



Poster session #3,
P03-037, Tuesday, Oct 10 2023, 9.30-11.00 AM

Distilling BlackBox to Interpretable models for Efficient Transfer Learning



Shantanu Ghosh¹, Ke Yu², Kayhan Batmanghelich¹

¹BU ECE, ²Pitt ISP



Neural networks fail to generalize



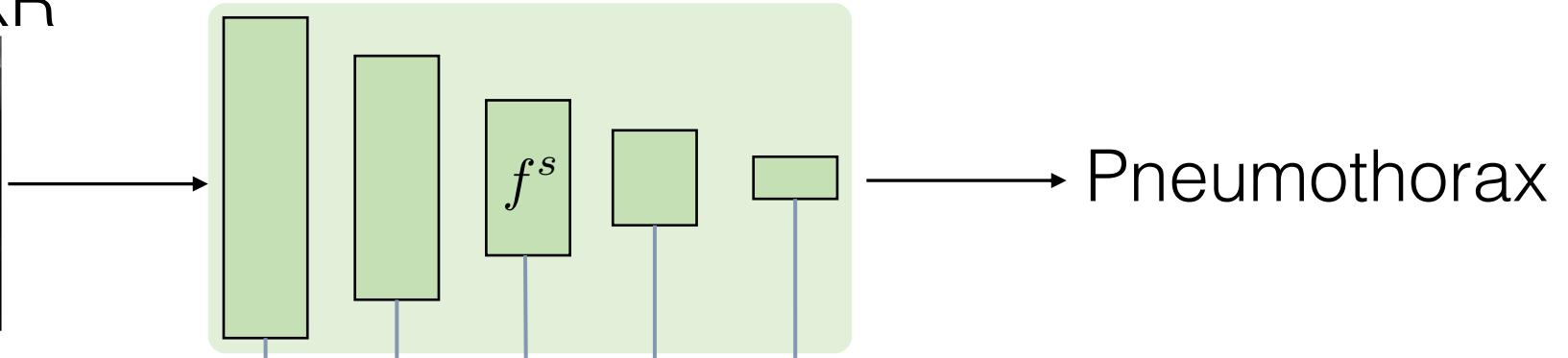
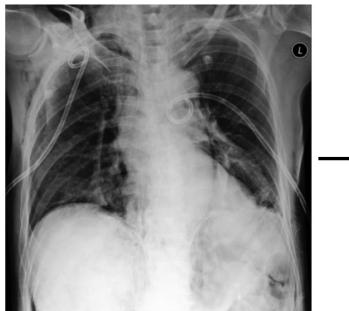
Impression:

Mild left costophrenic blunting, basilar pleural effusion, increased left suprahilar opacity, differential diagnosis includes increased volume loss and apical pleural fluid. Right costophrenic, right lung free of focal consolidation.

MTI Tags: Degenerative changes

Finetuning can solve the problem

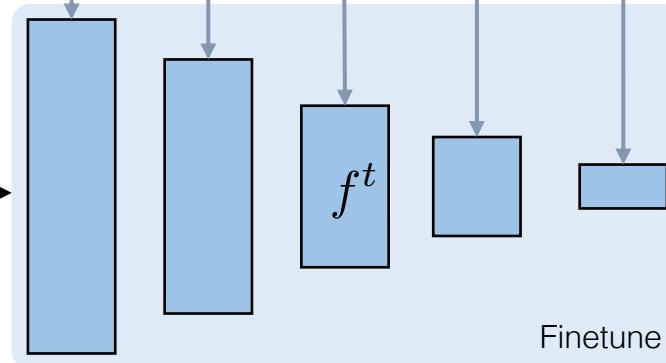
MIMIC-CXR



Stanford-CXR



Transfer learning



Data-efficient fine-tuning

MIMIC-CXR



?pneumothorax

Data and Computationally inefficient

Stanfo



?pneumothorax

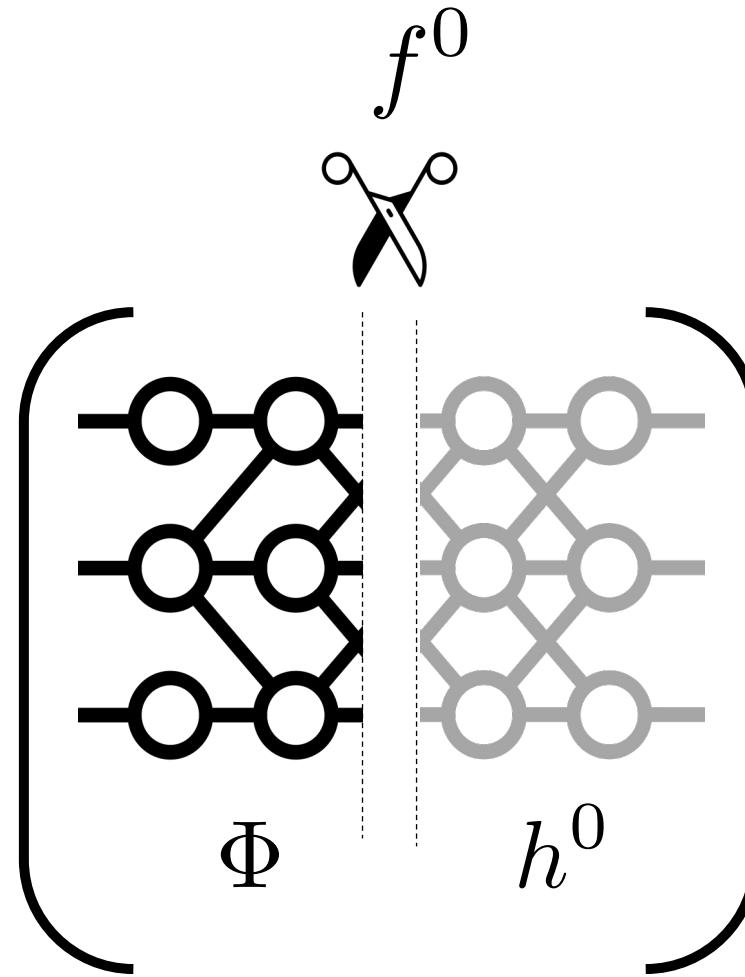
The clinical rules are “invariant”



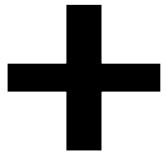
Finetune

So, we start with a Blackbox and carve out
interpretable models from the Blackbox
to extract
domain invariant rules.

Assumptions



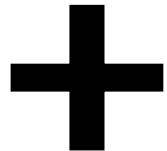
Assumptions



Report:

Right upper lobe consolidation with adjacent.
While this may be infectious in nature, a CT
scan is recommended for further clarification.

Assumptions



Report:

Right upper lobe consolidation with adjacent.
While this may be infectious in nature, a CT
scan is recommended for further clarification.

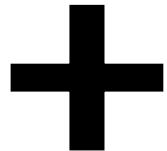
parse the reports to get the concepts

↓

C

right upper lobe
left lower lobe
heart size
....

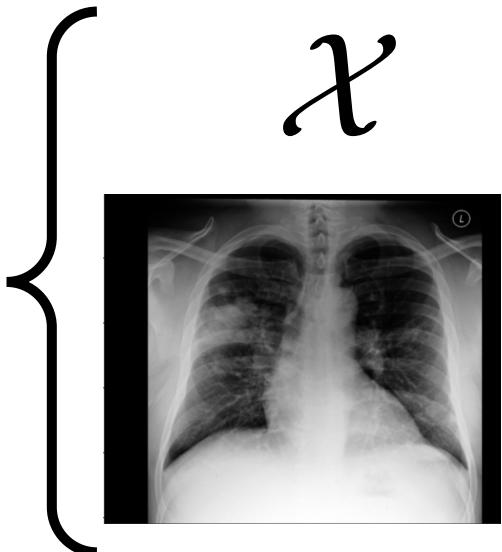
Assumptions



Report:

Right upper lobe **consolidation** with adjacent.
While this **may** be **infectious** in nature, a CT
scan is recommended for further clarification.

