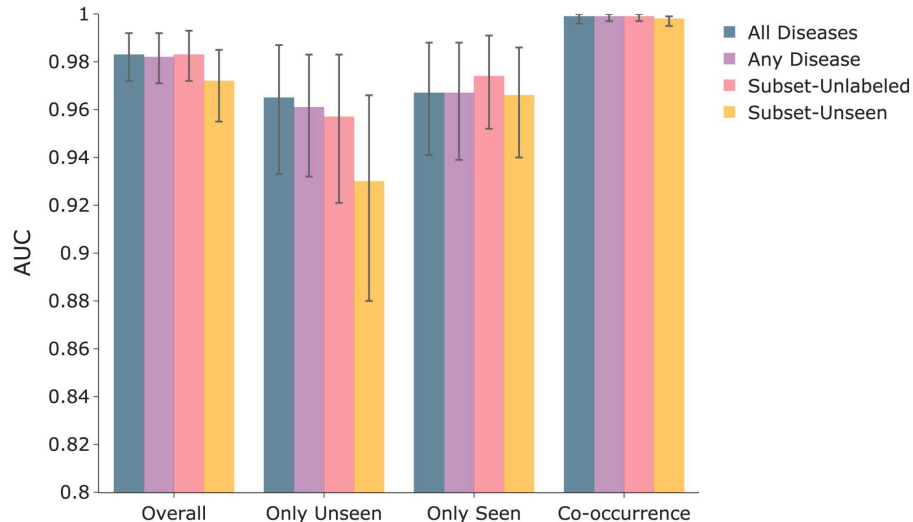


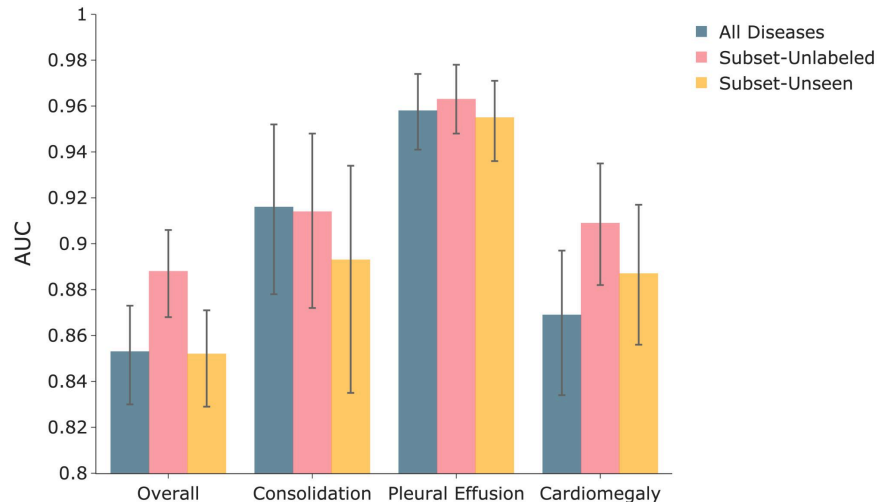
- A. Chest X-ray image labels included “No Disease”, three seen diseases, and six unseen diseases.
- B. Training data setup for the four multi-label models. The Subset-Unlabeled model was trained with all images in the train set, with the labels of “No Disease” and three seen diseases, while excluding the labels of six unseen diseases. The Subset-Unseen model was trained with only the images that do not have any of the six unseen diseases. The All Diseases model was trained with all images and all ten labels (“seen” diseases, “unseen” diseases and “no disease”), and served as a control. The Any Disease model was trained with all images for either having “any disease” or “no disease”, and served as another control.
- C. Three outputs from the models: final prediction layer, penultimate (intermediate) layer, and visualization map were used to train unseen disease classifiers, to predict the “unseen score” (whether an unseen disease was present during testing).

Can a model detect “no disease” and seen diseases in the presence of unseen diseases?

Models tend to falsely classify diseases outside of the subset (unseen diseases) as “no disease”.



Models are still able to detect seen diseases even when co-occurring with unseen diseases at a level comparable to the All Diseases model.



Can feature representations learned by models be used to detect the presence of unseen diseases given a small labeled set of unseen diseases?

- Classifiers trained with the penultimate layer output performed the best, followed by the final layer and the visualization map.
- The unseen scores from the Subset-Unlabeled model has higher performance than those from the Subset-Unseen model, likely because the former learns representations of the unlabeled diseases during training.

