Multimodal Representation Learning for Medical Image Analysis

Ruizhi "Ray" Liao

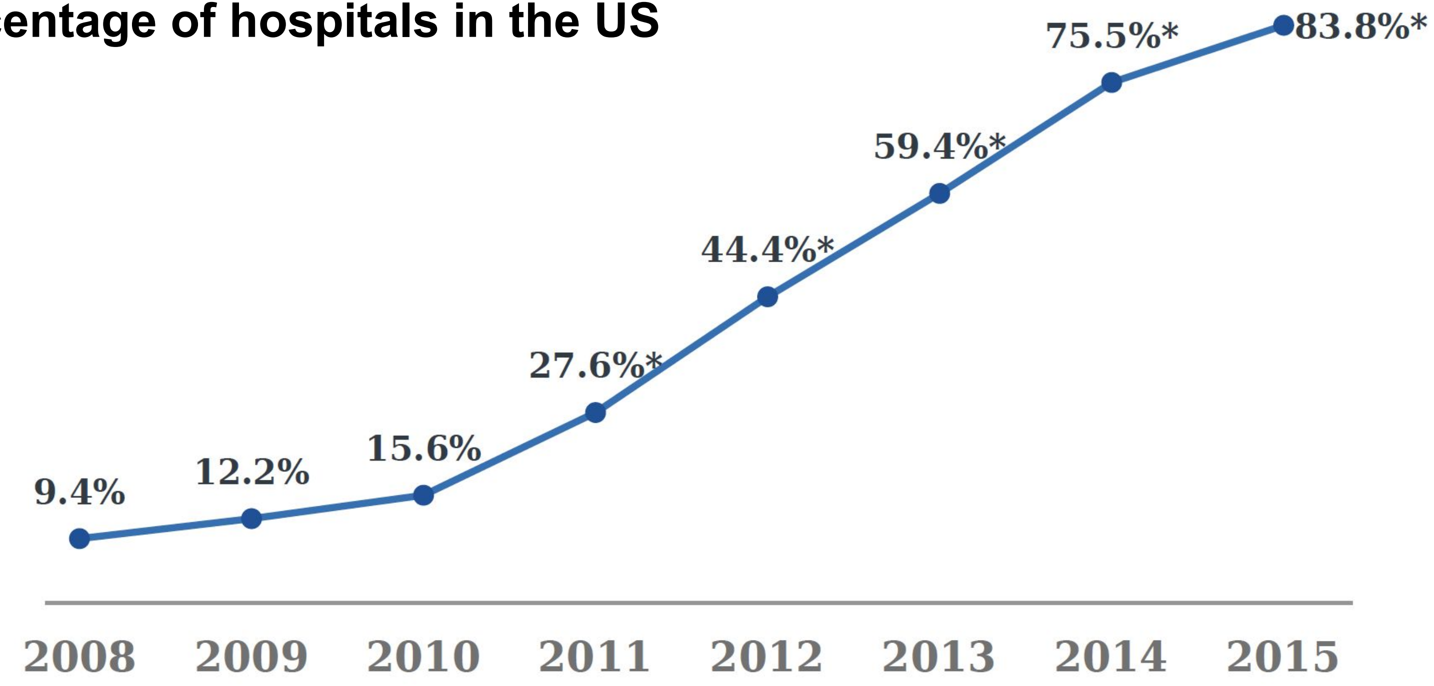
ruizhi@mit.edu

@rayruizhiliao

CSAIL

Adoption of electronic health records (EHR) has increased 9x in the US since 2008

Percentage of hospitals in the US

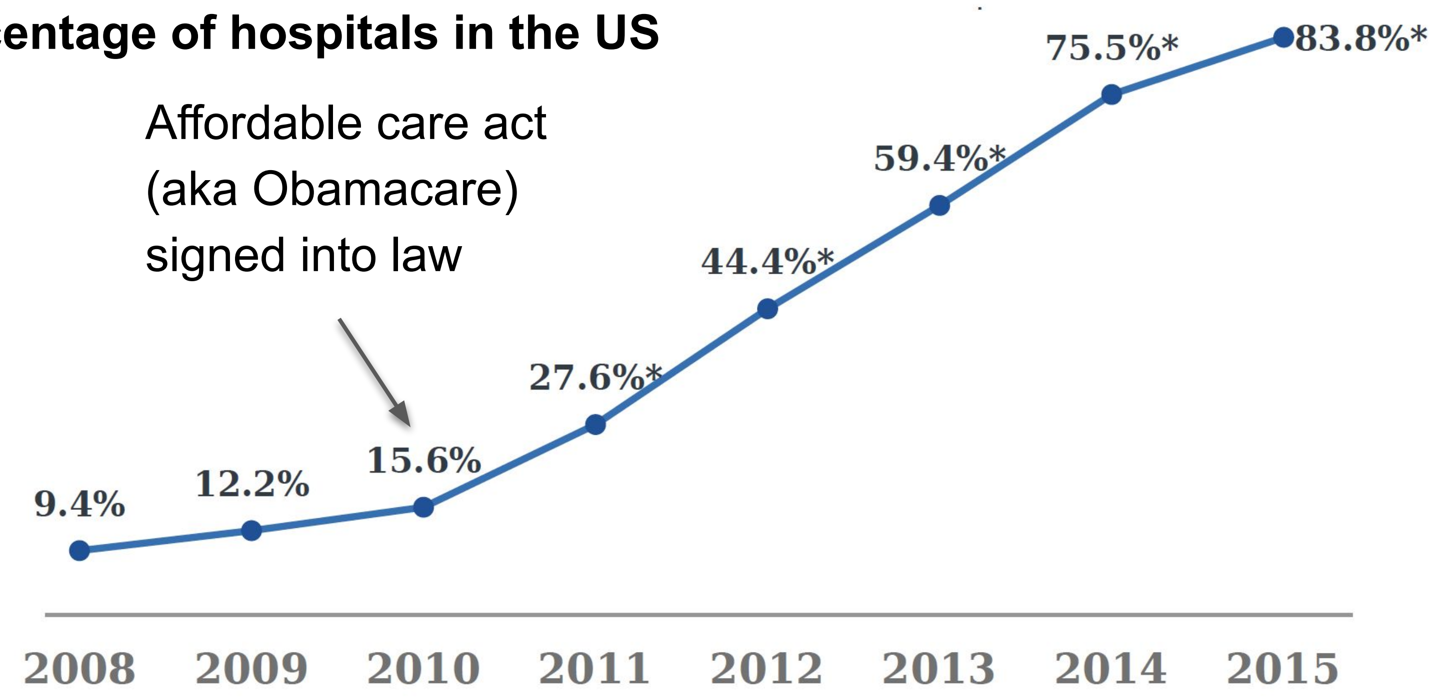


[Henry et al., ONC Data Brief, May 2016]

*Significantly different from previous year (p<0.05).

Adoption of electronic health records (EHR) has increased 9x in the US since 2008

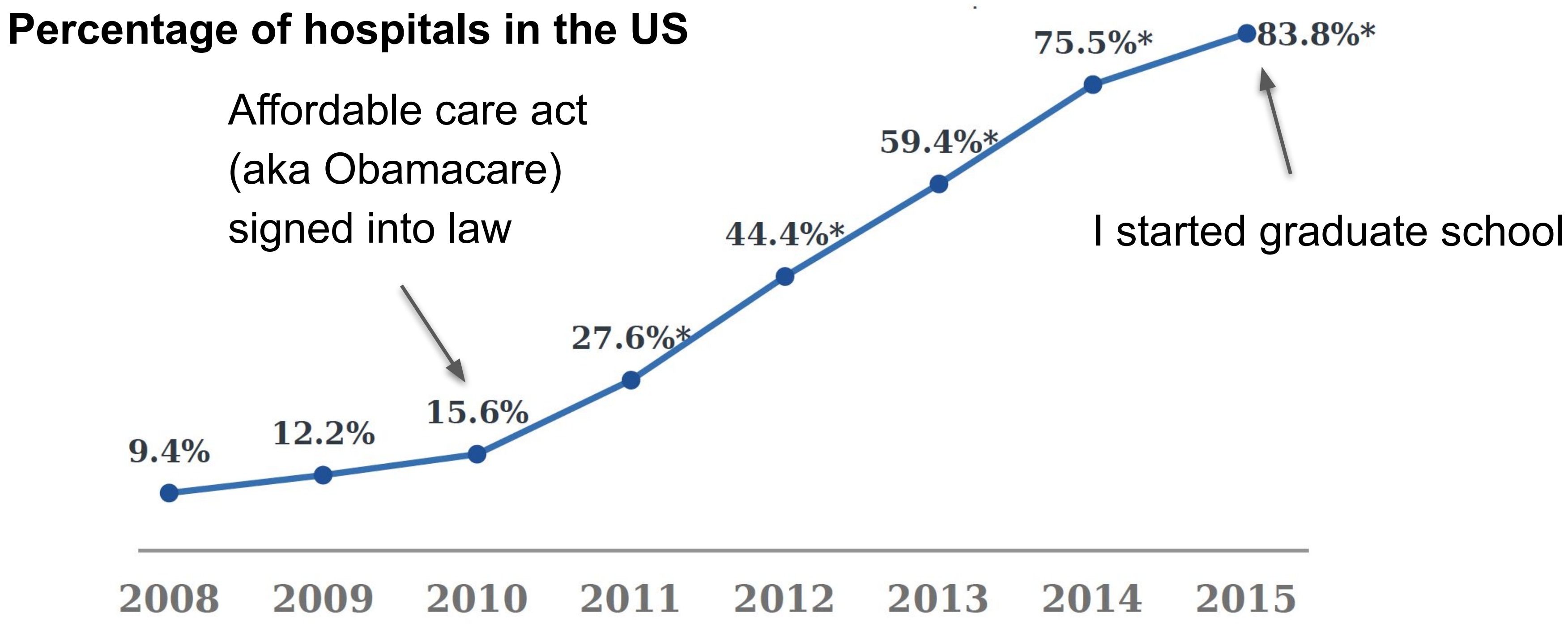
Percentage of hospitals in the US



[Henry et al., ONC Data Brief, May 2016]

*Significantly different from previous year (p<0.05).

Adoption of electronic health records (EHR) has increased 9x in the US since 2008



[Henry et al., ONC Data Brief, May 2016]

*Significantly different from previous year (p<0.05).

Clinical data is multimodal

Images x^{I}

```
                              FINAL REPORT
EXAMINATION:  CHEST (PORTABLE AP)

INDICATION:  ___ year old man with respiratory failure, ARDS  // Volume
overload?

TECHNIQUE:  Single frontal view of the chest

COMPARISON:  ___

IMPRESSION:

Moderate left pleural effusion decreased.  Large right pleural effusion is
probably unchanged.  Tracheostomy tube is in unchanged position.  Extensive
bilateral alveolar opacities have improved, consistent with improve severe
pulmonary edema.  Cardiac size is obscured by the pleural parenchymal
abnormalities
```

Text x^{R}

Numerical signals x^{N}

...

Multimodal clinical data reflect different yet correlated manifestations of a subject’s underlying physiological processes

FINAL REPORT

EXAMINATION: CHEST (PORTABLE AP)

INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

IMPRESSION:

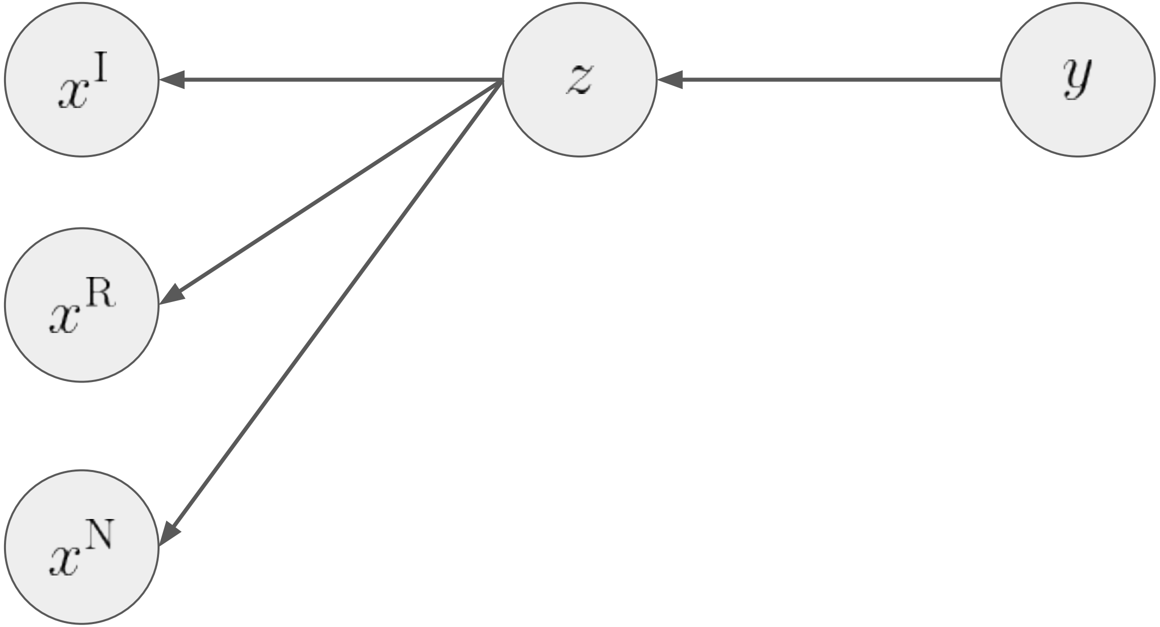
Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities

Images

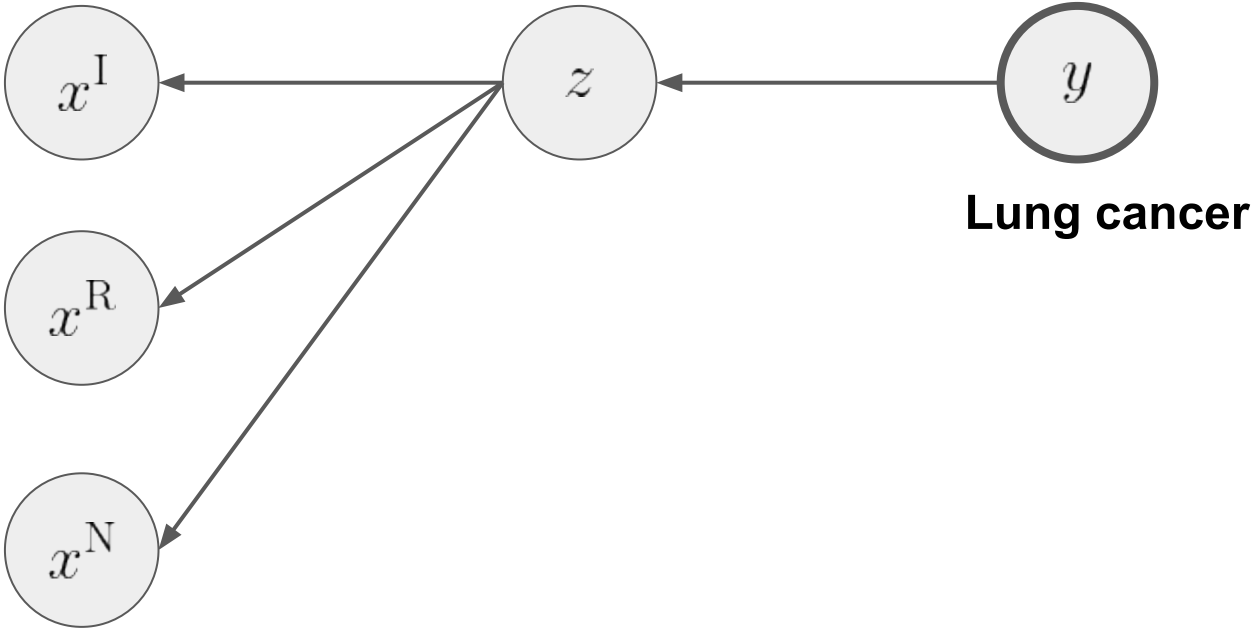
Text

Numerical signals

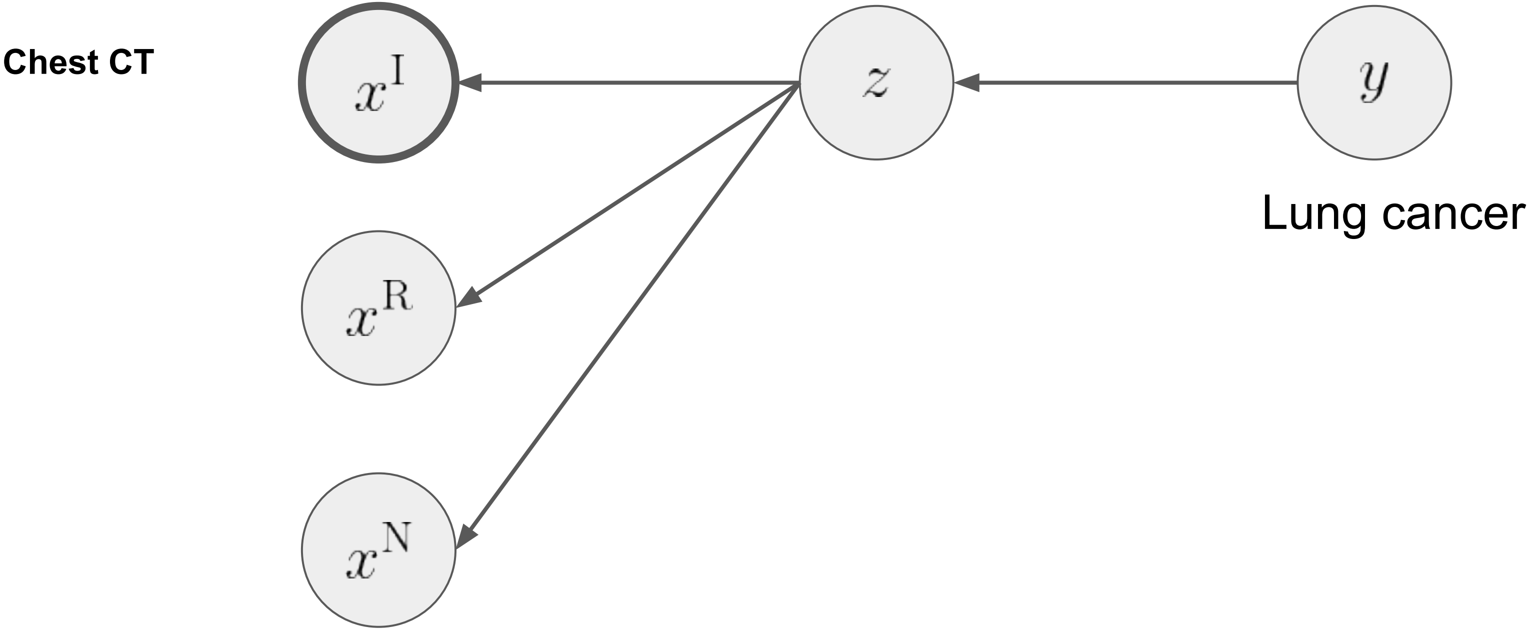
...