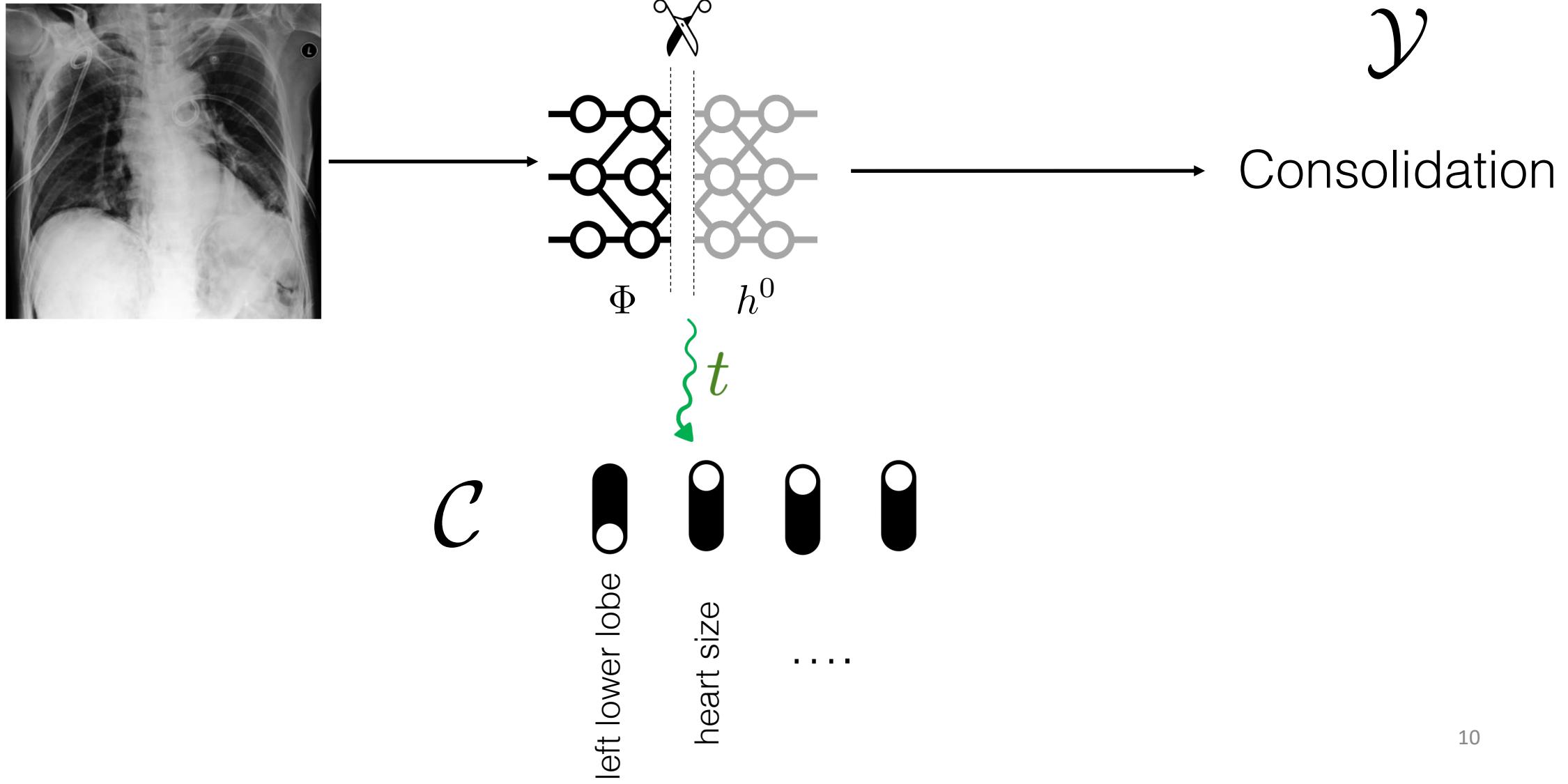
parse the reports to get the concepts



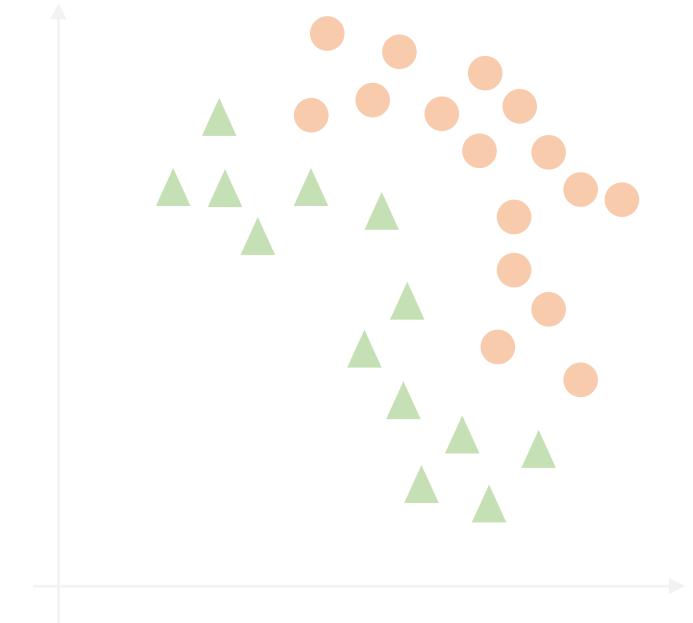
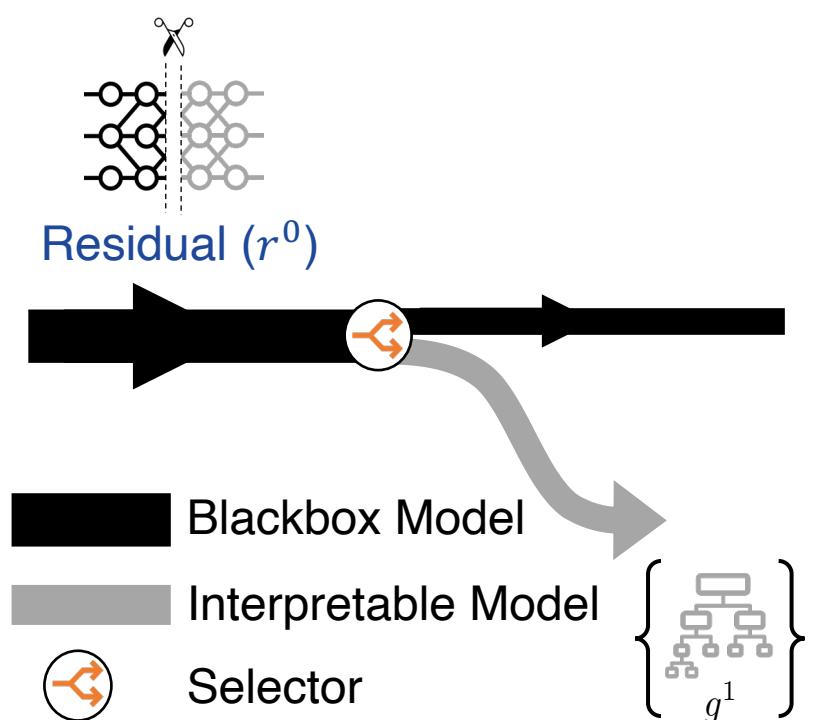
right upper lobe
left lower lobe
heart size
....

