

# DisCeRn: Disease-Contrastive Representations From Multi-Modal Medical Data

Vahid Balazadeh, Tom Ginsberg,  
Korbinian Koch, Vasudev Sharma

December 8<sup>th</sup> 2021



UNIVERSITY OF  
TORONTO

# Google Research changes the game for medical imaging with self-supervised learning



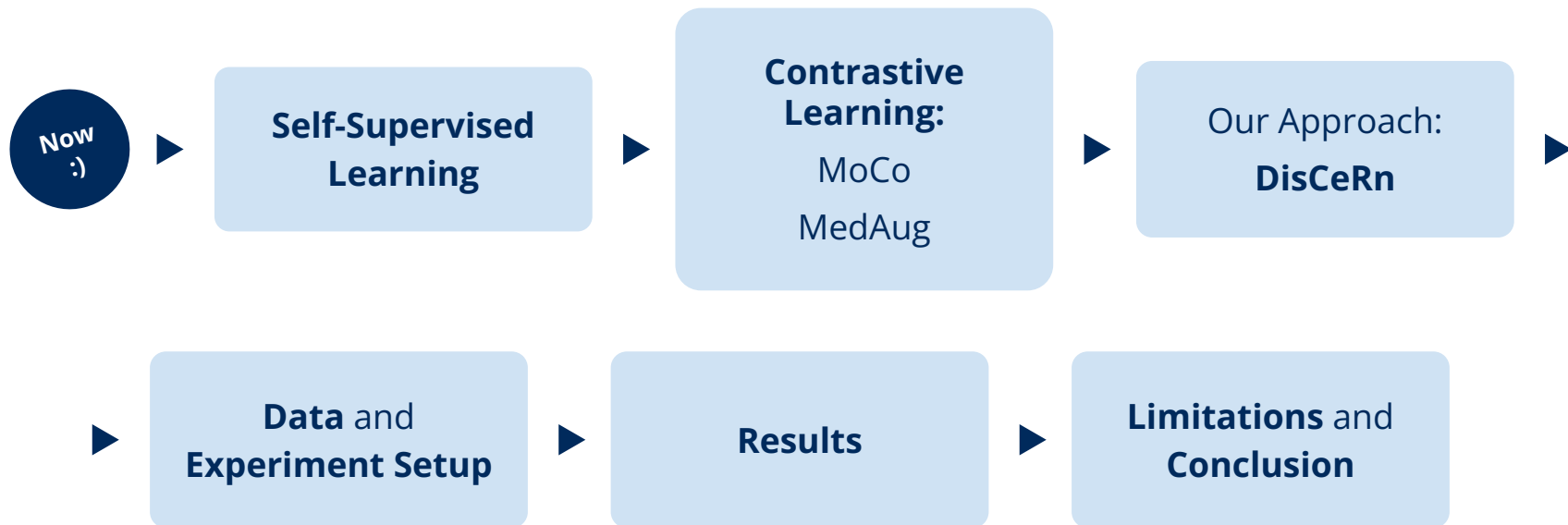
*Image Credit: MAD.vertise/Shutterstock*

