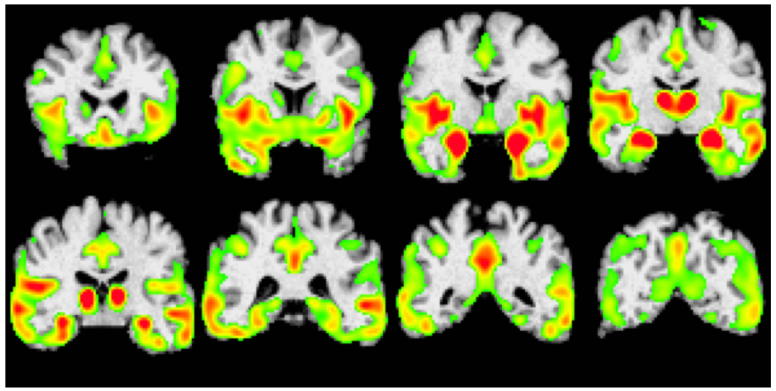
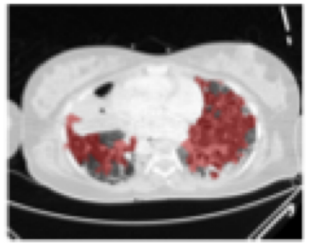
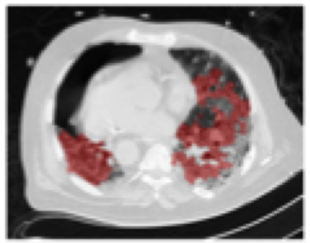
Generative-Discriminative Multi-Modal Learning

Kayhan Batmanghelich

Time

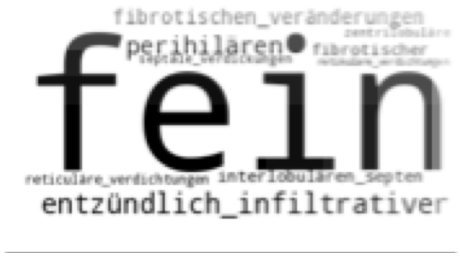


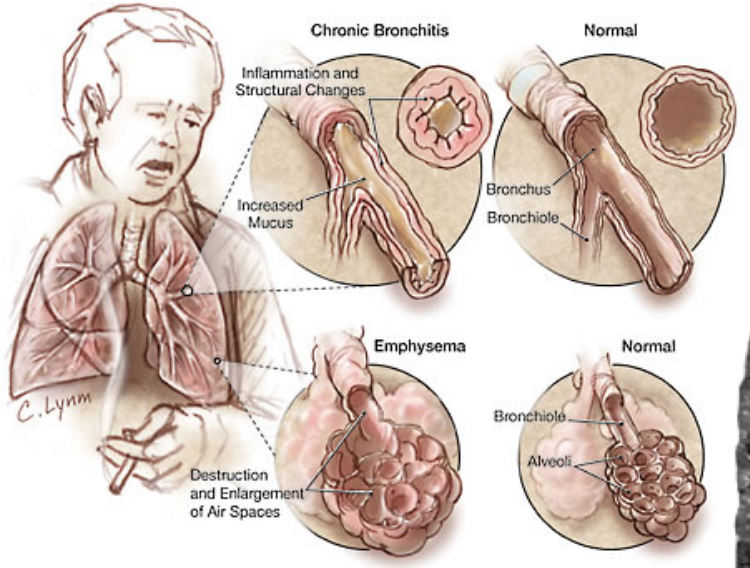


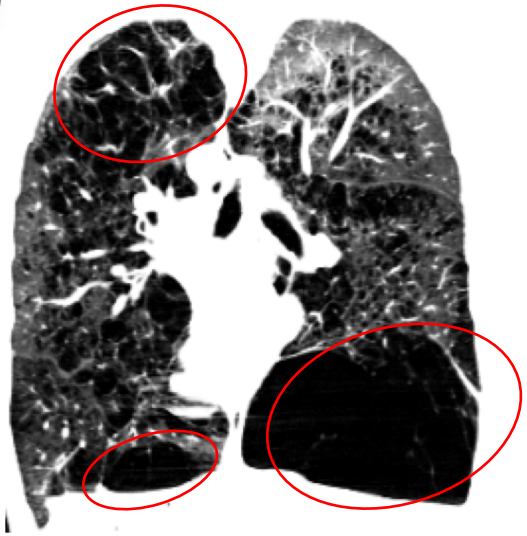


finely reticular

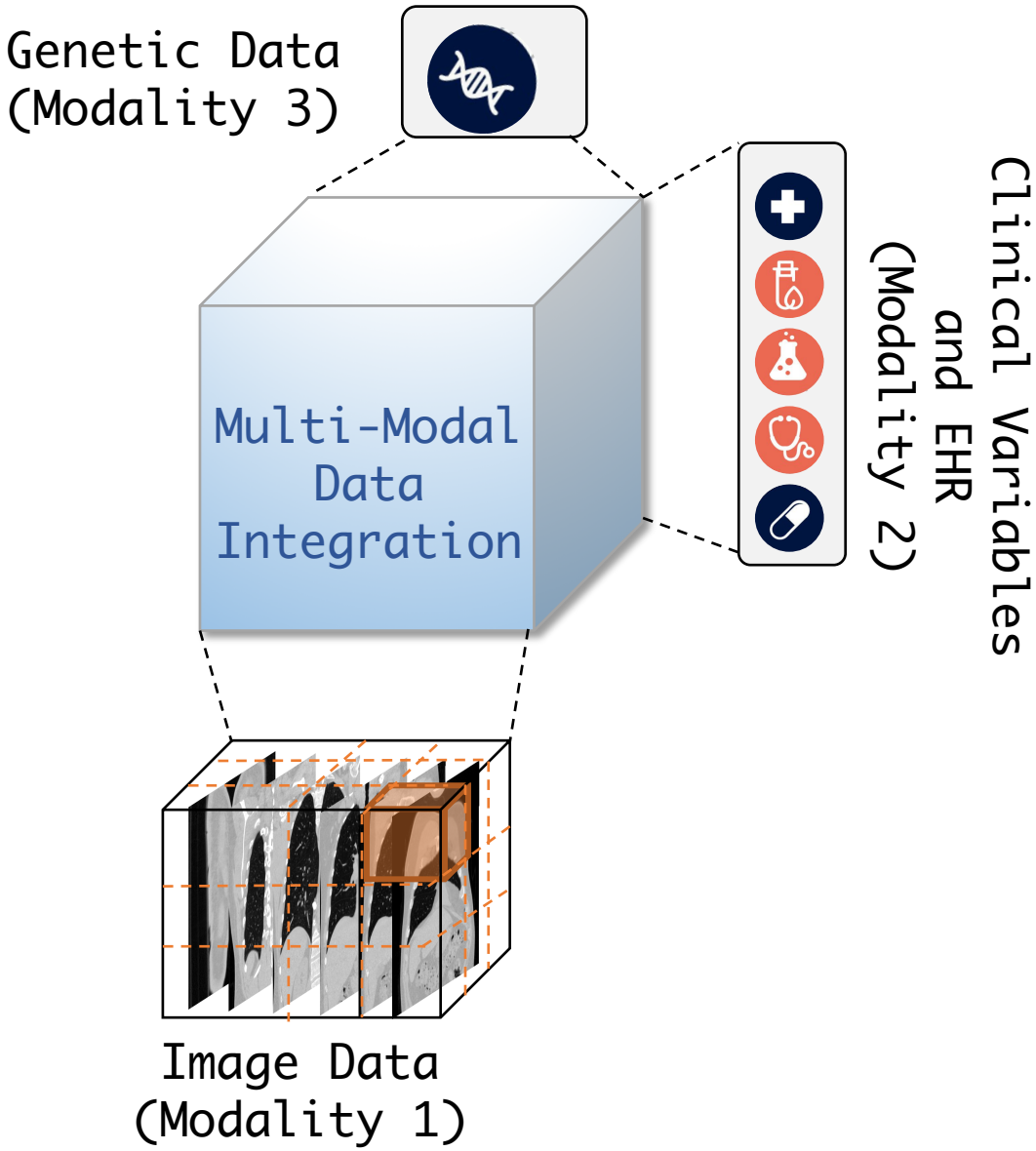
honeycombing
uip fibrotic lungdisease
areas





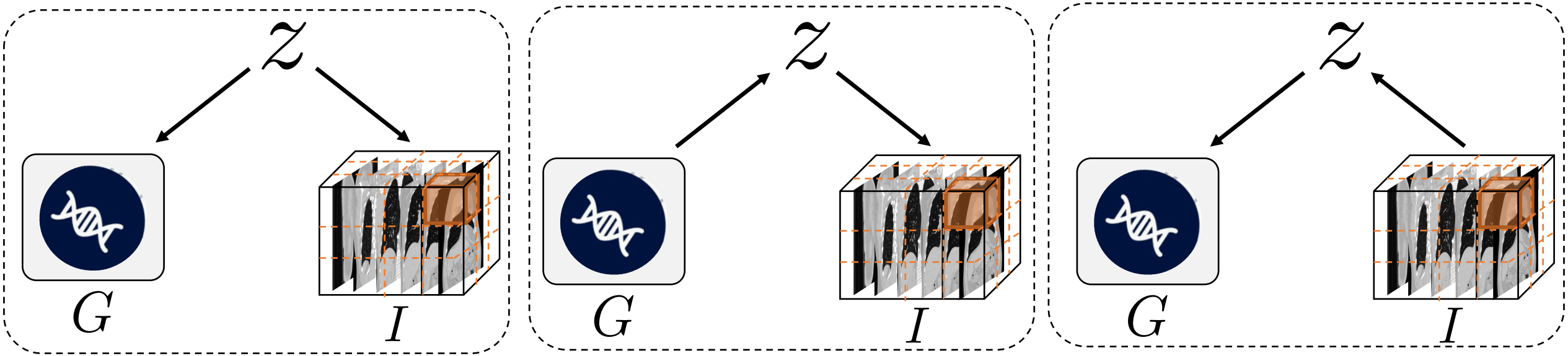


Multi-Modal Data Integration



Challenges

- How to model relationship between modalities:
  - Generative Approach
  - Conditional (Discriminative or Causal)



Challenges

- How to model relationship between modalities
- Interpretability

Model A (Acc=80.1%)

Model B (Acc=79%)

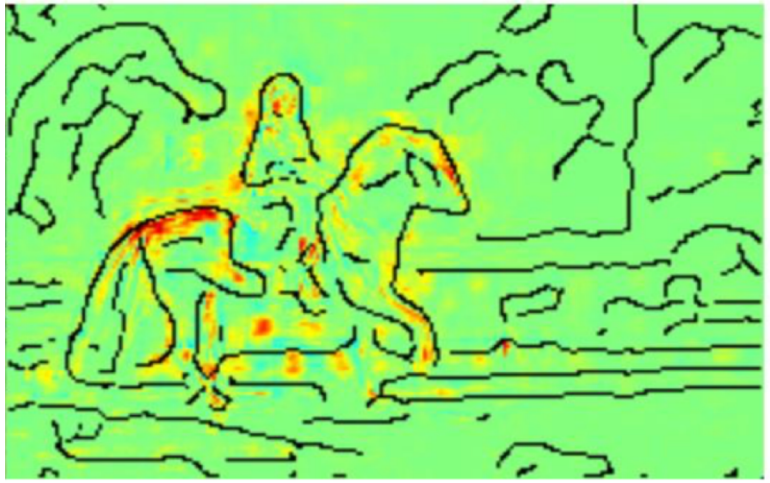


Figure credit: Klaus-Robert Müller, “Tutorial on Interpreting and Explaining Deep Models in Computer Vision ” c

Challenges

- How to model relationship between modalities
- Interpretability
- Robustness

Figure credit: https://the-modeling-agency.com/data-messy-dont-panic/

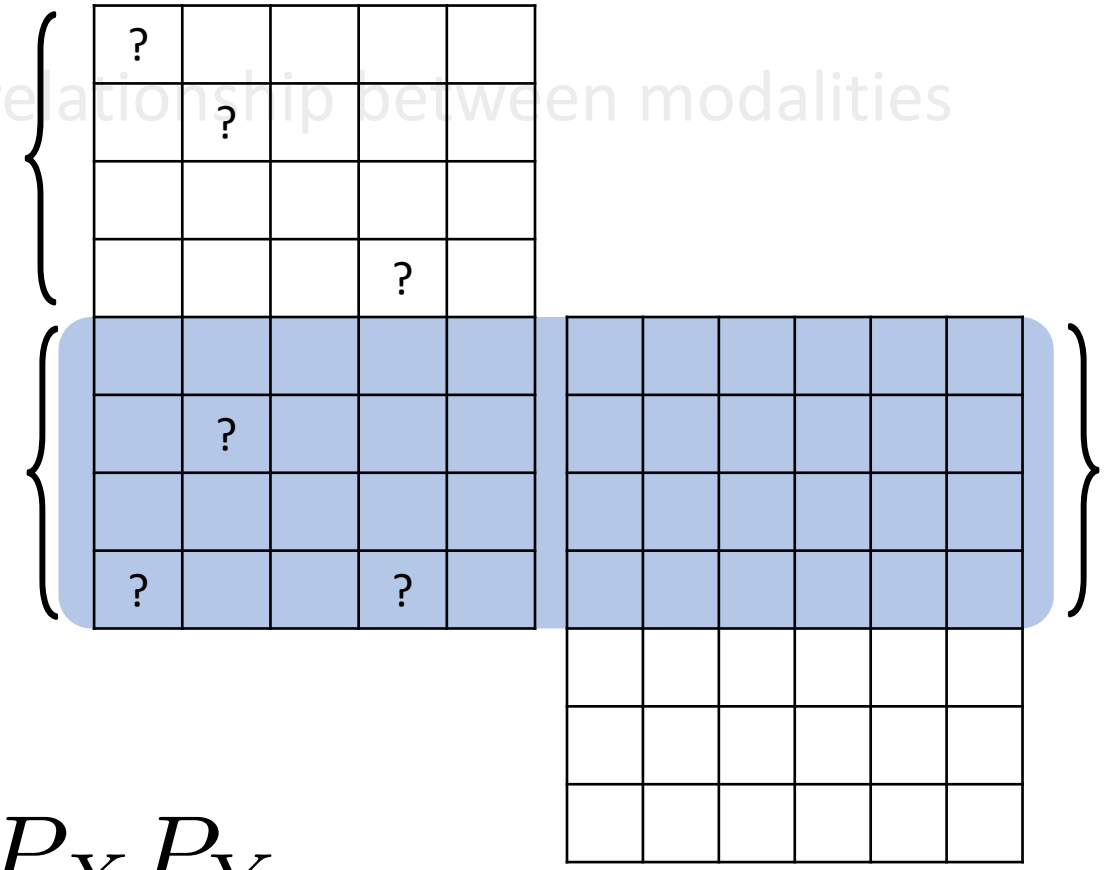
enges

- How to model relationship between modalities
- Interpretability
- Robustness

$X =$ 

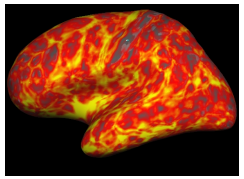
Hospital 2
(Domain 2)

Hospital 2
(Domain 2)



$$\mathcal{H}_0 : P_{XY} = P_X P_Y$$

$$\mathcal{H}_1 : P_{XY} \neq P_X P_Y$$

$Y =$ 

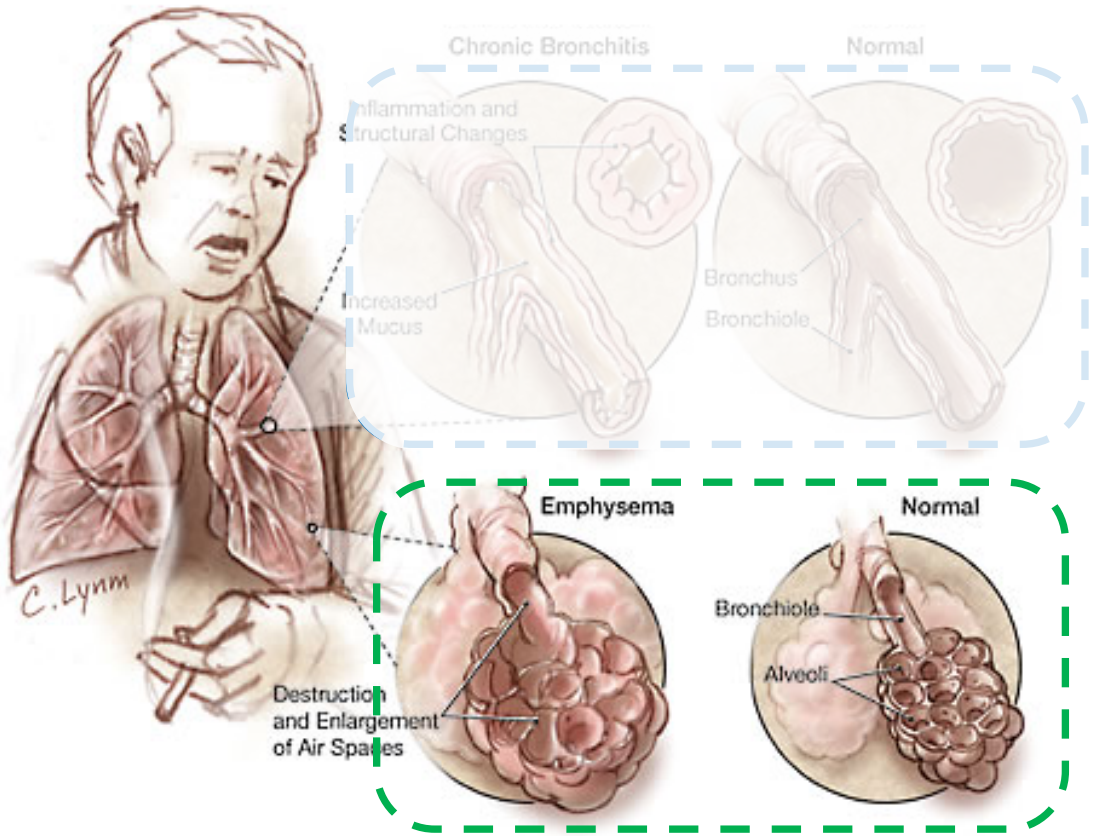
Outlines

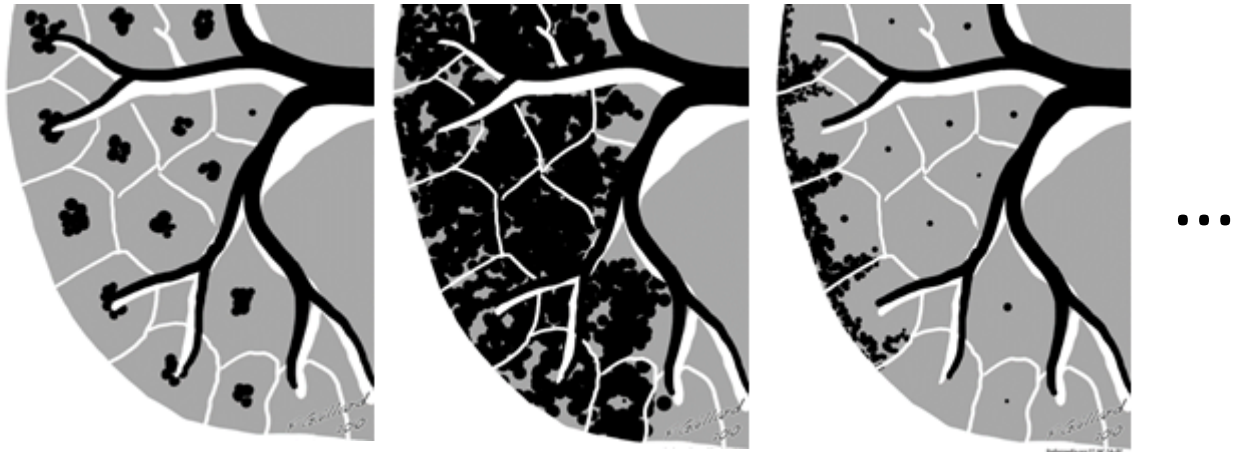
- Our previous research
  - Background
  - Pre-Deep Learning work
- Current Research
  - Hybrid Generative Discriminative Model
  - Some preliminaries on causal domain adaptation
- Future Directions

Outlines

- Our previous research
  - Background
  - Pre-Deep Learning work
- Current Research
  - Hybrid Generative Discriminative Model
  - Some preliminaries on causal domain adaptation
- Future Directions

Chronic Obstructive Pulmonary Disease (COPD)





Abnormal (Type 1) Abnormal (Type 2) Abnormal (Type 3) ...