γ
Consolidation

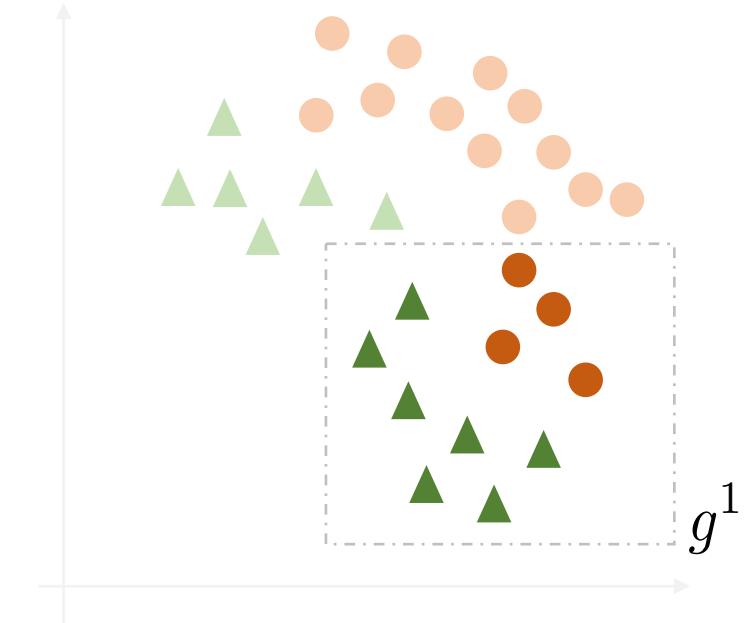
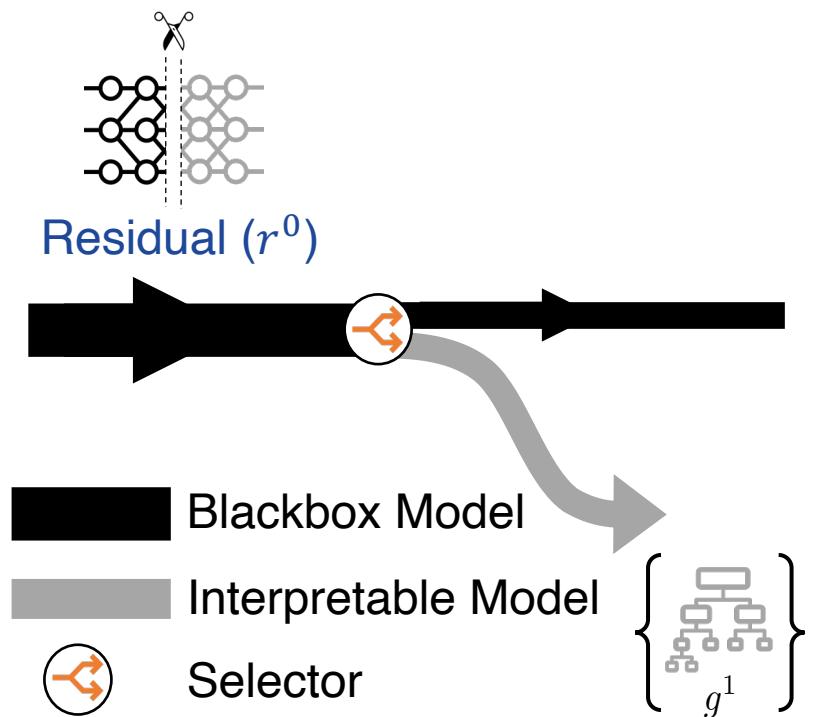
Discovering the hidden concepts



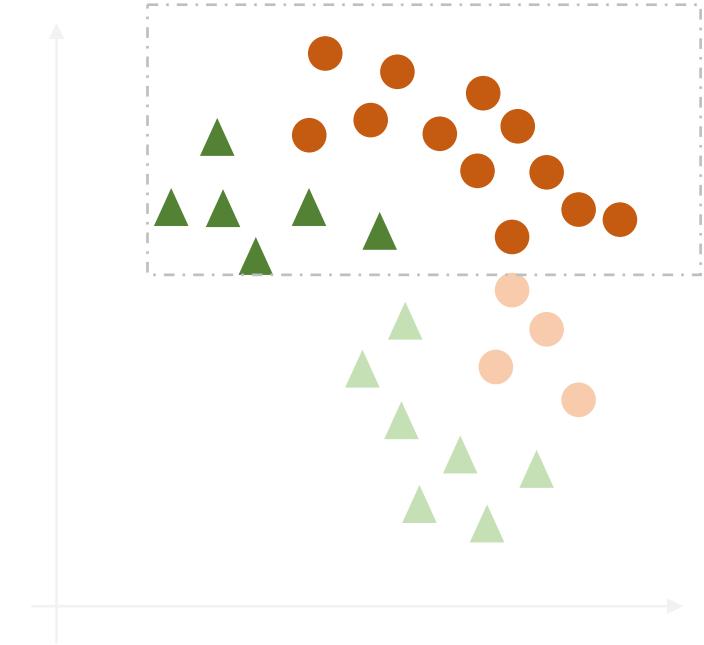
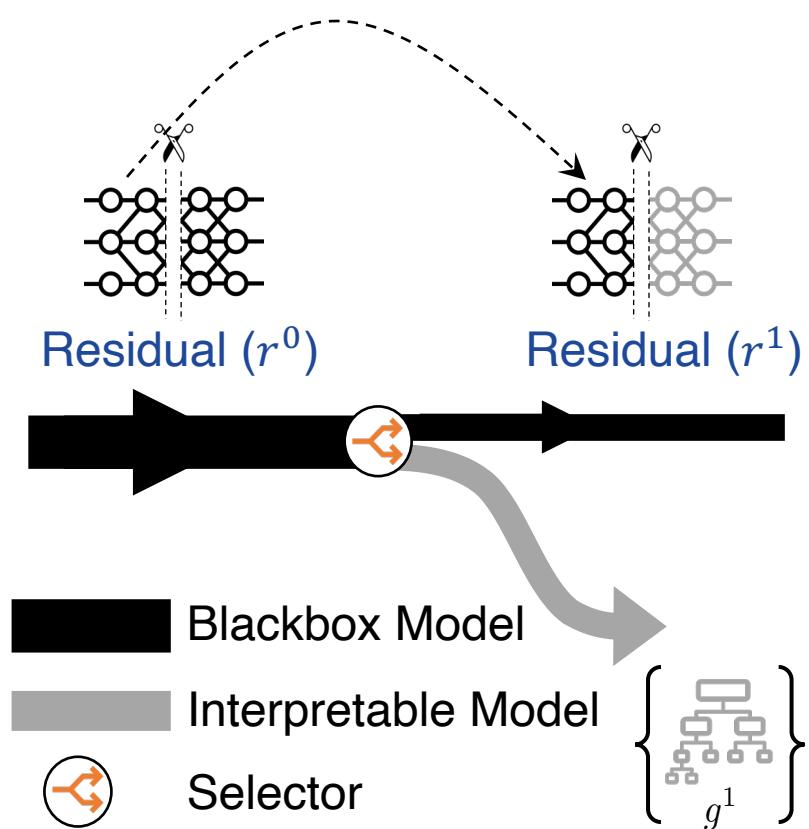
Carving out interpretable models