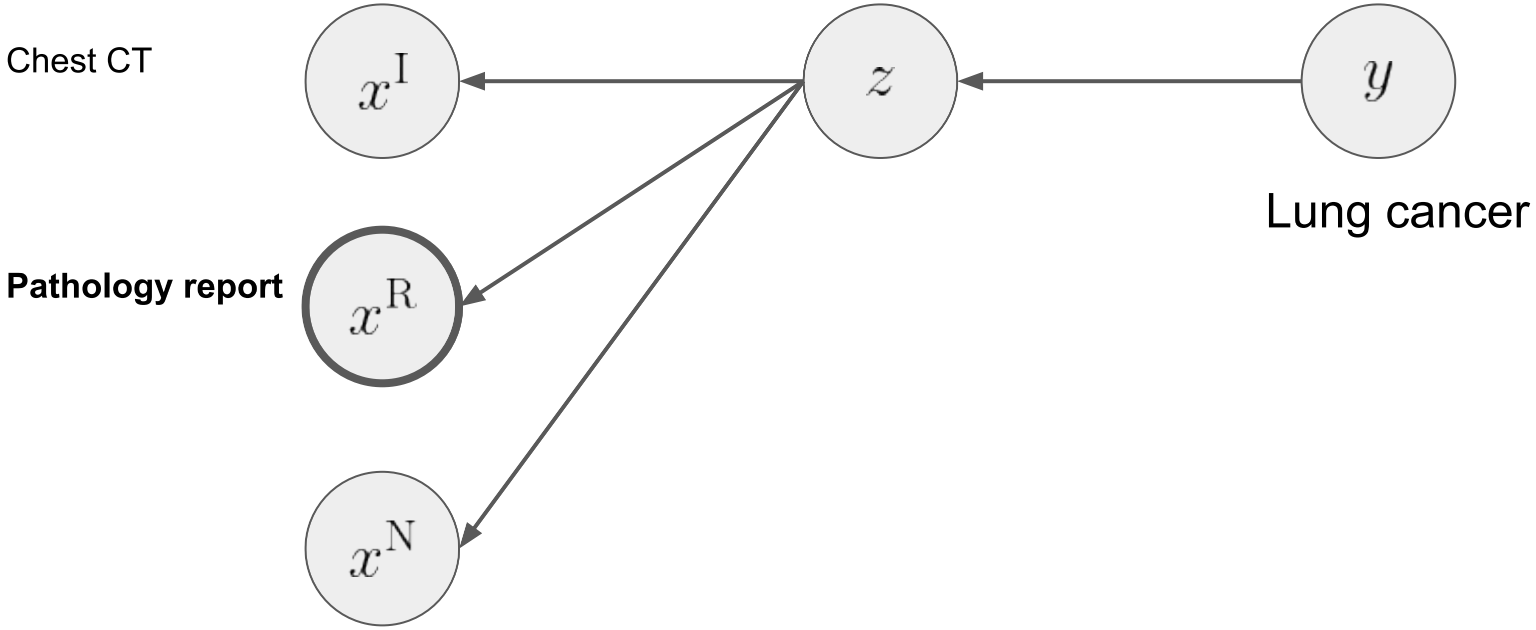
Multimodal clinical data reflect different yet correlated manifestations of a subject's underlying physiological processes



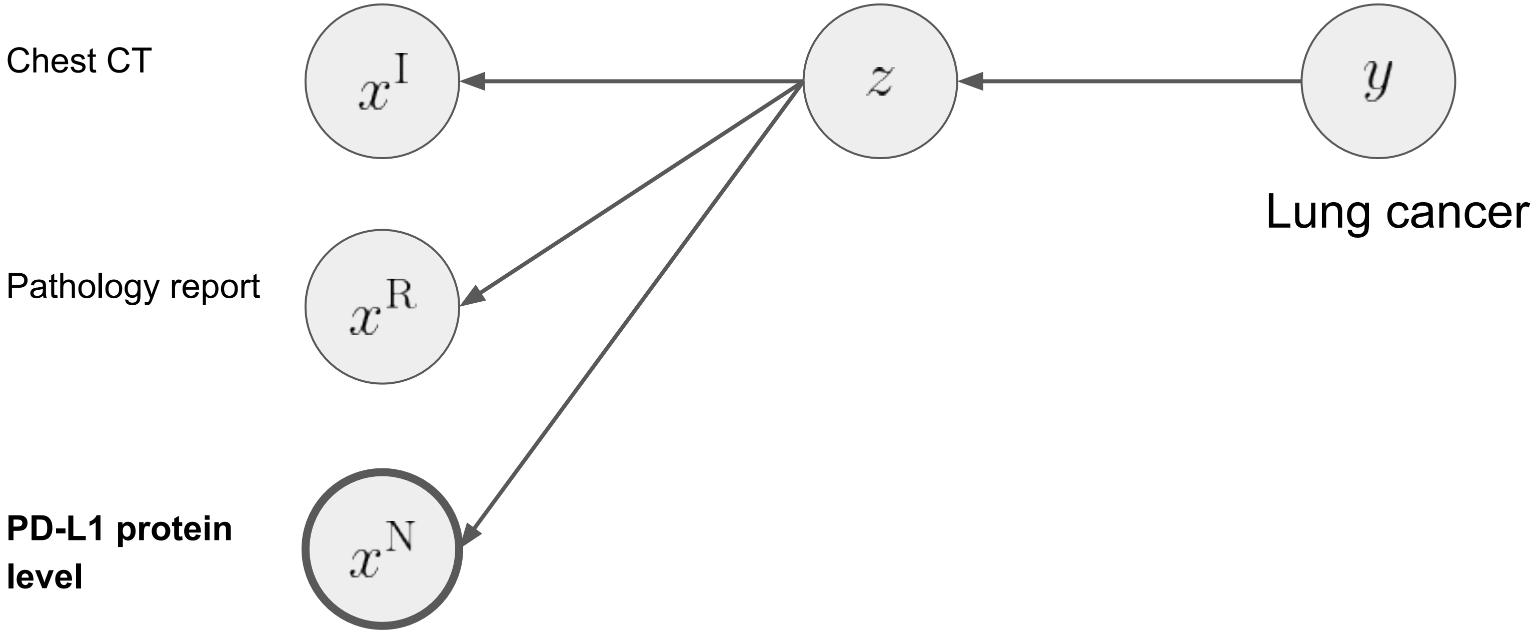
Multimodal clinical data reflect different yet correlated manifestations of a subject's underlying physiological processes



Multimodal clinical data reflect different yet correlated manifestations of a subject’s underlying physiological processes



Multimodal clinical data reflect different yet correlated manifestations of a subject's underlying physiological processes



Multimodal representation learning for medical image analysis

FINAL REPORT
EXAMINATION: CHEST (PORTABLE AP)

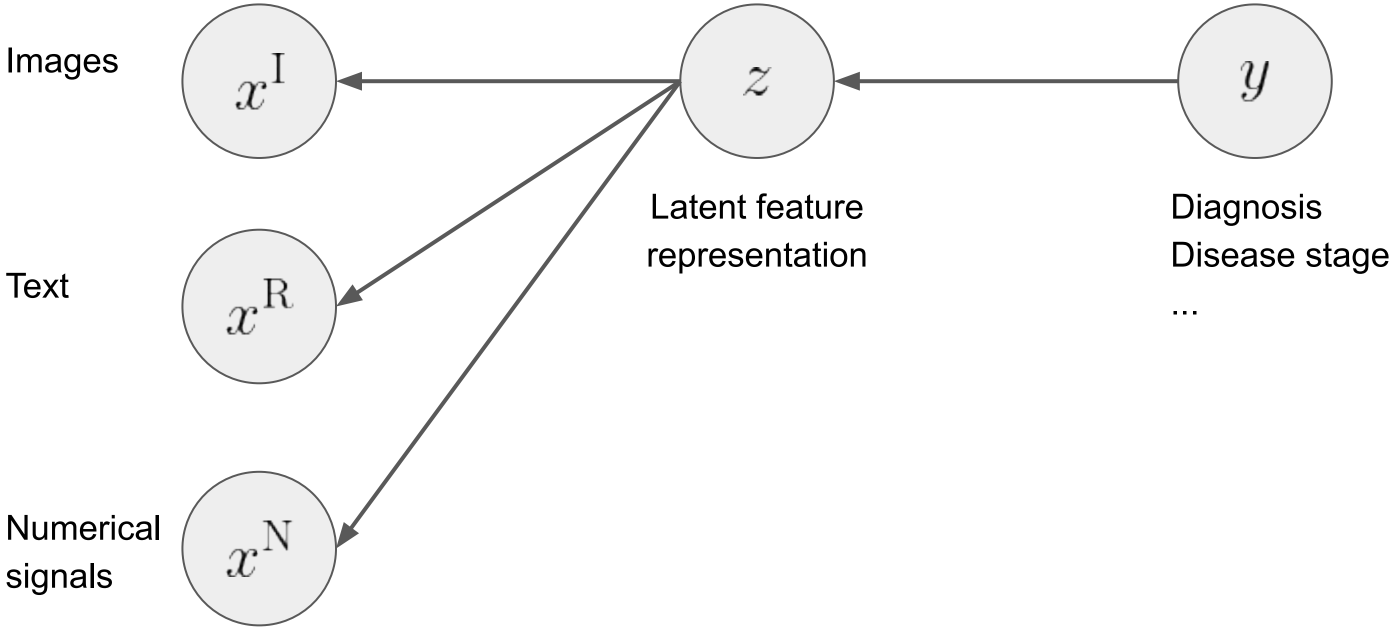
INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

IMPRESSION:

Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities



...

Multimodal representation learning for medical image analysis

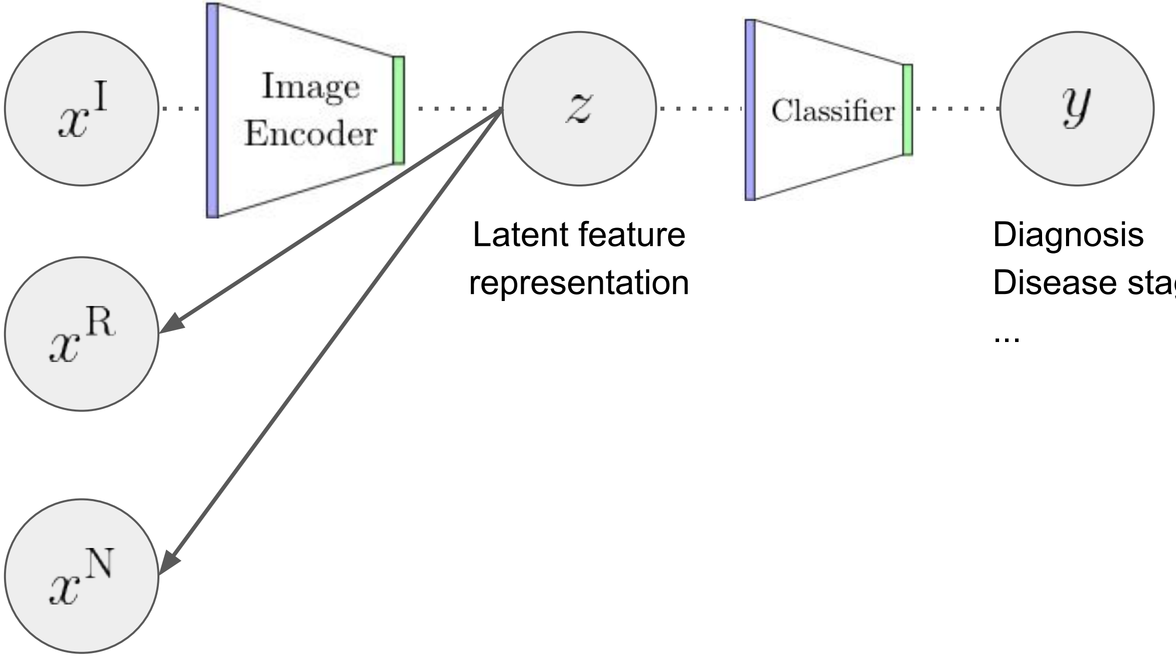
Images

Text

```
                        FINAL REPORT
EXAMINATION:  CHEST (PORTABLE AP)

INDICATION:  ___ year old man with respiratory failure, ARDS  // Volume
overload?

TECHNIQUE:  Single frontal view of the chest

COMPARISON:  ___

IMPRESSION:

Moderate left pleural effusion decreased.  Large right pleural effusion is
probably unchanged.  Tracheostomy tube is in unchanged position.  Extensive
bilateral alveolar opacities have improved, consistent with improve severe
pulmonary edema.  Cardiac size is obscured by the pleural parenchymal
abnormalities
```

Numerical signals

...

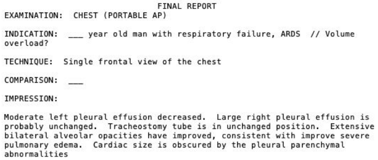


Latent feature representation

Diagnosis
Disease stage
...

Multimodal representation learning for medical image analysis

Images


FINAL REPORT
EXAMINATION: CHEST (PORTABLE AP)

INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

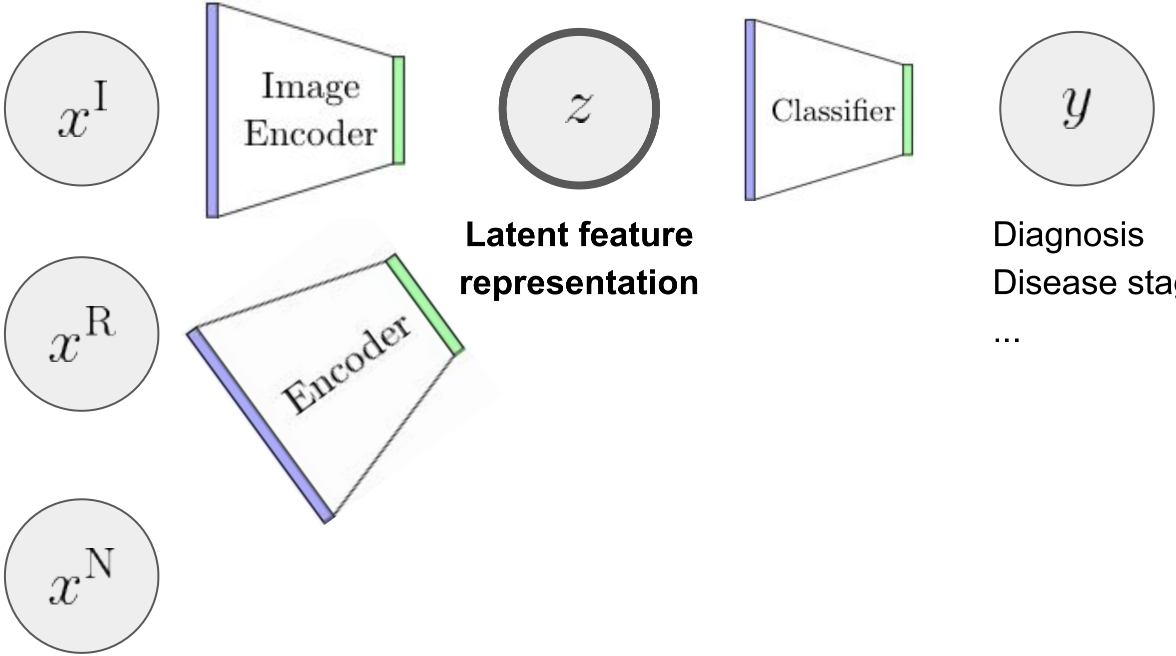
IMPRESSION:

Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities

Text

Numerical signals

...



Latent feature representation

Diagnosis
Disease stage
...

