# Anatomical Structure-Guided Medical Vision-Language Pre-training

Qingqiu Li, Xiaohan Yan, Jilan Xu, Runtian Yuan, Yuejie Zhang, Rui Feng, Quanli Shen, Xiaobo Zhang, and Shujun Wang

fengrui@fudan.edu.cn; zhangxiaobo0307@163.com

**Table 1.** The comparison of the Region-Sentence pairs provided by Chest ImaGenome Dataset and our ARSA. The sentences highlighted in purple indicate that they cannot be solely inferred from the current region, while the blue ones indicates which smaller regions can align with the sentence. It is obvious that our ARSA can provide more accurate and concise alignments.

| Region | Sentence |
|---|---|
| **Chest ImaGenome Dataset** | |
| Right lung | There is no vascular engorgement or pulmonary edema. Right basilar atelectasis has worsened. There is no pneumothorax. No evidence of vascular engorgement or cardiac decompensation. Worsening right basal atelectasis. |
| Right lower lung zone | Right basilar atelectasis has worsened. Worsening right basal atelectasis. |
| Right hilar structures | Swans-Ganz catheter, ETT and esophageal drainage tubes are unchanged in standard placements. There is no vascular engorgement or pulmonary edema. No evidence of vascular engorgement or cardiac decompensation. |
| Left lung | Left retrocardiac opacification minimally changed and probably reflects a combination of pleural effusion and atelectasis. There is no vascular engorgement or pulmonary edema. There is no pneumothorax. No evidence of vascular engorgement or cardiac decompensation. Unchanged left basal atelectasis. |
| Left lower lung zone | Left retrocardiac opacification minimally changed and probably reflects a combination of pleural effusion and atelectasis. Unchanged left basal atelectasis. |
| Left hilar structures | Swans-Ganz catheter, ETT and esophageal drainage tubes are unchanged in standard placements. There is no vascular engorgement or pulmonary edema. No evidence of vascular engorgement or cardiac decompensation. |
| Left costophrenic angle | Left retrocardiac opacification minimally changed and probably reflects a combination of pleural effusion and atelectasis. Swans-Ganz catheter, ETT and esophageal drainage tubes are unchanged in standard placements. |

| | | |
|---|---|---|
| Right atrium | | Left cardiac pacing/defibrillator device with transvenous right atrial, right ventricular and coronary sinus/left ventricular leads in unchanged position. |
| **Ours** | | |
| Cardiac silhouette | | no evidence of vascular engorgement or cardiac decompensation. left retrocardiac opacification minimally changed and probably reflects a combination of pleural effusion and atelectasis. |
| Right lower lung zone | | worsening right basal atelectasis. right basilar atelectasis has worsened. |
| Left lower lung zone | | unchanged left basal atelectasis. |
| Right atrium | | left cardiac pacing defibrillator device with transvenous right atrial right ventricular and coronary sinus left ventricular leads in unchanged position. |
| Merge Bbox | Lung | there is no vascular engorgement or pulmonary edema, there is no pneumothorax. |
| Split Sentence | Right lung | there is no vascular engorgement or right pulmonary edema. there is no pneumothorax of right lung. |
| | Left lung | there is no vascular engorgement or left pulmonary edema. there is no pneumothorax of left lung. |

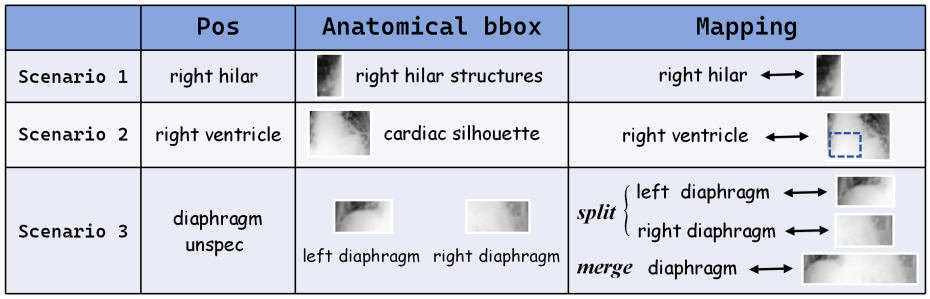


**Fig. 1.** Three scenarios of anatomical region-sentence alignment.

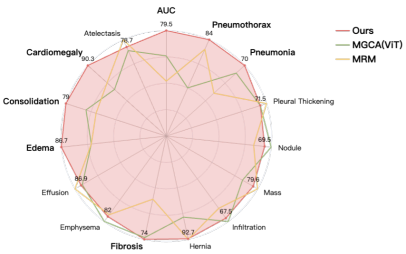


**Fig. 2.** Radar chart of NIH X-ray 14-class classification results.

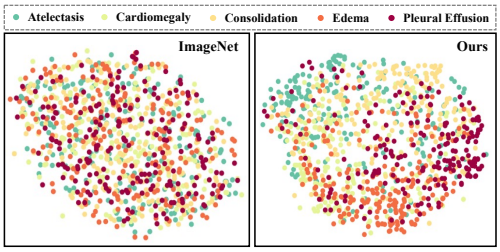


**Fig. 3.** t-SNE visualizations of encoded image representations on the MIMIC-5x200.