# Goal: Predicting Disease from X-Ray Images

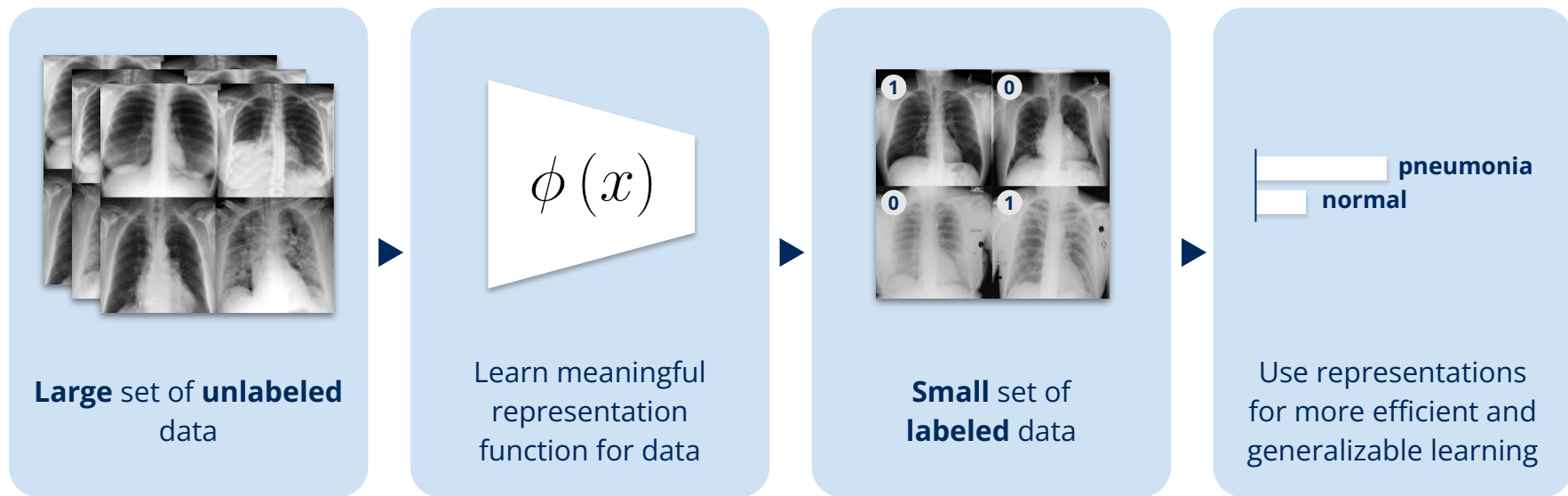


 pneumonia  
normal

# Outline



# Self Supervised Representation Learning (SSRL)



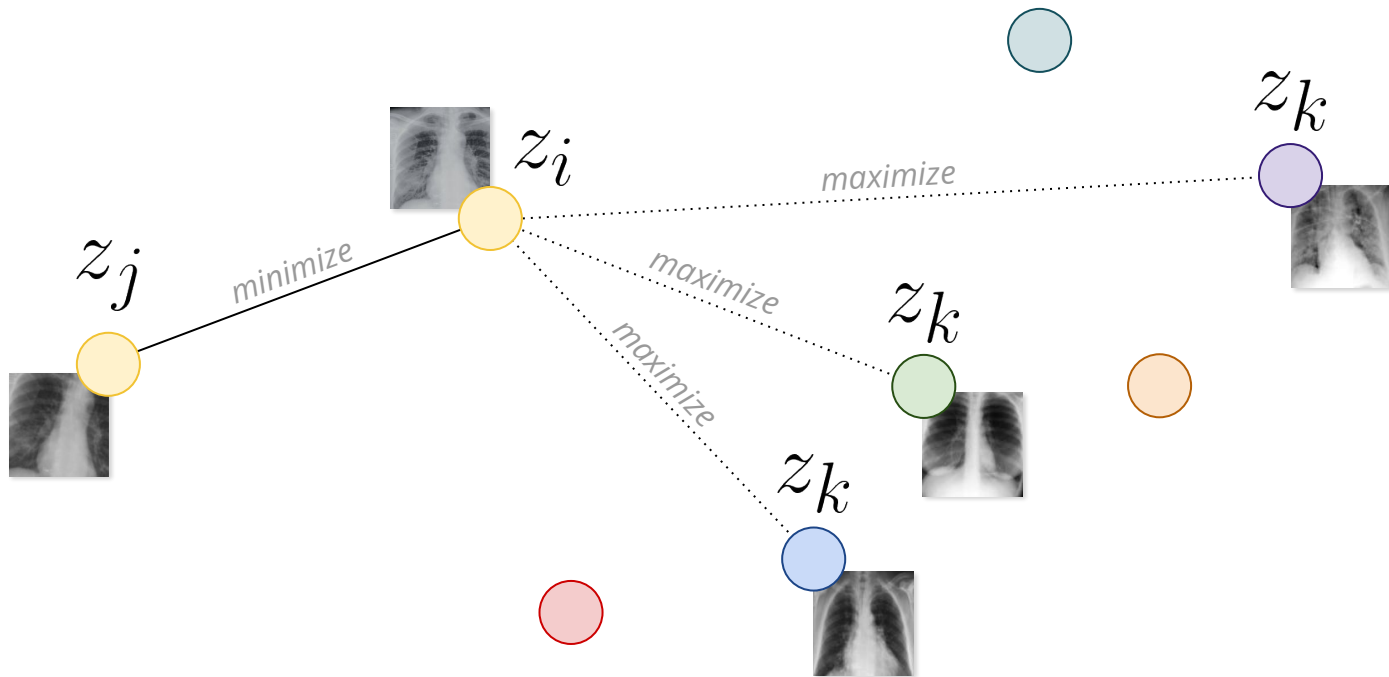
# SSRL Medical Images

Medical Datasets provide several unique characteristics that may be useful for SSL

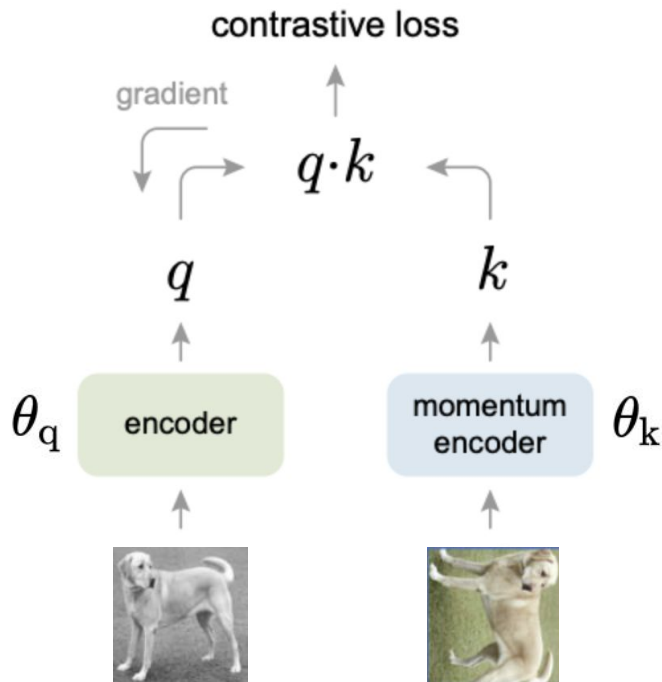
- Multiple images from various angles may be recorded for a single patient
  - *Can help learn more robust self supervised representations*