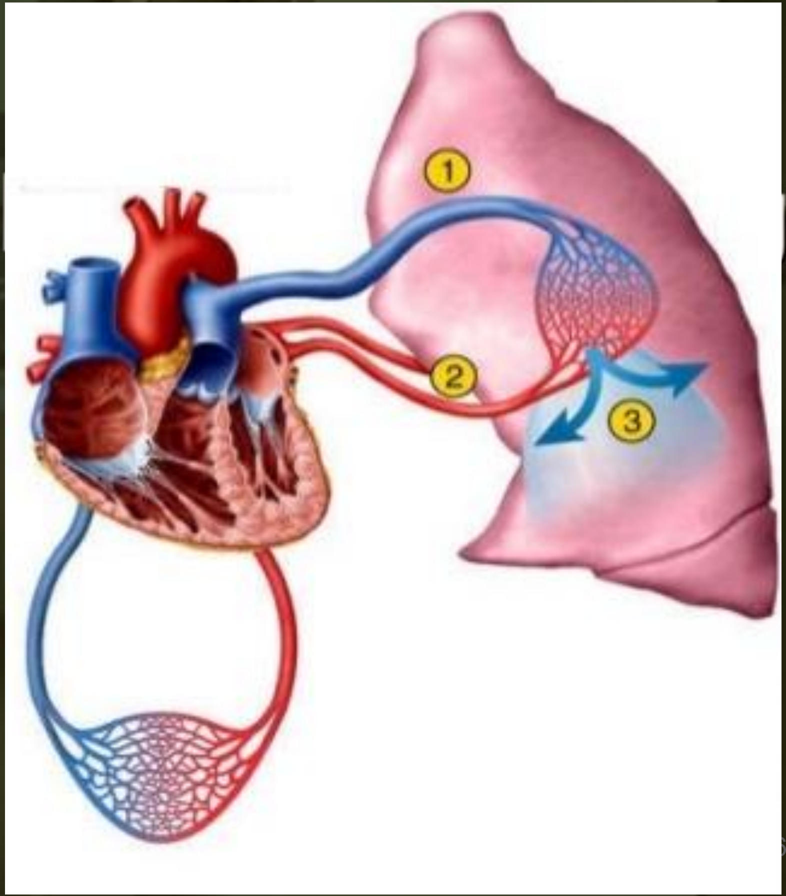
Outline

1. Motivating Clinical Problem
2. Image-based Model for Pulmonary Edema Assessment [Liao et al., 2019, Horng*, Liao* et al., 2021]
3. Joint Image-text Modeling [Chauhan*, Liao* et al., 2020]
4. Mutual Information for Representation Learning [Liao et al., 2021]
5. Conclusions

Outline

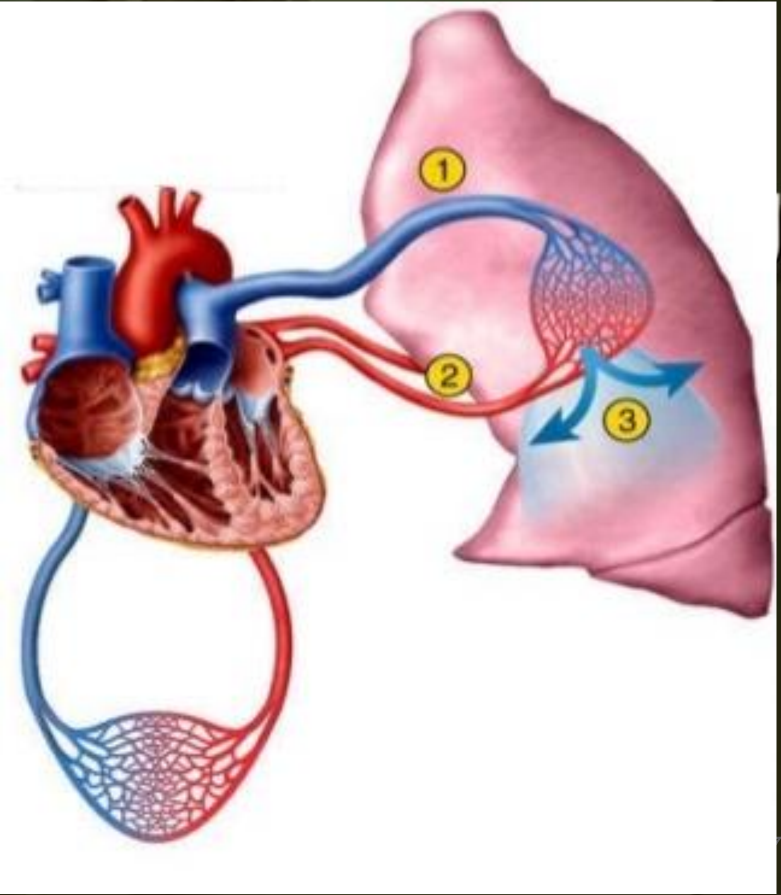
1. **Motivating Clinical Problem**
2. **Image-based Model for Pulmonary Edema Assessment [Liao et al., 2019, Horng*, Liao* et al., 2021]**
3. Joint Image-text Modeling [Chauhan*, Liao* et al., 2020]
4. Mutual Information for Representation Learning [Liao et al., 2021]
5. Conclusions

Most common cause of heart failure hospitalizations: pulmonary edema



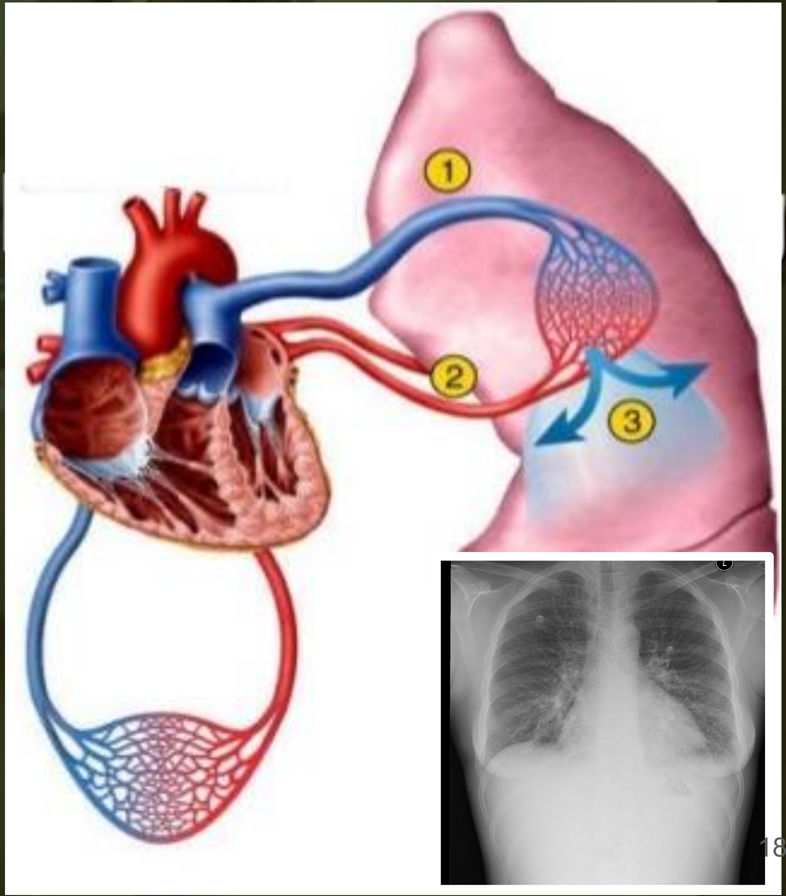
Heart failure is the leading cause of hospitalization in the US

- **1 million hospital stays** due to heart failure every year in the US (90% for pulmonary edema)
- **20% of heart failure patients readmitted** within 30 days of discharge
- **Roughly one out of eight US deaths** is caused at least in part by heart failure.

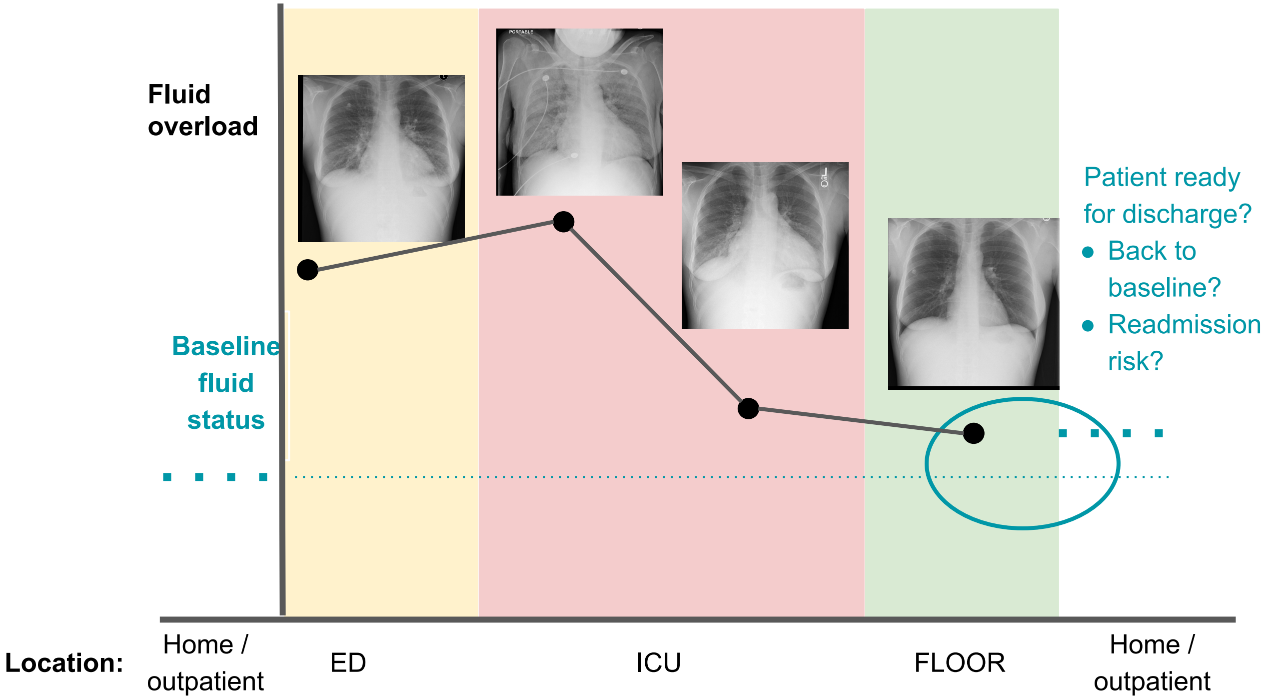


Chest x-ray is commonly performed to assess pulmonary edema

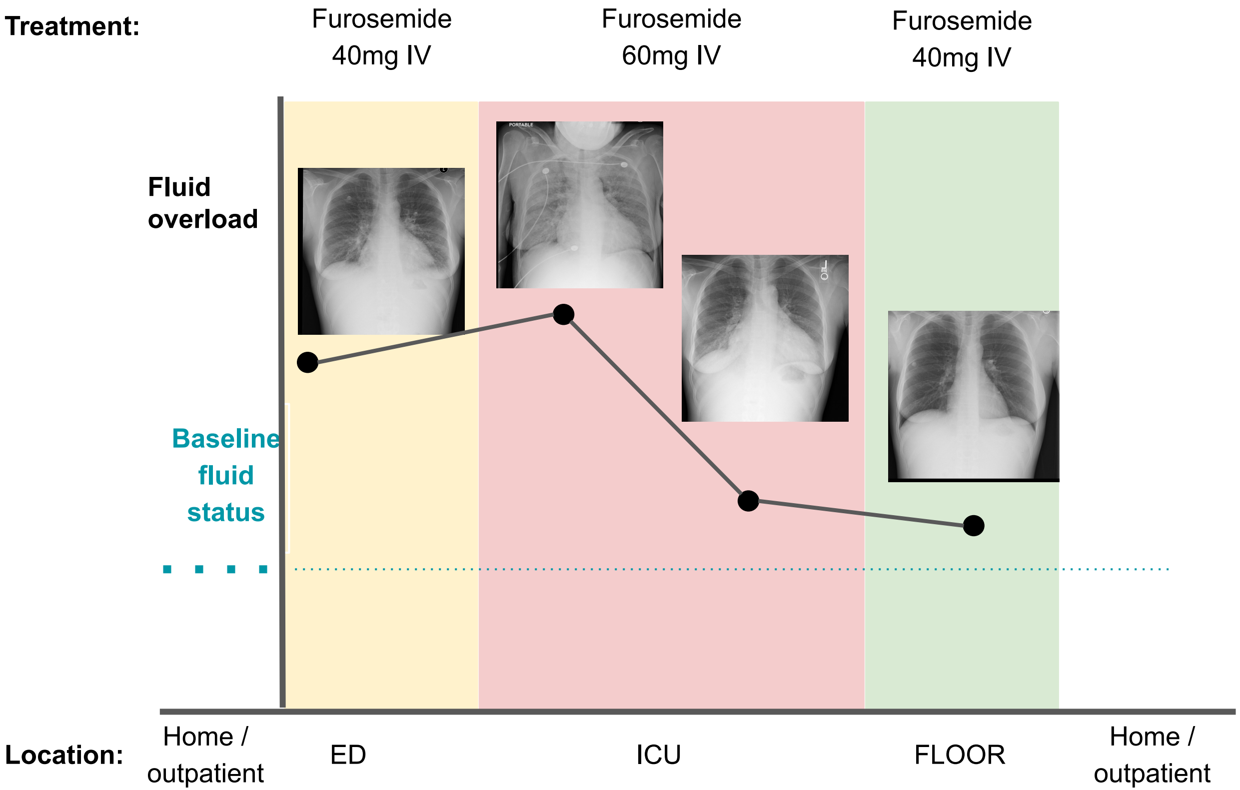
- **1 million hospital stays** due to heart failure every year in the US (90% for pulmonary edema)
- **20% of heart failure patients readmitted** within 30 days of discharge
- **Roughly one out of eight US deaths** is caused at least in part by heart failure.



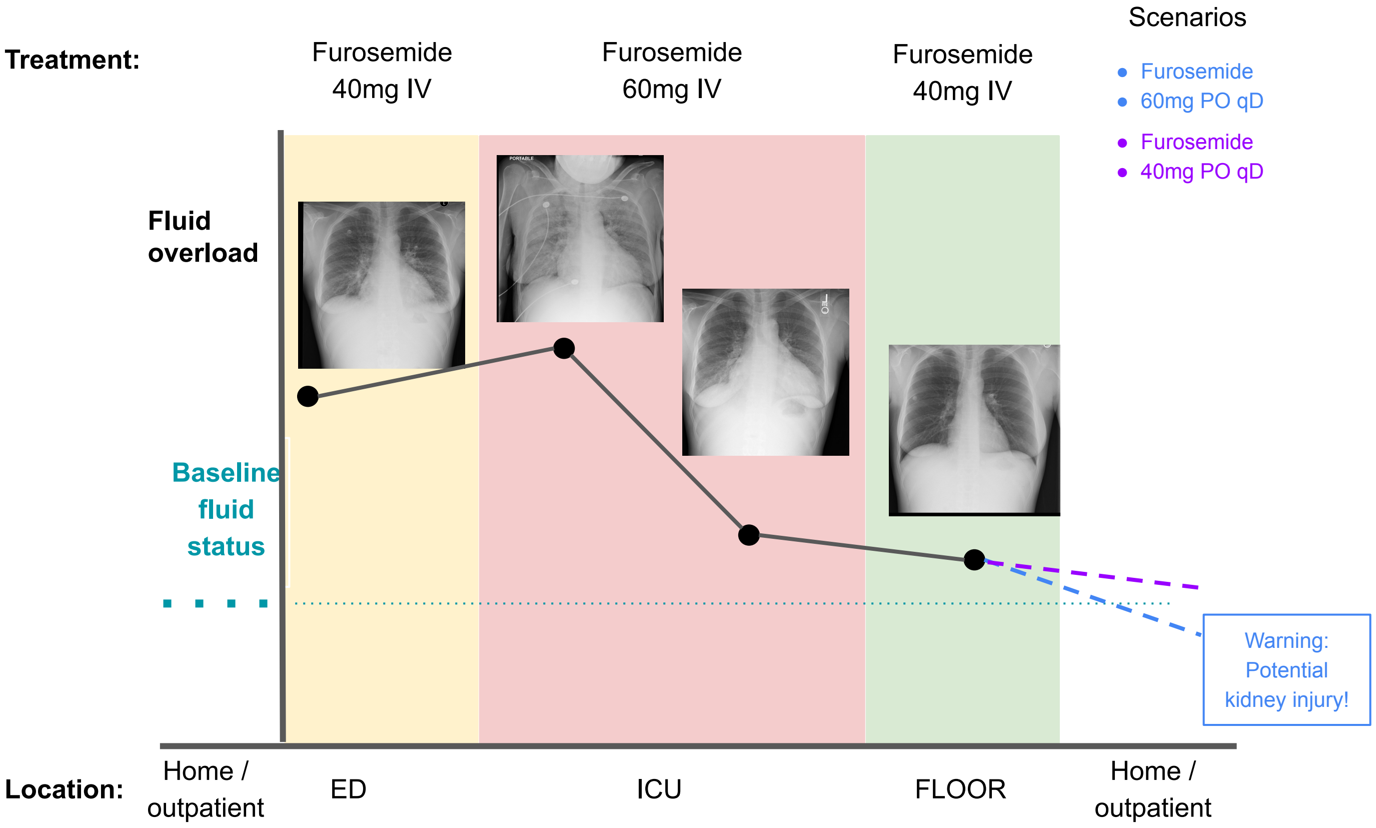
Retrospective clinical trajectory buried in the unstructured imaging data



Retrospective clinical trajectory buried in the unstructured imaging data



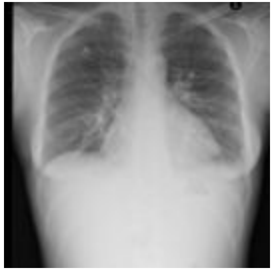
Retrospective clinical trajectory buried in the unstructured imaging data



I aim to develop computer vision models that assess pulmonary edema from chest x-rays

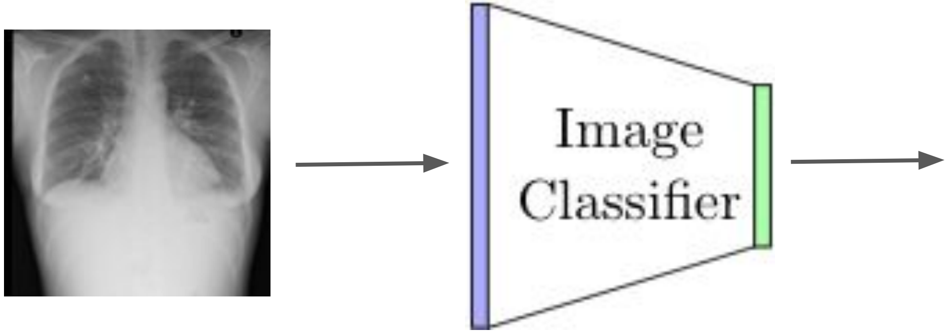
0: None

1: Vascular congestion (mild)



