

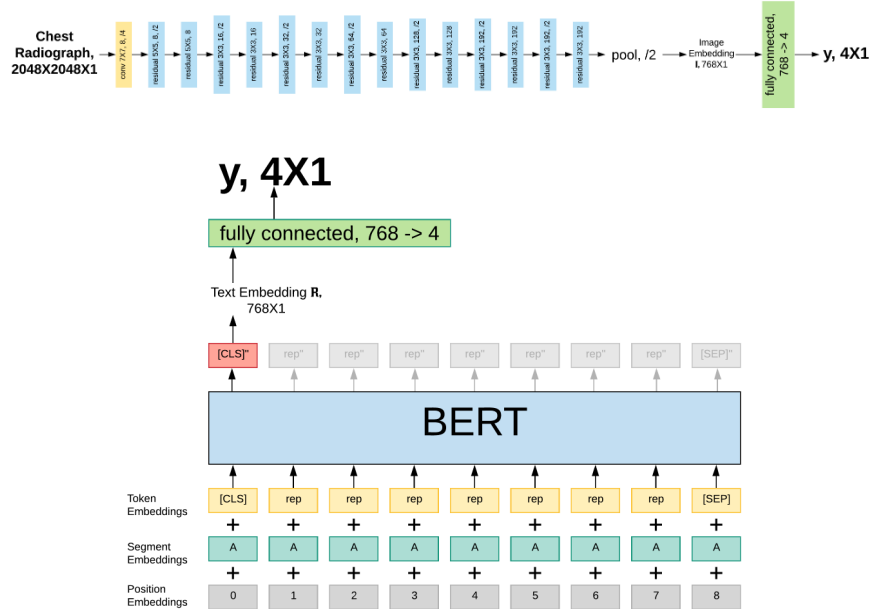
Supplementary Materials

Edema severity	Regex keyword terms	Number of reports	Accuracy
Overall	N/A	485	89.69%
Level 0 – none (n=216)	(no) pulmonary edema	222	88.74%
	(no) vascular congestion	43	100.00%
	(no) fluid overload	4	100.00%
	(no) acute cardiopulmonary process	115	98.27%
Level 1 – vascular congestion (n=98)	cephalization	17	94.12%
	pulmonary vascular congestion	96	98.96%
	hilar engorgement	3	100.00%
	vascular plethora	13	100.00%
	pulmonary vascular prominence	1	100.00%
Level 2 – interstitial edema (n=105)	pulmonary vascular engorgement	8	87.50%
	interstitial opacities	30	73.33%
	kerley	13	100.00%
	interstitial edema	92	94.57%
	interstitial thickening	6	66.67%
	interstitial pulmonary edema	21	100.00%
	interstitial marking	19	68.42%
	interstitial abnormality	10	70.00%
Level 3 – alveolar edema (n=66)	interstitial abnormalities	2	100.00%
	interstitial process	2	100.00%
	alveolar infiltrates	10	100.00%
	severe pulmonary edema	58	98.28%
	perihilar infiltrates	1	100.00%
	hilar infiltrates	1	100.00%
	parenchymal opacities	6	16.67%
	alveolar opacities	7	100.00%
	ill defined opacities	1	100.00%
ill-defined opacities	1	0.00%	
patchy opacities	10	10.00%	

Supplemental Table 1: Validation of regex keyword terms. The accuracy (positive predictive value) of the regular expression results for levels 0-3 based on the expert review results are 90.74%, 80.61%, 95.24%, and 90.91%, respectively. The total number of reports from all the keywords is more than 485 because some reports contain more than one keyword.

Hyperparameter	Setting
number-of-epochs (supervised)	50, 100, 150, 250
learning-rate	2e 5 , 5e-4, 1e-4, 1e-3
learning-rate-scheduler	warmup linear , reduce-on-plateau

Supplemental Table 2: Hyper-parameter search. Hyper-parameter settings were firstly experimented on the joint model in a supervised learning fashion. The experiments were performed on 5-fold cross validation within the training set, while holding out the test set. A learning rate of 2e-5 and the warmup-linear scheduler were chosen. Finally, the number of epochs was further experimented for the semi-supervised joint model learning with the 5-fold cross validation.



Supplemental Figure 1: Top: Image encoder and classifier architecture. Each residual block includes 2 convolutional layers. Bottom: Text encoder and classifier architecture using the BERT model. A full radiology report is encoded between [CLS] and [SEP] tokens; rep is the text associated with the report. Maximum input sequence length is set to 320.



Supplemental Figure 2: t-SNE visualization in 2 dimensions for image embeddings in the joint model (left) and the embeddings in the image-only model (right).