- Images are paired with useful metadata (age, heart-rate, blood pressure, lab tests)
  - **Available at test time** (*can be a model input*)
- Datasets much as MIMIC-CXR contain labels for various diseases and clinical findings
  - **Not available at test time** (*cannot be a model input*)
  - *Still adds information for learning representations while training*

# Contrastive Learning

$$l_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j))}{\sum_{k \neq i} \exp(\text{sim}(z_i, z_k))}$$



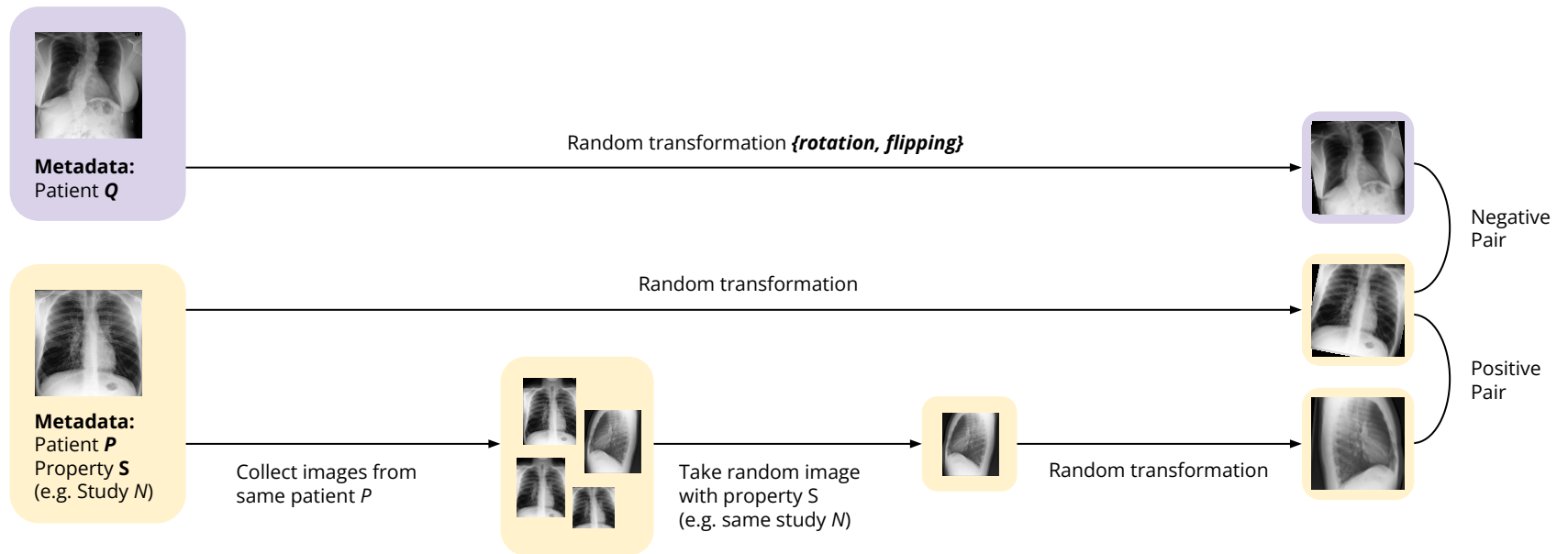
# Related Work: MOCO [He et al. 2020]