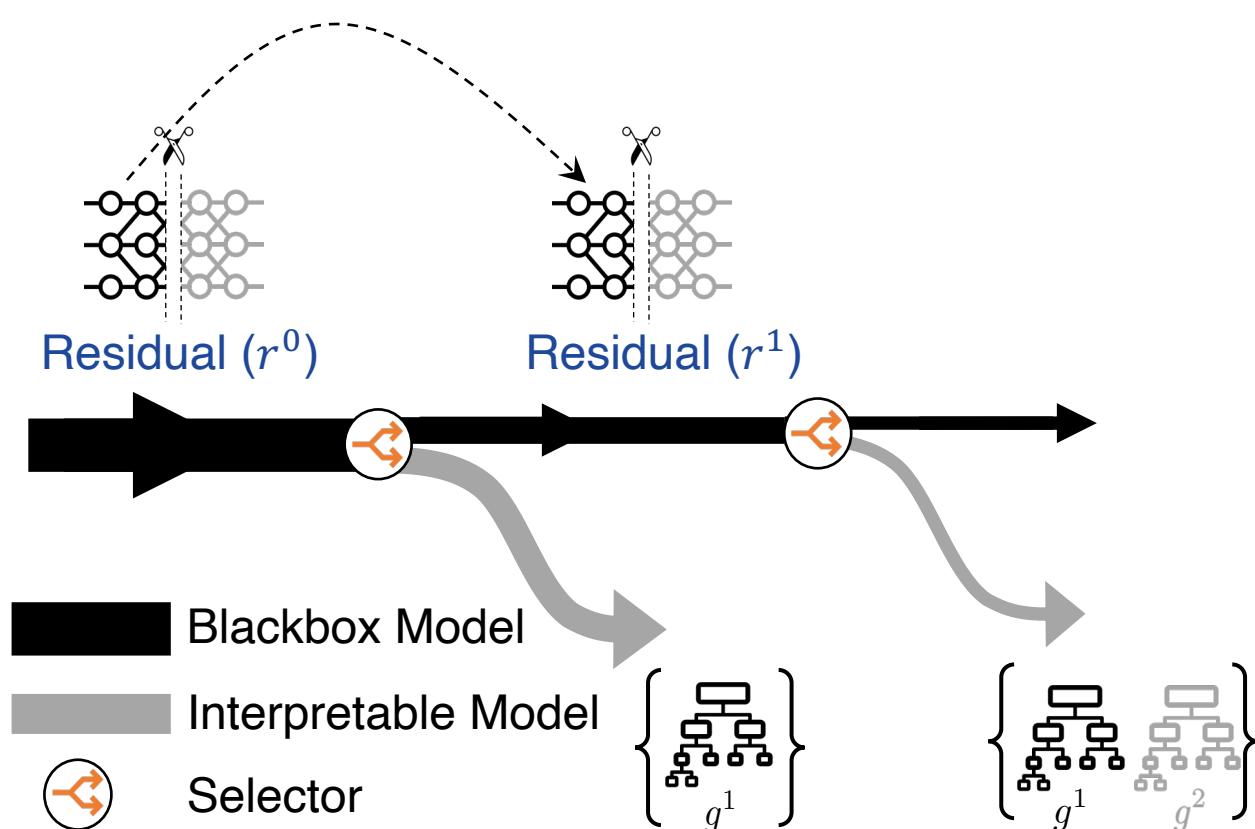
Carving out interpretable models



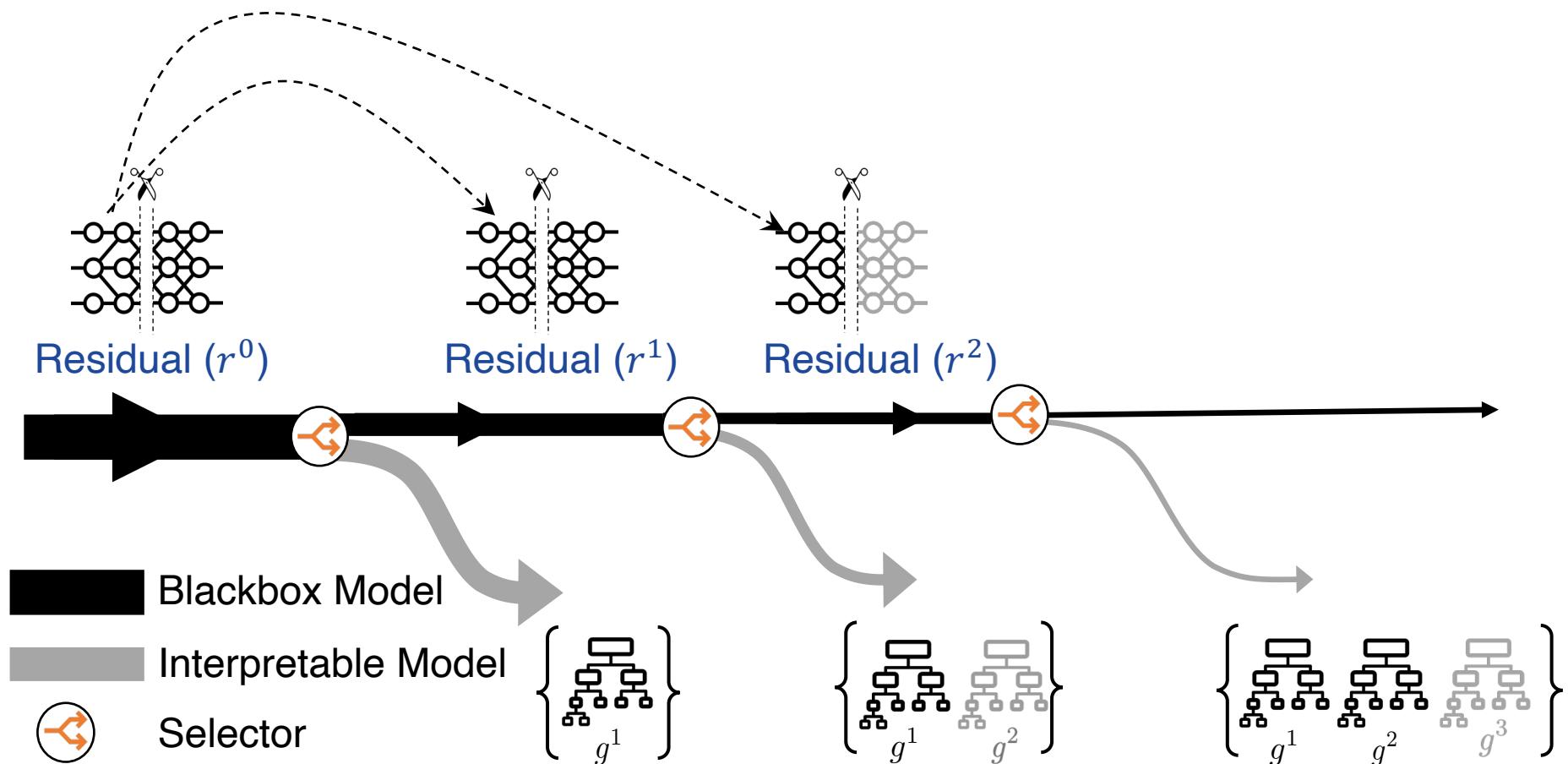
Carving out interpretable models



Carving out interpretable models