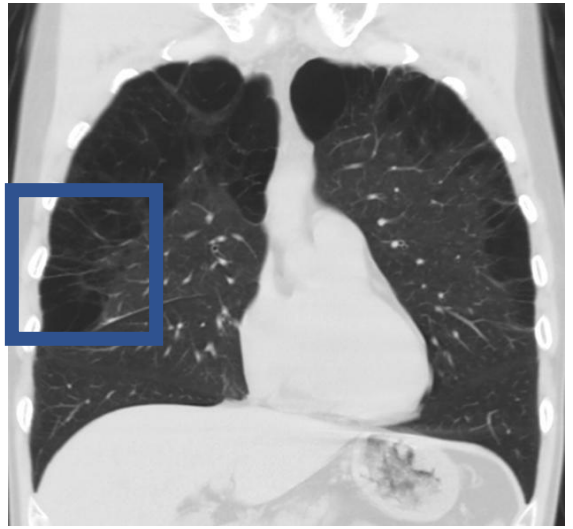
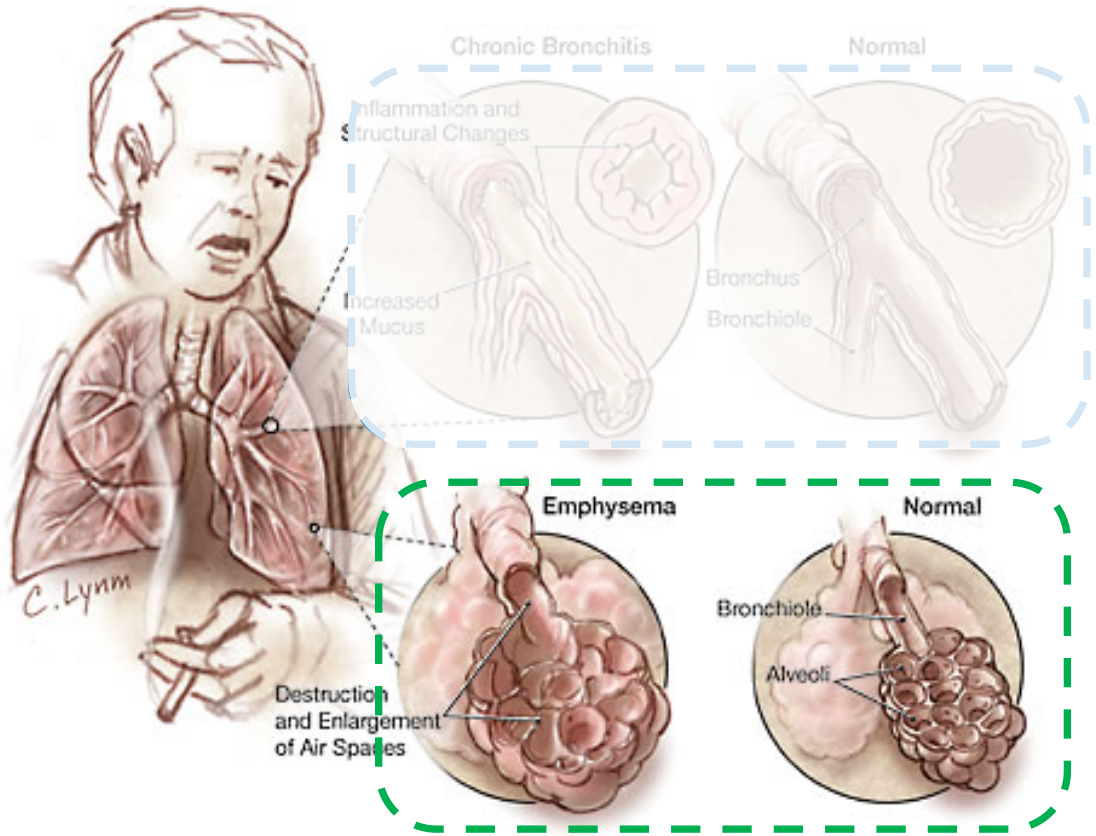
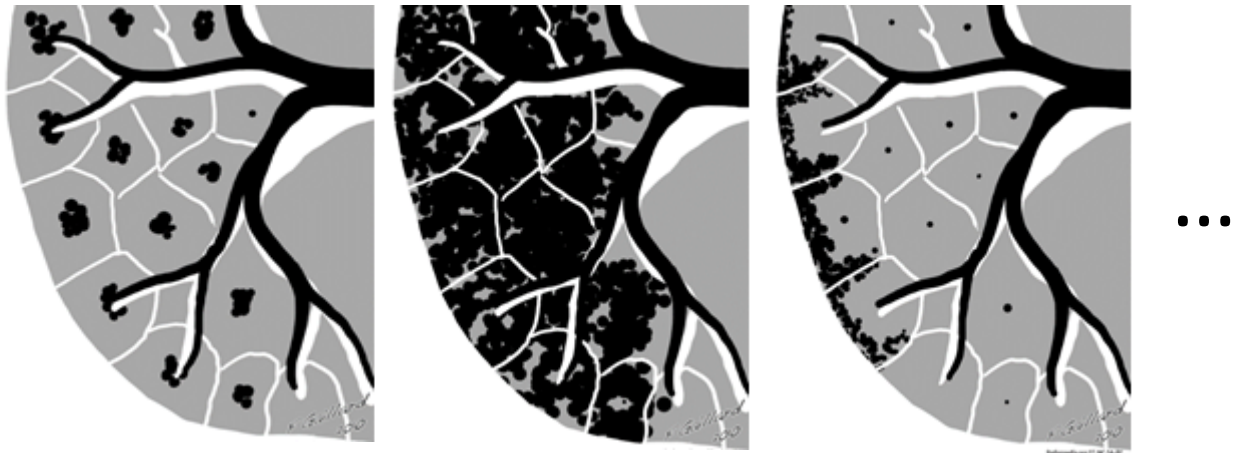


Image Reference: radiopedia.com

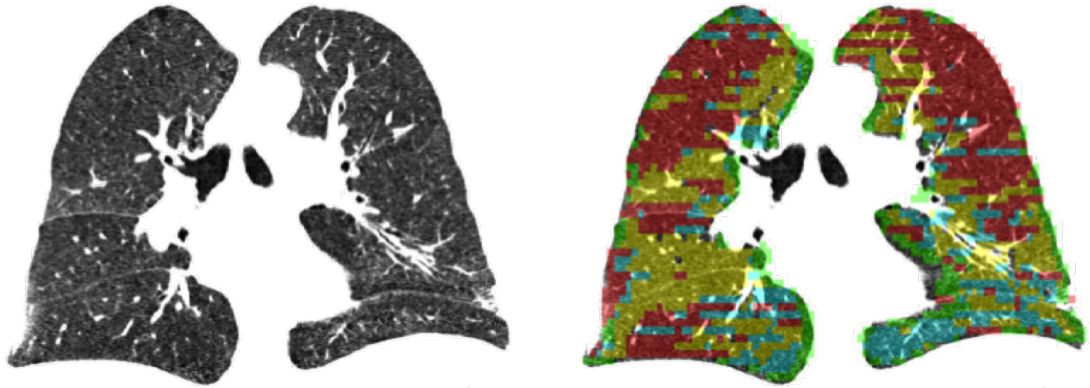
Chronic Obstructive Pulmonary Disease (COPD)





Abnormal (Type 1) Abnormal (Type 2) Abnormal (Type 3)

...



Normal Tissue
Abnormal (Type 1)
Abnormal (Type 2)
Abnormal (Type 3)
Abnormal (Type 4)

Carlos Mendoza, Raul San Jose Estepar et al. 2012

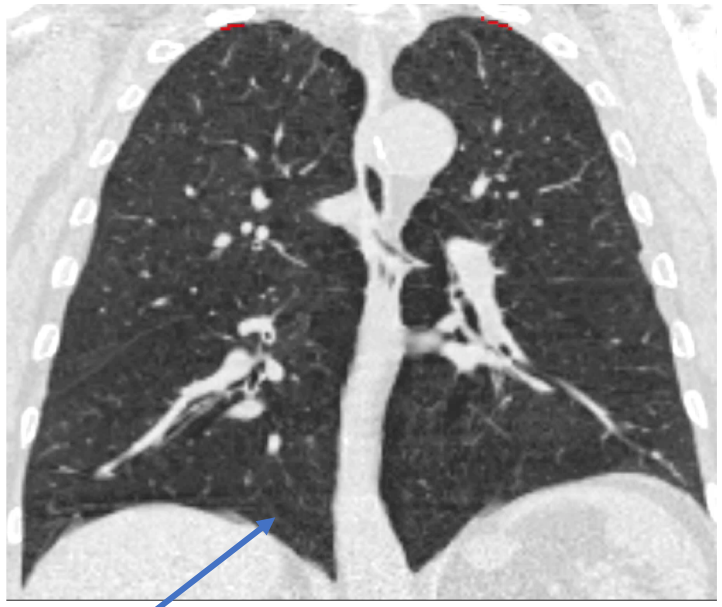
Image Reference: radiopedia.com

Outlines

- Our previous research
  - Background
  - Pre-Deep Learning work
- Current Research
  - Hybrid Generative Discriminative Model
  - Some preliminaries on causal domain adaptation
- Future Directions

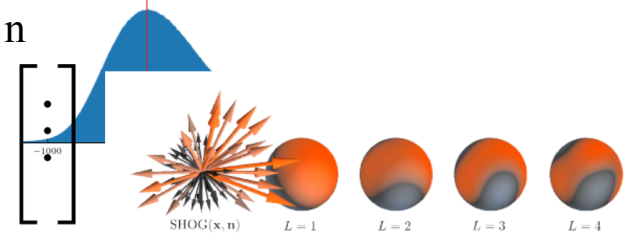
Imaging and Genetic Data

Patient s

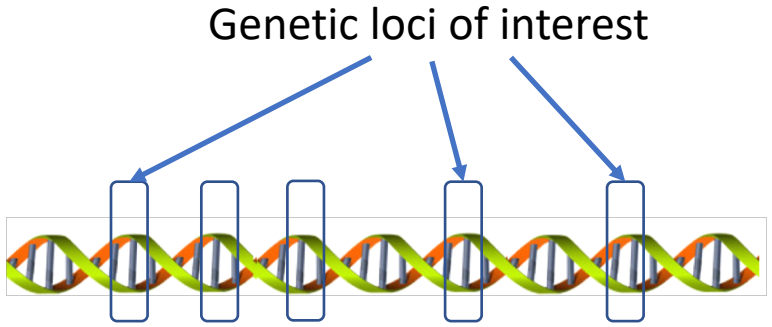


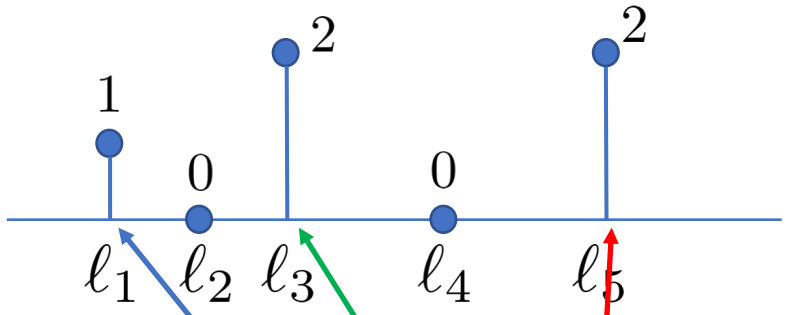
Imaging signature of Supervoxel n

$$I_{sn} =$$



Genetic loci of interest

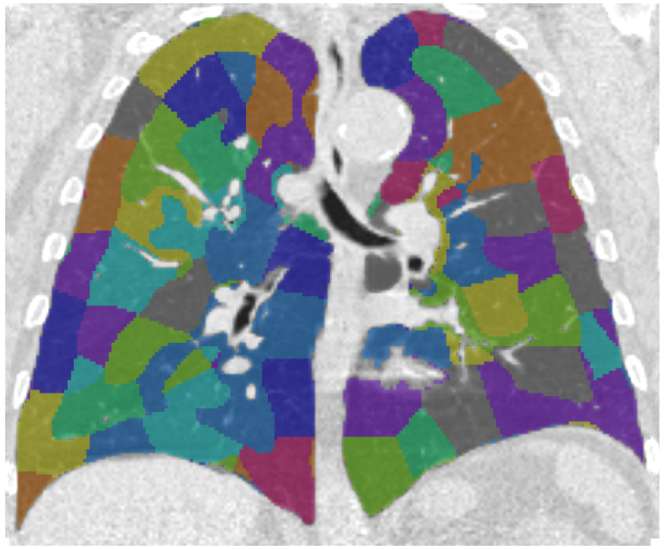




$$G_s = \{\ell_1, \ell_3, \ell_3, \ell_5, \ell_5\}$$

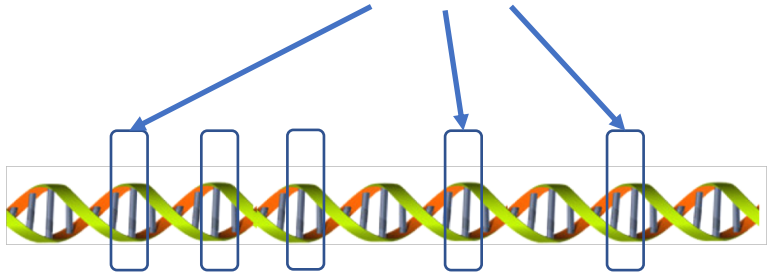
Bag of Words Model

Patient s



Visual Words (I_{sn})

Interesting Genetic loci



$$G_s = \{\ell_1, \ell_3, \ell_3, \ell_5, \ell_5\}$$

Genetic Words
(Genetic variants)

Patient ⟷ Document

A Simplified Model

Simplified version of our model

$$p_i(x) = \sum_{k=1}^{K} \pi_k(i) g(x; \theta_k)$$

Probability density of the patterns in patient i

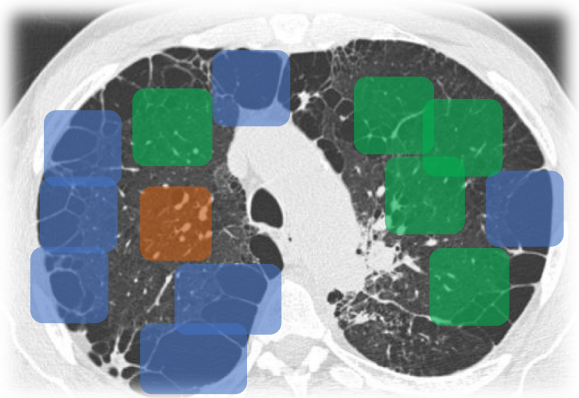
Should be chosen based on the local image descriptor

Patient Representation

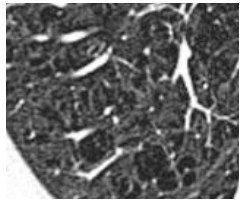
Pattern 1

Pattern 2

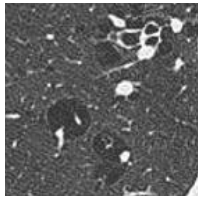
Pattern 3



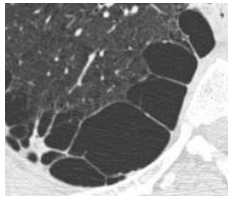
Pattern 1 (θ_1)



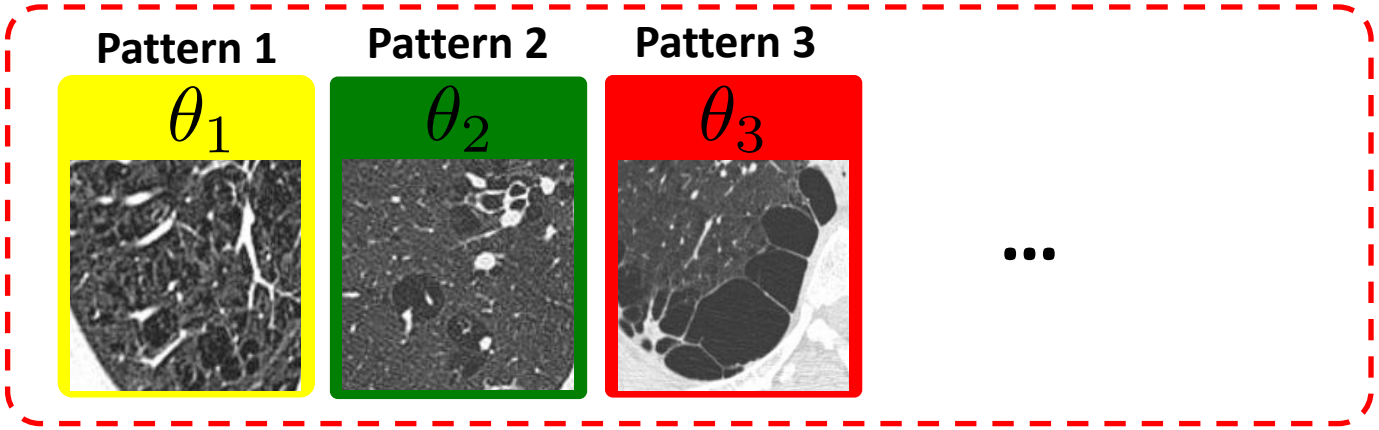
Pattern 2 (θ_2)