$$\theta_k \leftarrow m\theta_k + (1 - m)\theta_q$$

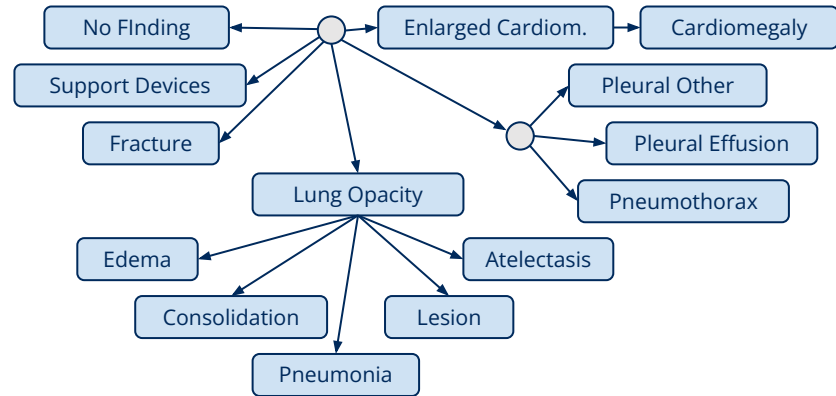


# Related Work: MedAug [Yen et al. 2021]

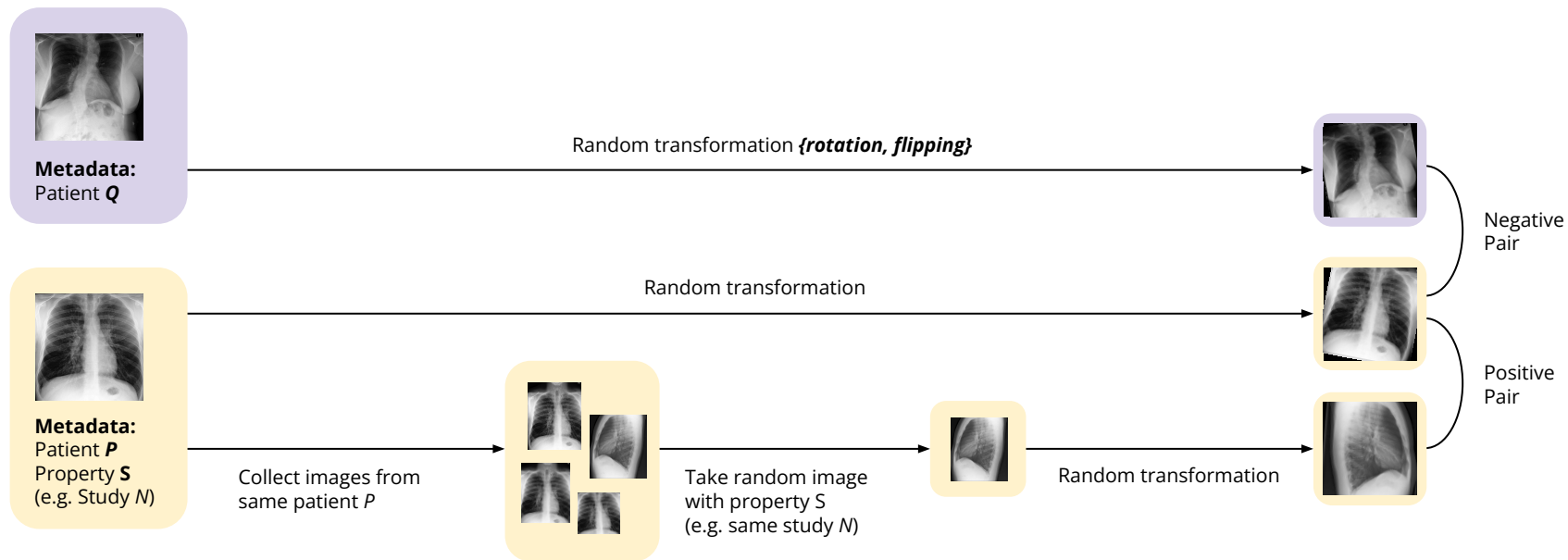


# MedAug → DisCeRn (Disease-Contrastive Representations)

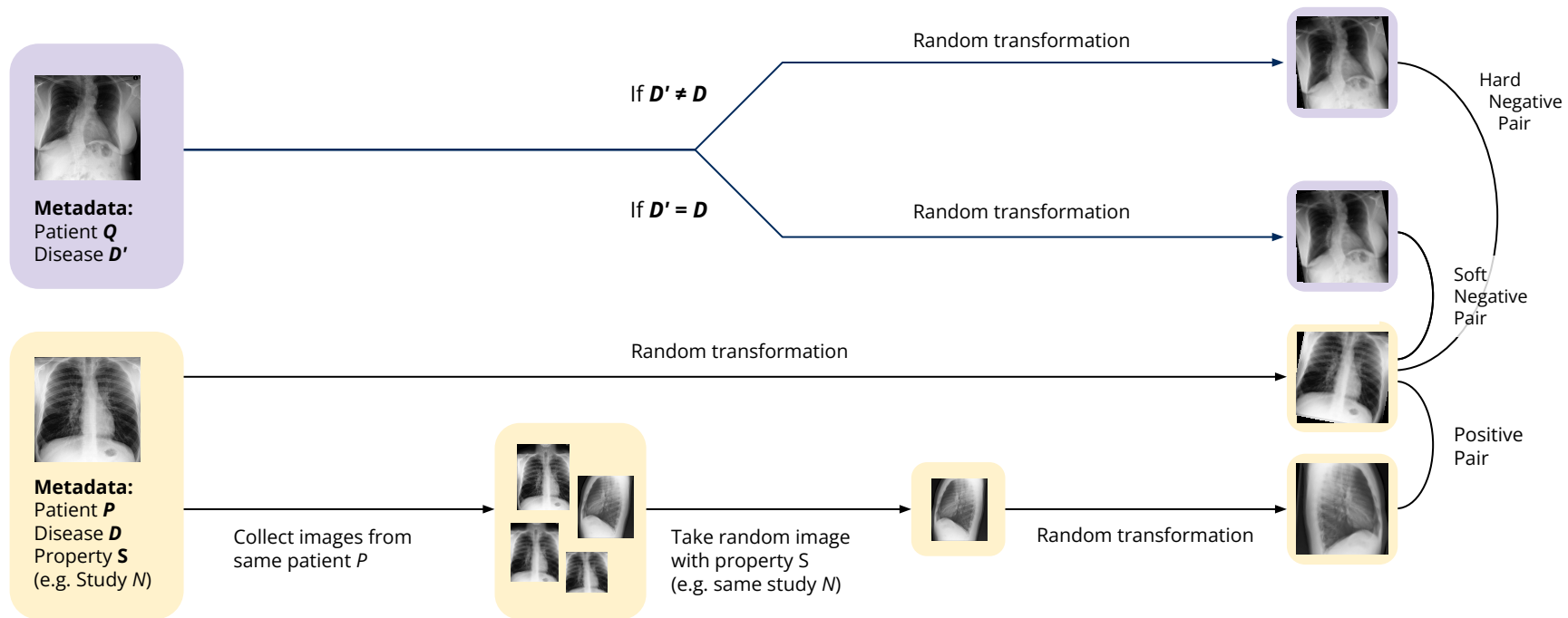
- *MedAug/MoCo* treat all negative samples (= different patients) equal
- This means they don't express any preference on how the samples should be organized in the representation space
- Does it help if we maximize the distance to some negative samples stronger than others?
- This blurs the lines between self-supervision and finetuning, but may help even for unrelated tasks



# Related Work: MedAug [Yen et al. 2021]



# Our Approach: DisCeRn



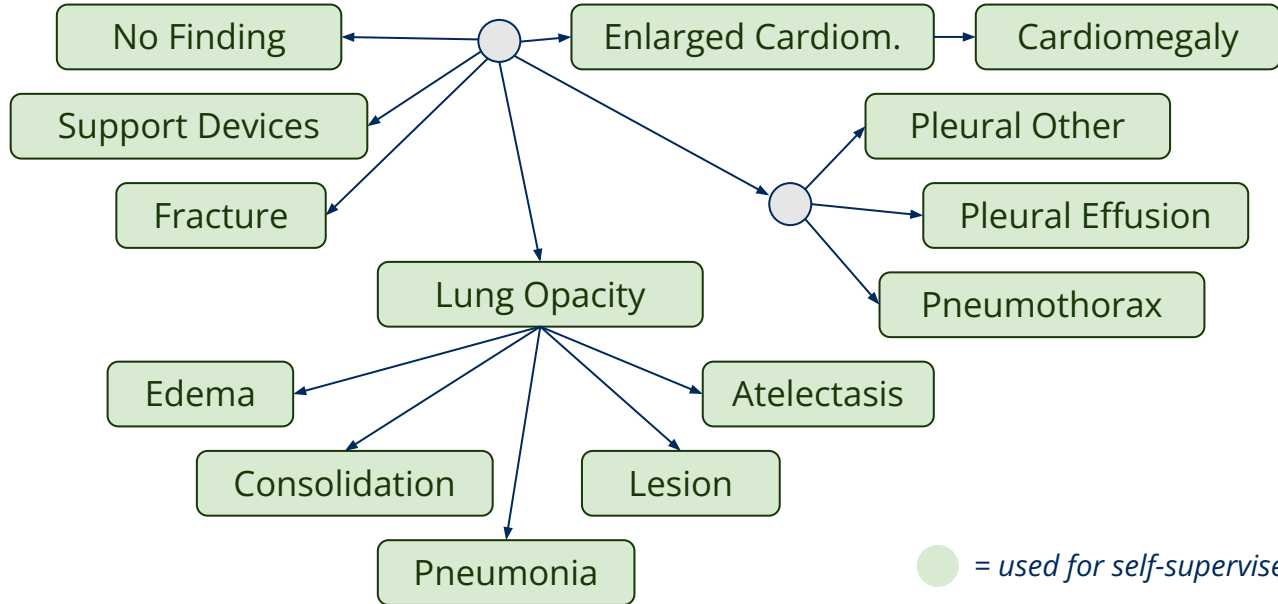
# MedAug → DisCeRn (Our Approach)

There are more diseases that can be classified using chest x-rays than our 14:

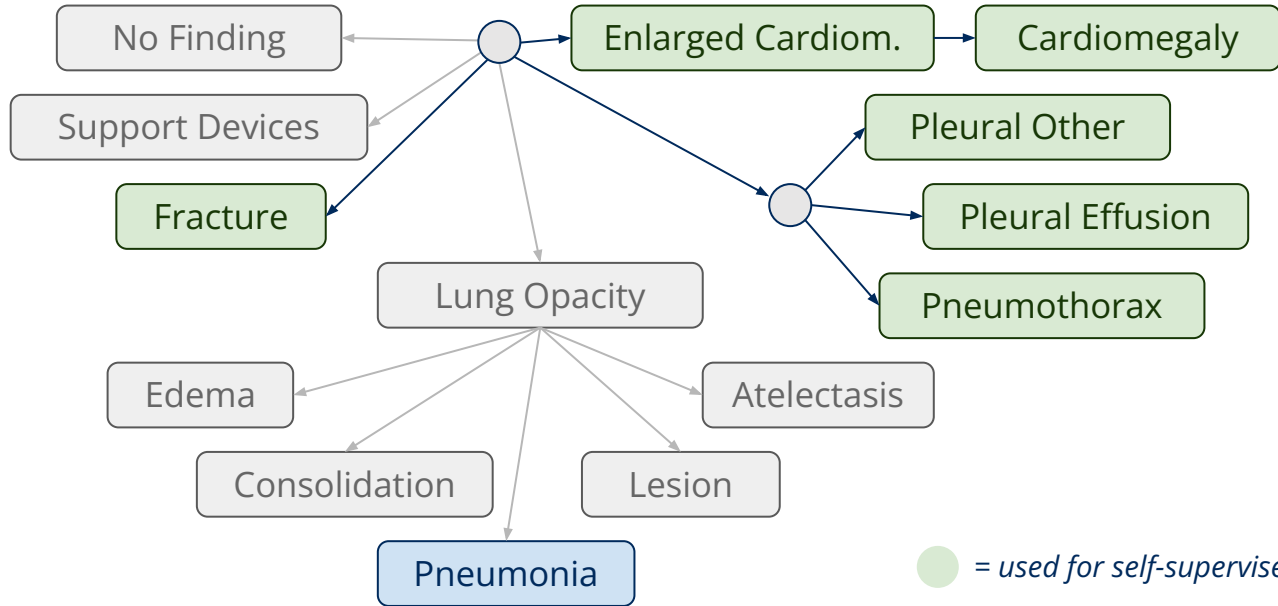
- *COVID-19*
- *tumors*
- *aortic aneurysm*
- *lung emphysema*
- *diaphragmatic hernia*

⇒ **Potential problem: does the promoted cluster organisation obscure useful information for the classification of other diseases?**

# MedAug → DisCeRn (Our Approach)



# MedAug → DisCeRn (Our Approach)



*used for supervised evaluation*

# MedAug → DisCeRn (Our Approach)

- At inference time, we have more information available than just the X-ray image
- Does it help for downstream tasks if we append information such as O2 saturation or lab values to the image representation or are these sufficiently captured by the pretrained model already?

 $\phi(x)$ **[0.234, -0.857, ..., 0.462, 0.947]**

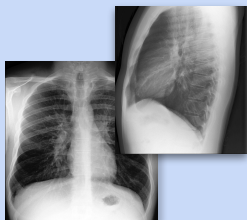
vs.

 $\phi(x)$ **[0.234, -0.857, ..., 0.462, 0.947]** +

	O2 Sat	resprate	Lab		
P1	94	17.5	33	15.5	4.03



# Datasets



## Metadata:

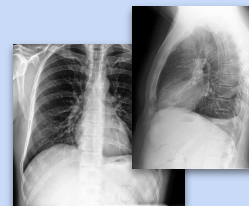
Patient **P1**  
Disease **D1**  
Study **S1**

**MIMIC-CXR-JPG**

*377,110 images*  
*65,379 patients*

	O2 Sat	resprate	Lab	
P1	94	17.5	33	15.5
P2	98	16	35	12

**MIMIC-IV**



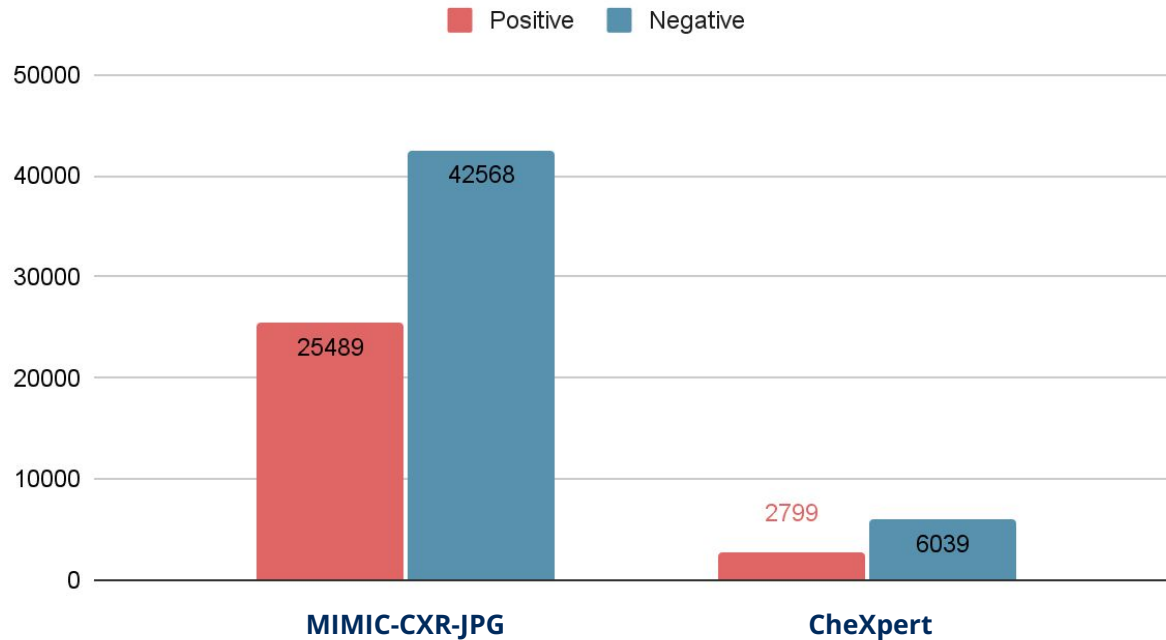
## Metadata:

Patient **P1**  
Disease **D1**

**CheXpert**

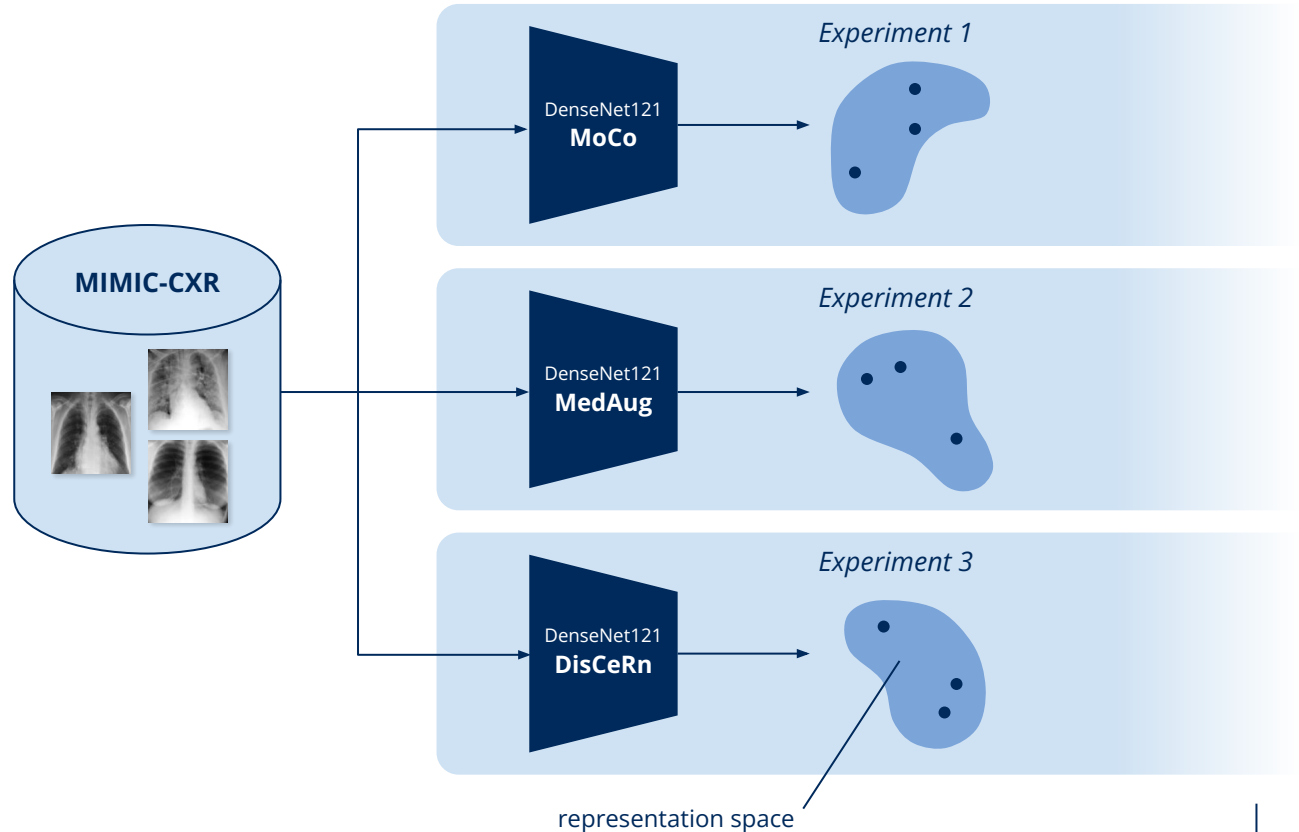
*224,316 images*  
*65,240 patients*

# Subsets Used for Supervised Training (Only Labels 1 and 0 for Pneumonia)



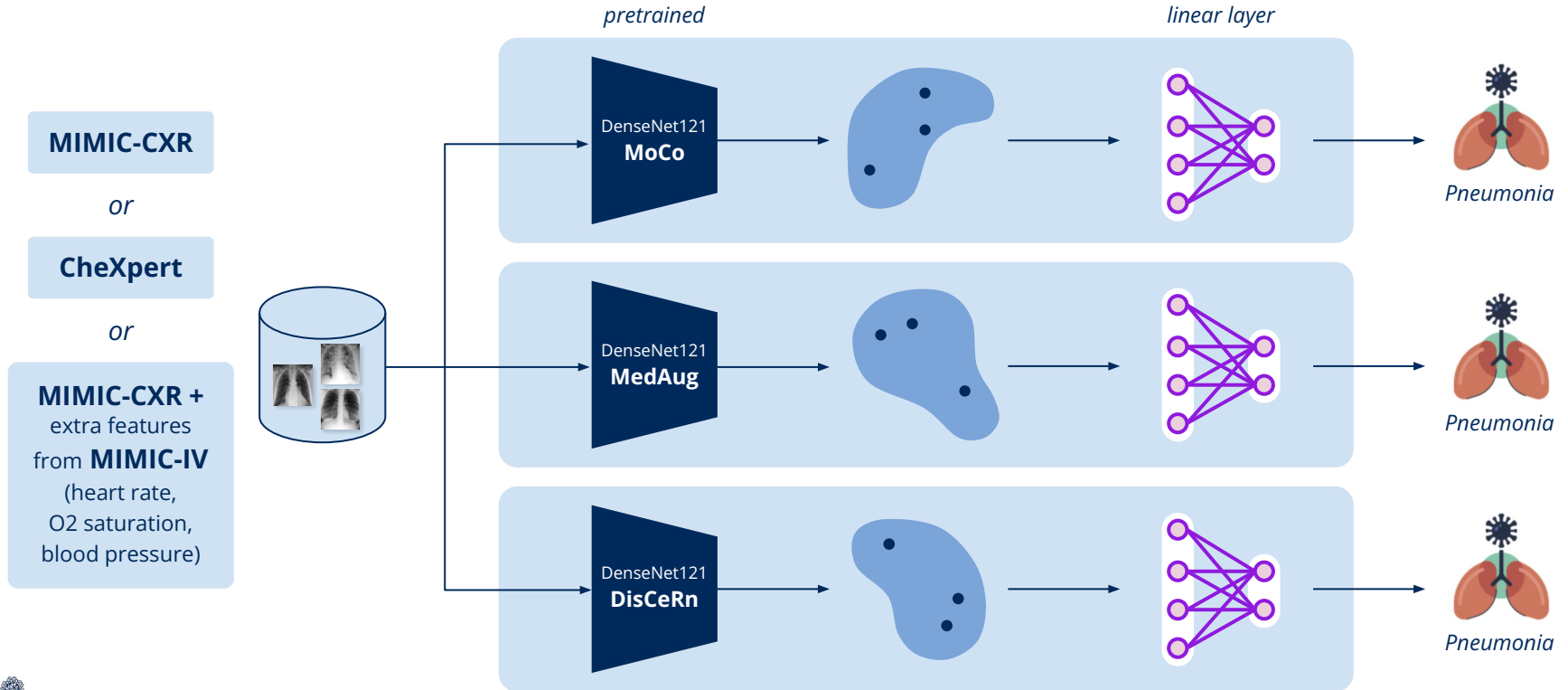
# Experimental Setup: Pre-training phase