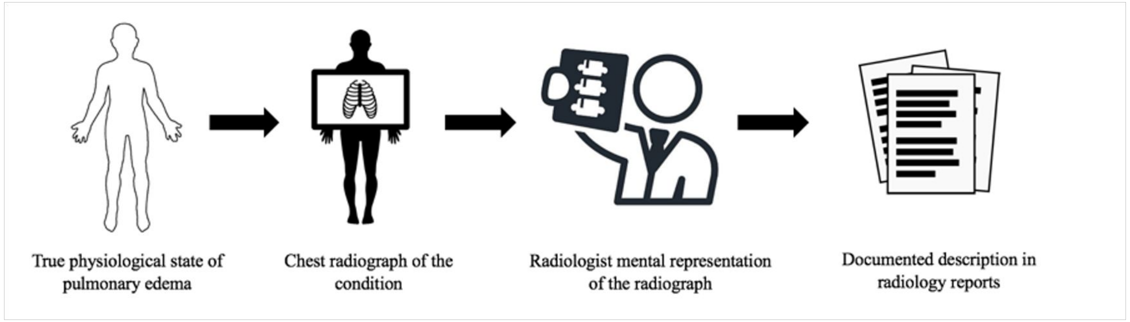
2: Interstitial edema (moderate)

3: Alveolar edema (severe)



2: Interstitial edema

MIMIC-CXR consists of 370K chest radiographs associated with radiology reports



```
                              FINAL REPORT
EXAMINATION:  CHEST (PORTABLE AP)

INDICATION:  ___ year old man with respiratory failure, ARDS  // Volume
overload?

TECHNIQUE:  Single frontal view of the chest

COMPARISON:  ___

IMPRESSION:

Moderate left pleural effusion decreased.  Large right pleural effusion is
probably unchanged.  Tracheostomy tube is in unchanged position.  Extensive
bilateral alveolar opacities have improved, consistent with improve severe
pulmonary edema.  Cardiac size is obscured by the pleural parenchymal
abnormalities
```

[https://physionet.org/content/mimic-cxr/2.0.0/]

Labeling from radiology reports

- Regular expression (regex) labeled 6,710 reports. **[Training]**
- Three experts labeled 485 radiology reports. **[Validation/Test]**

```
                         FINAL REPORT
 INDICATION:  Evaluation for interval change in
a patient status post core valve.

 COMPARISON:  ___ through ___.

 FINDINGS:  Portable AP semi-upright view of
the chest is reviewed and compared to the prior
study.  An aortic core valve projects over the
heart and a transvenous right internal jugular
pacer follows the expected course and is
unchanged in position.  Interstitial
abnormality is unchanged since ___, but
increased since ___, probably due to edema,
exaggerated by low post operatived lung
volumes. There is no significant pleural
effusion or pneumothorax. The cardiomediastinal
silhouette, reflecting mild cardiomegaly, are
unchanged. Elevation of the left hemidiaphragm
is chronic.

 IMPRESSION:

 1.  Mild interstitial edema stable since ___,
increased since ___.
```

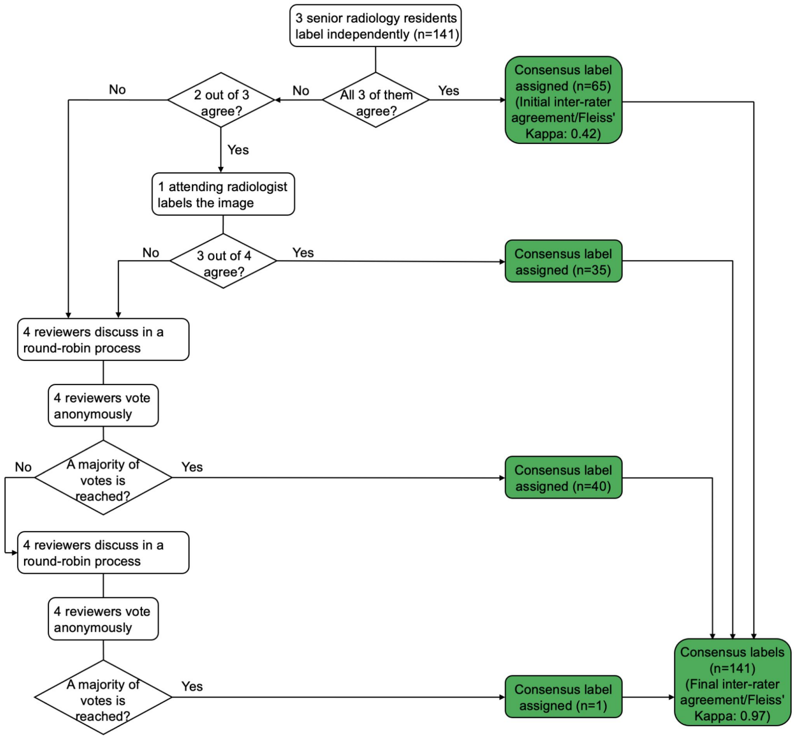
Labeled as level 2: interstitial edema

[Liao et al., 2019, Horng*, Liao* et al., 2021]

Consensus labeling (modified Delphi process) from chest x-ray images

- Regular expression (regex) labeled 6,710 reports. **[Training]**
- Three experts labeled 485 radiology reports. **[Validation/Test]**
- Four radiologists labeled 141 chest x-ray images. **[Test]**

Labels released on PhysioNet!

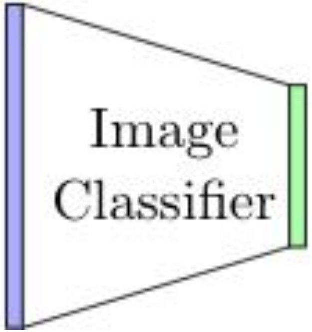


[Liao et al., 2019, Horng*, Liao* et al., 2021]

We have limited numerical labels

(370K chest x-ray images)

x^{I}



y

$$y \in \{0, 1, 2, 3\}$$

(7K labels for training,
<1K labels for evaluation)

x^{R}

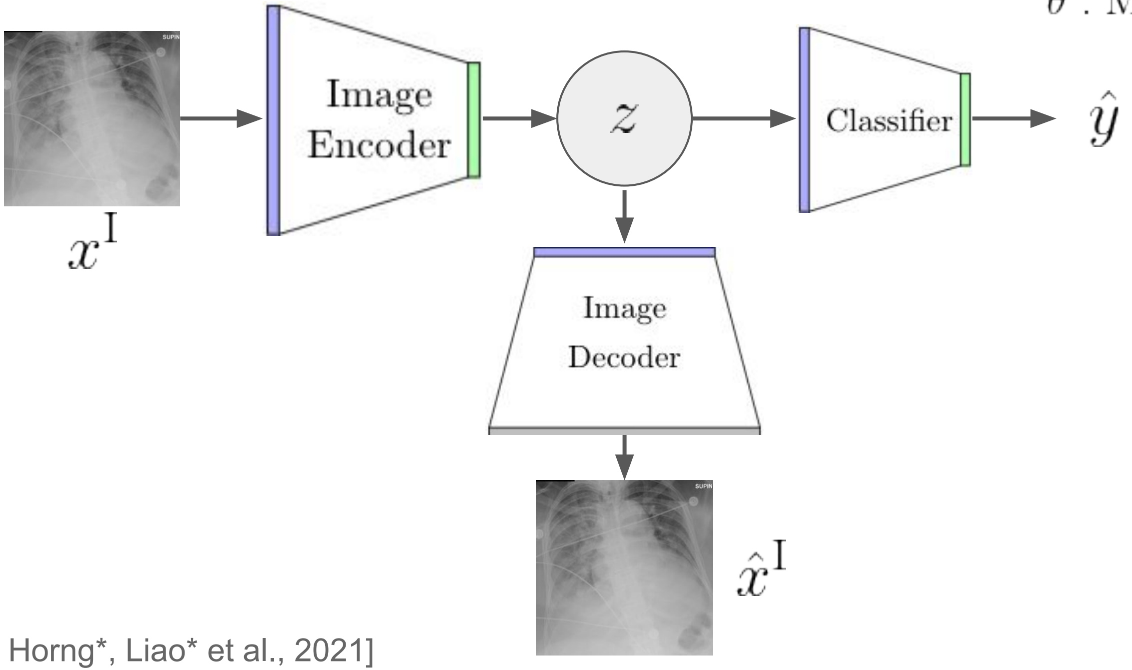
(230K radiology reports)

Semi-supervised learning to utilize unlabeled images

$$\max \log p(\mathbf{x}^{\mathrm{I}}, \mathbf{y}; \theta) = \sum_{i=N_{\mathrm{L}}+1}^{N} \log p(\mathrm{x}_i^{\mathrm{I}}; \theta) + \sum_{i=1}^{N_{\mathrm{L}}} \log p(\mathrm{x}_i^{\mathrm{I}}, \mathrm{y}_i; \theta)$$

N_{L} : Number of labeled images

θ : Model parameters



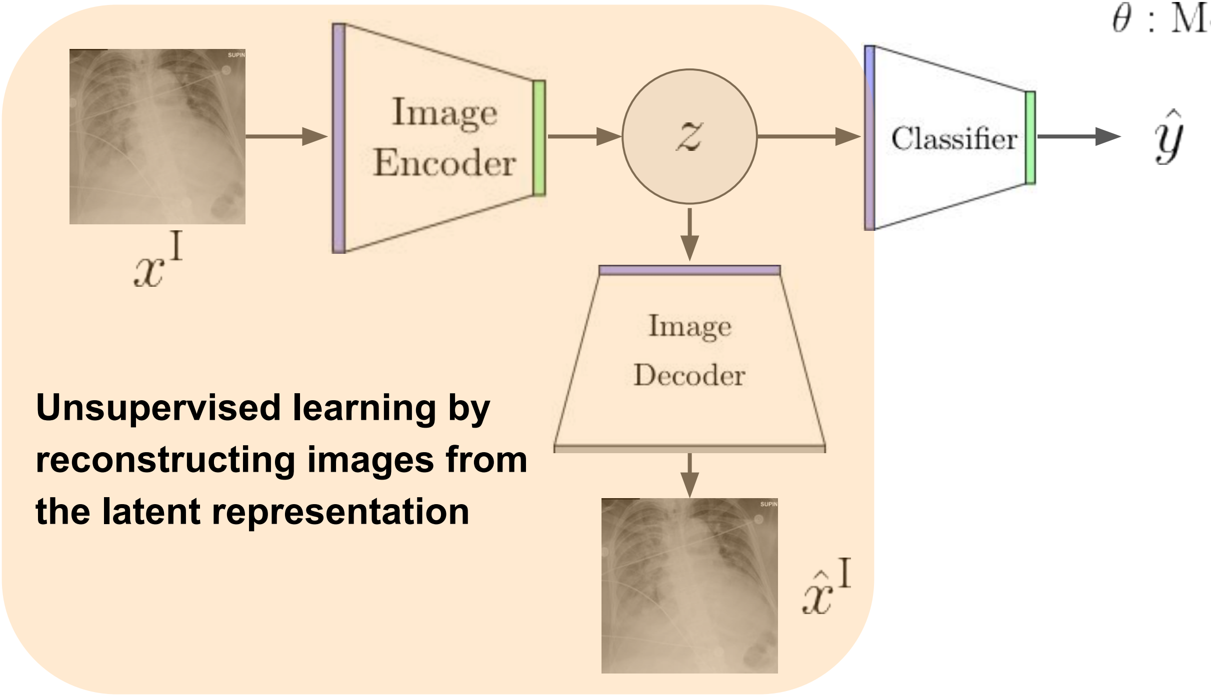
[Liao et al., 2019, Horng*, Liao* et al., 2021]

Semi-supervised learning to utilize unlabeled images

$$\max \log p(\mathbf{x}^{\mathrm{I}}, \mathbf{y}; \theta) = \sum_{i=N_{\mathrm{L}}+1}^{N} \log p(\mathrm{x}_i^{\mathrm{I}}; \theta) + \sum_{i=1}^{N_{\mathrm{L}}} \log p(\mathrm{x}_i^{\mathrm{I}}, \mathrm{y}_i; \theta)$$

N_{L} : Number of labeled images

θ : Model parameters

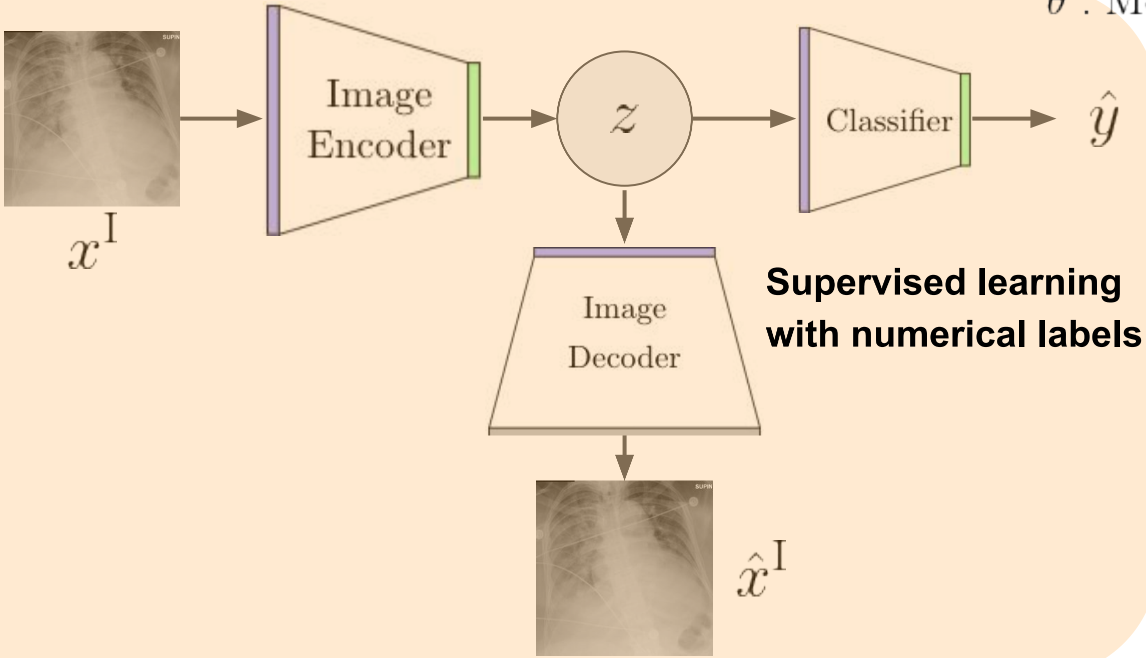


Semi-supervised learning to utilize unlabeled images

$$\max \log p(\mathbf{x}^{\mathrm{I}}, \mathbf{y}; \theta) = \sum_{i=N_{\mathrm{L}}+1}^{N} \log p(\mathrm{x}_i^{\mathrm{I}}; \theta) + \sum_{i=1}^{N_{\mathrm{L}}} \log p(\mathrm{x}_i^{\mathrm{I}}, \mathrm{y}_i; \theta)$$

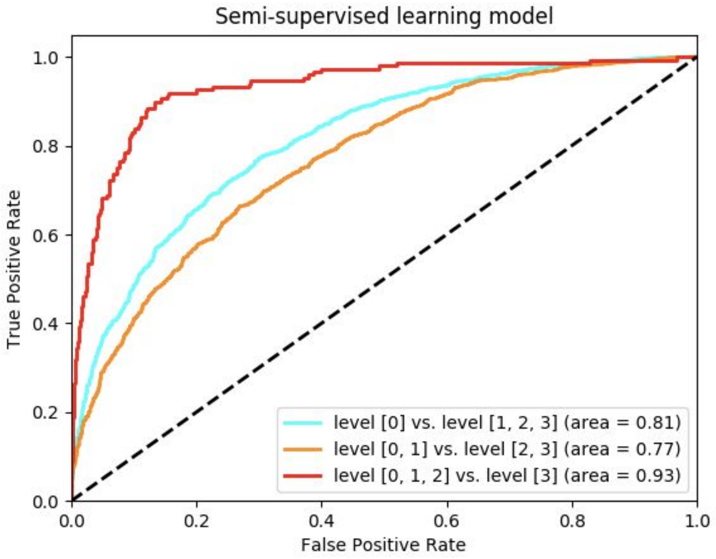
N_{L} : Number of labeled images

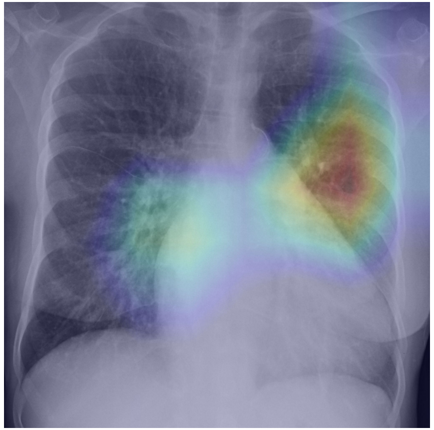
θ : Model parameters



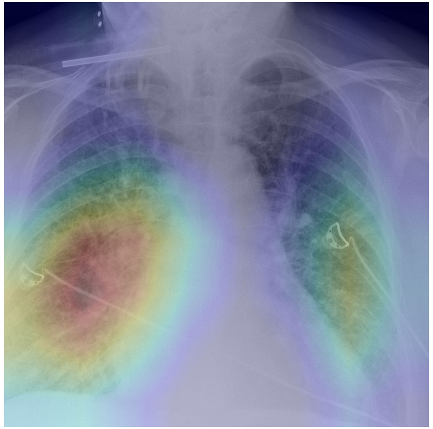
Supervised learning with numerical labels

Semi-supervised learning trained with regex labels and evaluated on consensus labels





1: Vascular congestion



3: Alveolar edema

[Liao et al., 2019, Horng*, Liao* et al., 2021]

Outline

FINAL REPORT
EXAMINATION: CHEST (PORTABLE AP)

INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

IMPRESSION:

Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities

Images x^{I} z y

Latent feature representation

Diagnosis
Disease stage
...

Text x^{R}

Numerical signals x^{N}

...

