Explaining Chest X-ray Pathology Models using Textual Concepts

Using Foundation Models for Conceptual Counterfactual Explanations (CoCoX)

Vijay Sadashivaiah[1], Pingkun Yan[2], James A. Hendler[1]

[1]Department of Computer Science, [2]Department of Biomedical Engineering

Summary

Topic: Improving interpretability in medical imaging via concept-based explanations.

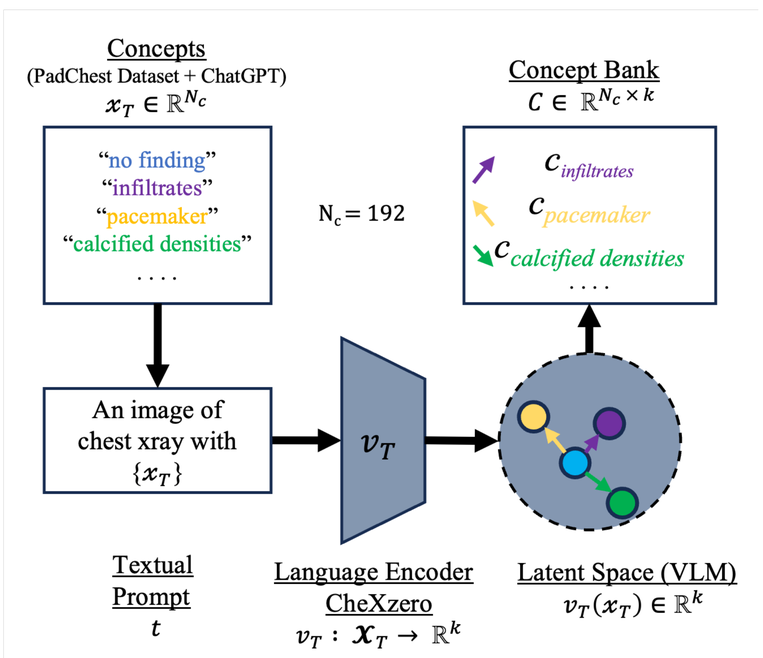
Problem: Existing methods require large, manually annotated datasets, which are scarce in the medical domain.

Solution: CoCoX leverages pre-trained vision-language models to explain black-box classifier outcomes without annotated data.

Application: Explaining cardiothoracic pathologies in chest X-rays.