- Pretraining:  
(per experiment)  
**2 GPU weeks** on  
RTX A600 48GB
- Dataset:  
MIMIC-CXR (**337K  
images**)
- Augmentations:  
**Rotate, Crop, Flip**



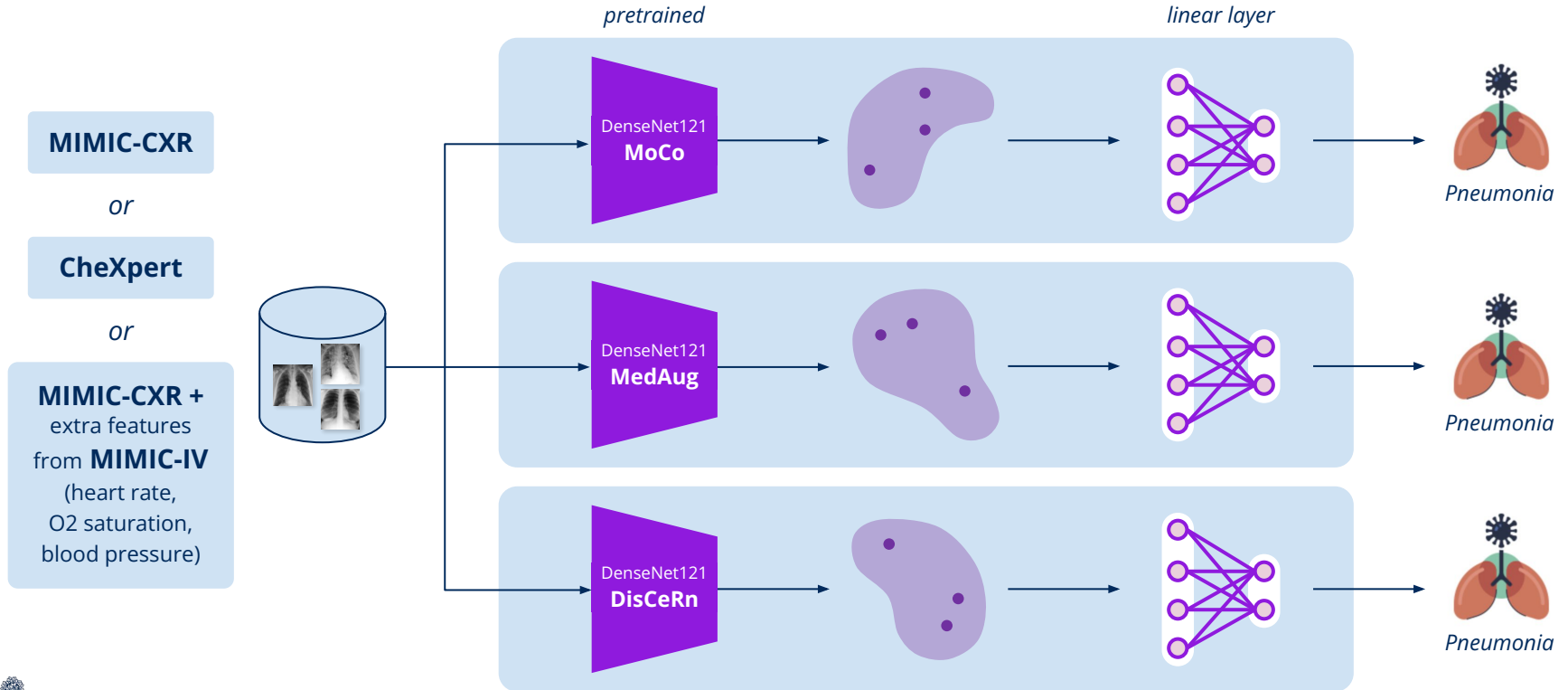
# Experimental Setup: Fine-tuning phase (linear)

● = trainable



# Experimental Setup: Fine-tuning phase (end-to-end)

● = trainable



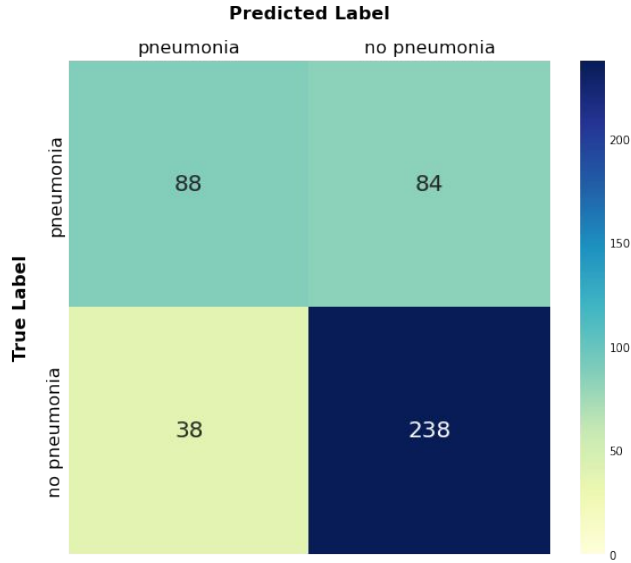
# Expected Results

- MoCo < MedAug < DisCeRn
- DisCeRn < DisCeRn on MIMIC-CXR + MIMIC-IV
- Linear < End-End
- All perform better than ImageNet baseline