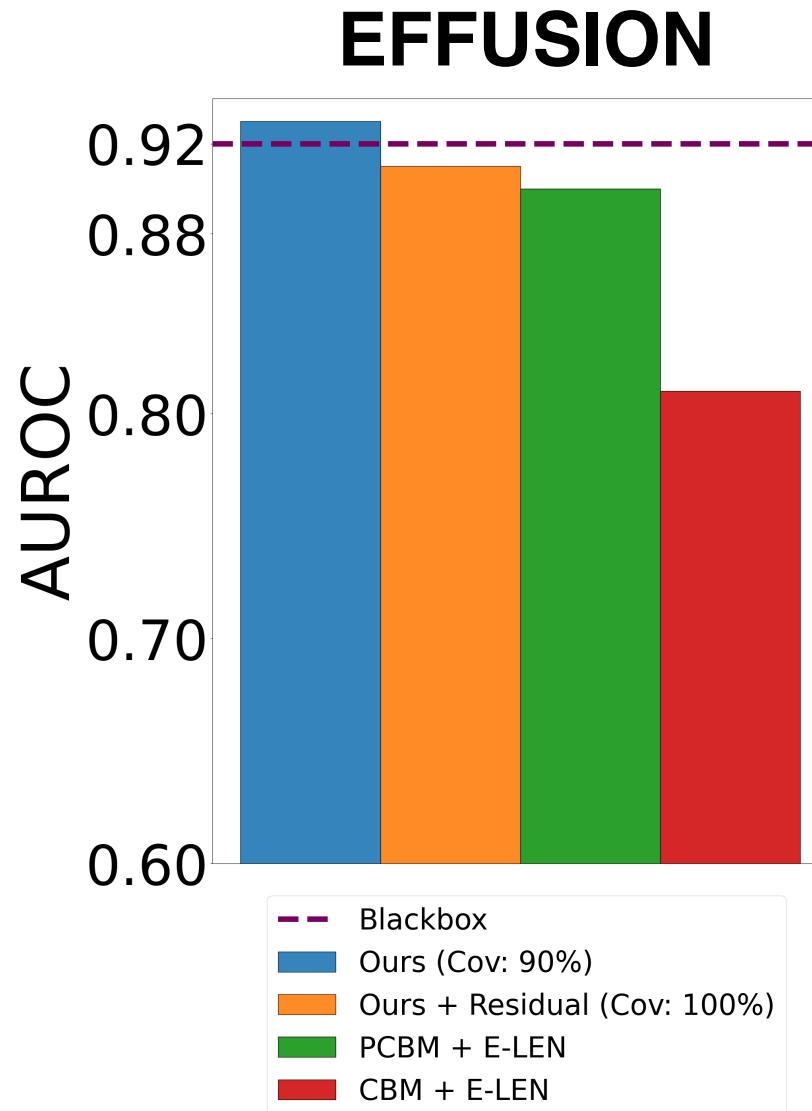


Carving out interpretable models



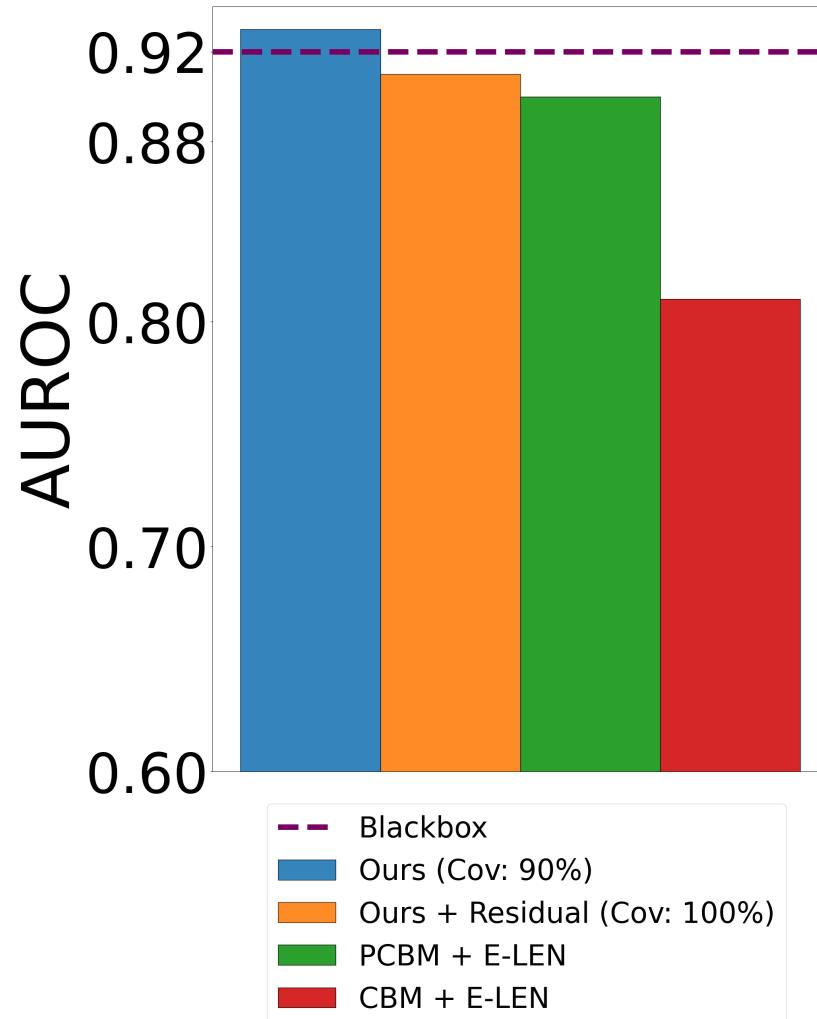
Each g is Entropy based logical neural networks (Barberio et al. AAAI 2022).