Outline - Joint Image-text Modeling [Chauhan*, Liao* et al., 2020]

```
                              FINAL REPORT
EXAMINATION:  CHEST (PORTABLE AP)

INDICATION:  ___ year old man with respiratory failure, ARDS  // Volume
overload?

TECHNIQUE:  Single frontal view of the chest

COMPARISON:  ___

IMPRESSION:

Moderate left pleural effusion decreased.  Large right pleural effusion is
probably unchanged.  Tracheostomy tube is in unchanged position.  Extensive
bilateral alveolar opacities have improved, consistent with improve severe
pulmonary edema.  Cardiac size is obscured by the pleural parenchymal
abnormalities
```

Images x^{I}

Text x^{R}

Numerical signals x^{N}

...

z

Latent feature representation

y

Diagnosis
Disease stage
...

Outline - Joint Image-text Modeling [Chauhan*, Liao* et al., 2020]

FINAL REPORT
EXAMINATION: CHEST (PORTABLE AP)

INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

IMPRESSION:

Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities

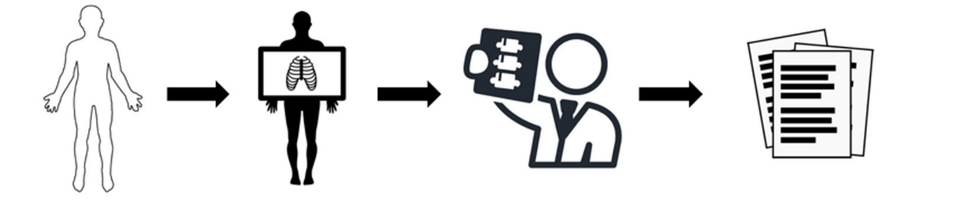
Images x^{I}

Text x^{R}

Latent feature representation z

Diagnosis
Disease stage
... y

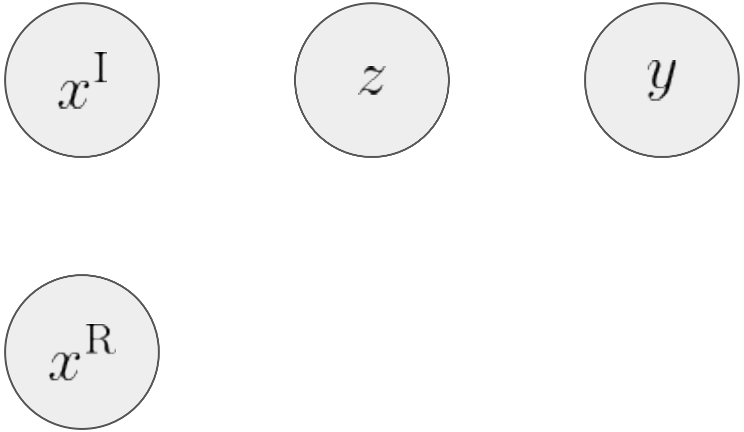
Numerical signals x^{N}



...

Prior work in image-text modeling

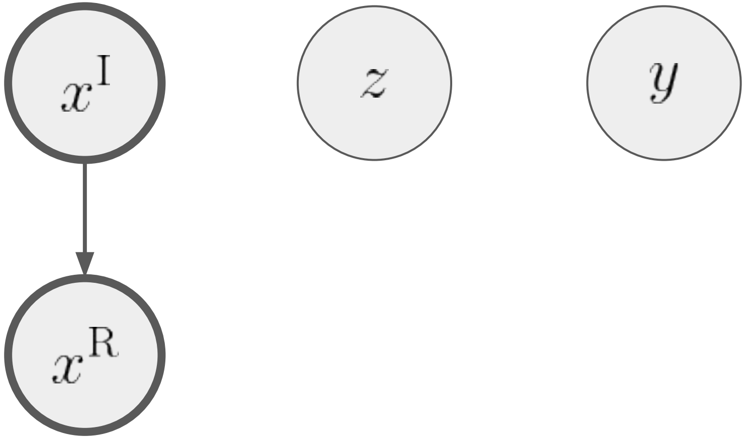
- Image captioning
  - The model generates text from an image
- Visual question answering
  - Training based on both images and text
  - Inference performed on an image-text pair
- Joint representation learning
  - Training based on both images and text
  - Inference performed on an image



[Anderson et al. 2018, Antol et al. 2015, Lu et al. 2016, Jing et al. 2017, Plummer et al. 2017, Vasudevan et al. 2017, Xu et al. 2015]

Prior work in image-text modeling

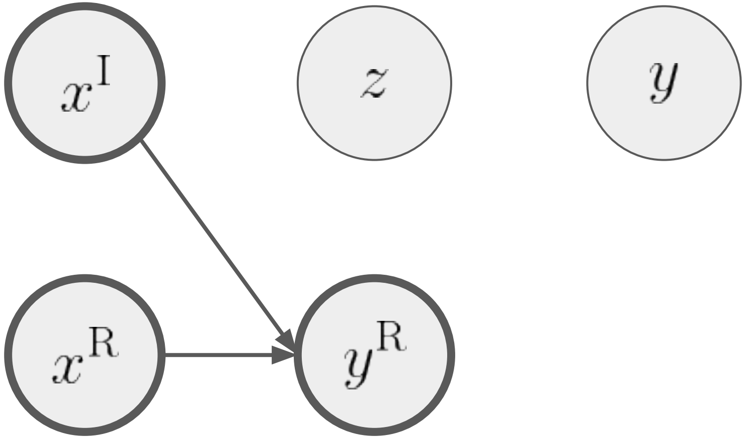
- **Image captioning**
  - **The model generates text from an image**
- Visual question answering
  - Training based on both images and text
  - Inference performed on an image-text pair
- Joint representation learning
  - Training based on both images and text
  - Inference performed on an image



[Anderson et al. 2018, Antol et al. 2015, Lu et al. 2016, Jing et al. 2017, Plummer et al. 2017, Vasudevan et al. 2017, Xu et al. 2015]

Prior work in image-text modeling

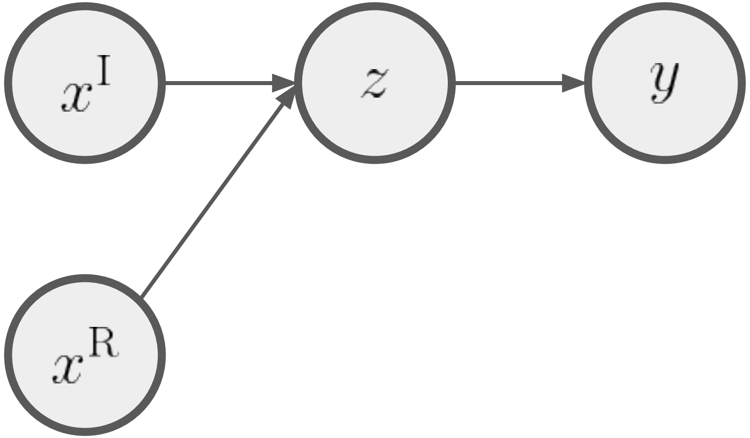
- Image captioning
  - The model generates text from an image
- **Visual question answering**
  - **Training based on both images and text**
  - **Inference performed on an image-text pair**
- Joint representation learning
  - Training based on both images and text
  - Inference performed on an image



[Anderson et al. 2018, Antol et al. 2015, Lu et al. 2016, Jing et al. 2017, Plummer et al. 2017, Vasudevan et al. 2017, Xu et al. 2015]

Prior work in image-text modeling

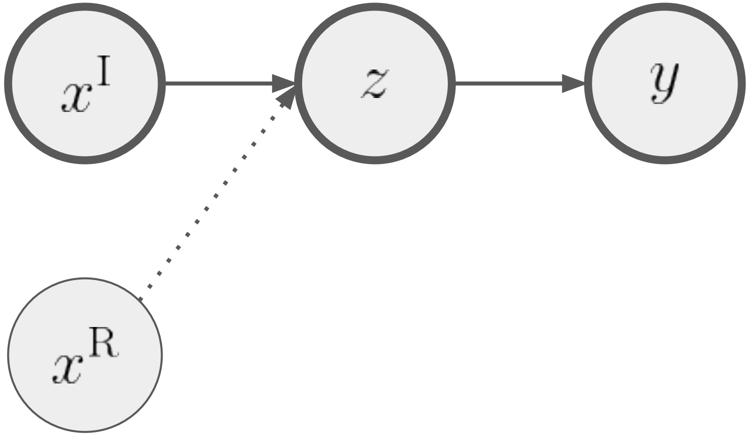
- Image captioning
  - The model generates text from an image
- Visual question answering
  - Inference performed on an image-text pair
- **Joint representation learning**
  - **Training based on both images and text**
  - **Inference performed on an image**



[Anderson et al. 2018, Antol et al. 2015, Lu et al. 2016, Jing et al. 2017, Plummer et al. 2017, Vasudevan et al. 2017, Xu et al. 2015]

Prior work in image-text modeling

- Image captioning
    - The model generates text from an image
- Visual question answering
    - Inference performed on an image-text pair
- **Joint representation learning**
    - **Training based on both images and text**
    - **Inference performed on an image**



[Anderson et al. 2018, Antol et al. 2015, Lu et al. 2016, Jing et al. 2017, Plummer et al. 2017, Vasudevan et al. 2017, Xu et al. 2015]

Training: joint image-text representation learning model

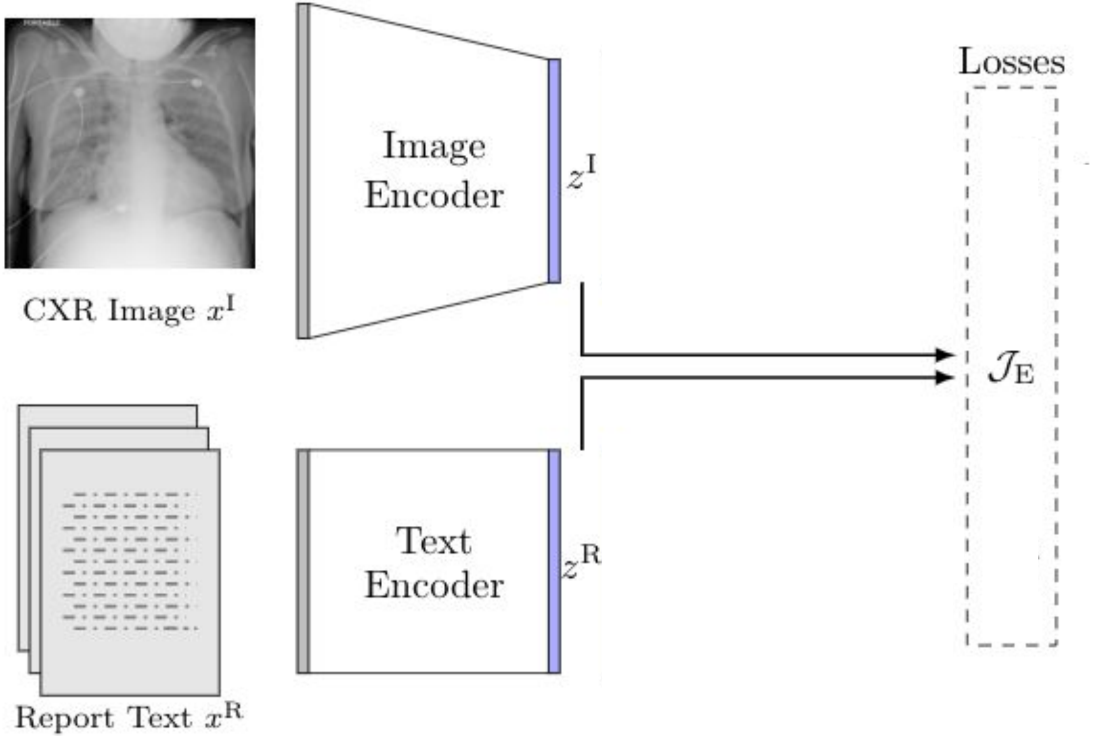
CXR Image x^{I}

y

Report Text x^{R}

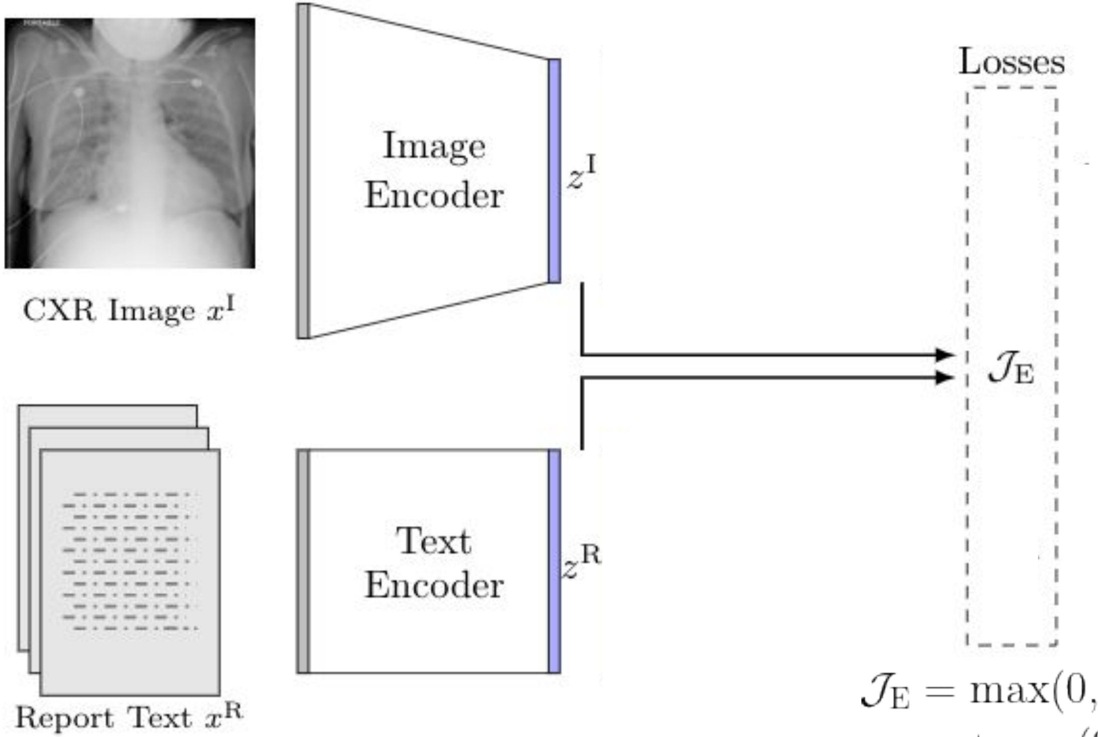
[Chauhan*, Liao* et al., 2020]

Training: joint image-text representation learning model



[Chauhan*, Liao* et al., 2020]

Training: joint image-text representation learning model



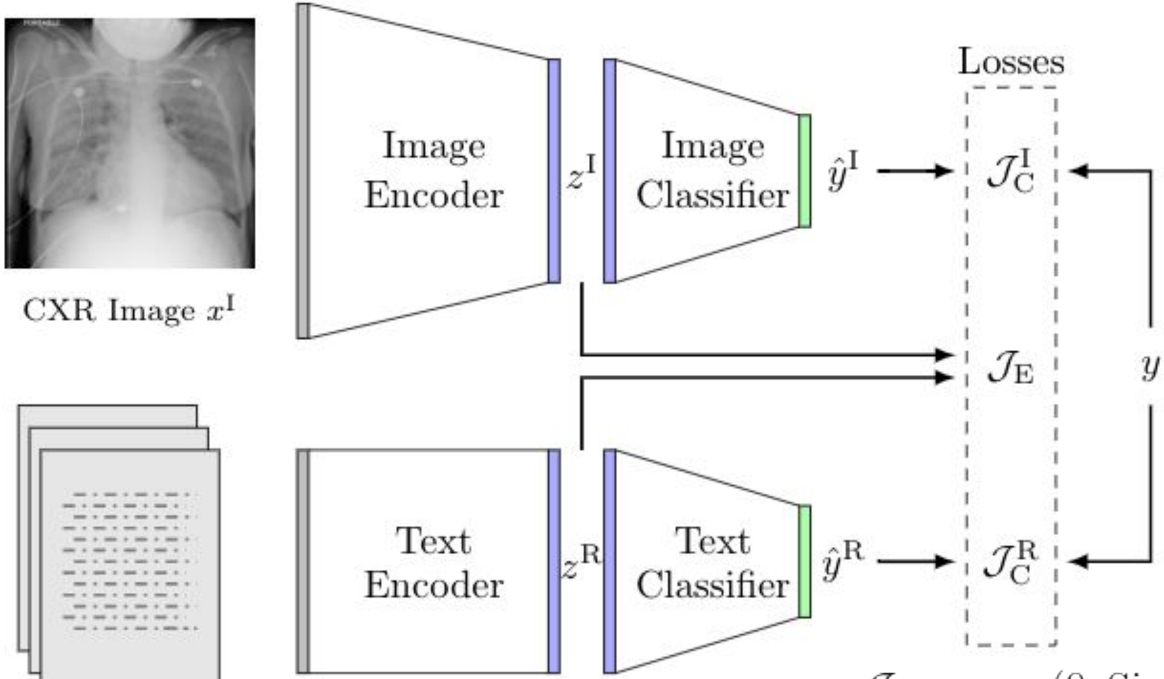
Joint embedding loss

(contrastive loss)

$$\mathcal{J}_{\mathrm{E}} = \max(0, \mathrm{Sim}(z_j^{\mathrm{I}}, z_{s(j)}^{\mathrm{R}}) - \mathrm{Sim}(z_j^{\mathrm{I}}, z_j^{\mathrm{R}}) + \eta)$$
$$+\max(0, \mathrm{Sim}(z_{s(j)}^{\mathrm{I}}, z_j^{\mathrm{R}}) - \mathrm{Sim}(z_j^{\mathrm{I}}, z_j^{\mathrm{R}}) + \eta)$$

[Chauhan*, Liao* et al., 2020]

Training: joint image-text representation learning model



Report Text x^{R}

Image classification loss

(cross-entropy loss)

Joint embedding loss

(contrastive loss)