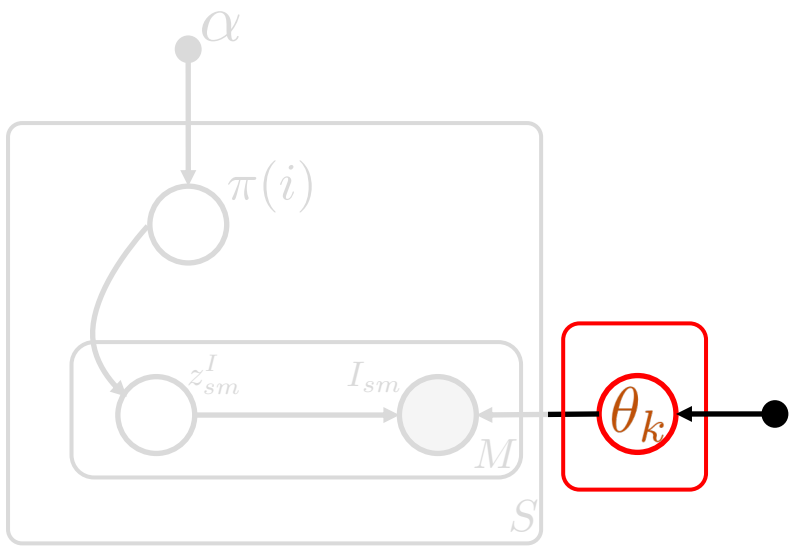


Pattern 3 (θ_3)



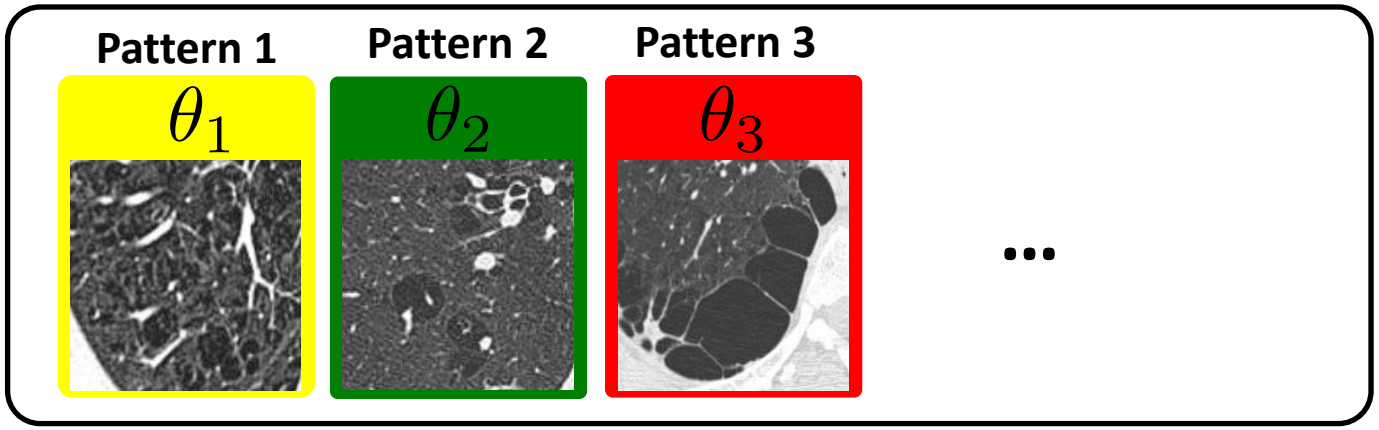
Probabilistic Modeling

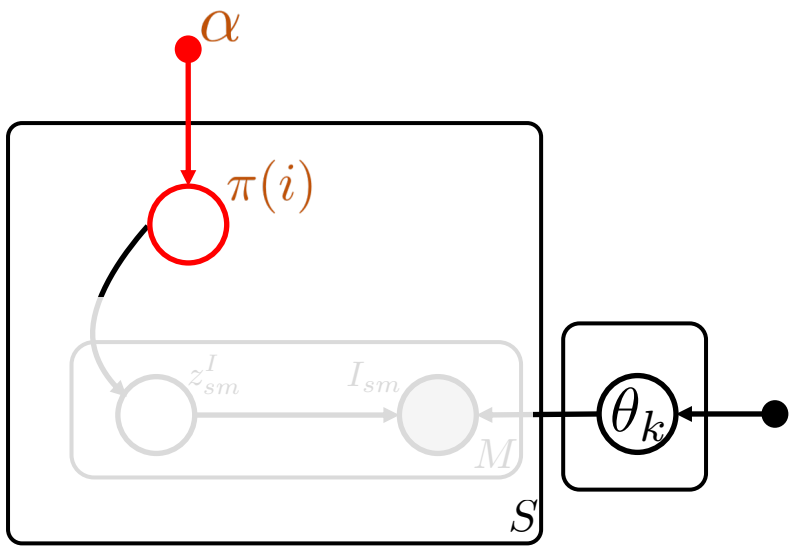


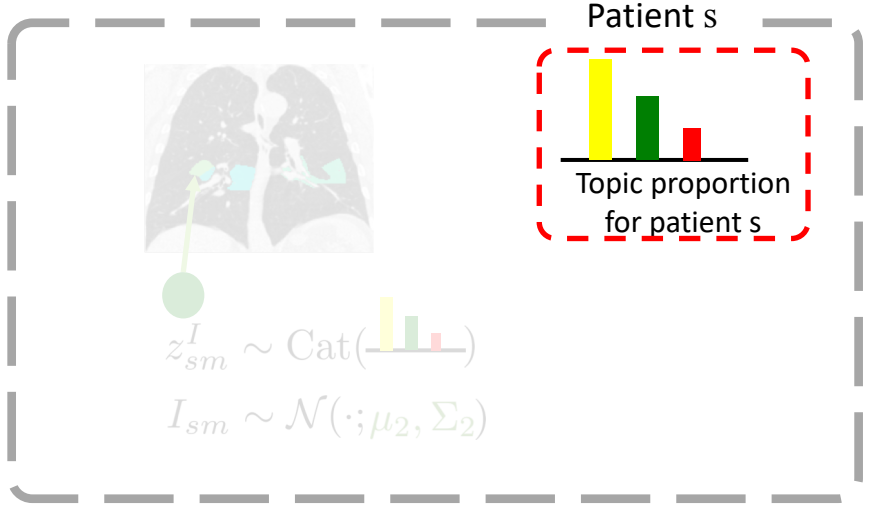


$$p_i(x) = \sum_{k=1}^{K} \pi_k(i) g(x; \theta_k)$$

Probabilistic Modeling

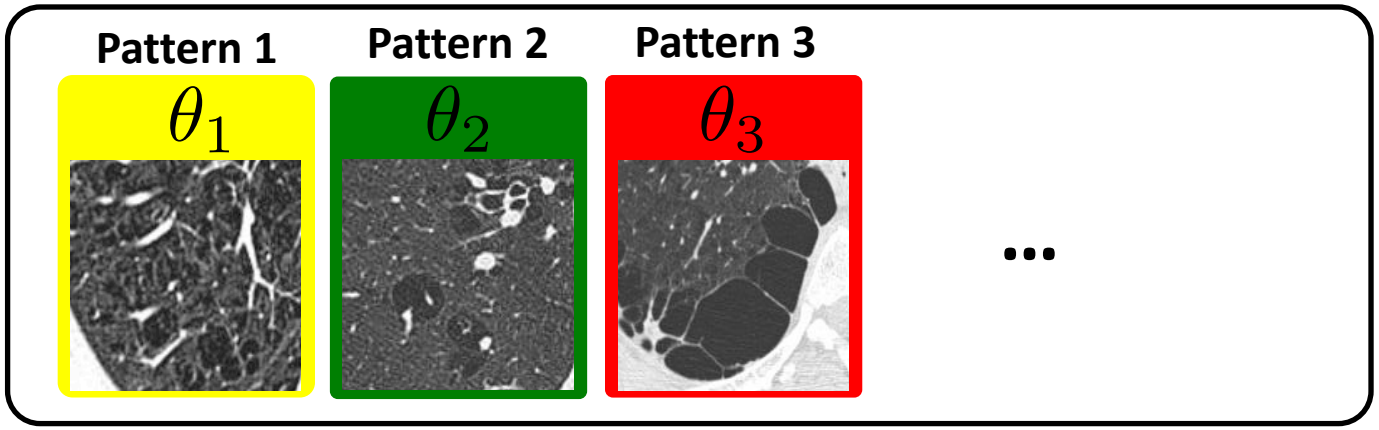


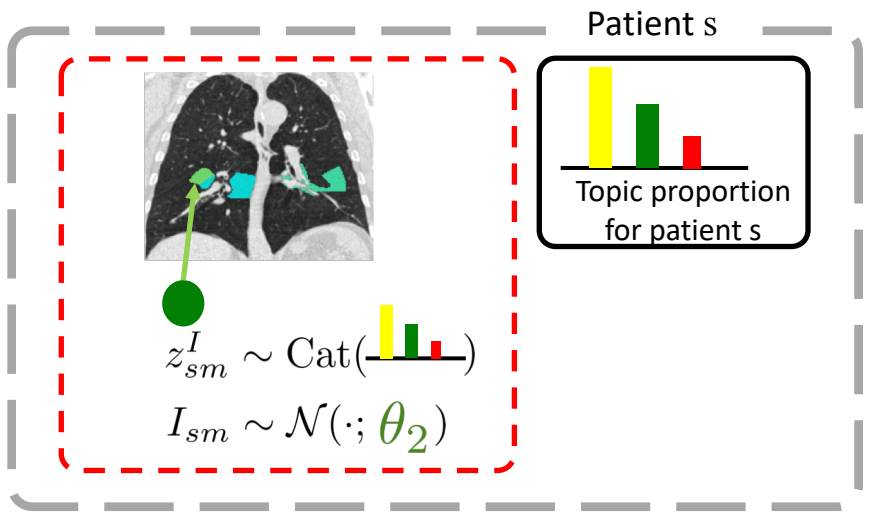


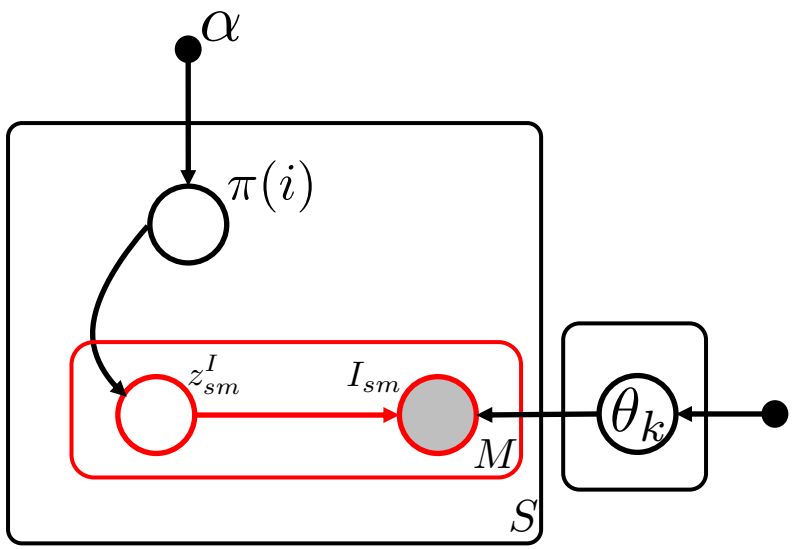


$$p_i(x) = \sum_{k=1}^{K} \pi_k(i) g(x; \theta_k)$$

Probabilistic Modeling





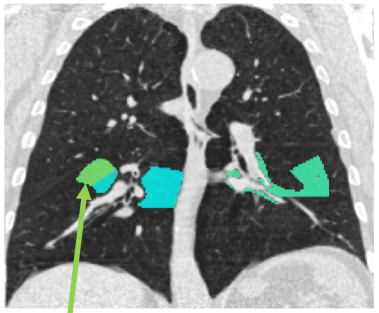


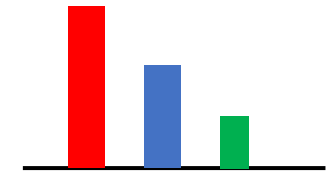
Extending to Multi-Modal

Probabilistic Model

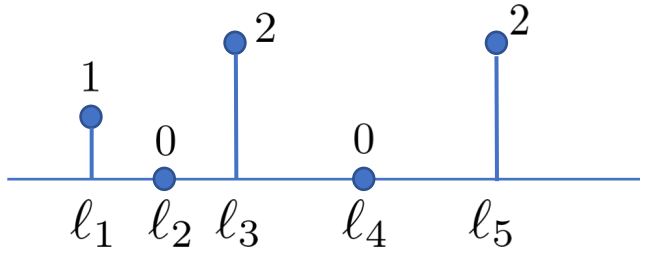
$(\mu_1, \Sigma_1), \beta_1$ $(\mu_2, \Sigma_2), \beta_2$ $(\mu_3, \Sigma_3), \beta_3$ • • • $(\mu_K, \Sigma_K), \beta_K$

Patient s





Topic proportion for patient s



$$G_s = \{\ell_1, \ell_3, \ell_3, \ell_5, \ell_5\}$$

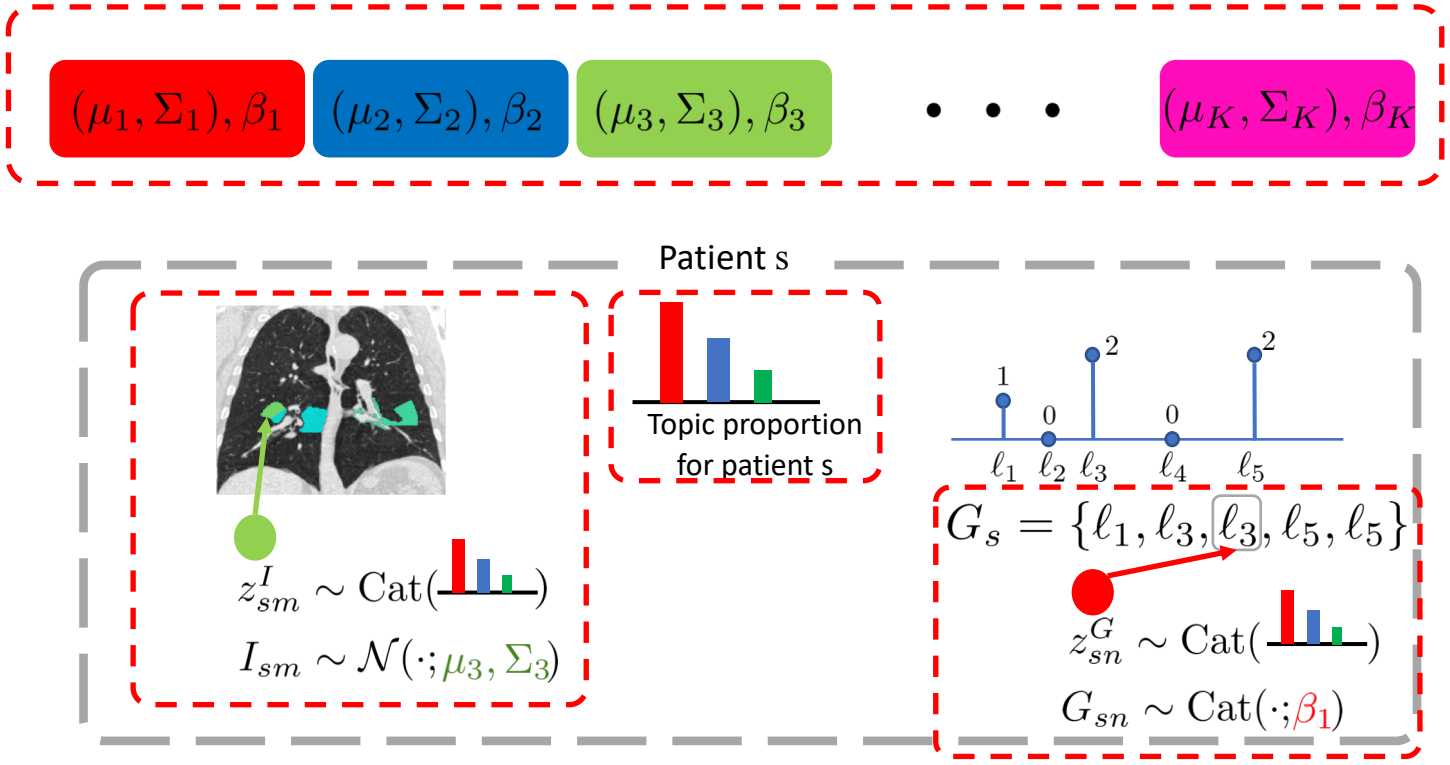
$$z^I_{sm} \sim \text{Cat}(\quad)$$

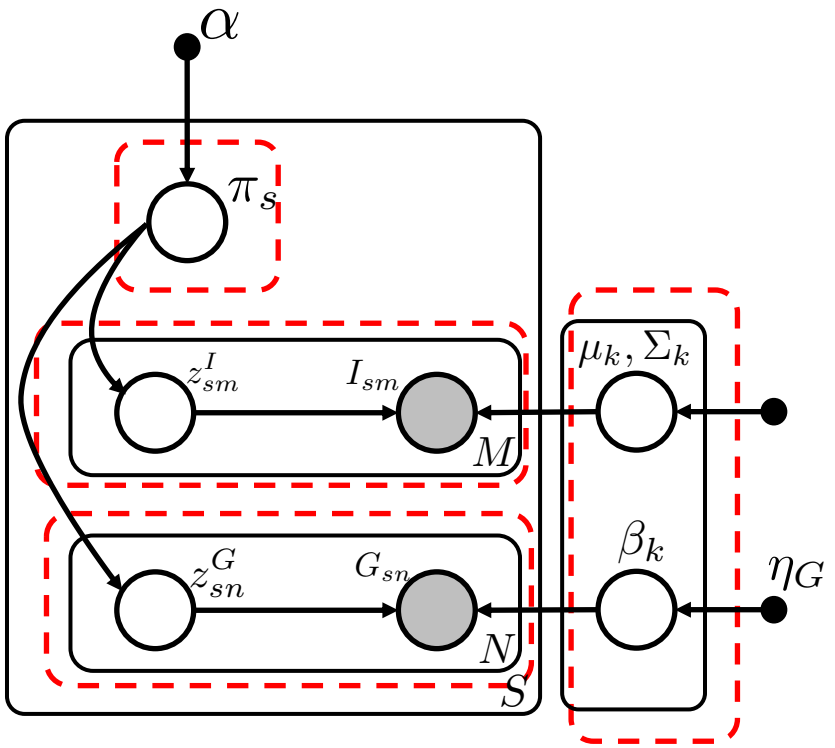
$$I_{sm} \sim \mathcal{N}(\cdot; \mu_3, \Sigma_3)$$

$$z^G_{sn} \sim \text{Cat}(\quad)$$

$$G_{sn} \sim \text{Cat}(\cdot; \beta_1)$$

Graphical Model

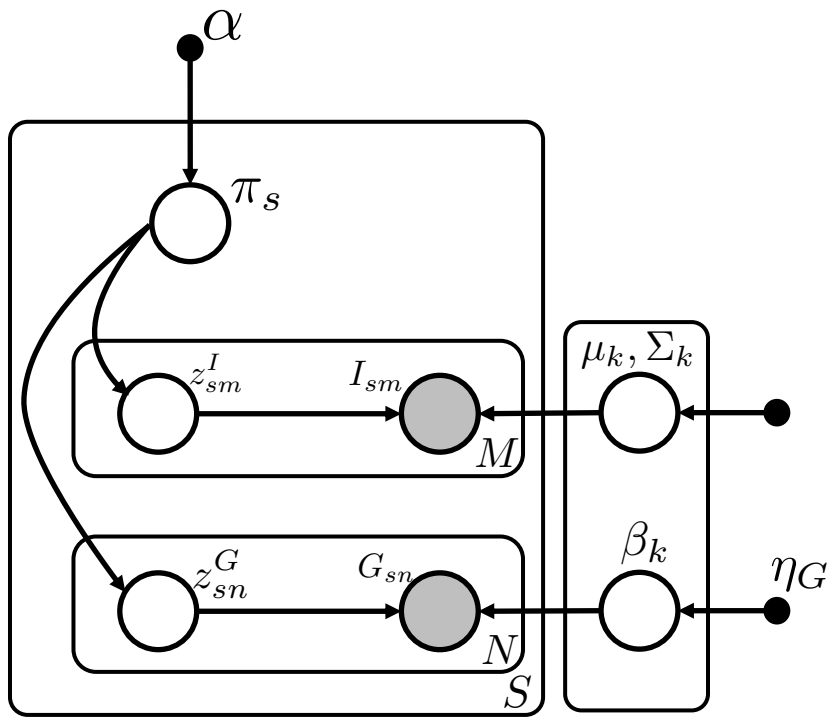




$$(\mu_k, \Sigma_k) \sim \text{NIW}(\eta^I)$$
$$\beta_k \sim \text{Dir}(\eta^G)$$