Does not compromise performance

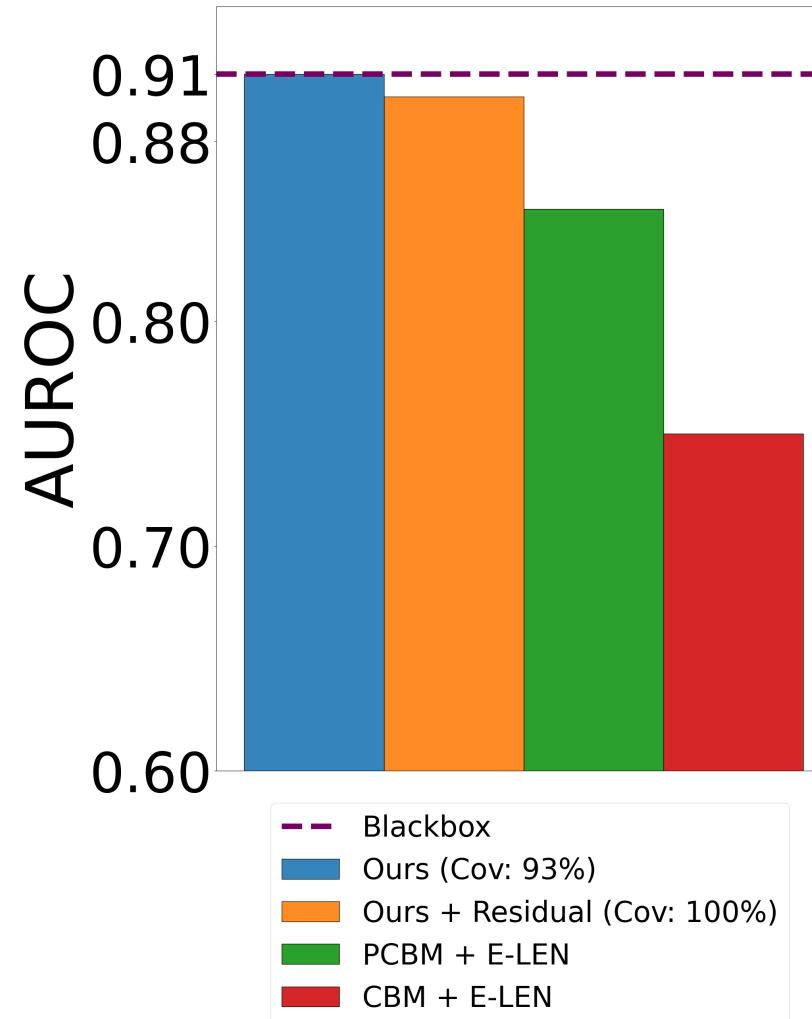


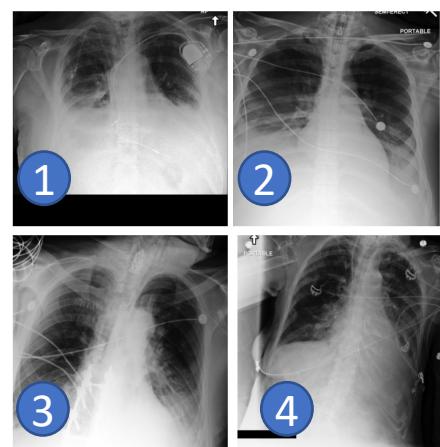
Does not compromise performance

EFFUSION



PNEUMOTHORAX

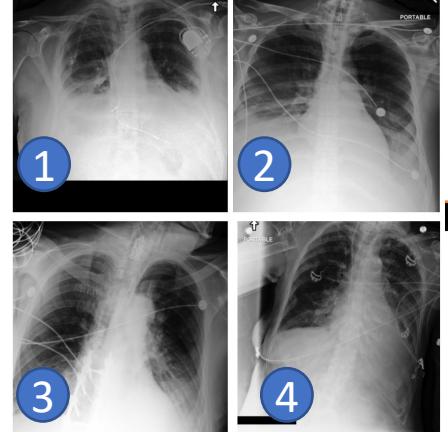




Diversity in local explanations

Pleural unspec is “unspecified pleural effusion” referred to as “hydrothorax”.

Hydrothorax is a noninflammatory collection of serous fluid within the pleural cavities



Diversity in local explanations



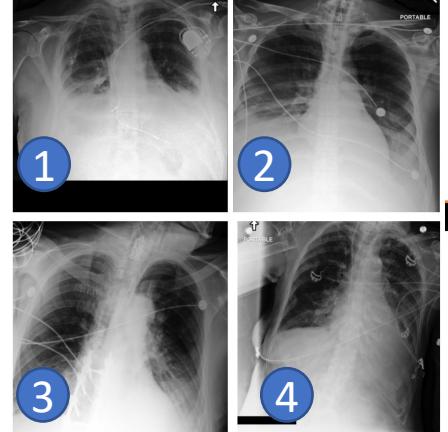
Expert 1

Effusion ↔
left_pleural
^ right_pleural
^ pleural_unspec

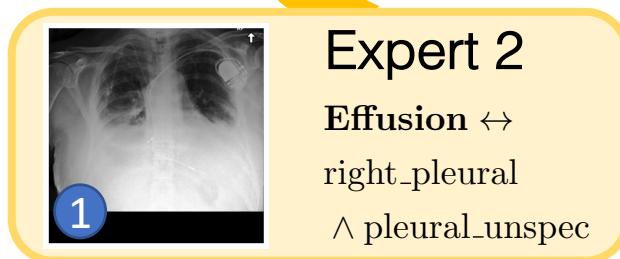
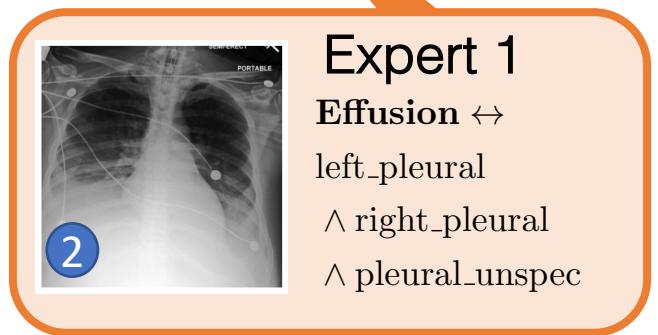


Pleural_unspec is “unspecified pleural effusion” referred to as “hydrothorax”.

Hydrothorax is a noninflammatory collection of serous fluid within the pleural cavities

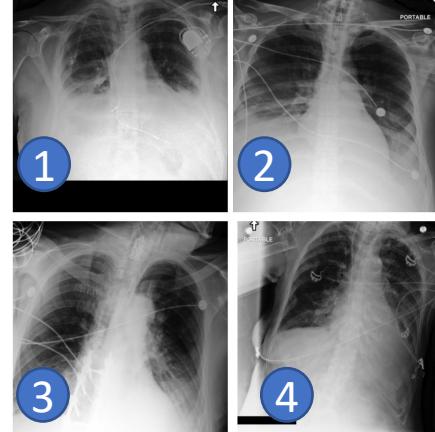


Diversity in local explanations

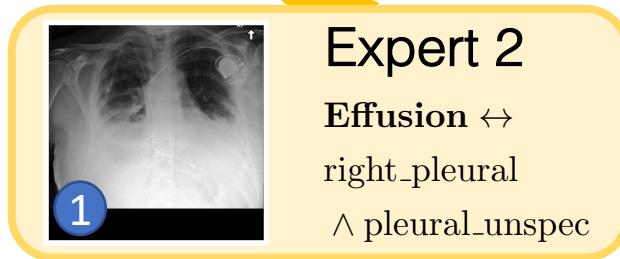
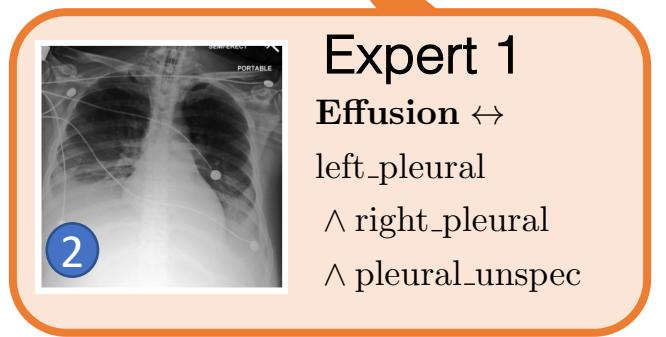


Pleural_unspec is “unspecified pleural effusion” referred to as “hydrothorax”.

Hydrothorax is a noninflammatory collection of serous fluid within the pleural cavities



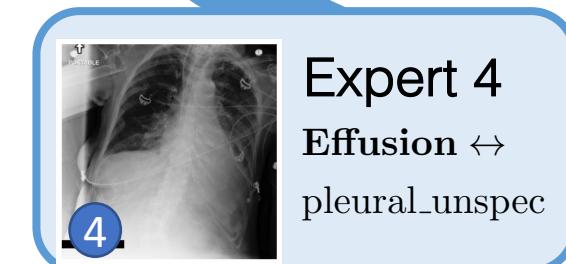
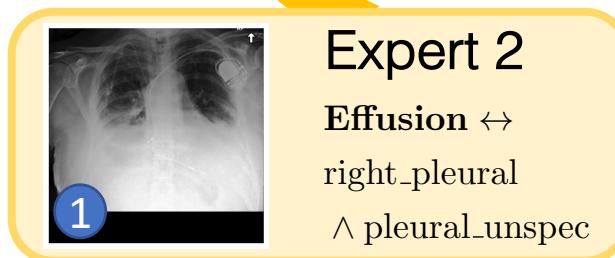
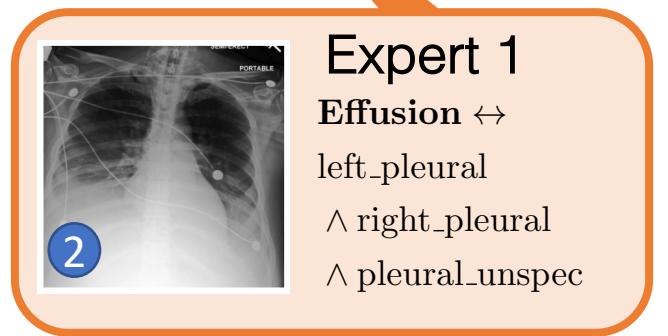
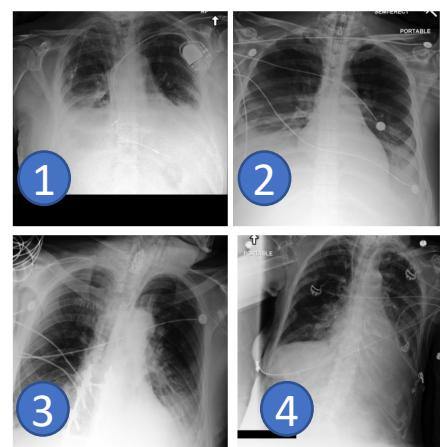
Diversity in local explanations