Text classification loss

(cross-entropy loss)

$$\mathcal{J}_{\mathrm{E}} = \max(0, \mathrm{Sim}(z_j^{\mathrm{I}}, z_{s(j)}^{\mathrm{R}}) - \mathrm{Sim}(z_j^{\mathrm{I}}, z_j^{\mathrm{R}}) + \eta) \\ + \max(0, \mathrm{Sim}(z_{s(j)}^{\mathrm{I}}, z_j^{\mathrm{R}}) - \mathrm{Sim}(z_j^{\mathrm{I}}, z_j^{\mathrm{R}}) + \eta)$$

[Chauhan*, Liao* et al., 2020]

Inference: image classification

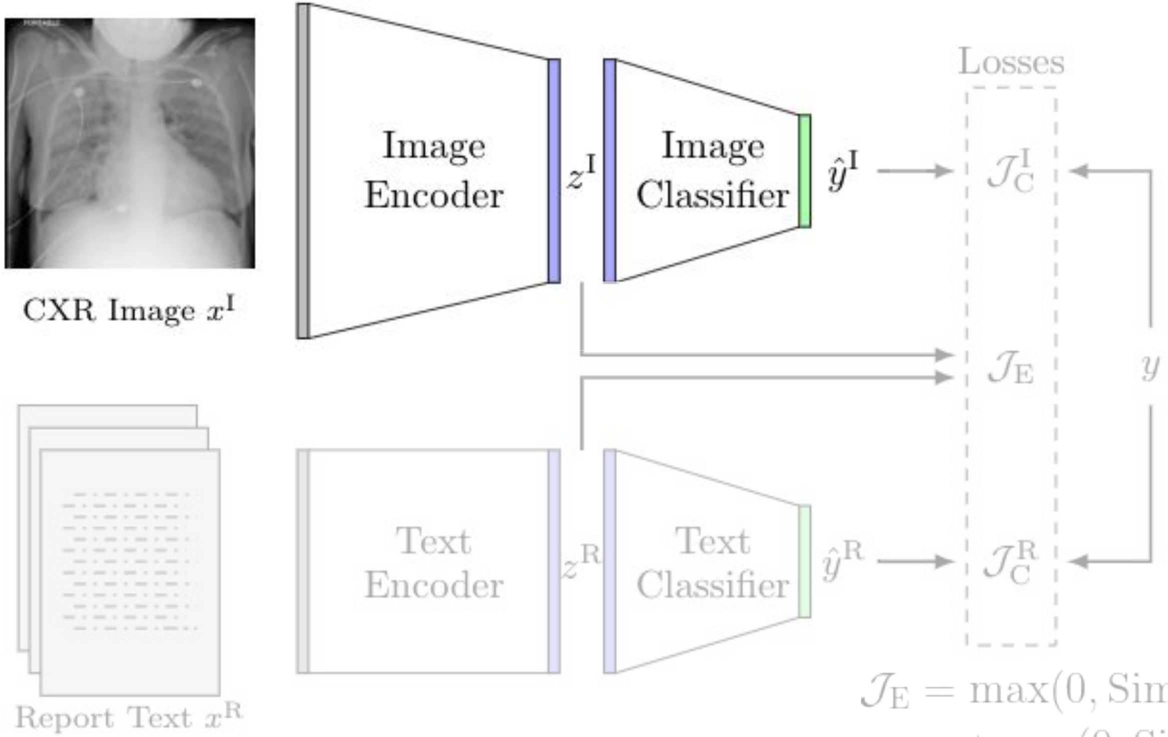


Image classification loss

(cross-entropy loss)

Joint embedding loss

(contrastive loss)

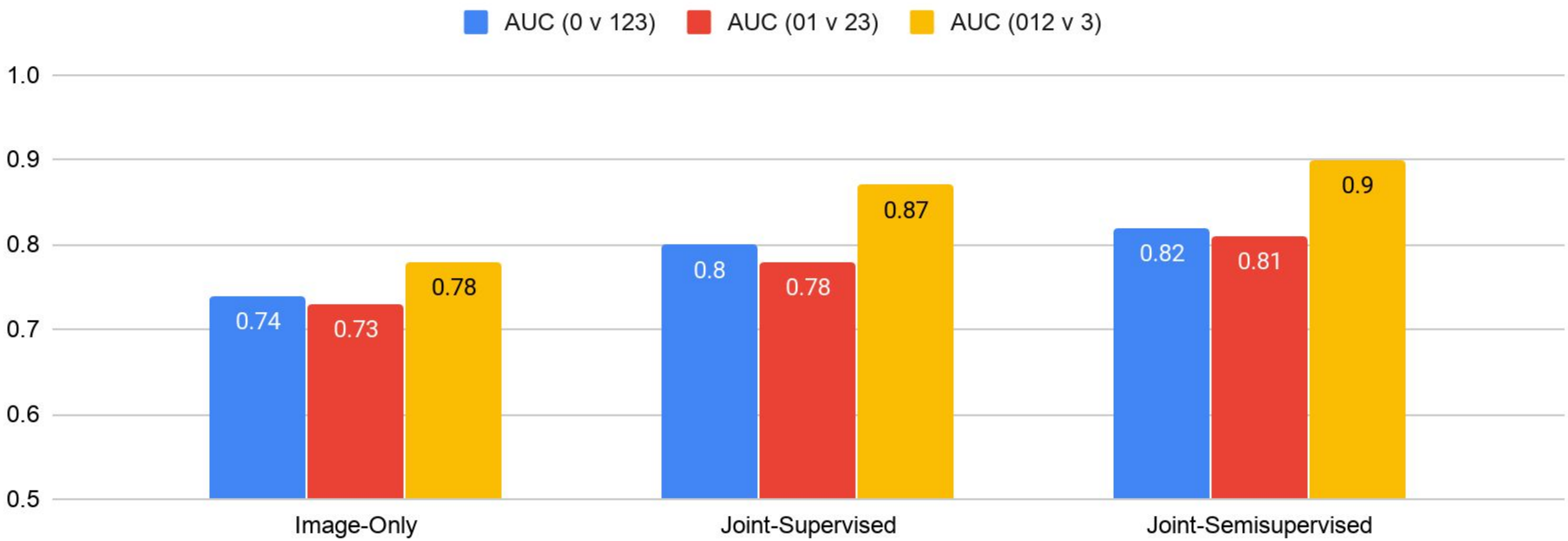
Text classification loss

(cross-entropy loss)

$$\mathcal{J}_{\mathrm{E}} = \max(0, \mathrm{Sim}(z_j^{\mathrm{I}}, z_{s(j)}^{\mathrm{R}}) - \mathrm{Sim}(z_j^{\mathrm{I}}, z_j^{\mathrm{R}}) + \eta) + \max(0, \mathrm{Sim}(z_{s(j)}^{\mathrm{I}}, z_j^{\mathrm{R}}) - \mathrm{Sim}(z_j^{\mathrm{I}}, z_j^{\mathrm{R}}) + \eta)$$

[Chauhan*, Liao* et al., 2020]

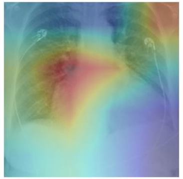
Results: Leveraging free-text radiology reports improves the image model performance



[Chauhan*, Liao* et al., 2020]

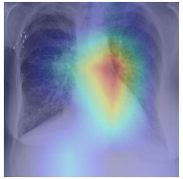
Results: Image model *interpretation* with free text

Level 1



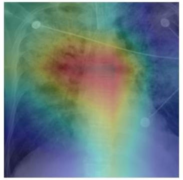
[CLS] frontal and lateral radiographs of the chest demonstrates slight decrease in size of the severely enlarged cardiac sil hou ette . persistent small bilateral pleural effusion s . probable small hi atal hernia . there is persistent mild pulmonary vascular congestion . clear lungs . no pneum othorax . decrease in severe enlargement of the cardiac sil hou ette likely due to decrease in peric ardial effusion with persistent small effusion s and pulmonary vascular congestion . no pneumonia [SEP]

Level 2



[CLS] surgical clips are again present in the right axi ll a . the cardiac , mediast inal and hil ar contours appear unchanged . upward tent ing of the medial right hem idia ph rag m is very similar . there is a persistent small - to - moderate pleural effusion on the right wit and a small number on the left . fiss ures are mildly thick ened . sub ple ural thickening at the right lung apex appears stable . there is a new mild interstitial abnormality including ker ley b lines and peri bro nc hi al cuff ing suggesting mild - to - moderate interstitial pulmonary edema . however , there is no definite new focal opacity . bony structures are unre mark able . findings most consistent with pulmonary edema . [SEP]

Level 3



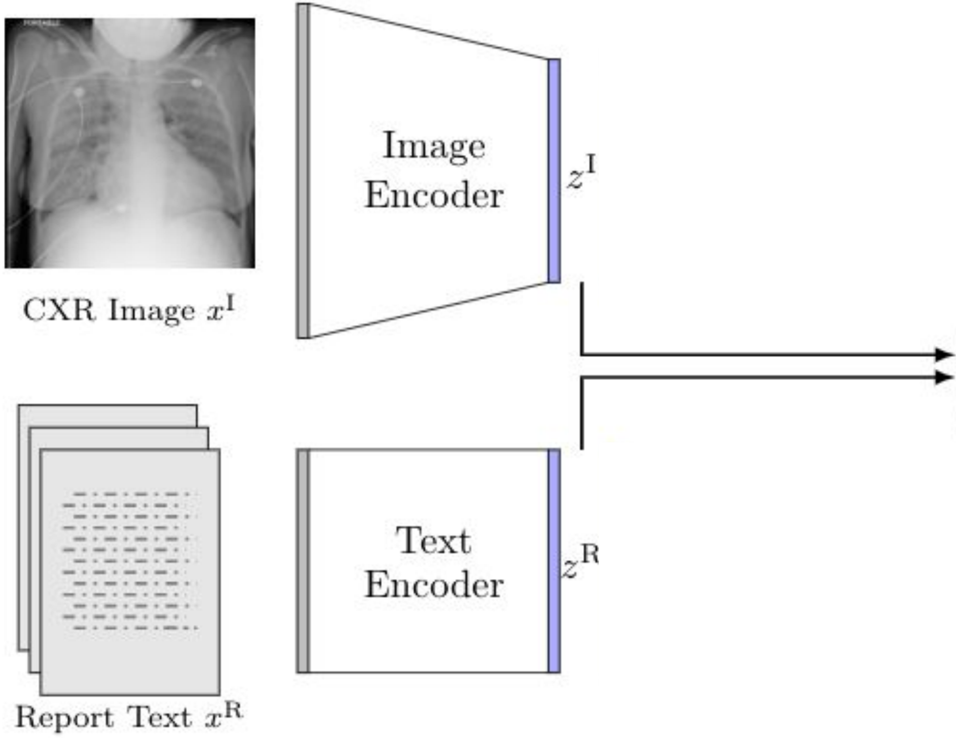
[CLS] a trache ostomy and left - side d pic c are stable in position . widespread alveolar op aci ties have increased from are less significant in extent compared to . this likely reflects a combination of increasing edema and persistent multif ocal infection . no pleural effusion or pneum othorax is identified . the cardio media sti nal and hil ar contours are within normal limits . widespread alveolar op aci ties are increased from the most recent prior exam consistent with increasing edema in the setting of persistent multif ocal infection . [SEP]

[Chauhan*, Liao* et al., 2020]

Outline

1. Motivating Clinical Problem
2. Image-based Model for Pulmonary Edema Assessment [Liao et al., 2019, Horng*, Liao* et al., 2021]
3. Joint Image-text Modeling [Chauhan*, Liao* et al., 2020]
4. **Mutual Information for Representation Learning [Liao et al., 2021]**
5. Conclusions

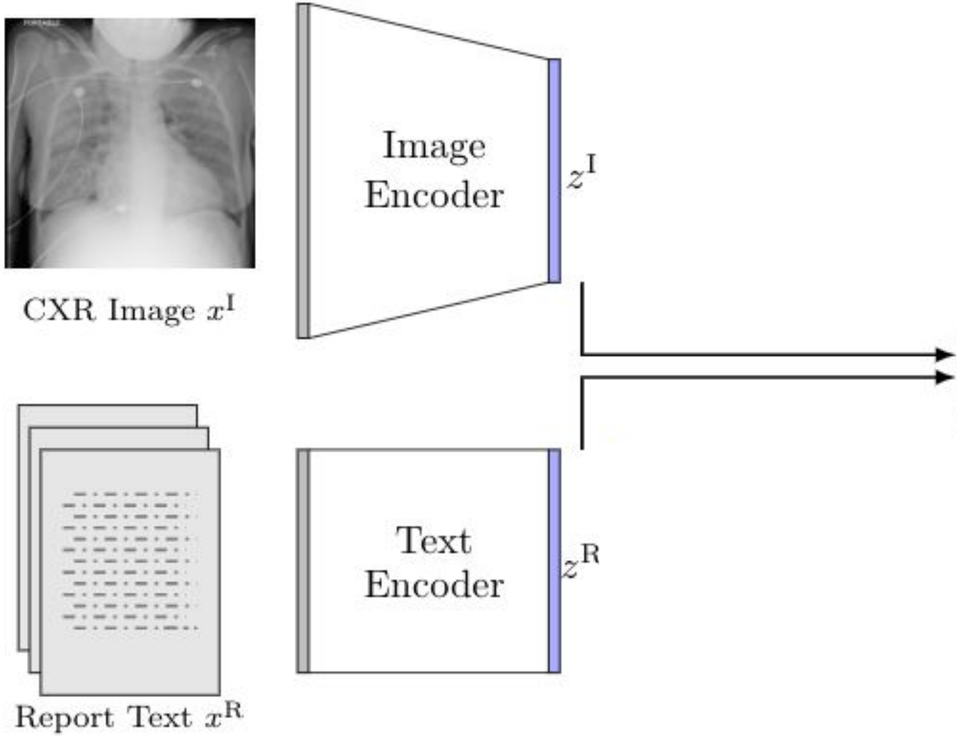
Mutual information (MI) quantifies statistical dependencies between two random variables



Maximize mutual information

$$I(z^{\mathrm{I}}, z^{\mathrm{R}})$$

Mutual information (MI) quantifies statistical dependencies between two random variables



Maximize mutual information

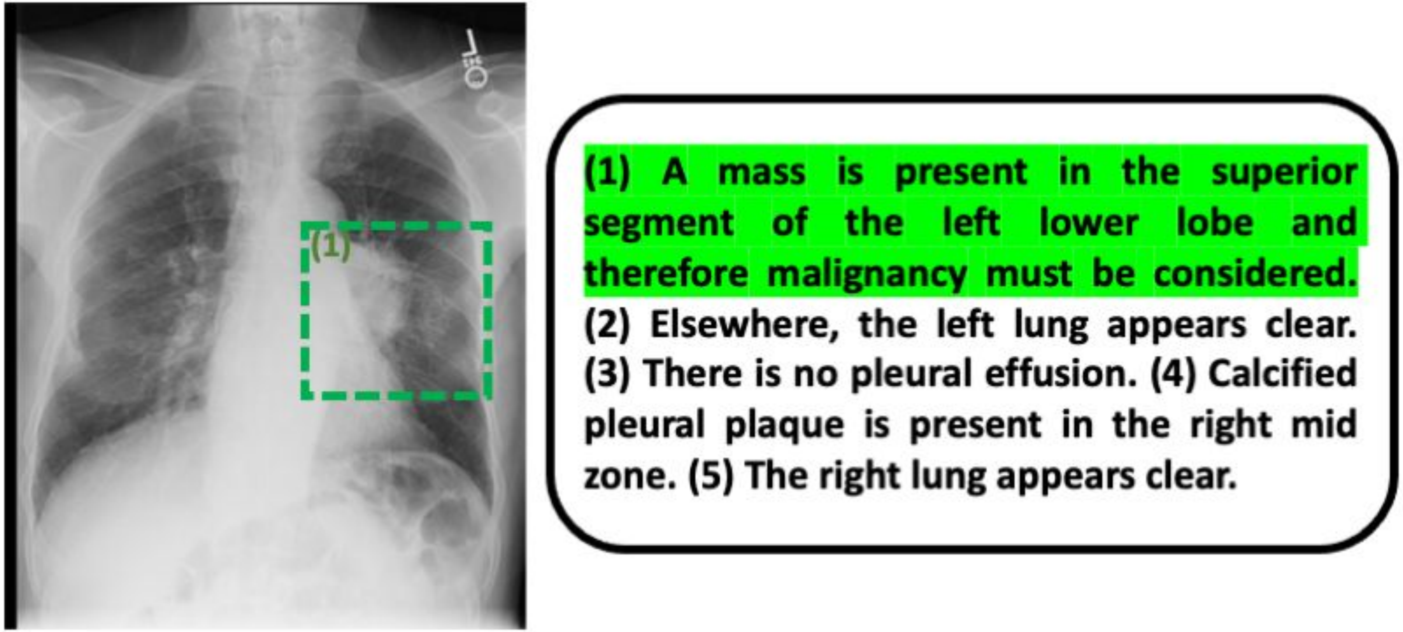
$$I(z^{\mathrm{I}}, z^{\mathrm{R}}) = H(z^{\mathrm{I}}) - H(z^{\mathrm{I}} | z^{\mathrm{R}})$$

Leveraging local correspondences: each sentence in the report describes the findings in a particular region of the image

(1) A mass is present in the superior segment of the left lower lobe and therefore malignancy must be considered. (2) Elsewhere, the left lung appears clear. (3) There is no pleural effusion. (4) Calcified pleural plaque is present in the right mid zone. (5) The right lung appears clear.