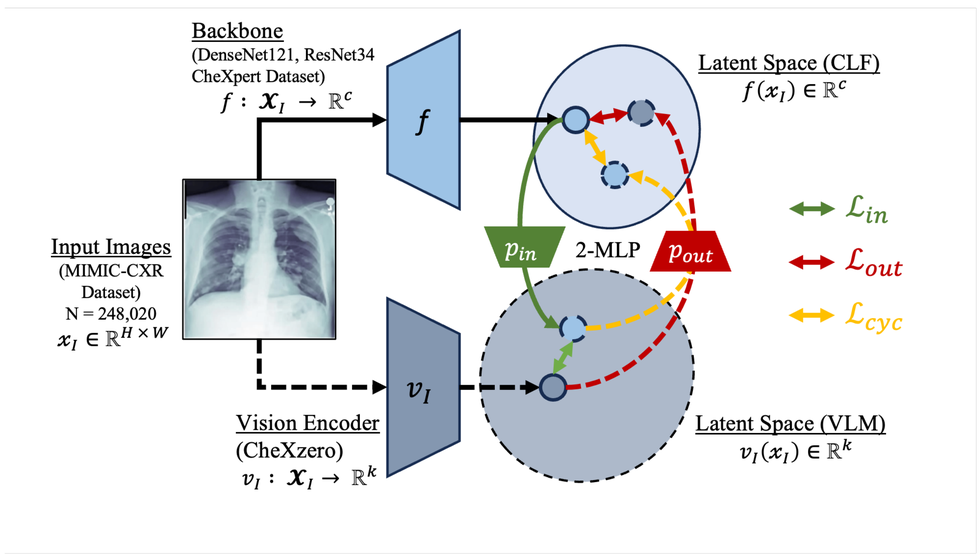
Methods

Step 1: Constructing Concept Bank for Chest X-ray Images



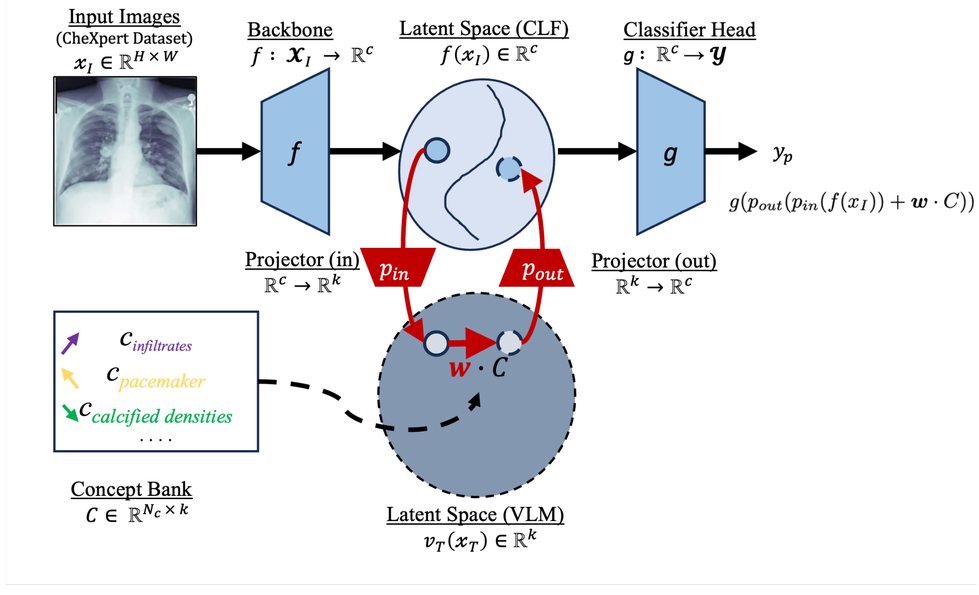
$$c = \frac{v_T(t_s) - v_T(t_n)}{||v_T(t_s) - v_T(t_n)||} \tag{1}$$

Step 2: Learning projection functions



$$\mathcal{L}_{total} = \mathcal{L}_{in} + \mathcal{L}_{out} + \mathcal{L}_{cyc} = ||p_{in}(f(x_I)) - v_I(x_I)|| + ||p_{out}(v_I(x_I)) - f(x_I)|| + ||p_{out}(p_{in}(f(x_I))) - f(x_I)|| \tag{2}$$

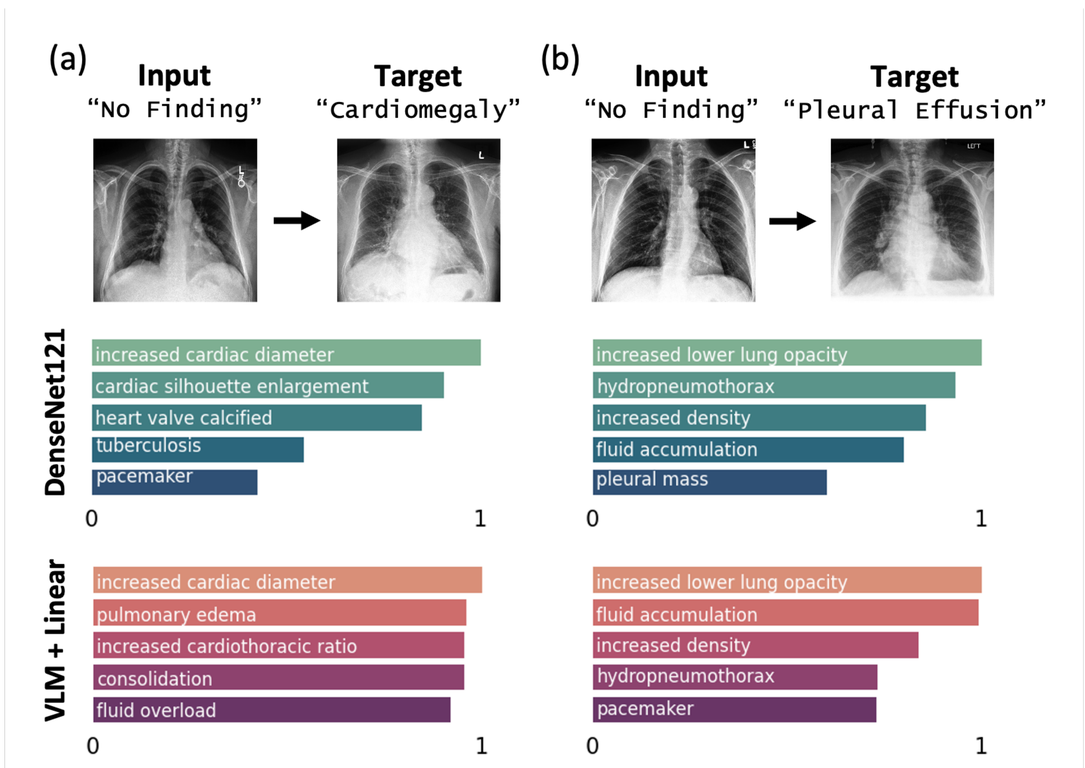
Step 3: Learning conceptual perturbations