Pleural_unspec is “unspecified pleural effusion” referred to as “hydrothorax”.

Hydrothorax is a noninflammatory collection of serous fluid within the pleural cavities

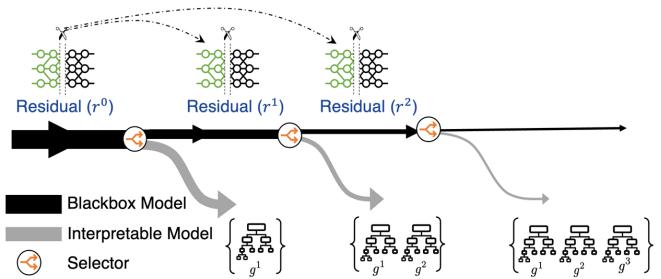
Diversity in local explanations



Fine-tune to a New Domain

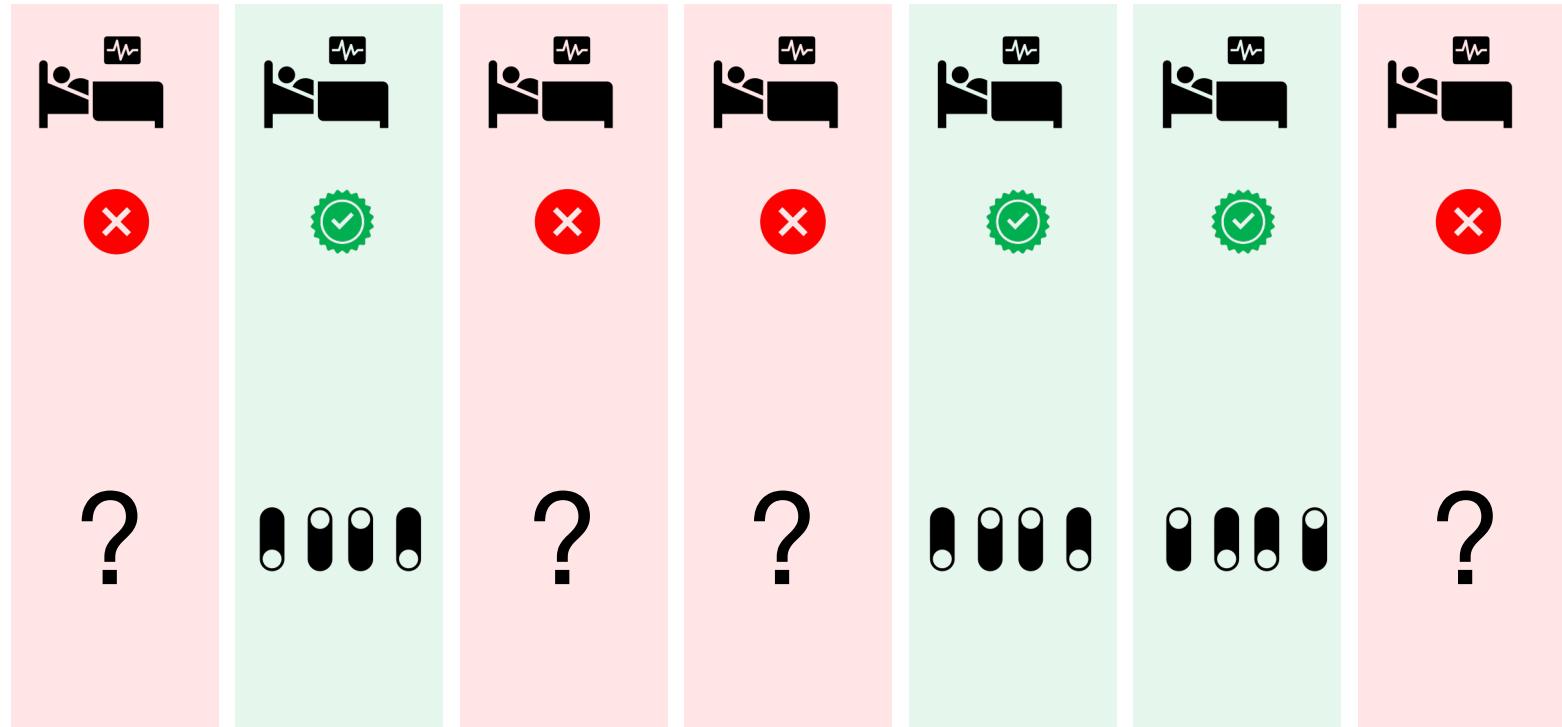
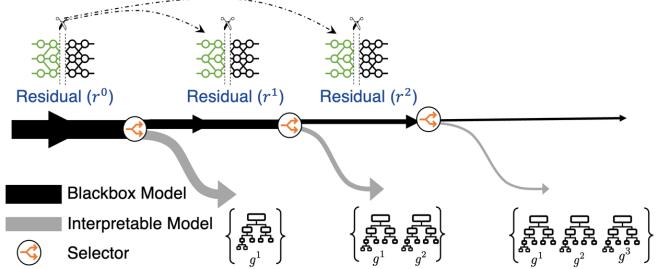
1