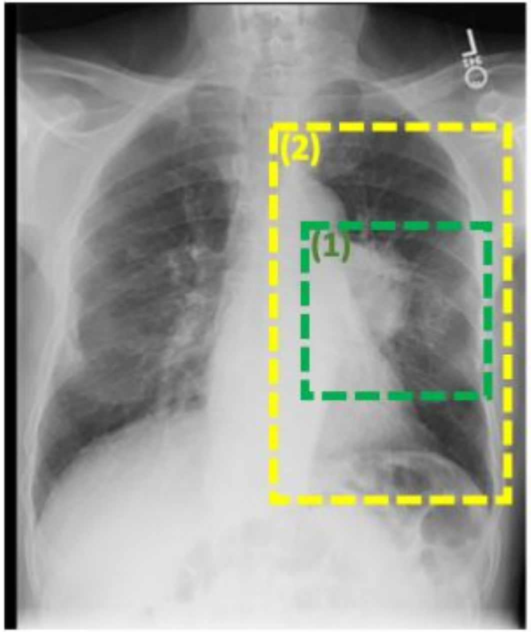
[Liao et al., 2021]

Leveraging local correspondences: each sentence in the report describes the findings in a particular region of the image



[Liao et al., 2021]

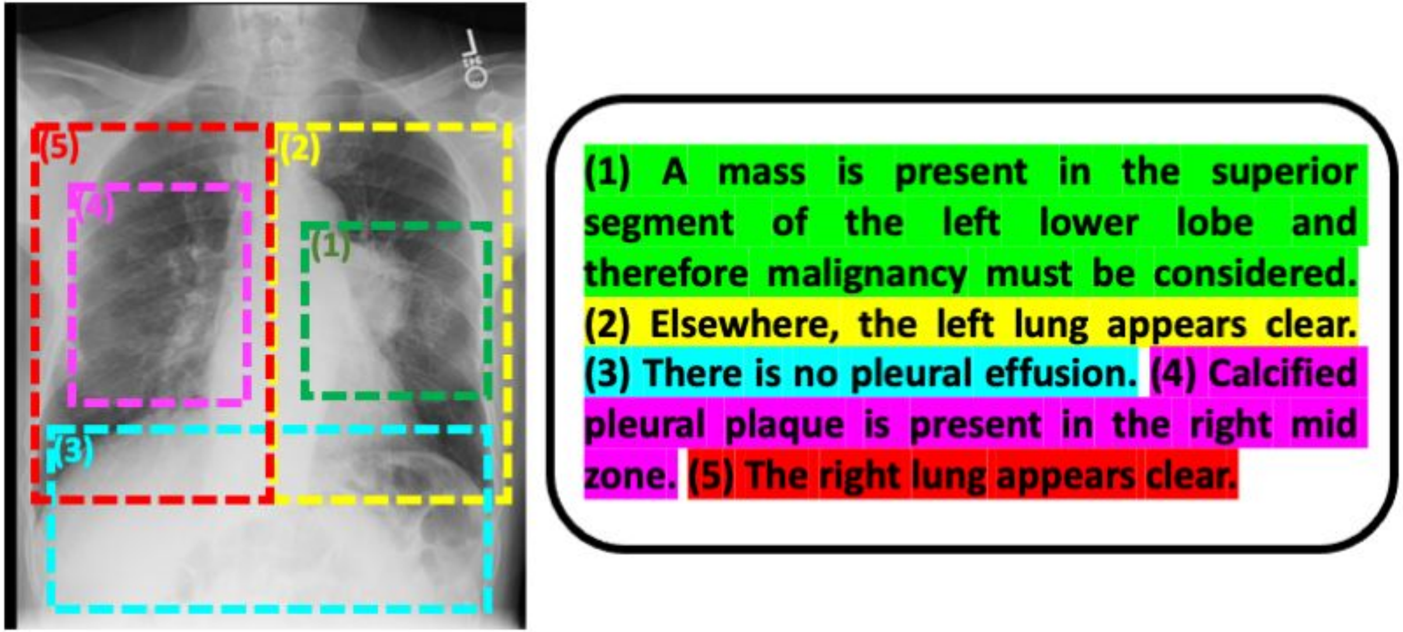
Leveraging local correspondences: each sentence in the report describes the findings in a particular region of the image



(1) A mass is present in the superior segment of the left lower lobe and therefore malignancy must be considered. (2) Elsewhere, the left lung appears clear. (3) There is no pleural effusion. (4) Calcified pleural plaque is present in the right mid zone. (5) The right lung appears clear.

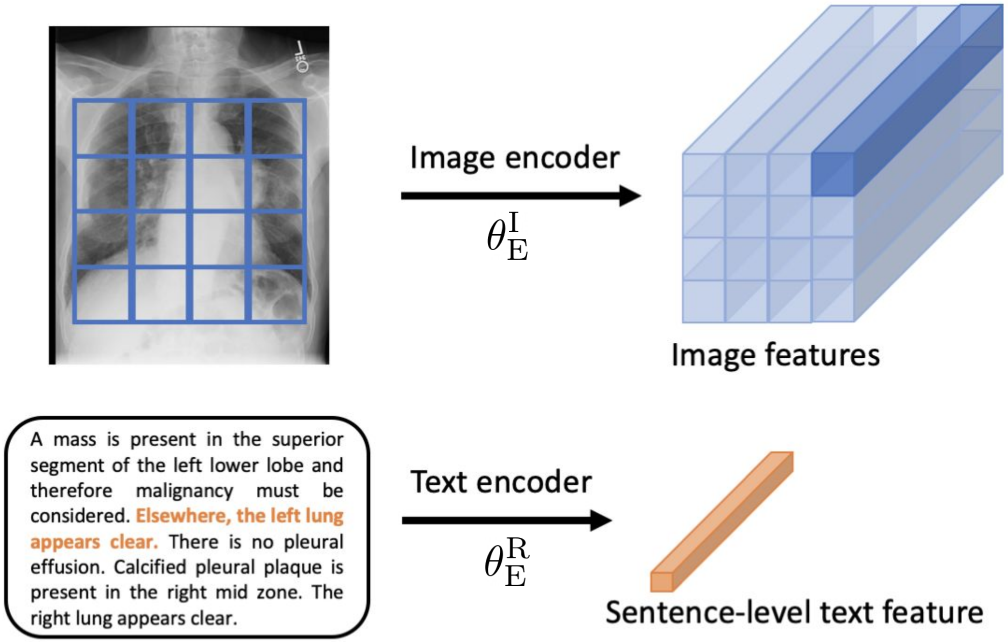
[Liao et al., 2021]

Leveraging local correspondences: each sentence in the report describes the findings in a particular region of the image

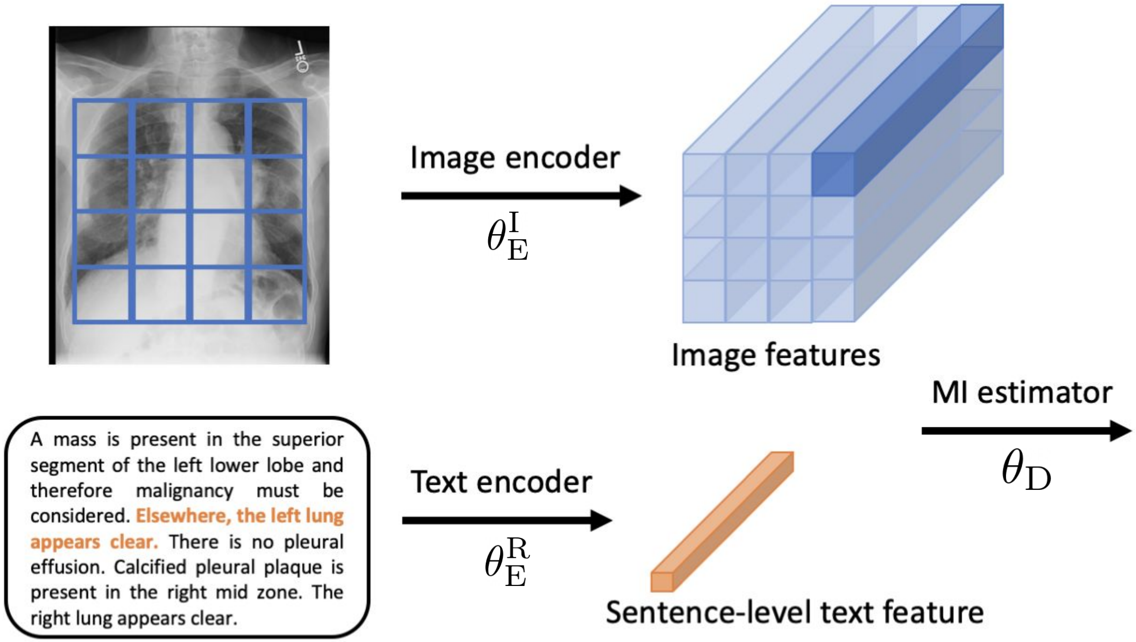


[Liao et al., 2021]

Maximize local mutual information

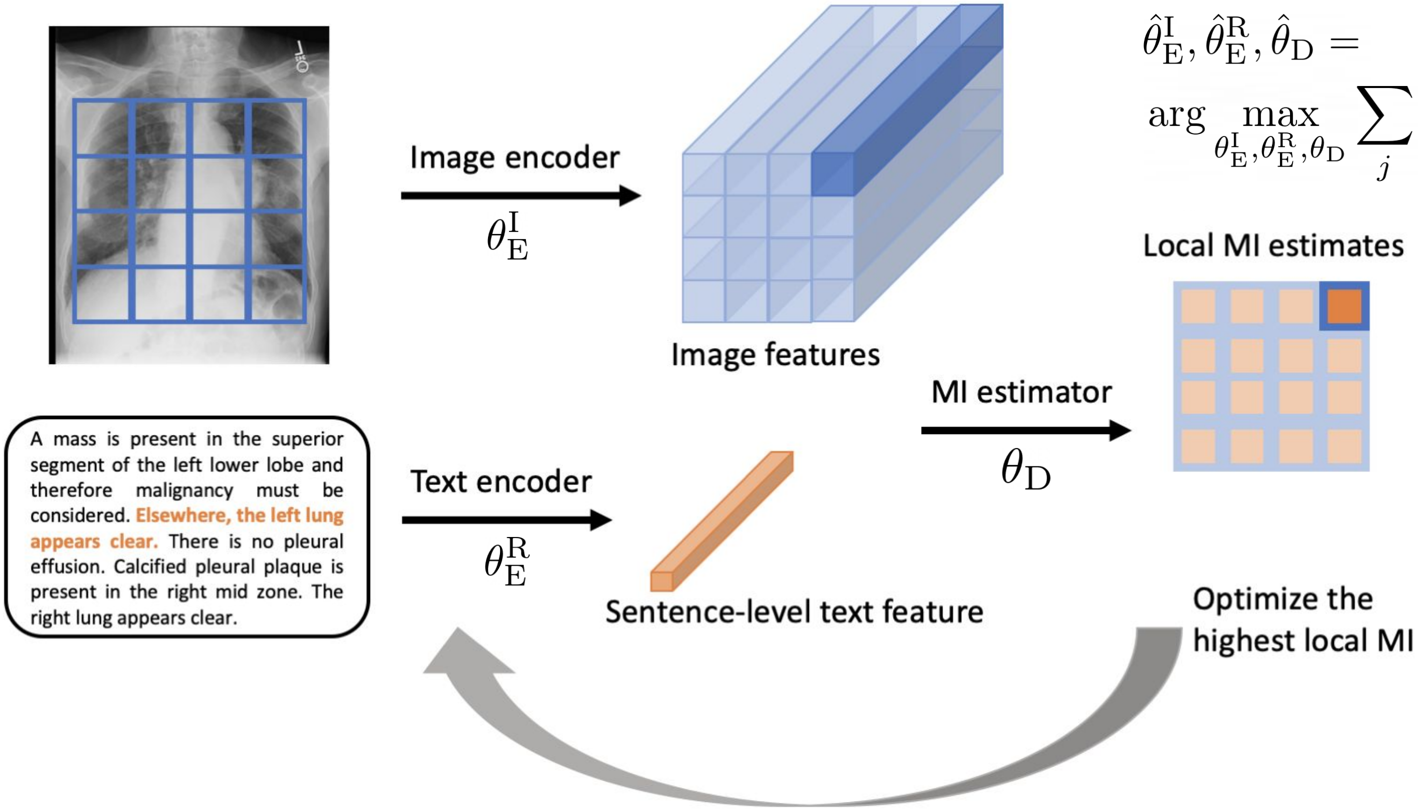


Maximize local mutual information



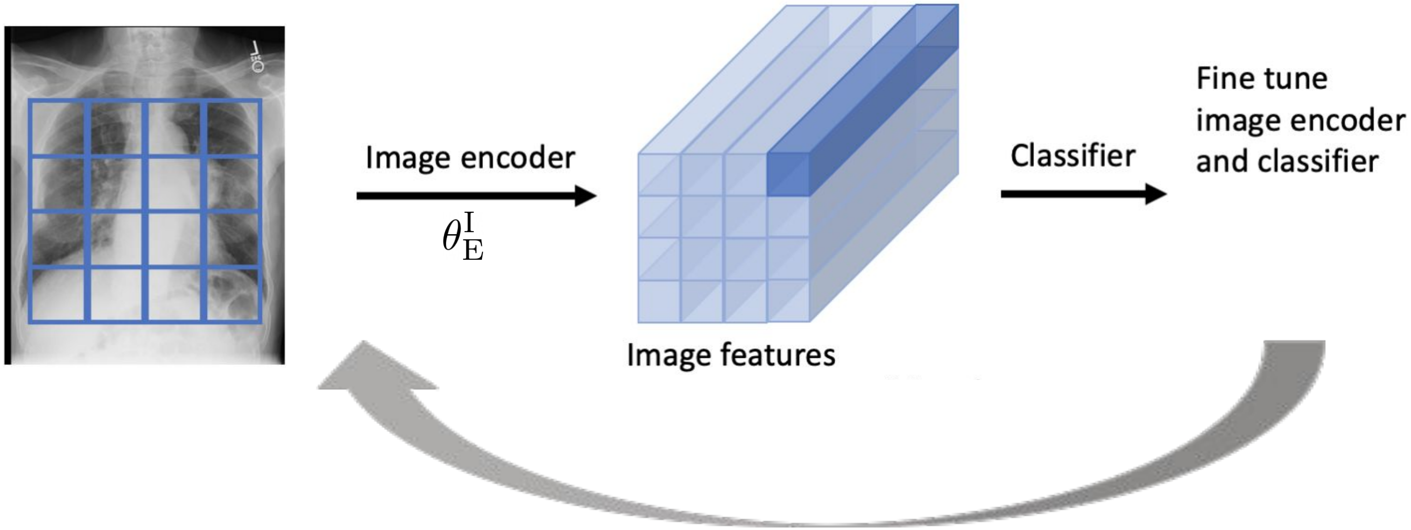
- **Mutual Information Neural Estimation (MINE)**: Donsker-Varadhan (DV) representation for the KL divergence as the lower bound
- **Contrastive Predictive Coding (CPC/infoNCE)**: Approximating the lower bound of the likelihood ratio

Maximize local mutual information



$$\hat{\theta}_{\mathrm{E}}^{\mathrm{I}}, \hat{\theta}_{\mathrm{E}}^{\mathrm{R}}, \hat{\theta}_{\mathrm{D}} = \arg\max_{\theta_{\mathrm{E}}^{\mathrm{I}}, \theta_{\mathrm{E}}^{\mathrm{R}}, \theta_{\mathrm{D}}} \sum_{j} \sum_{m} \max_{n} \hat{I}(\mathrm{z}_{j,(n)}^{\mathrm{I}}, \mathrm{z}_{j,(m)}^{\mathrm{R}})$$

Fine tune for downstream image classification



Results: local MI to learn joint image-text representation leads to the best performance of downstream image classification

Method	*Re-train* Encoder?	Level 0 vs 1,2,3		Level 0,1 vs 2,3		Level 0,1,2 vs 3	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.80		0.71		0.90	
global-mi	**frozen**	0.81	0.83	0.77	0.78	0.93	0.89
global-mi	**tuned**	0.81	0.82	0.79	0.81	0.93	0.93
local-mi	**frozen**	0.77	0.76	0.72	0.76	0.75	0.86
local-mi	**tuned**	**0.87**	0.83	0.83	**0.85**	**0.97**	0.93

Method	*Re-train* Encoder?	Atelectasis		Cardiomegaly		Consolidation	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.76		0.71		0.78	
global-mi	**frozen**	0.65	0.63	0.79	0.79	0.67	0.65
global-mi	**tuned**	0.74	0.77	0.81	0.81	0.81	0.82
local-mi	**frozen**	0.74	0.61	0.73	0.77	0.65	0.65
local-mi	**tuned**	0.73	**0.86**	0.82	**0.84**	**0.83**	**0.83**
–	–	Edema		Lung Opacity		Pleural Effusion	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	**0.89**		0.86		0.69	
global-mi	**frozen**	0.81	0.81	0.69	0.68	0.74	0.74
global-mi	**tuned**	0.87	0.88	0.83	0.84	0.90	0.90
local-mi	**frozen**	0.78	0.80	0.66	0.69	0.69	0.72
local-mi	**tuned**	**0.89**	**0.89**	0.82	**0.88**	**0.92**	**0.92**
–	–	Pneumonia		Pneumothorax		Support Devices	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.75		0.65		0.72	
global-mi	**frozen**	0.71	0.70	0.65	0.66	0.70	0.68
global-mi	**tuned**	0.75	0.76	0.75	0.77	0.77	0.79
local-mi	**frozen**	0.61	0.66	0.70	0.67	0.72	0.74
local-mi	**tuned**	0.78	**0.79**	**0.79**	0.76	**0.87**	0.81

[Liao et al., 2021]

Local MI to learn joint image-text representation leads to the best performance of downstream image classification

Method	*Re-train* Encoder?	Level 0 vs 1,2,3		Level 0,1 vs 2,3		Level 0,1,2 vs 3	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.80		0.71		0.90	
global-mi	**frozen**	0.81	0.83	0.77	0.78	0.93	0.89
global-mi	**tuned**	0.81	0.82	0.79	0.81	0.93	0.93
local-mi	**frozen**	0.77	0.76	0.72	0.76	0.75	0.86
local-mi	**tuned**	**0.87**	0.83	0.83	**0.85**	**0.97**	0.93