$$\mathcal{L}_{final} = CE(y_p, y_t) + \alpha||\boldsymbol{w}||_1 + \beta||\boldsymbol{w}||_2, \tag{3}$$

Results

Conceptual counterfactuals generated by CoCoX



- Top 5 concepts changing y_p from "No Finding" to Target
- These concepts are medically relevant for underlying pathology

Comparing against Radiologists' evaluation

Table 1: Comparison of recall (R) at k concepts to radiologists' evaluation.

Pathology	Finding	DenseNet121		VLM + Linear		ResNet34	
		R@5	R@10	R@5	R@10	R@5	R@10
Cardiomegaly	*Primary(1)*	0.97	**1.00**	**1.00**	**1.00**	0.98	**1.00**
	Secondary(4)	0.35	0.48	0.50	0.50	0.32	**0.52**
Pleural Effusion	*Primary(2)*	0.46	0.78	0.50	**1.00**	0.43	0.81
	Secondary(3)	0.17	0.29	**0.33**	0.33	0.19	0.25
Atelectasis	*Primary(2)*	0.55	0.69	0.50	**0.85**	0.48	0.73
	Secondary(3)	0.26	0.61	0.33	**0.66**	0.21	0.59

Primary (P) and Secondary (S) medical concepts annotated by radiologists in CheXplaining in Style paper.
Cardiomegaly: (P) Increased cardiothoracic ratio, (S) Reduced lung tissue opacity, Pleural Effusion, Pacemaker, Older patients
Pleural Effusion: (P) Obstruction of the pleural recessus, Opaque lower lungs, (S) Increased cardiac diameter, Fluid overload, Pneumonia
Atelectasis: (P) Mediastinal shift, Wide barrel-like thorax, (S) Pleural Effusion, Infiltration, Older patients

- CoCoX successfully recalls primary concepts in each pathology
- Lower scores for secondary since there are more concepts to recall

Conclusions and Future Work

- Investigated conceptual counterfactual explanations to improve explainability in medical-image classifiers, focusing on chest X-rays.
- Developed a concept bank from radiology reports, using latent embedding manipulation for counterfactual generation.
- Plan to extend the method to other medical imaging domains and improve concept bank creation and evaluation metrics.

References

[1] M. Atad, V. Dmytrenko, Y. Li, X. Zhang, M. Keicher, J. Kirschke, B. Wiestler, A. Khakzar, and N. Navab. Chexplaining in style: Counterfactual explanations for chest x-rays using stylegan. *arXiv preprint arXiv:2207.07553*, 2022.

[2] A. Bustos, A. Pertusa, J.-M. Salinas, and M. De La Iglesia-Vaya. Padchest: A large chest x-ray image dataset with multi-label annotated reports. *Medical image analysis*, 66:101797, 2020.

[3] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597, 2019.

[4] A. E. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, R. G. Mark, and S. Horng. Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific data*, 6(1):317, 2019.

Acknowledgements

I would like to thank my advisors James A. Hendler, Pingkun Yan and thoughtful discussions with Dr. Mannudeep Kalra at Massachusetts General Hospital.