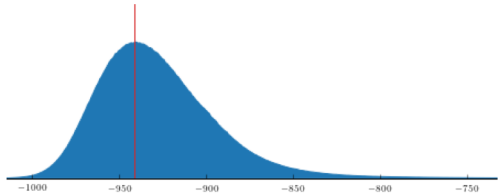
Inference

$(\mu_1, \Sigma_1), \beta_1$ $(\mu_2, \Sigma_2), \beta_2$ $(\mu_3, \Sigma_3), \beta_3$ • • • $(\mu_K, \Sigma_K), \beta_K$



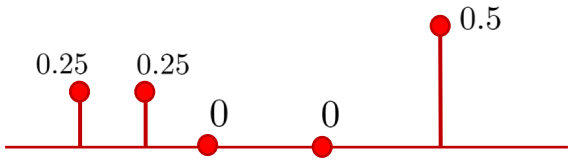
Topic pairs

$p(\mu_k | \{I_{sm}\}, \{G_{sn}\}; \pi)$



(Imaging signature)

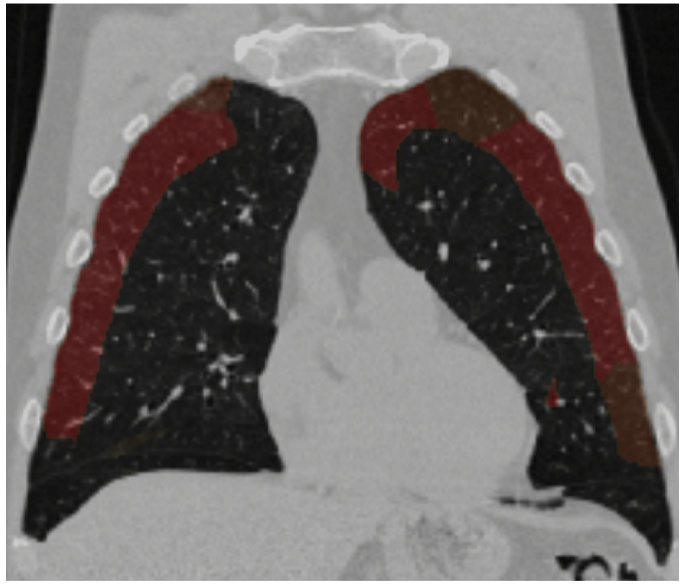
$p(\beta_k | \{I_{sm}\}, \{G_{sn}\}; \pi)$

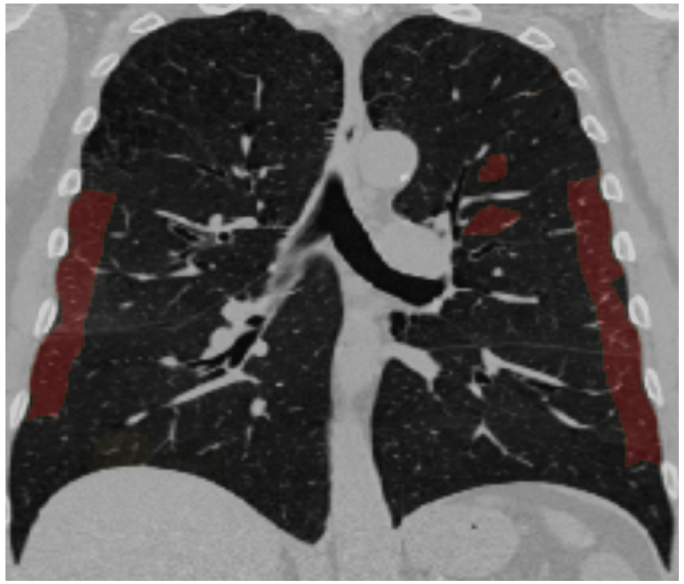


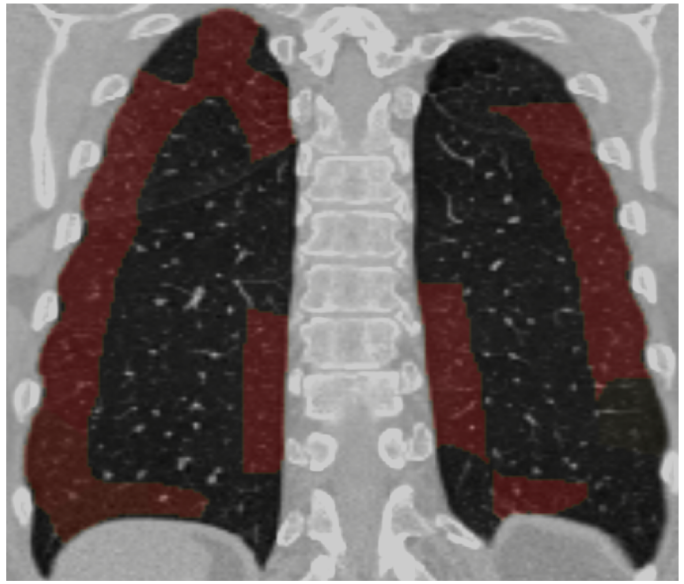
(Different Ranking SNPs)

We use stochastic vibrational inference to infer the posterior densities.

Interpretation on the Patient Level

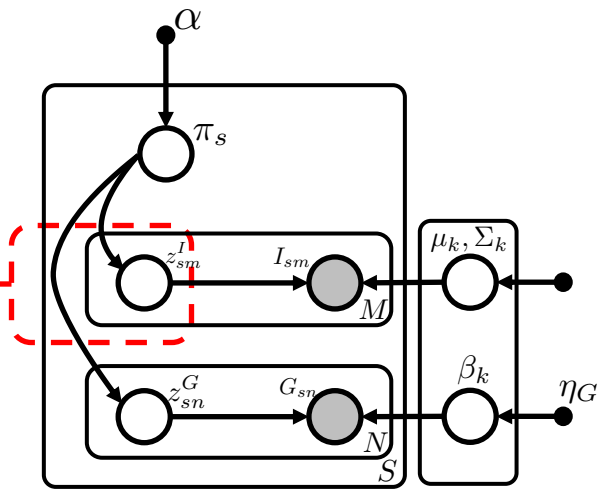






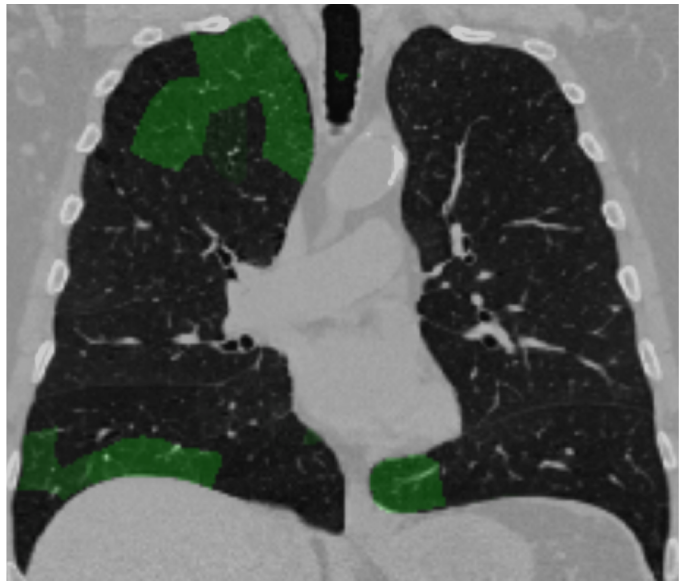
Rind Topic

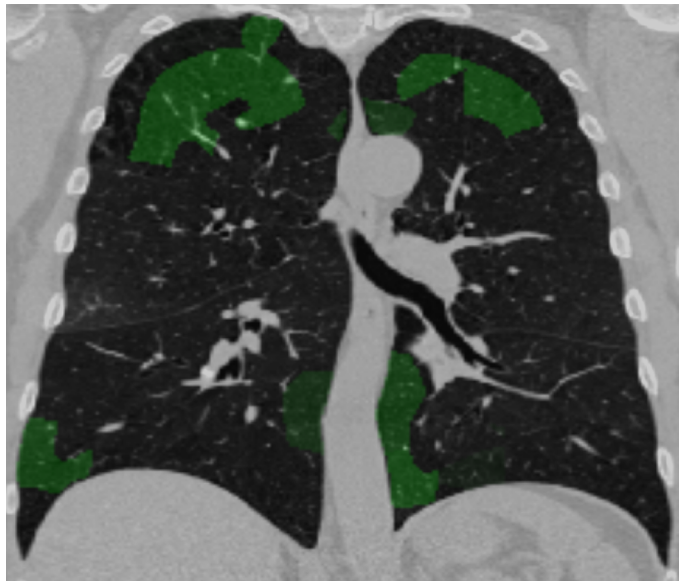
$$p(z_s^I = 1|\mathcal{D})$$

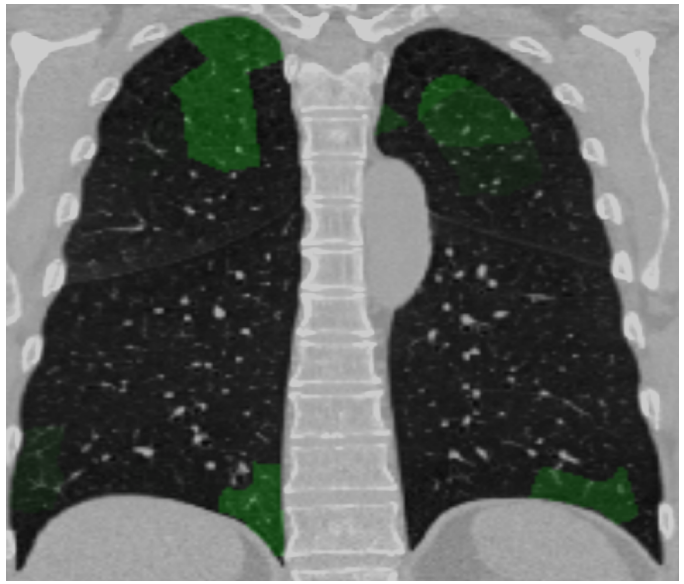


Nakano Y. et al, Core to Rind distribution of severe emphysema predicts outcome of lung volume reduction surgery , AJRCCM 2001