# Experimental Results (AUROC)

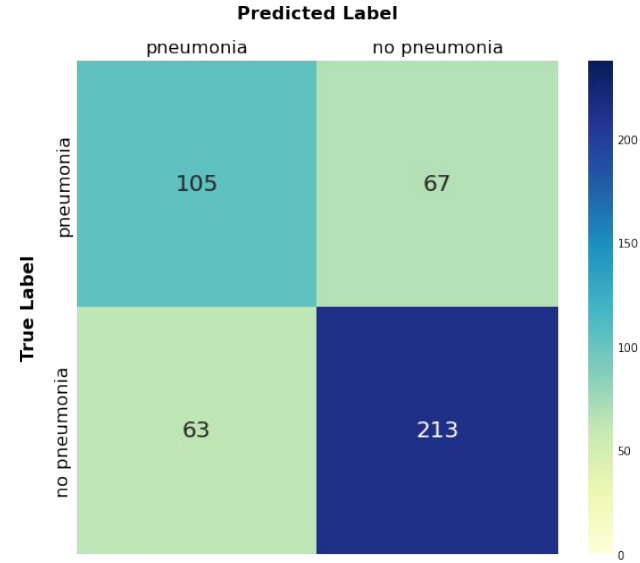
Method	CheXpert <i>Train: 8838   Test: 234</i>	MIMIC-CXR <i>Train: 68057   Test: 495</i>	MIMIC-CXR + MIMIC-IV <i>Train: 68057   Test: 495</i>
MoCo (Linear)	84.98%	79.14%	71.82%
MoCo (End-to-End)	88.42%	79.63%	73.36%
MedAug (Linear)	88.11%	78.74%	73.99%
MedAug (End-to-End)	85.10%	<b>80.88%</b>	<b>78.00%</b>
DisCeRn (Linear)	86.25%	78.64%	72.76%
<b>DisCeRn (End-to-End)</b>	<b>89.03%</b>	<b>80.51%</b>	75.59%
Baseline (ImageNet Init / End-to-End)	87.57%	76.91%	68.48%

# Experimental Results (MIMIC-CXR)



**MedAug**

F1-score: 0.69



**DisCeRn + MIMIC-IV**

F1-score: 0.69



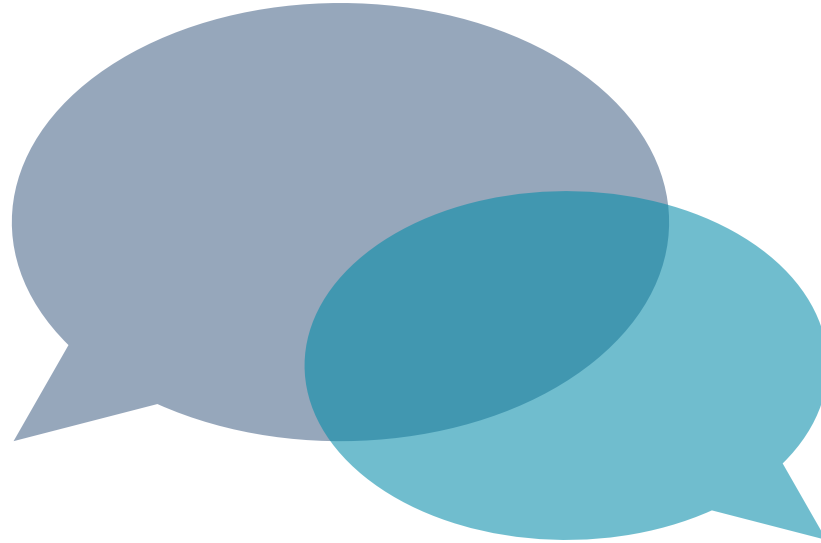
# Limitations

- Adding features from MIMIC-IV severely punishes performance  
⇒ *Better data fusion methods besides concatenation should be explored*
- While extended medaug appears to outperform SOTA, small test set sizes provide low statistical confidence on this improvement  
⇒ *Further statistical testing (i.e random permutation test) is required*
- Moco is not required to produce embeddings that are linearly separable  
⇒ *More downstream layers may be required for classification*
- Minimal hyperparameter tuning was performed (architecture, optimizer, scheduler, ...)
- No exploration of how test prediction quality changes with smaller # of training samples
- Lack of comprehensive evaluations for different pathology labels

# Conclusions

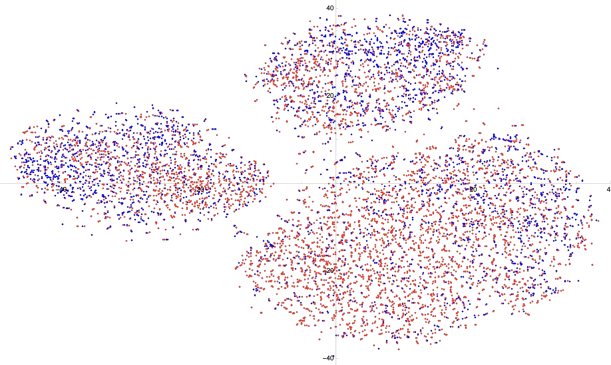
- Using domain knowledge related to medical imaging shows promise to improve self supervised learning (SSL)
- **Labels for diseases marginally related to a target disease can be helpful for improving SSL**
  - ⇒ *i.e a label for Pleural effusion might be useful while pre training to learn Pneumonia*

# QUESTIONS

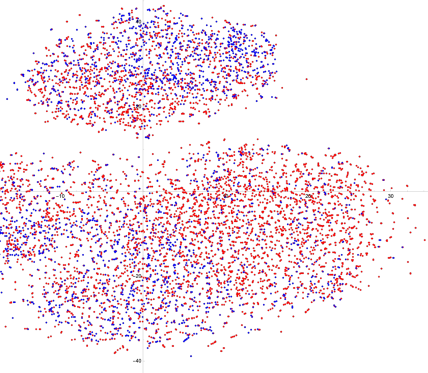


# Experimental Results

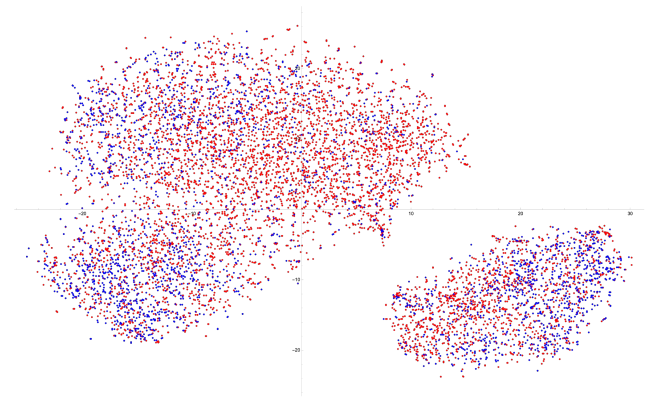
- Pneumonia
- No Pneumonia



**Moco**



**Medaug**



**Extended  
Medaug**