Method	*Re-train* Encoder?	Atelectasis		Cardiomegaly		Consolidation	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.76		0.71		0.78	
global-mi	**frozen**	0.65	0.63	0.79	0.79	0.67	0.65
global-mi	**tuned**	0.74	0.77	0.81	0.81	0.81	0.82
local-mi	**frozen**	0.74	0.61	0.73	0.77	0.65	0.65
local-mi	**tuned**	0.73	**0.86**	0.82	**0.84**	**0.83**	**0.83**
–	–	Edema		Lung Opacity		Pleural Effusion	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	**0.89**		0.86		0.69	
global-mi	**frozen**	0.81	0.81	0.69	0.68	0.74	0.74
global-mi	**tuned**	0.87	0.88	0.83	0.84	0.90	0.90
local-mi	**frozen**	0.78	0.80	0.66	0.69	0.69	0.72
local-mi	**tuned**	**0.89**	**0.89**	0.82	**0.88**	**0.92**	**0.92**
–	–	Pneumonia		Pneumothorax		Support Devices	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.75		0.65		0.72	
global-mi	**frozen**	0.71	0.70	0.65	0.66	0.70	0.68
global-mi	**tuned**	0.75	0.76	0.75	0.77	0.77	0.79
local-mi	**frozen**	0.61	0.66	0.70	0.67	0.72	0.74
local-mi	**tuned**	0.78	**0.79**	**0.79**	0.76	**0.87**	0.81

[Liao et al., 2021]

Local MI to learn joint image-text representation leads to the best performance of downstream image classification

Method	*Re-train* Encoder?	Level 0 vs 1,2,3		Level 0,1 vs 2,3		Level 0,1,2 vs 3	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.80		0.71		0.90	
global-mi	**frozen**	0.81	0.83	0.77	0.78	0.93	0.89
global-mi	**tuned**	0.81	0.82	0.79	0.81	0.93	0.93
local-mi	**frozen**	0.77	0.76	0.72	0.76	0.75	0.86
local-mi	**tuned**	**0.87**	0.83	0.83	**0.85**	**0.97**	0.93

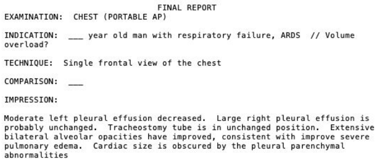
Method	*Re-train* Encoder?	Atelectasis		Cardiomegaly		Consolidation	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.76		0.71		0.78	
global-mi	**frozen**	0.65	0.63	0.79	0.79	0.67	0.65
global-mi	**tuned**	0.74	0.77	0.81	0.81	0.81	0.82
local-mi	**frozen**	0.74	0.61	0.73	0.77	0.65	0.65
local-mi	**tuned**	0.73	**0.86**	0.82	**0.84**	**0.83**	**0.83**
–	–	Edema		Lung Opacity		Pleural Effusion	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	**0.89**		0.86		0.69	
global-mi	**frozen**	0.81	0.81	0.69	0.68	0.74	0.74
global-mi	**tuned**	0.87	0.88	0.83	0.84	0.90	0.90
local-mi	**frozen**	0.78	0.80	0.66	0.69	0.69	0.72
local-mi	**tuned**	**0.89**	**0.89**	0.82	**0.88**	**0.92**	**0.92**
–	–	Pneumonia		Pneumothorax		Support Devices	
–	–	CPC	MINE	CPC	MINE	CPC	MINE
image-only	N/A	0.75		0.65		0.72	
global-mi	**frozen**	0.71	0.70	0.65	0.66	0.70	0.68
global-mi	**tuned**	0.75	0.76	0.75	0.77	0.77	0.79
local-mi	**frozen**	0.61	0.66	0.70	0.67	0.72	0.74
local-mi	**tuned**	0.78	**0.79**	**0.79**	0.76	**0.87**	0.81

Advantages of local MI

- Better fit to image-text structure
- Better optimization landscape
- Better representation fit to downstream tasks

[Liao et al., 2021]

Conclusions



FINAL REPORT
EXAMINATION: CHEST (PORTABLE AP)

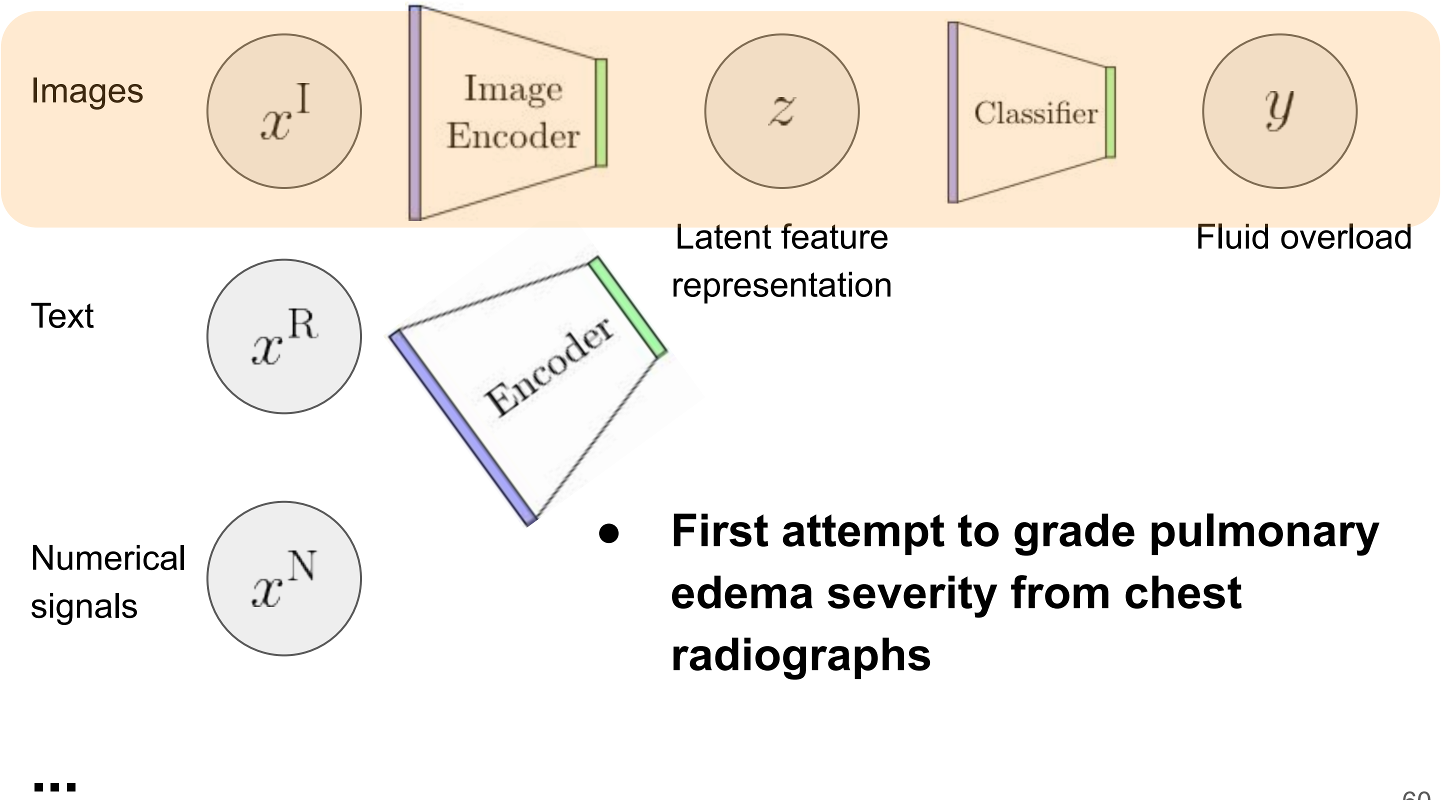
INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

IMPRESSION:

Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities



- **First attempt to grade pulmonary edema severity from chest radiographs**

[Liao et al., 2019, Horng*, Liao* et al., 2021]

Conclusions

FINAL REPORT

EXAMINATION: CHEST (PORTABLE AP)

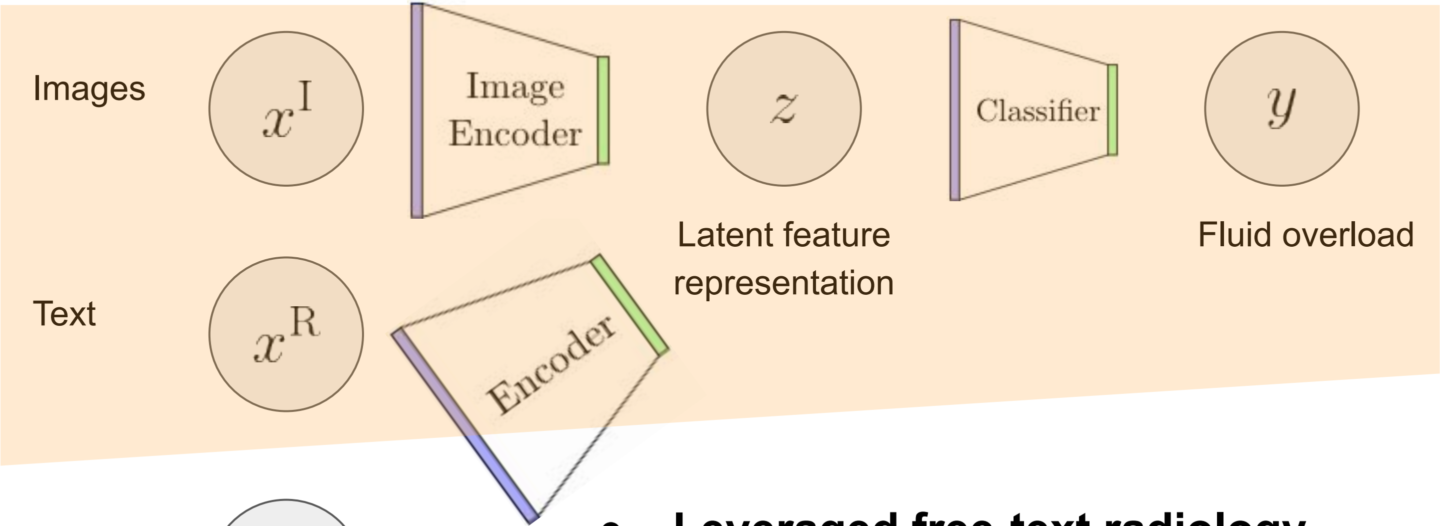
INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

IMPRESSION:

Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities



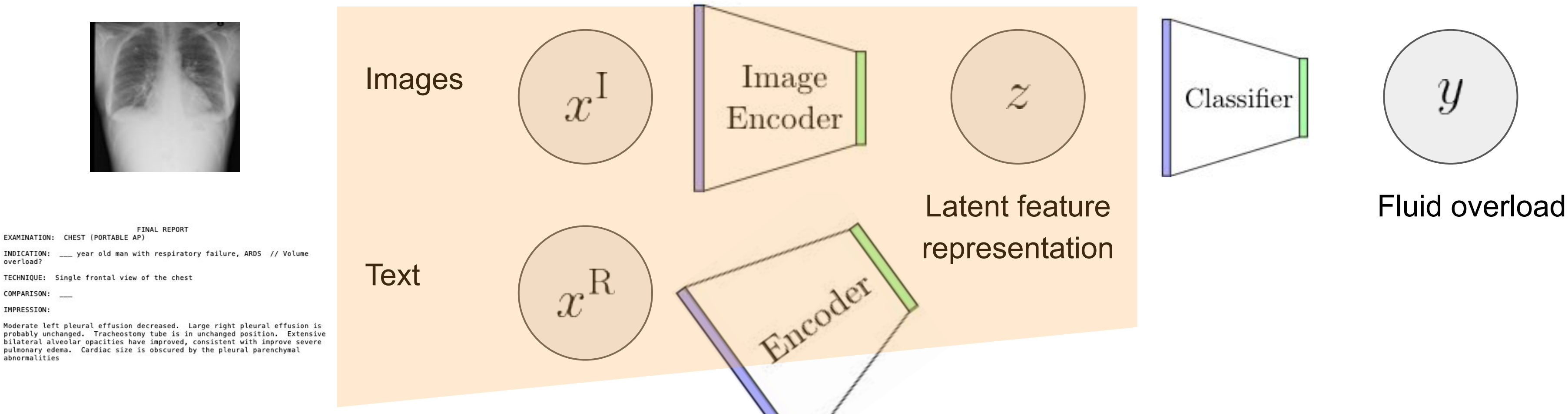
Numerical signals

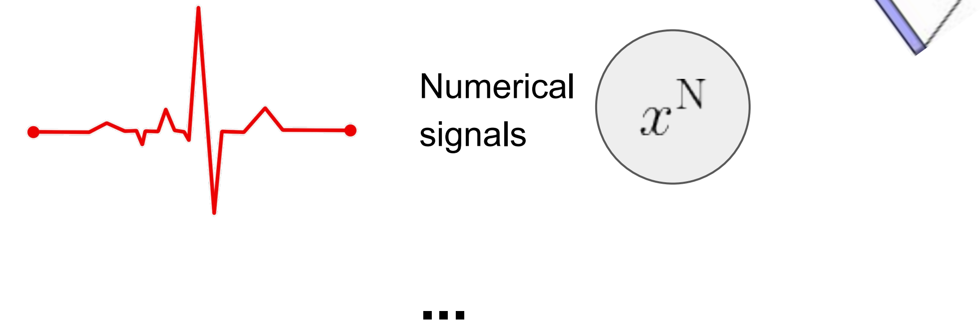
...

- **Leveraged free-text radiology reports to improve the image model performance**

[Chauhan*, Liao* et al., 2020]

Conclusions





- **Developed a novel multimodal representation learning approach**

[Liao et al., 2020, Liao et al., 2021]

Conclusions

FINAL REPORT
EXAMINATION: CHEST (PORTABLE AP)

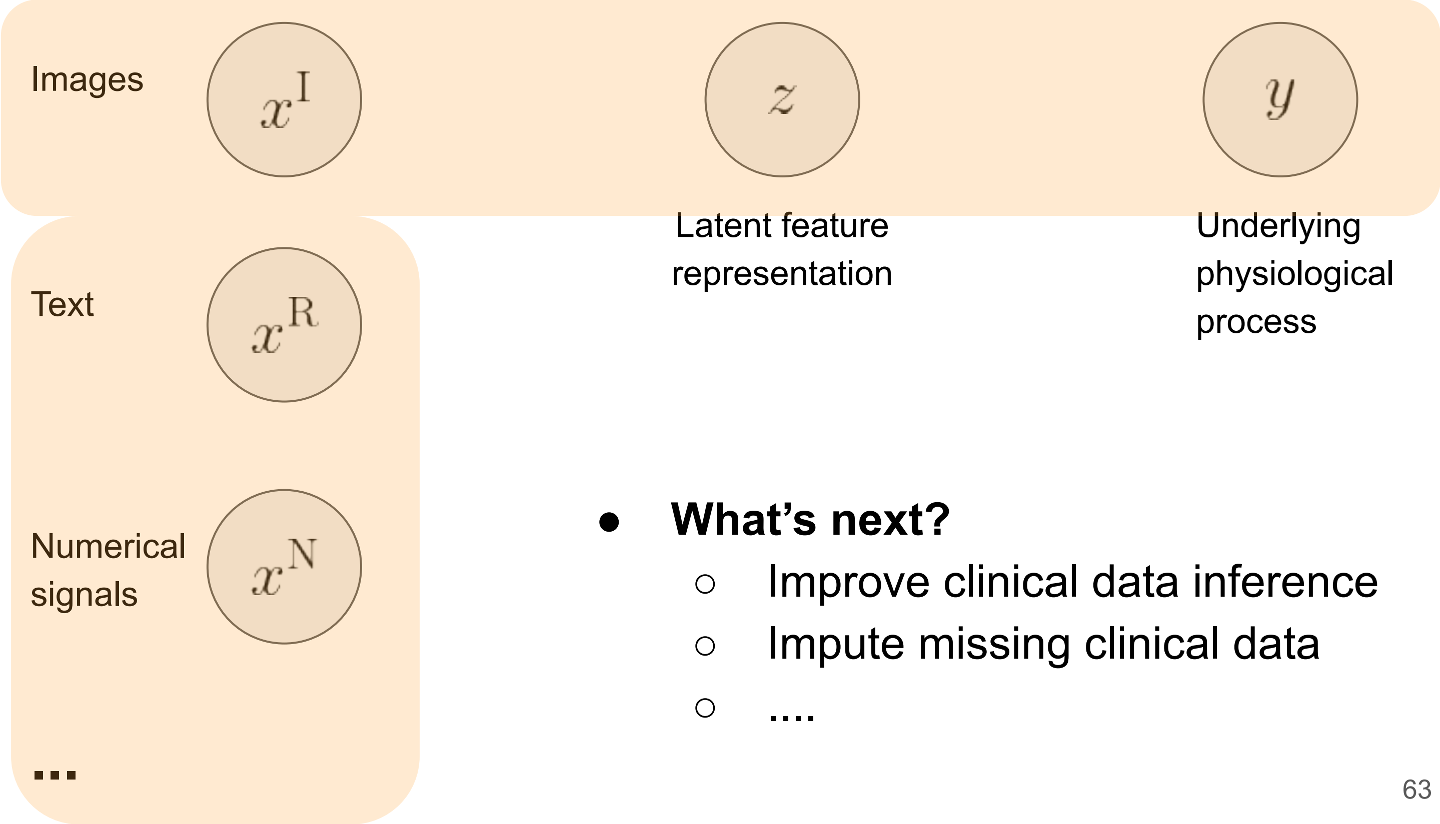
INDICATION: ___ year old man with respiratory failure, ARDS // Volume overload?

TECHNIQUE: Single frontal view of the chest

COMPARISON: ___

IMPRESSION:

Moderate left pleural effusion decreased. Large right pleural effusion is probably unchanged. Tracheostomy tube is in unchanged position. Extensive bilateral alveolar opacities have improved, consistent with improve severe pulmonary edema. Cardiac size is obscured by the pleural parenchymal abnormalities



- **What's next?**
  - Improve clinical data inference
  - Impute missing clinical data
  - ....

Multimodal Representation Learning for Medical Image Analysis

Ruizhi "Ray" Liao

ruizhi@mit.edu

@rayruizhiliao