Interpretation on the Patient Level

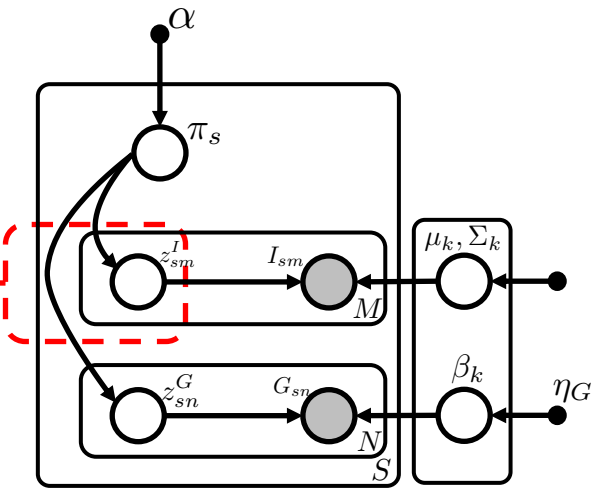




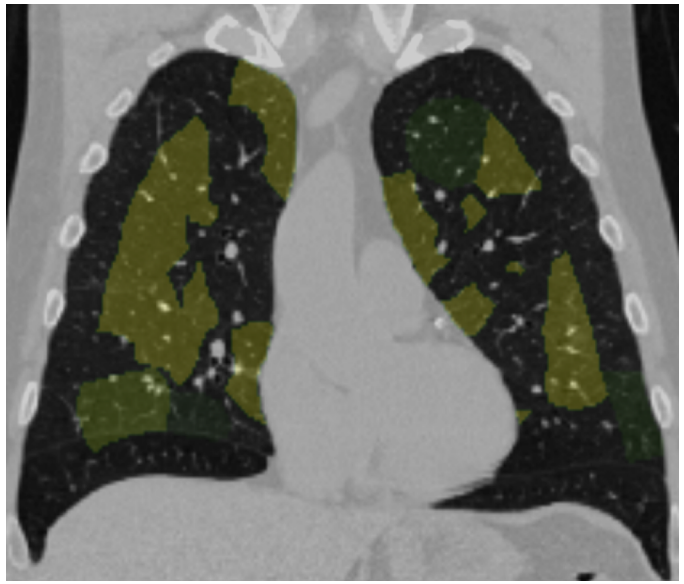


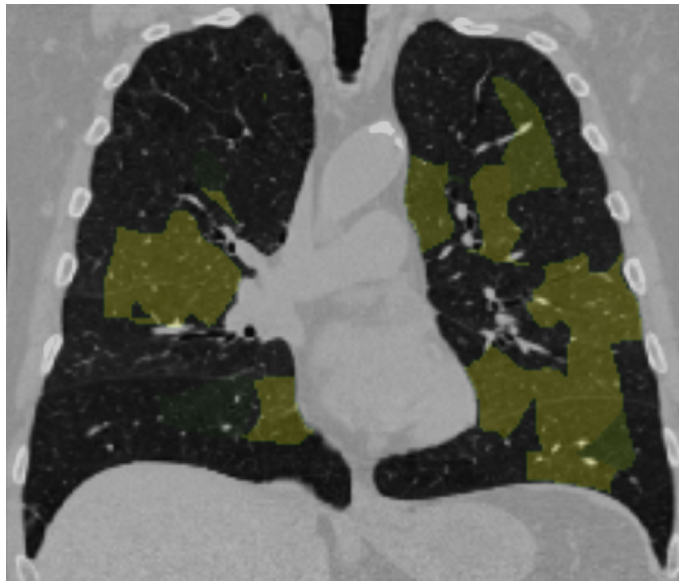
Apical Topic

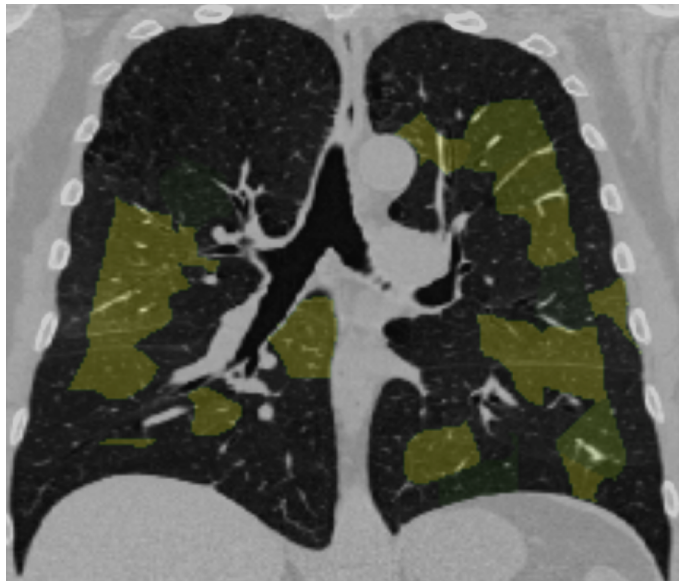
$$p(z_s^I = 2|\mathcal{D})$$



One Subject Example

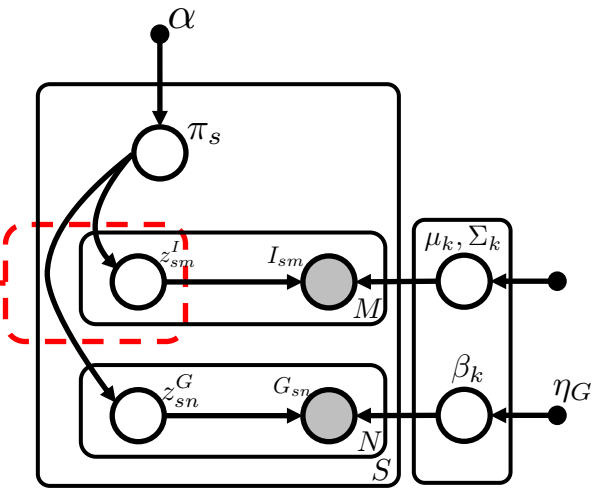




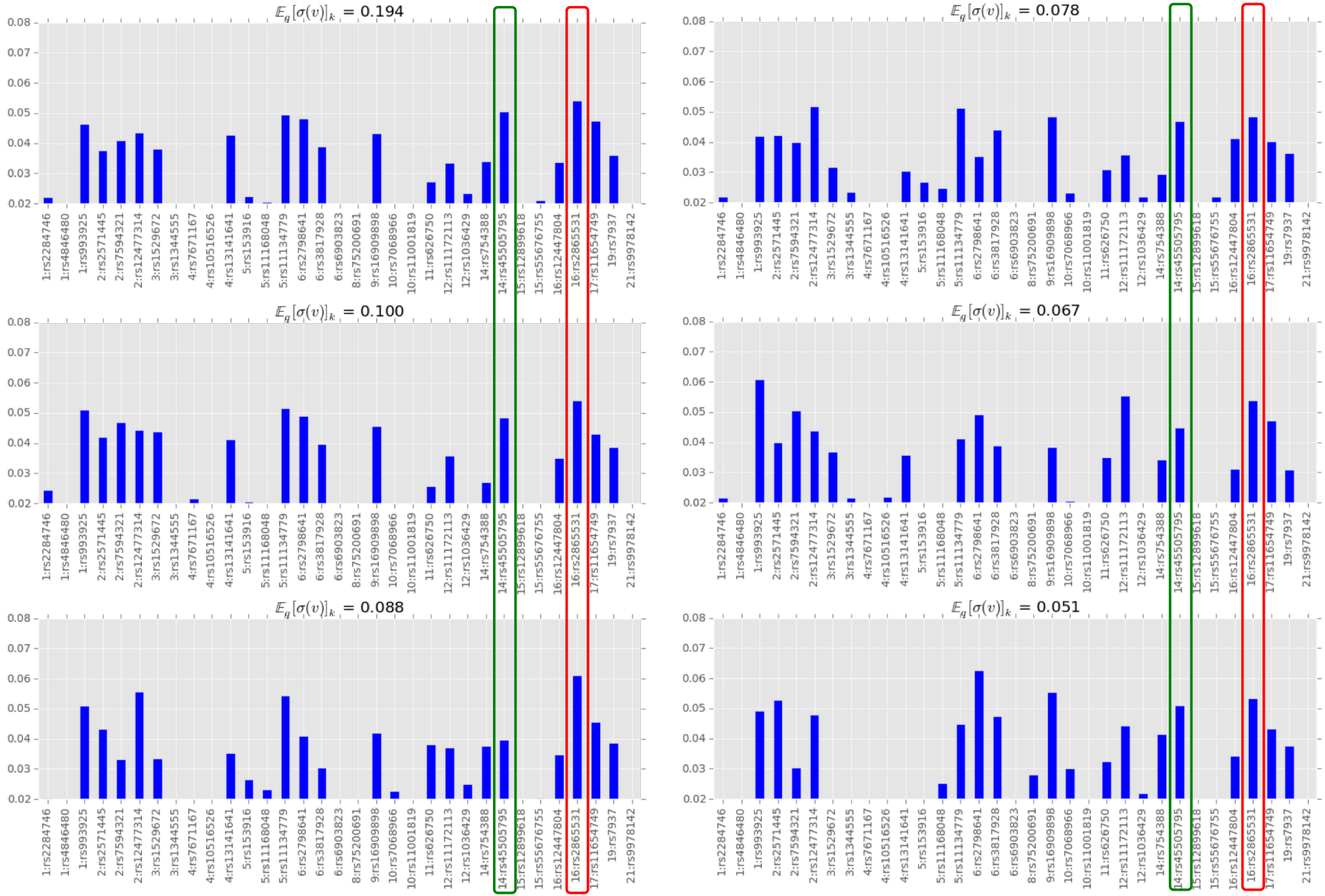


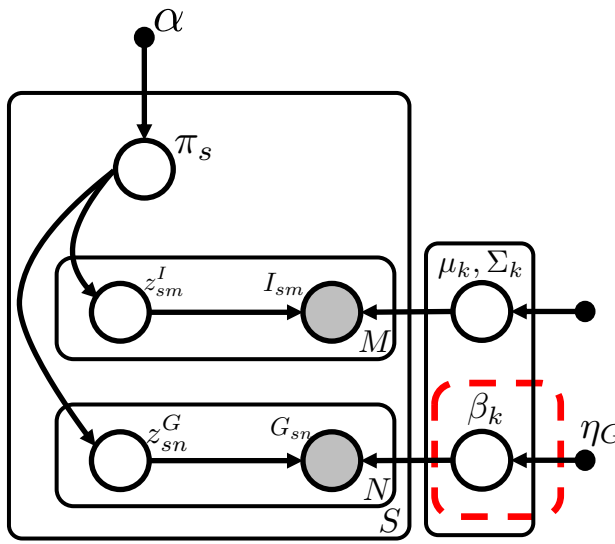
Normal Appearing

$$p(z_s^I = 4|\mathcal{D})$$



Genetic Topics



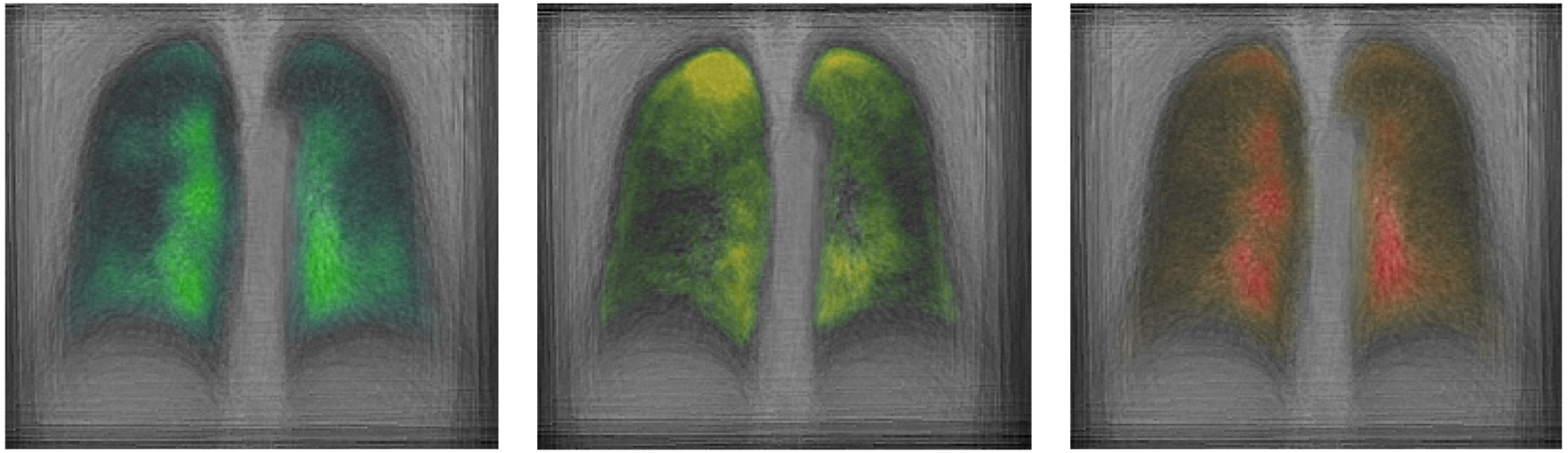


14:CFDP1

16:Alpha-1 Antitrypsin Deficiency

Population Average Distribution

- Data driven way to define anatomical distributions of the disease subtypes.



- What is missing?
  - ❑ Which image features (local descriptors) are good for sub-typing the disease?
  - ❑ Can we improve it using Deep Learning?

Which Features Should be Used?

Simplified version of our model

$$p_i(x) = \sum_{k=1}^{K} \pi_k(i) g(x; \theta_k)$$

Revisiting our goals: Interpretation vs. Prediction

When do I need to have a fully generative model?

Prediction

- **Supervised :** Objectively using image data to define disease subtypes that are indicative of the disease severity.

It turns out that we can side-step the inference of the subtypes.

Interpretation

Unsupervised: Relating image patterns to the side information (such as genetic data).

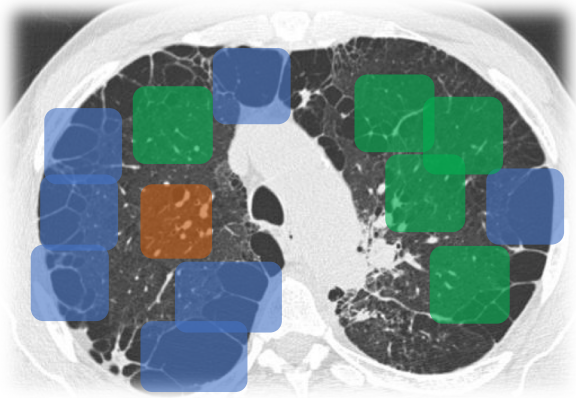
Discriminative Approach: Focusing on Prediction

If We Only Care about Prediction

Pattern 1

Pattern 2

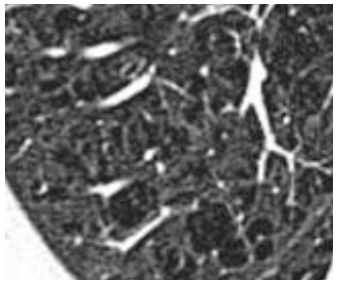
Pattern 3



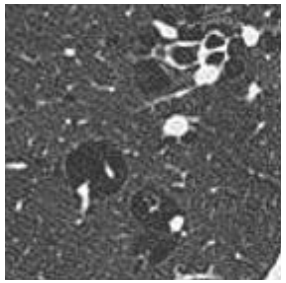
Simplified version of our model

$$p_i(x) = \sum_{k=1}^{K} \pi_k(i) g(x; \theta_k)$$

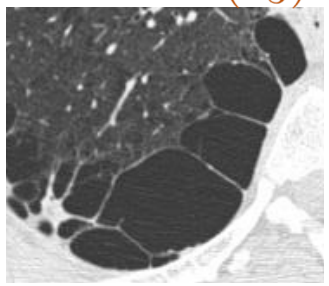
Pattern 1 (θ_1)



Pattern 2 (θ_2)



Pattern 3 (θ_3)



What if we can estimate the dissimilarities directly without an explicit parametric assumption?

$$\mathcal{D}(p_i, p_j)$$

Can We Avoid the Inference?

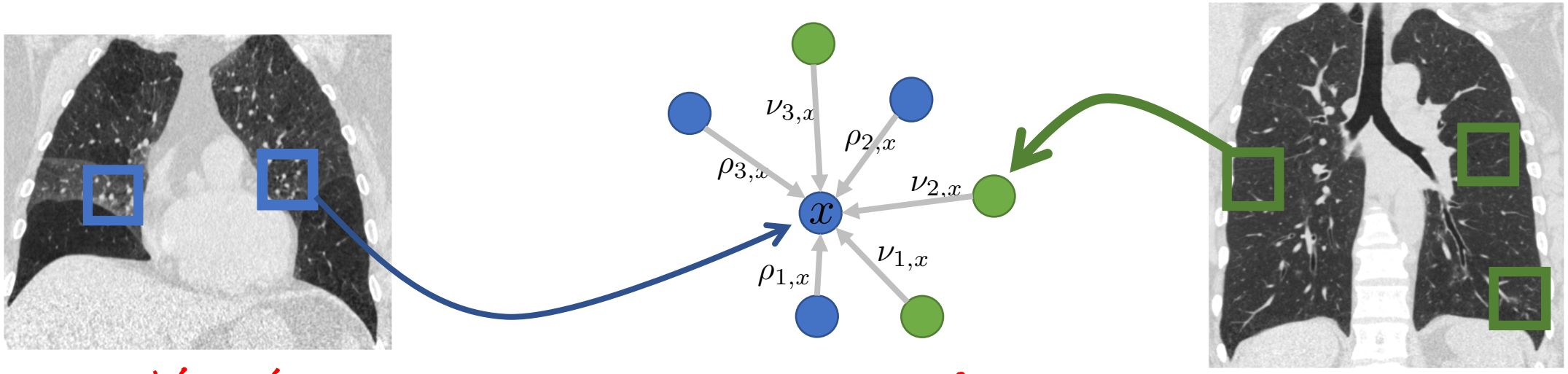
$$\mathcal{D}(p_i, p_j) \equiv \mathrm{KL}(p_i \| p_j) = \mathbb{E}_{x \sim p_i}\left[\log \frac{p_i}{q_j}\right]$$

- The KL distance can be estimated without any parametric assumption on p_i's !
- The estimator is unbiased and consistent.
- Other choices of distance between distributions are possible such as Maximum Mean Discrepancy (MMD).

Poczos et al, "Nonparametric Divergence Estimation with Applications to Machine Learning on Distributions ," UAI 2011

Can We Avoid the Inference?

$$\mathcal{D}(p_i, p_j) \equiv \mathrm{KL}(p_i \| p_j) = \mathbb{E}_{x \sim p_i}\left[\log \frac{p_i}{q_j}\right]$$



Subject i

Feature space of image patches

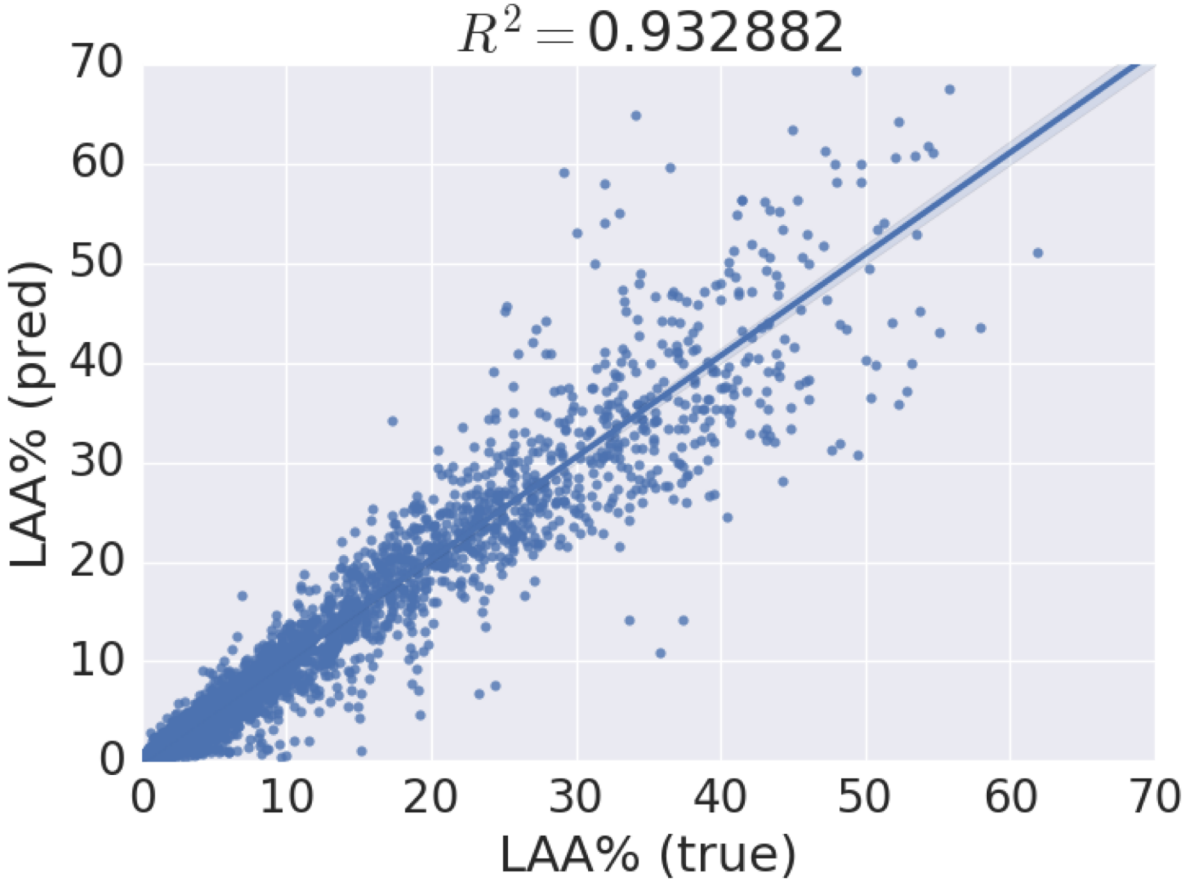
Subject j

$X_i = \{x_{i1}, \cdots, x_{iN}\}$

$X_j = \{x_{j1}, \cdots, x_{jM}\}$

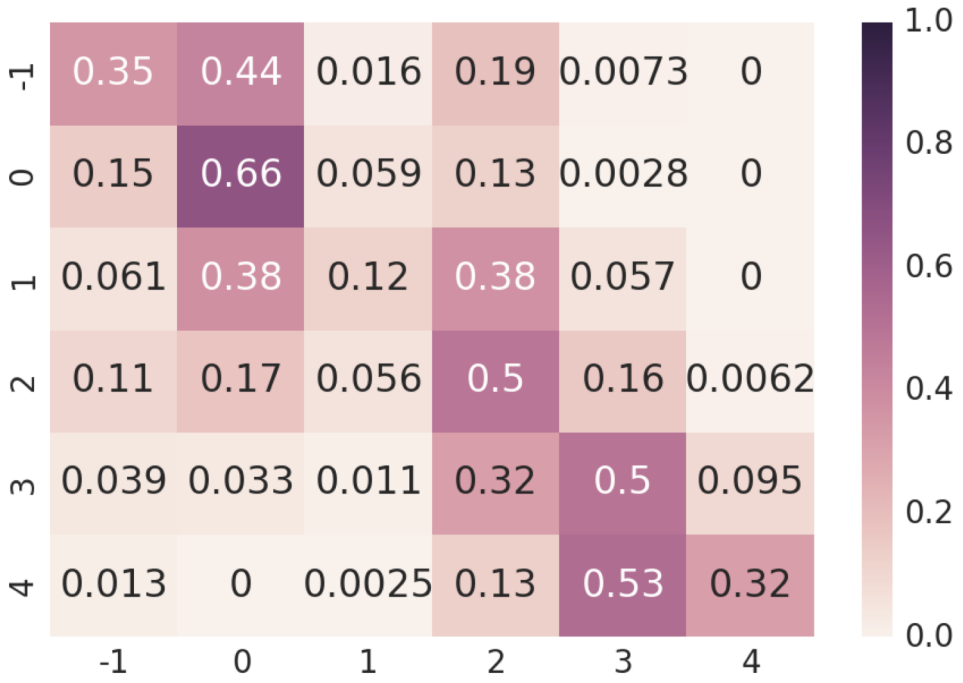
An Easy Task

- Predicting clinical image measurement.



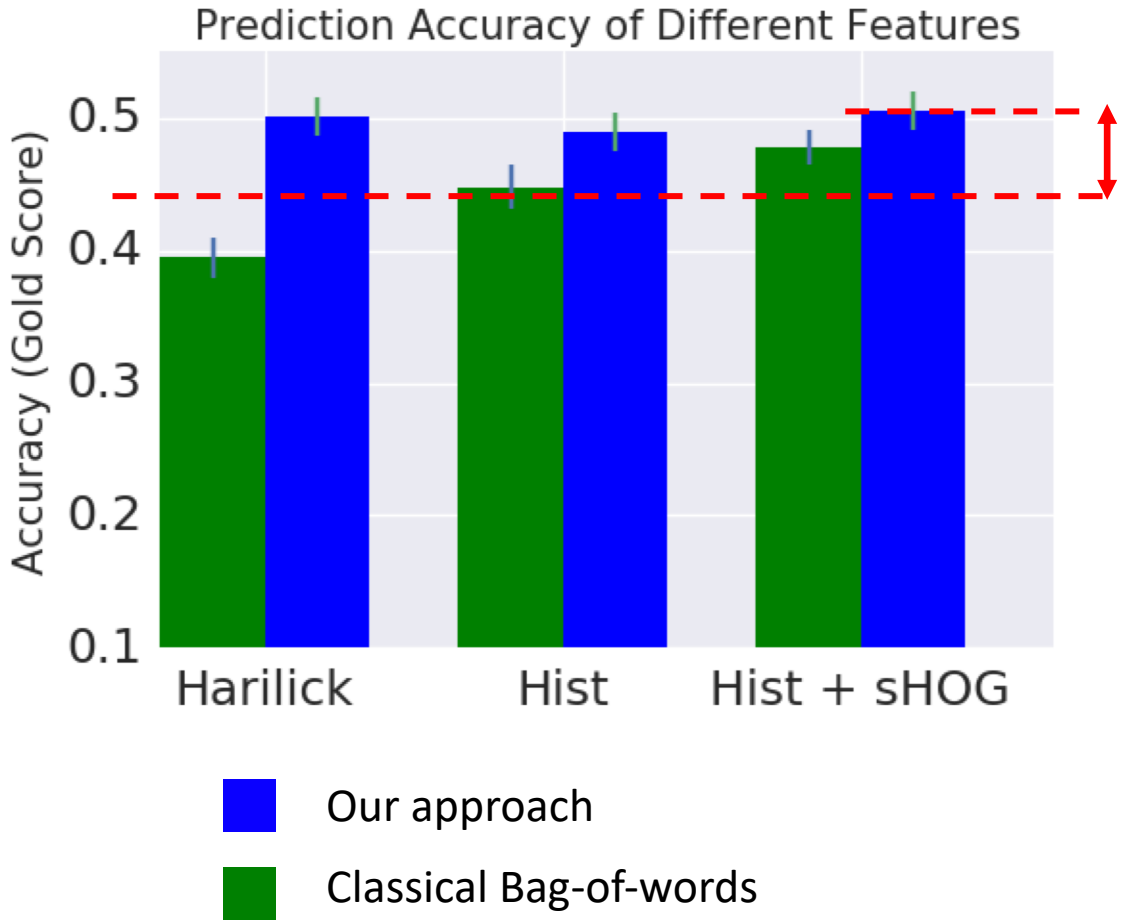
Predicting GOLD

- Predicting clinical image measurement.
- **GOLD score**: 5+1 grade characterizing the severity of disease using respiratory signal.



