC	0.786	0.774	0.786	0.794	0.783
$\lambda = *$	Acc / Rec	0.624	0.623	0.630	0.614	0.604
	Prec	0.624	0.624	0.628	0.614	0.607
	F1	0.622	0.623	0.628	0.613	0.605
	AUC	0.797	0.801	0.795	0.782	0.775

REFERENCES

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3462–3471.

[2] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei, "Thoracic disease identification and localization with limited supervision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8290–8299.

[3] S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, and D. Comaniciu, "Learning to recognize abnormalities in chest x-rays with location-aware dense networks," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 23rd Iberoamerican Congress, CIARP 2018, Madrid, Spain, November 19-22, 2018, Proceedings 23*. Springer, 2019, pp. 757–765.

[4] C. Ma, H. Wang, and S. C. Hoi, "Multi-label thoracic disease image classification with cross-attention networks," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*. Springer, 2019, pp. 730–738.

[5] X. Ouyang, S. Karanam, Z. Wu, T. Chen, J. Huo, X. S. Zhou, Q. Wang, and J.-Z. Cheng, "Learning hierarchical attention for weakly-supervised chest x-ray abnormality localization and diagnosis," *IEEE transactions on medical imaging*, vol. 40, no. 10, pp. 2698–2710, 2020.

[6] S. Kabir, L. Farrokhvar, and A. Dabouei, "A weakly supervised approach for thoracic diseases detection," *Expert Systems with Applications*, vol. 213, p. 118942, 2023.

[7] M. Yang, H. Tanaka, and T. Ishida, "Performance improvement in multi-label thoracic abnormality classification of chest x-rays with noisy labels," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 1, pp. 181–189, 2023.