Comparing Different Image Features

- **GOLD score**: 5+1 grade characterizing the severity of disease using respiratory signal.
- Predicting clinical image measurement.
- Quick Comparison between image features.



Summary so far …

What we have:

- ✓ The **Generative** approach provides good interpretation while the **Discriminative** approach provides good prediction.
- ✓ A **Likelihood free** approach to compute patient-patient similarity avoiding an explicit parameterization.
- ✓ We can still have a **population-level** interpretation.

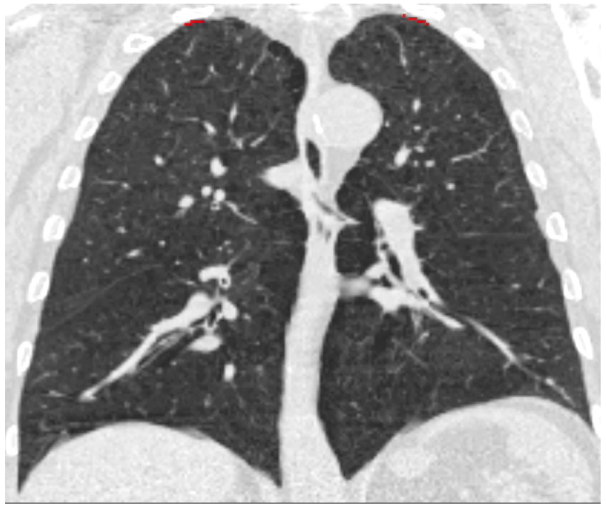
What is missing:

- ❑ Can we consolidate **Generative** and **Discriminative** views?
- ❑ Can **Deep Learning** improve the construction of the image feature in data-driven way?
- ❑ What about the **patient-level** interpretation?

Outlines

- Our previous research
  - Background
  - Pre-Deep Learning work
- Current Research
  - Hybrid Generative Discriminative Model
  - Some preliminaries on causal domain adaptation
- Future Directions

So far …



Patient i

$$X_i = \{x_{i1}, \cdots, x_{iN}\}$$

The Generative Model:

$$\pi_i, \qquad \{\theta_k\}_k$$

Patient-level representation

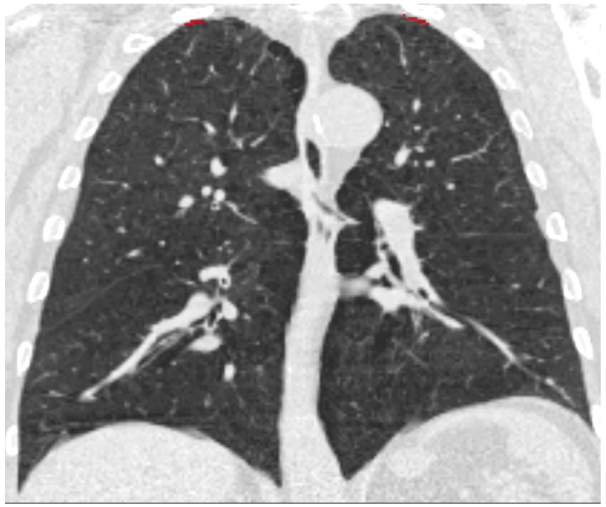
Population-level representation

The Likelihood-free Model:

$$\mathcal{D}(X_i, X_j)$$

Patient-Patient Similarity

Learning from a Set



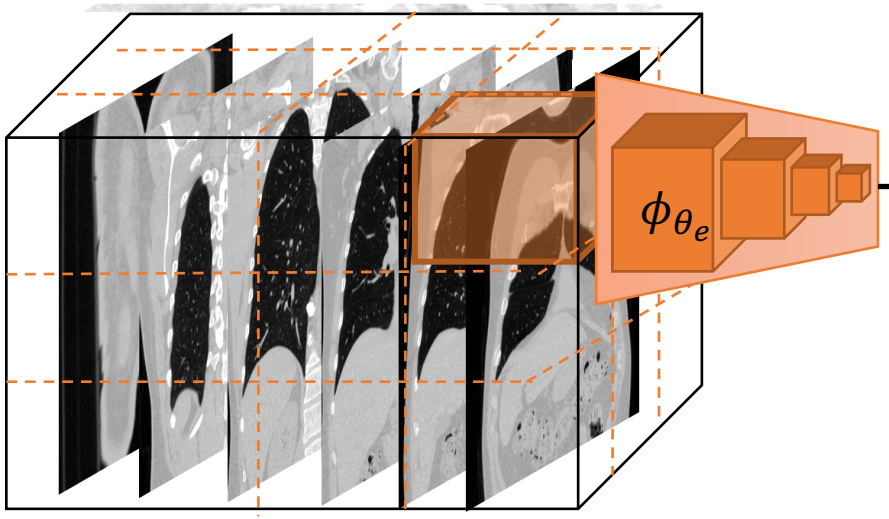
Patient i

- Can we construct a Discriminative model that directly predict a desire target from the content of a "set"?
- Goal: mapping from a variable length representation (i.e, a "set") to a fixed-length representation.

$$X_i = \{x_{i1}, \cdots, x_{iN}\}$$

Zaheer et al, "Deep Sets," NIPS 2017

Deep Sets



$\phi(x_{i1})$

Patient i

$$X_i = \{x_{i1}, \cdots, x_{iN}\}$$

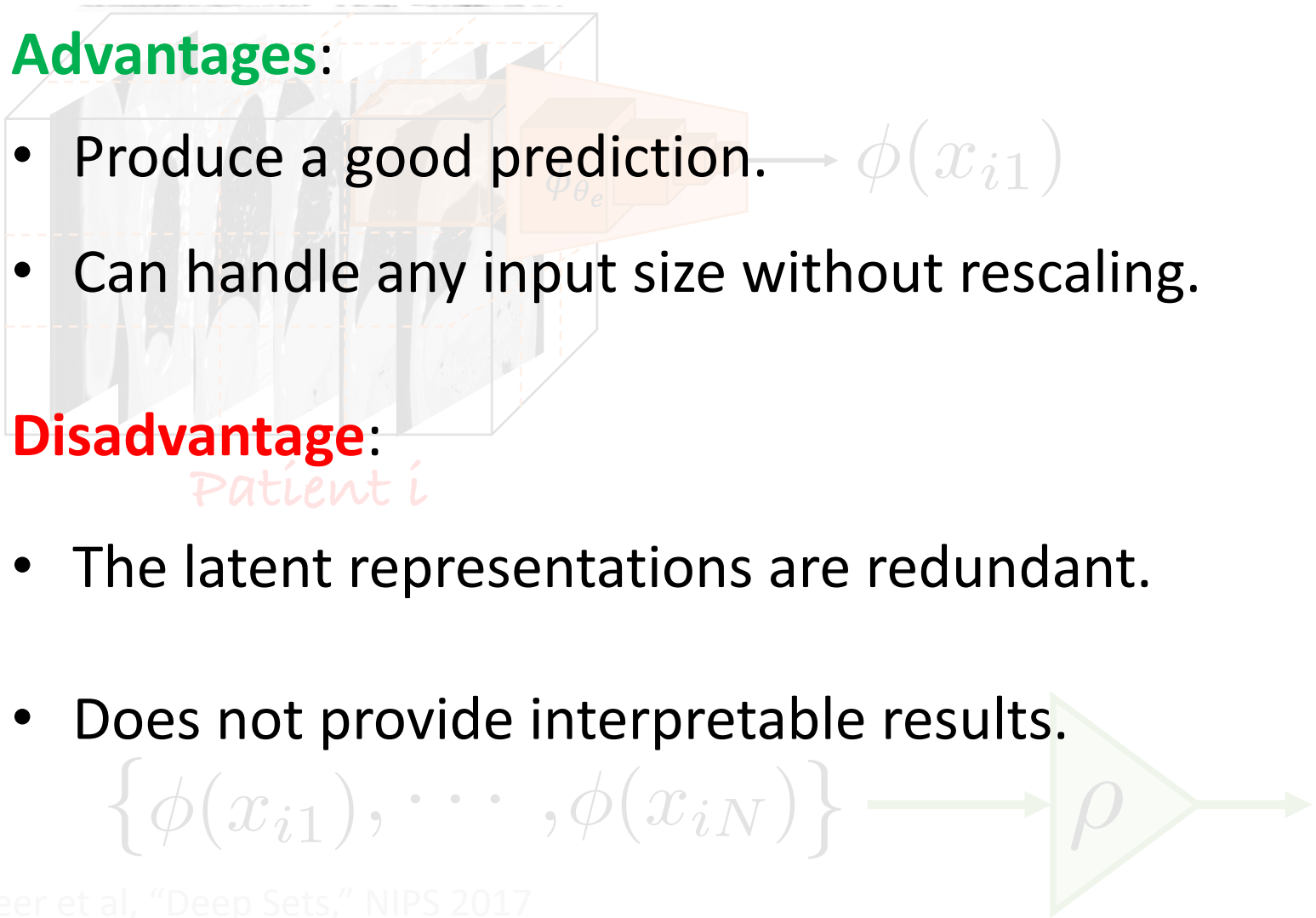
$$\{\phi(x_{i1}), \cdots, \phi(x_{iN})\}$$

Permutation Invariant Function

ρ

Zaheer et al, "Deep Sets," NIPS 2017

Deep Sets



Advantages:

- Produce a good prediction.
- Can handle any input size without rescaling.

Disadvantage:

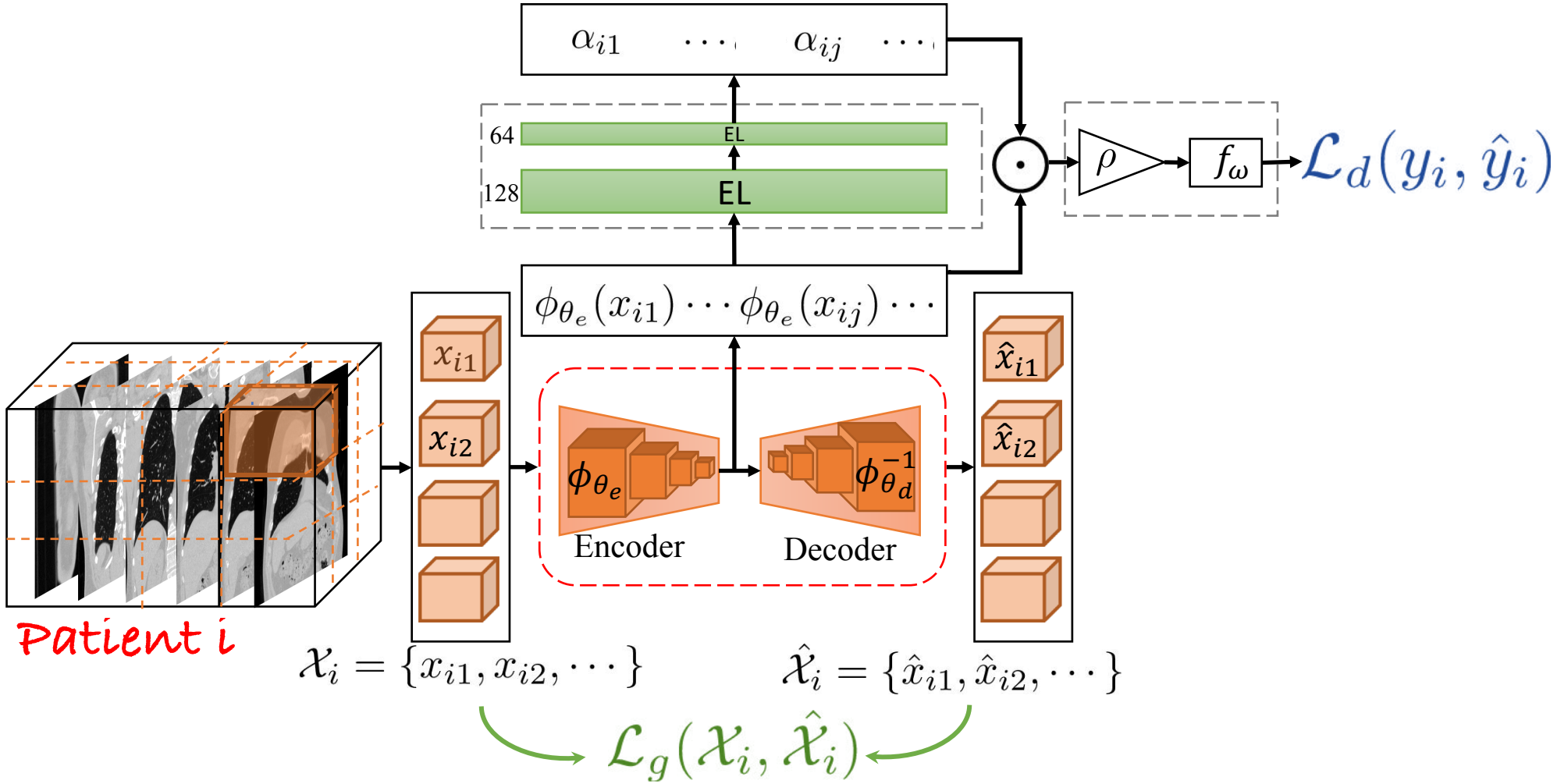
- The latent representations are redundant.
- Does not provide interpretable results.

Regularization with a Generative Network

Attention Network

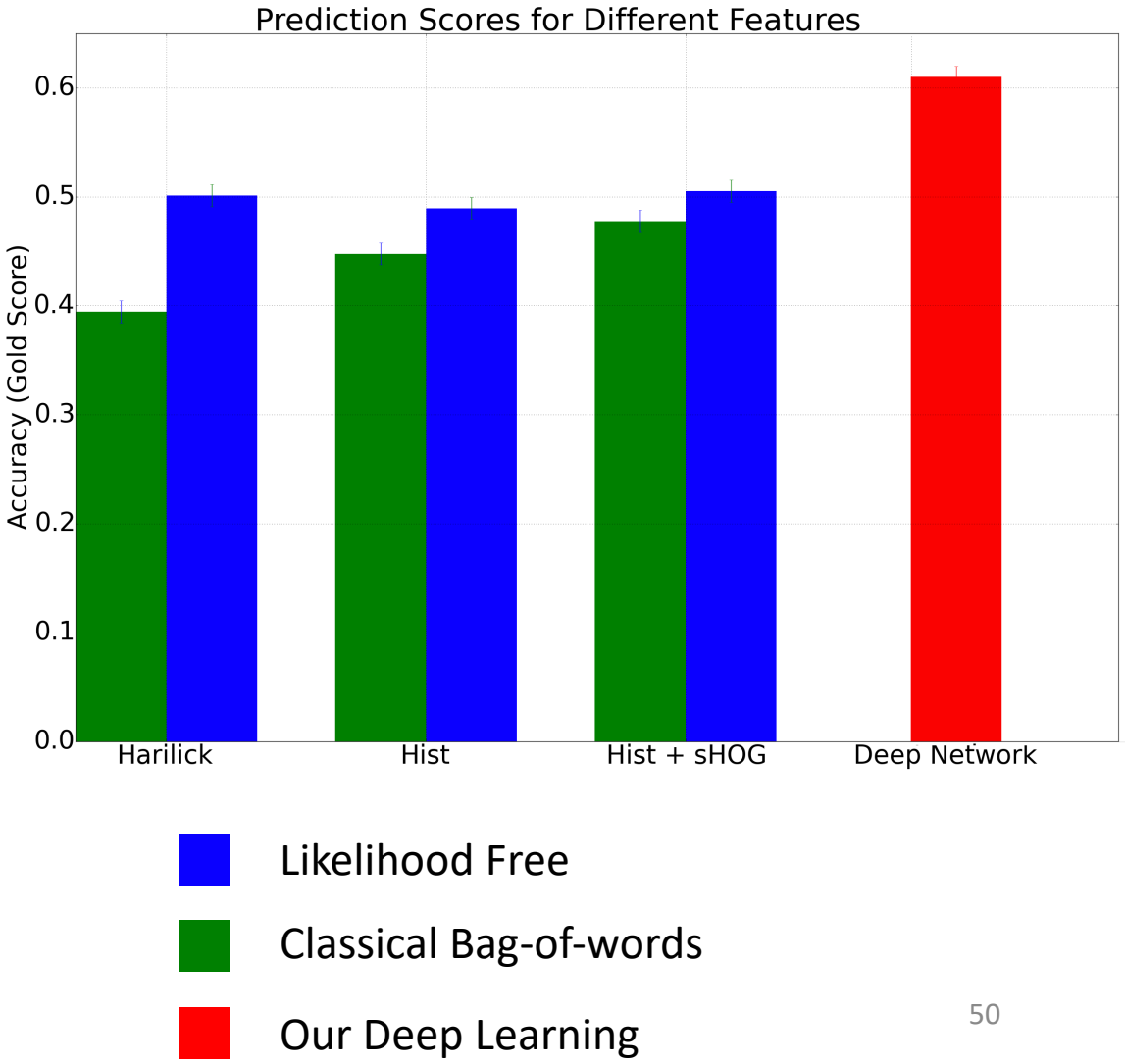
Zaheer et al, "Deep Sets," NIPS 2017

Our Hybrid Generative-Discriminative Model



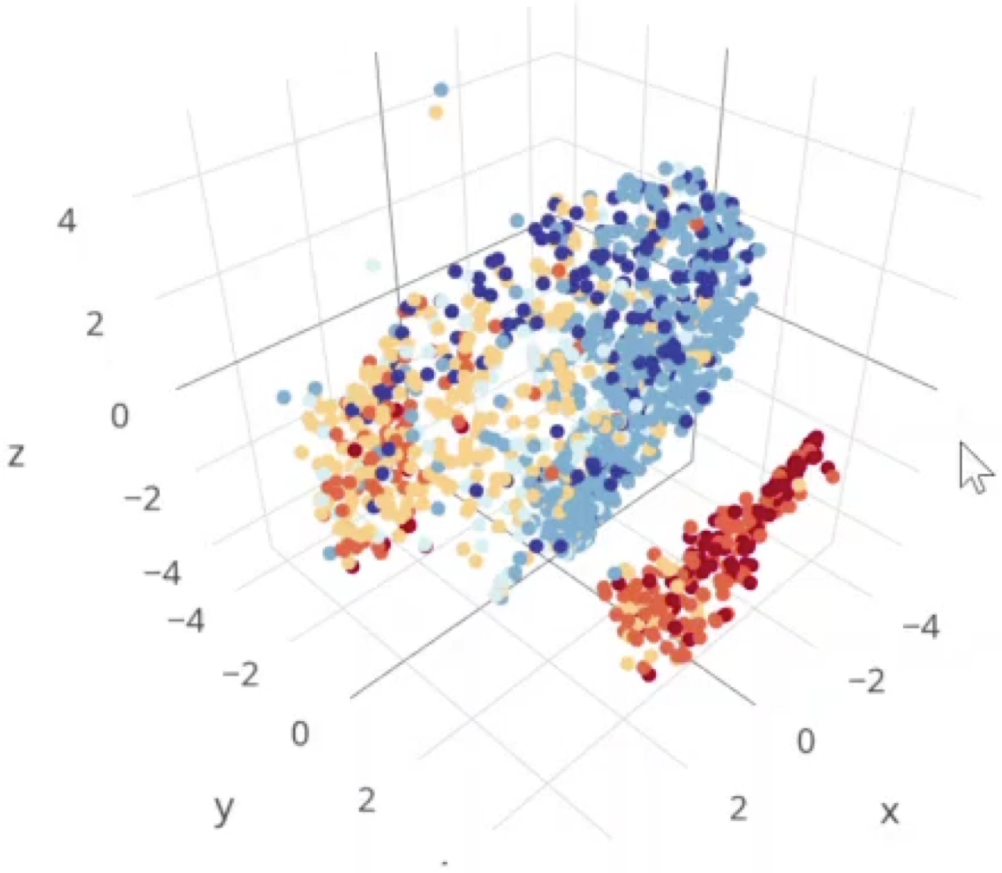
Predicting Clinical Measurements

- **GOLD score**: 5+1 grade characterizing the severity of disease using respiratory signal.

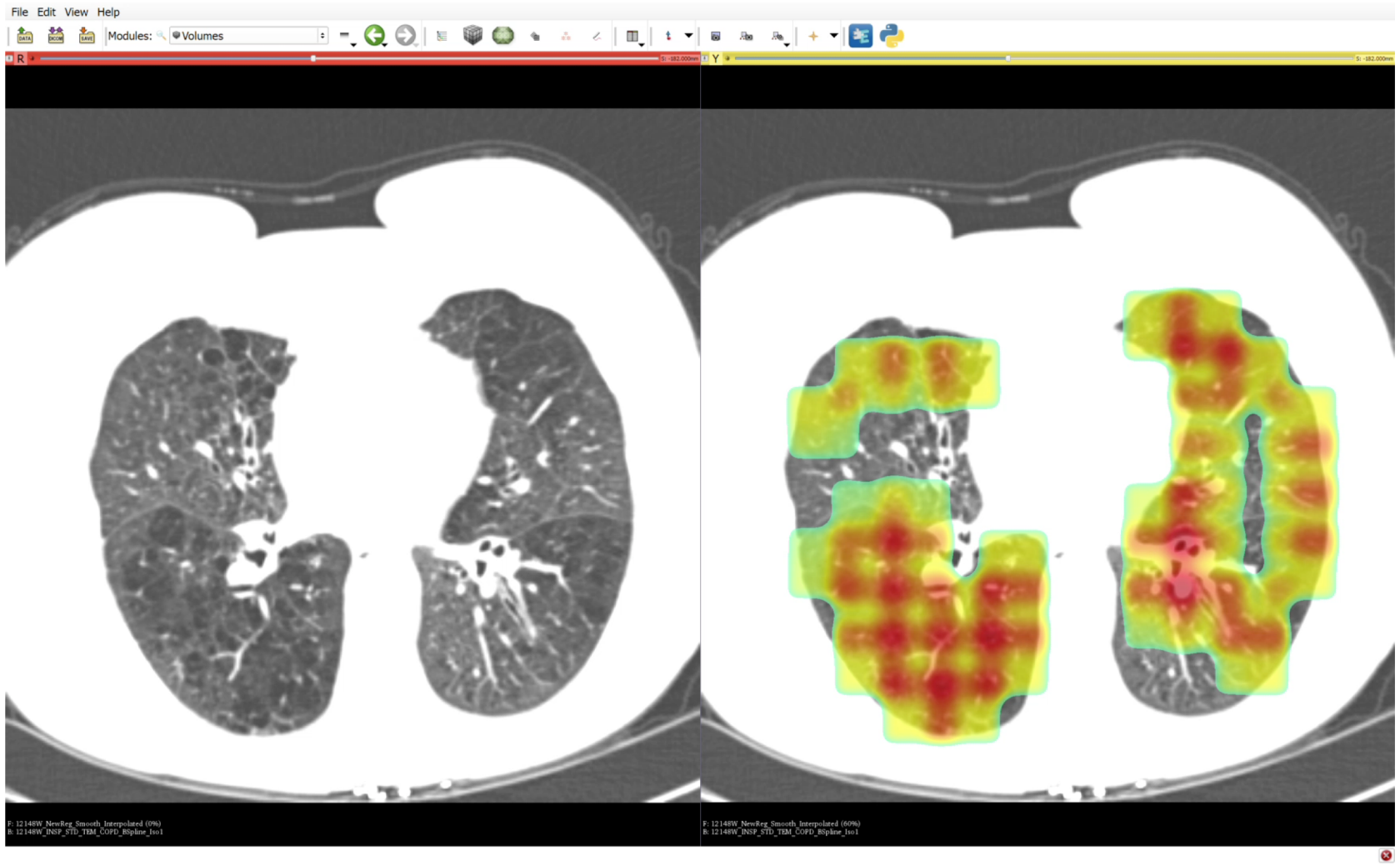


Predicting Clinical Measurements

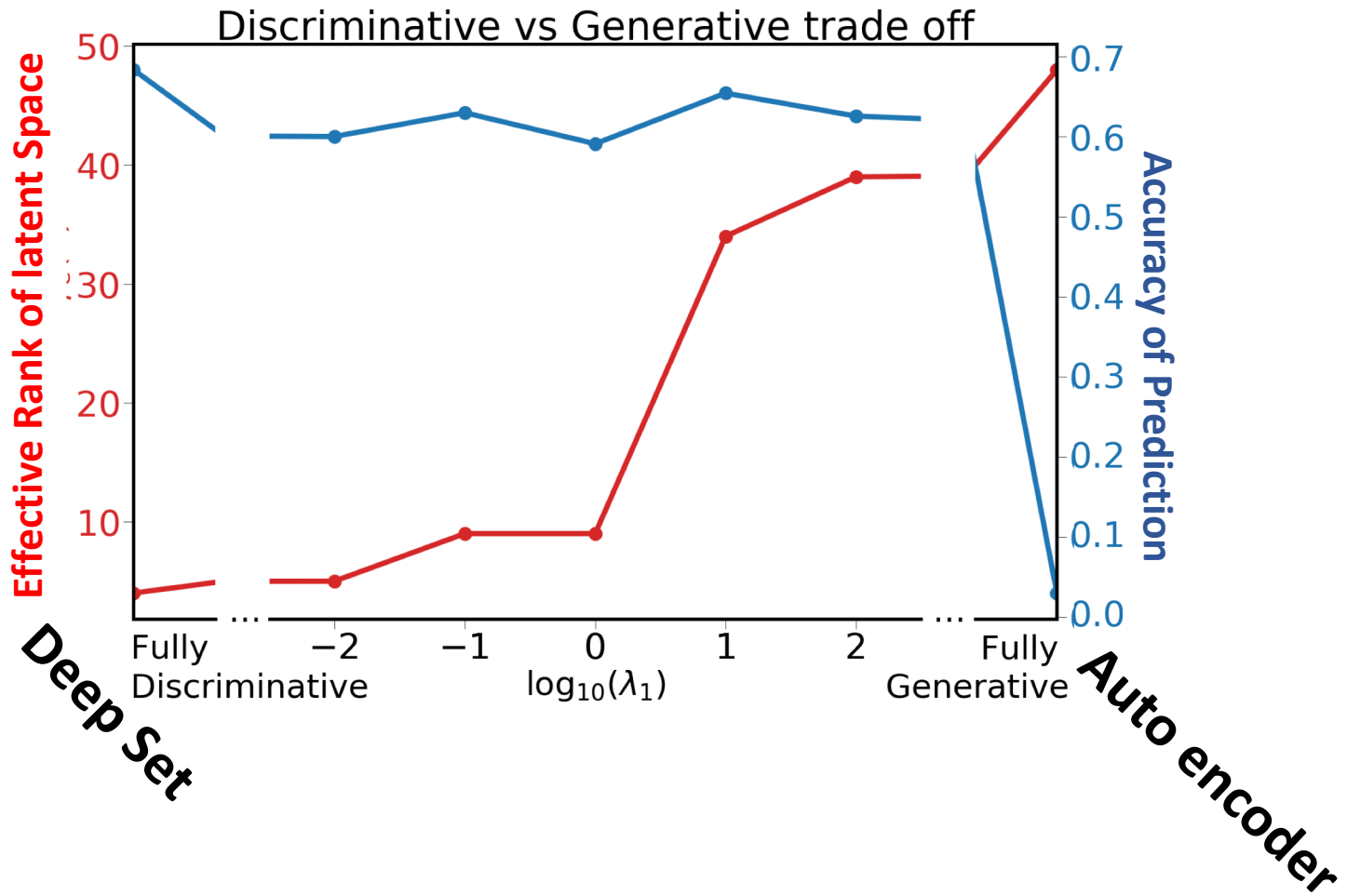
- **GOLD score**: 5+1 grade characterizing the severity of disease using respiratory signal.
- Visualizing patients with respect to the entire population.



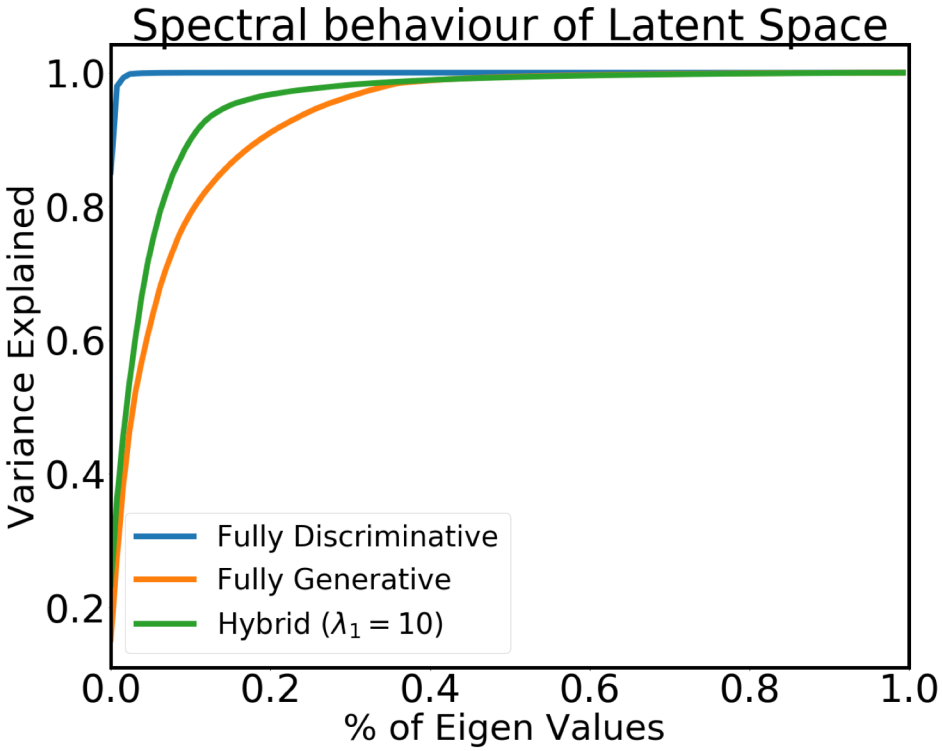
The Attention Map

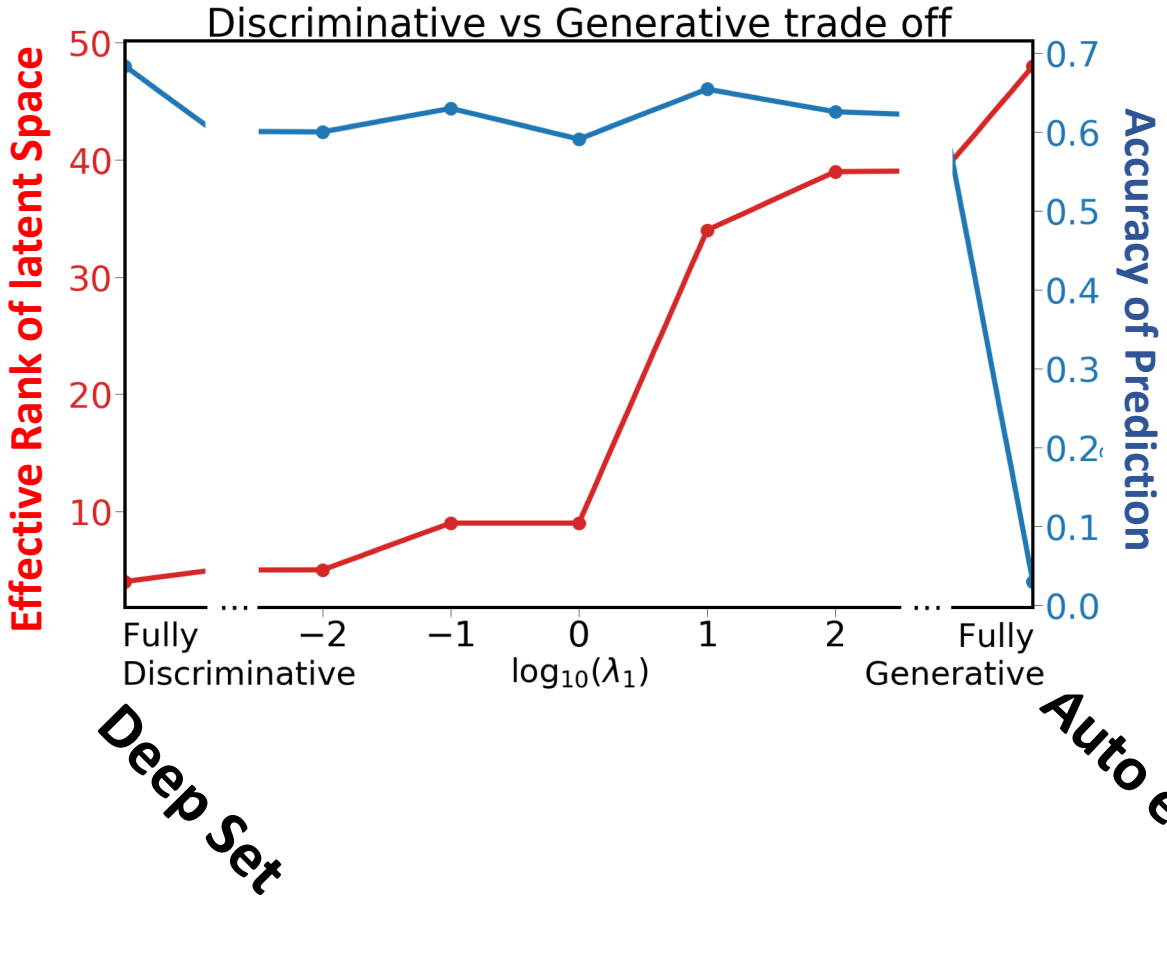


Comparison with Deep Sets



Comparison with Deep Sets





Comparing with Other Methods

Method	FEV1	FEV1/FVC	GOLD exact	GOLD one-off
Our method ($\lambda_1 = 0$)	**0.68**	**0.71**	**61.17 %**	**87.64 %**
Our method ($\lambda_1 = 10$)	**0.64**	0.70	**55.60 %**	**84.57%**
CNN [6]	0.53	—	51.1 %	74.9 %
Non-Parametric [17]	0.58	0.70	50.47 %	—
K-Means [17]	0.54	0.67	48.23 %	—
Baseline	0.52	0.69	49.06 %	—

6. González, G., Ash, S.Y., V S, G.: Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. AJRCCM pp. 201705–0860 (2017)

17. Schabdach, J., Wells, W.M., Cho, M., Batmanghelich, K.N.: A likelihood-free approach for characterizing heterogeneous diseases in large-scale studies. In: IPMI. vol. 10265 LNCS, pp. 170–183 (2017)

Summary so far …

What we have:

- ✓ The **Deep Learning** approach provides good characterization of the disease.
- ✓ The **Attentions** network provides a patient-level interpretation.
- ✓ The latent space representation yields an insight on the **population-level**.

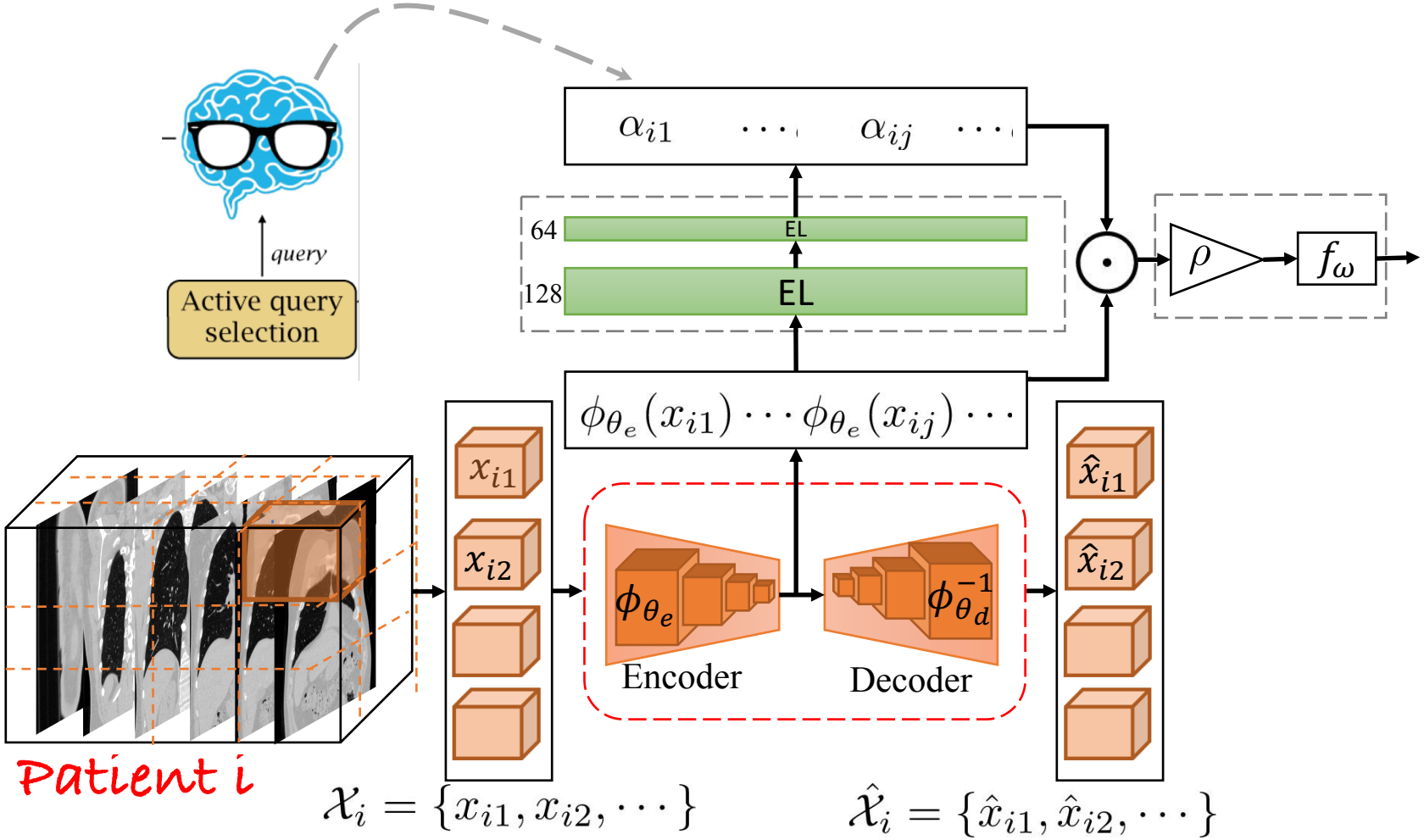
What is missing:

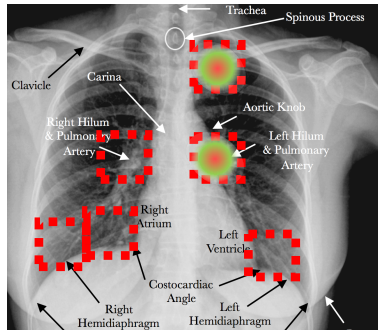
- ❑ Can we consolidate **Generative** and **Discriminative** views?
- ❑ The Deep Learning approach did not include the **genetic** data.

Outlines

- Our previous research
  - Background
  - Pre-Deep Learning work
- Current Research
  - Hybrid Generative Discriminative Model
  - Some preliminaries on causal domain adaptation
- Future Directions

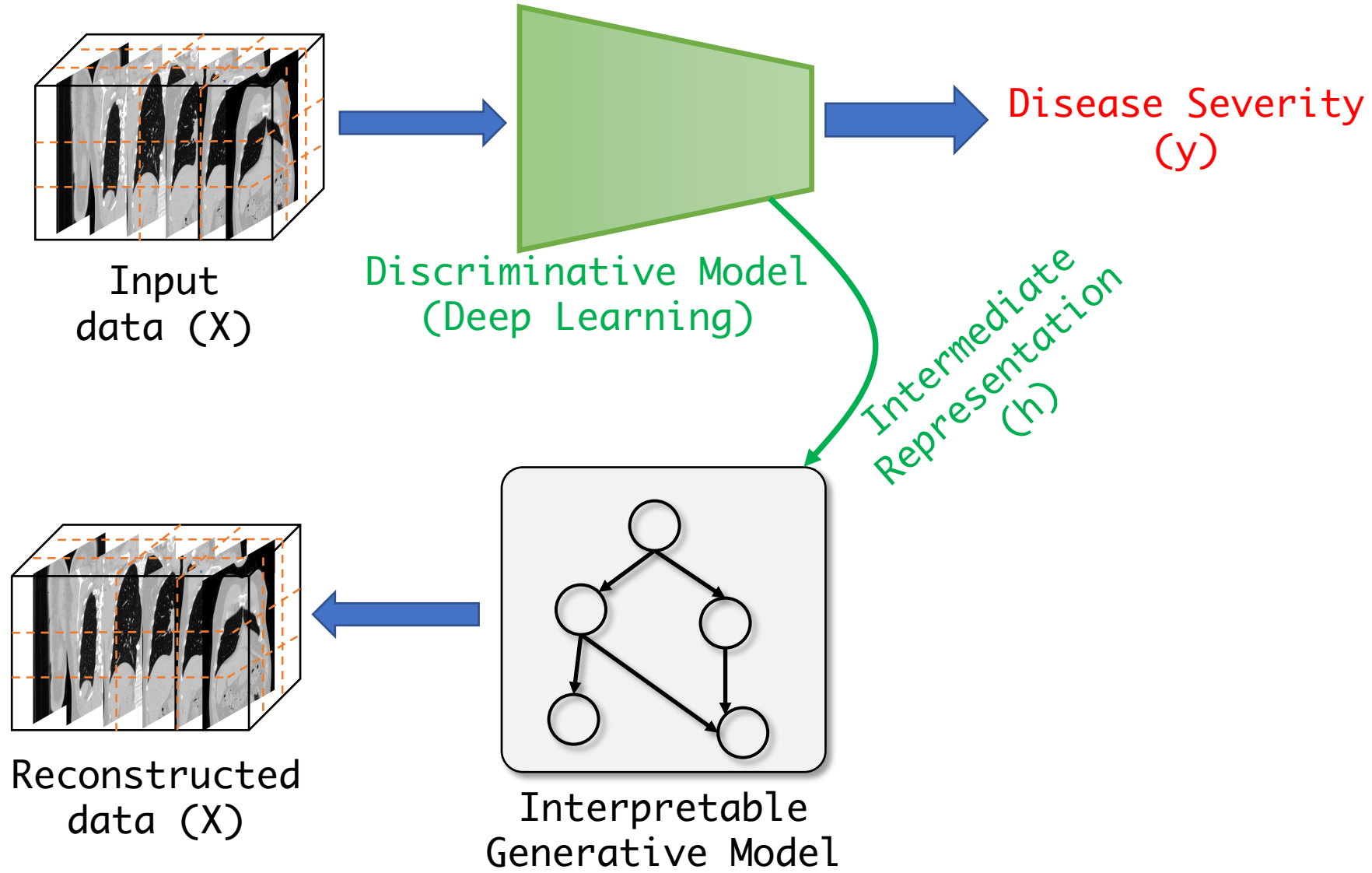
Interactive Learning with Image and Text



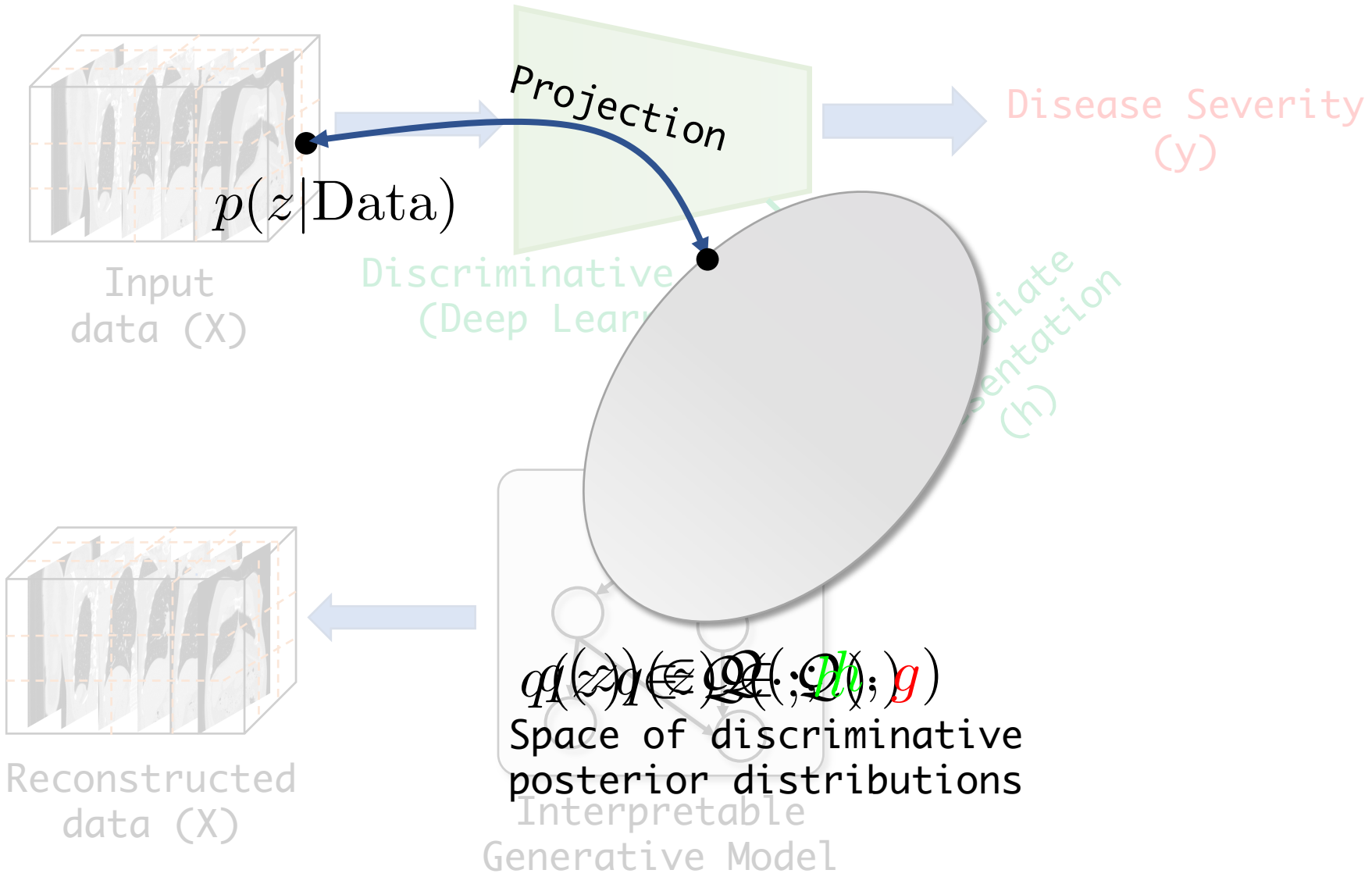


… pneumothorax is observable in the right lung …there is a significant emphysema in the right lobe….

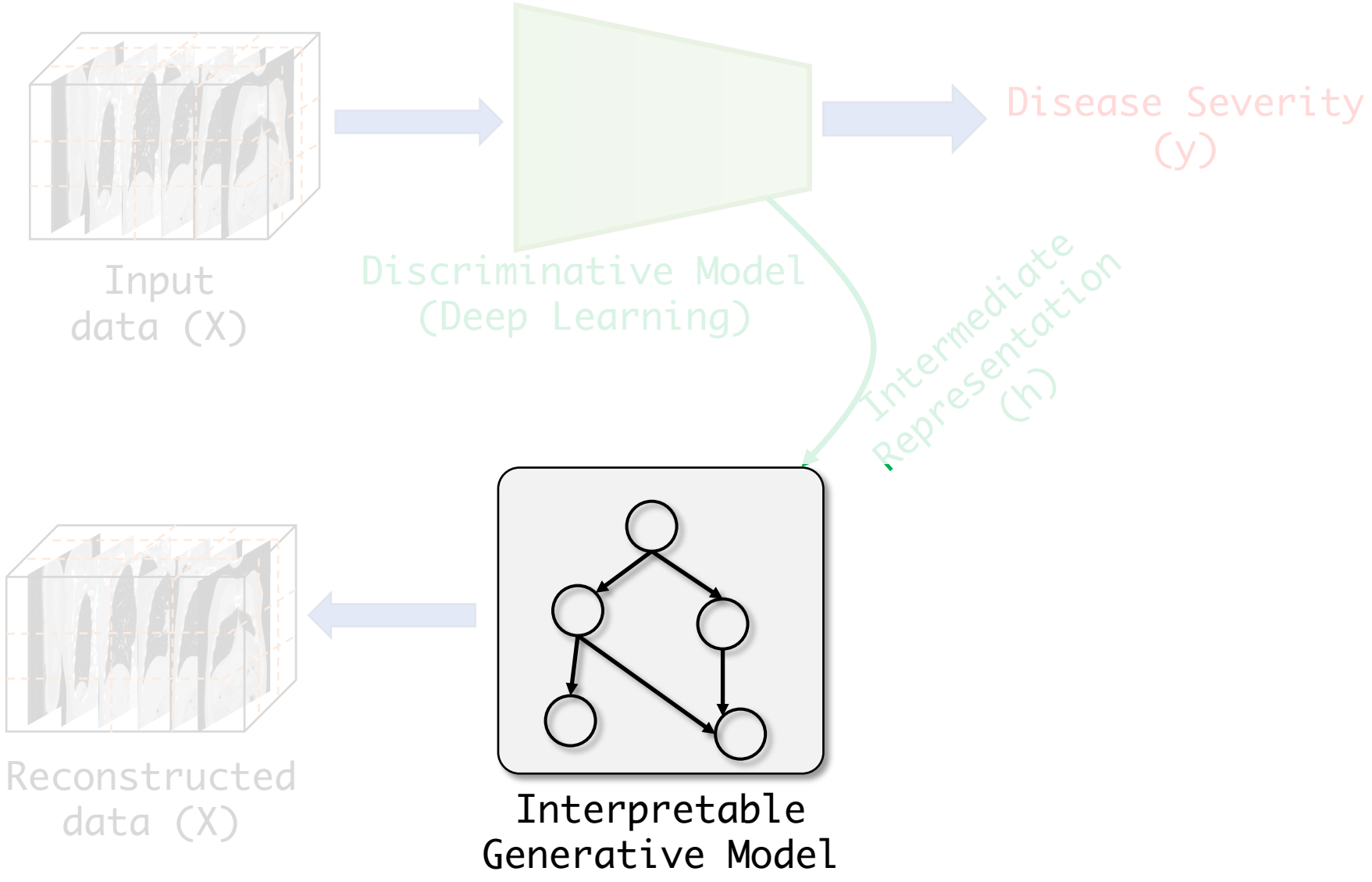
Developing Interpretable High-Dimensional Phenotype



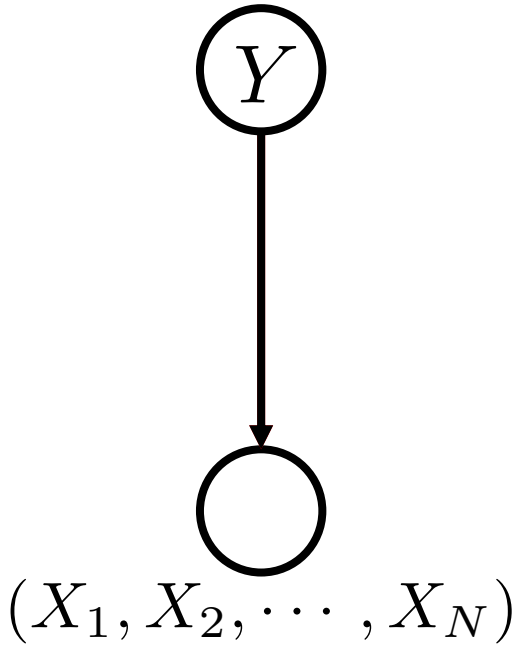
Posterior Regularization

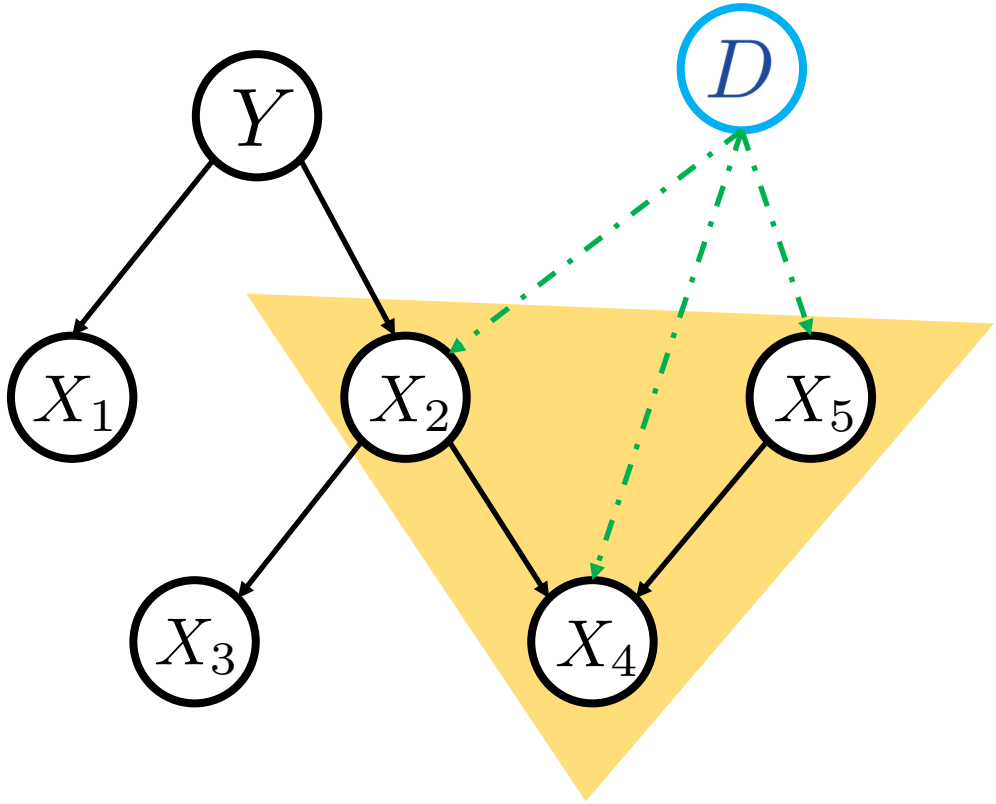


Developing Interpretable High-Dimensional Phenotype



Generative Domain Adaptation





M. Gong, K. Zhang, B Huang, C Glymour, D Tao, **K Batmanghelich**, "Causal Generative Domain Adaptation Networks, " arXiv:1804.04333

Acknowledgment

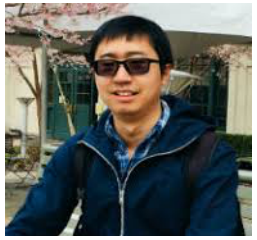
Collaborators:

Barnabas Pocsoz

Suvrit Sra

Group Members:

Payman Yadollahpour



Mingming Gong

Sumedha Singla

Javad Rahimikollu

Funding's:

- **SAP**: Deep Multi-Domain Learning: A Framework to Incorporate Weak Labels to the Attention Models
- **NIH 1R01HL141813-01:** An Integrative Radiogenomic Approach to Design Genetically-Informed Image Biomarker for Characterizing COPD.
- **Pfizer**: Developing Statistical Method to Jointly Model Genotype and High Dimensional Imaging Endophenotype.
- **CMRF**: Competitive Medical Research Fund Grant Application