Apply source model



Fine-tune to a New Domain

1 Apply source model

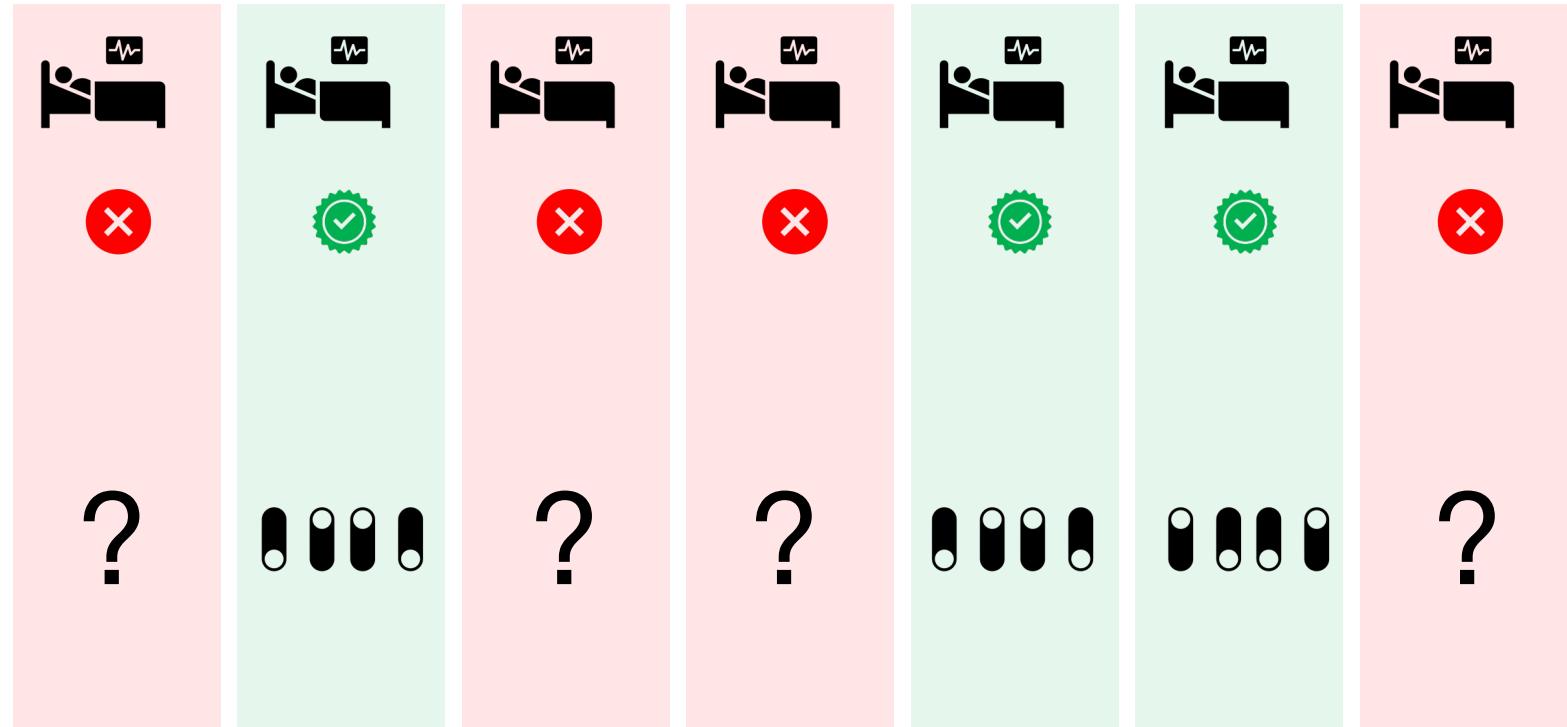
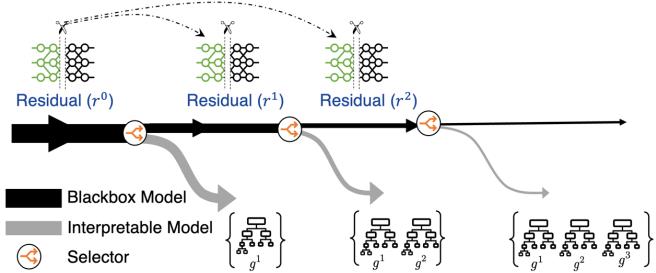


2 Use concepts from matching patients

C

Fine-tune to a New Domain

1 Apply source model



2 Use concepts from matching patients

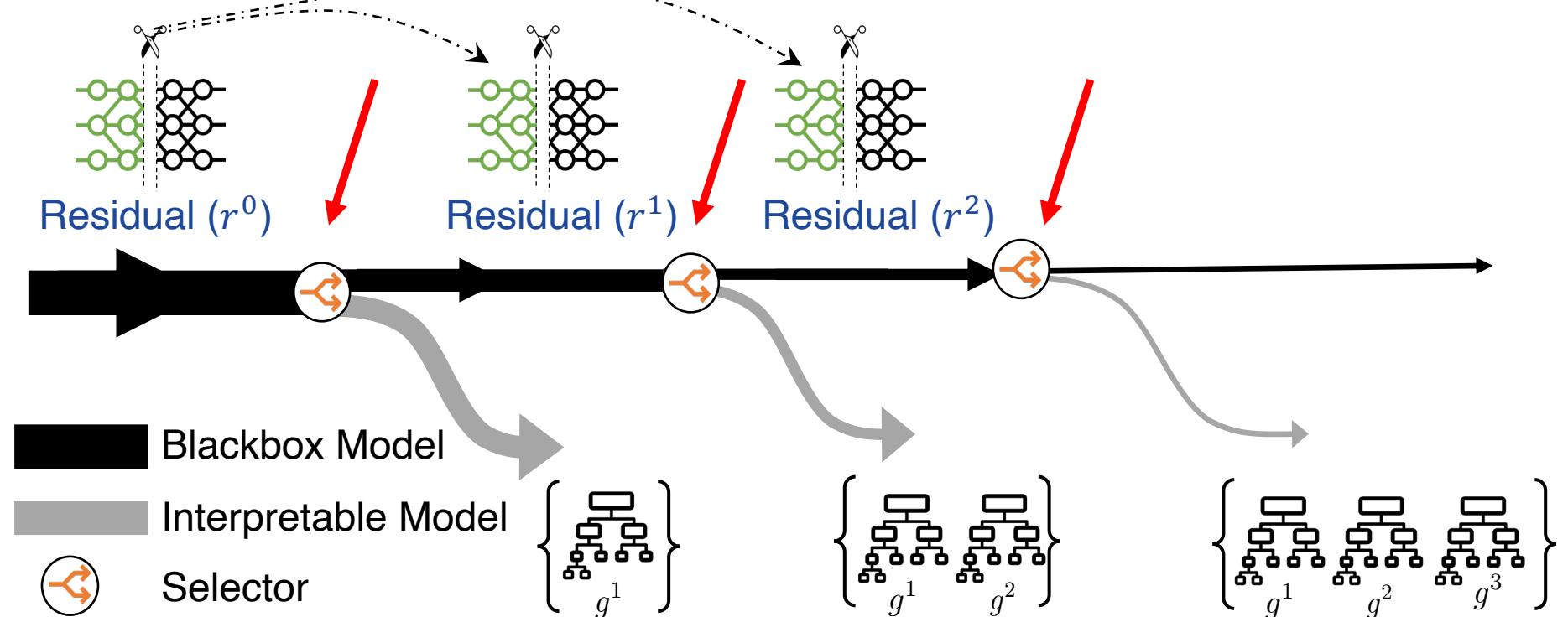
C

3 Propagate the concepts and update the concept extractor

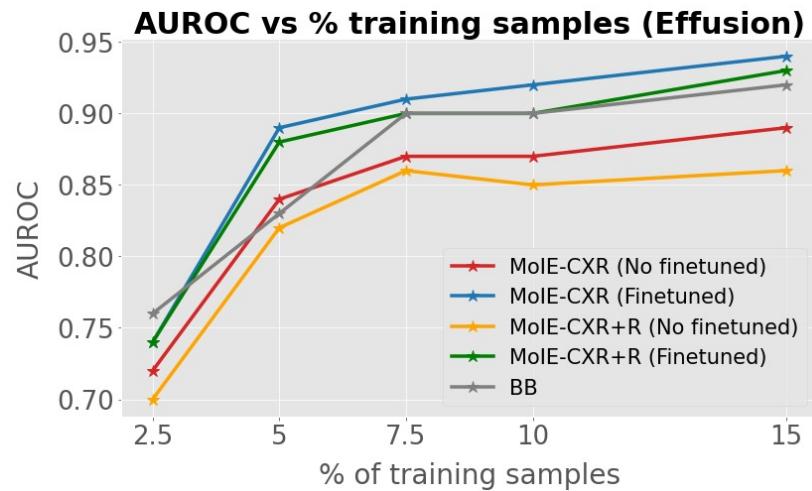
Fine-tune to a New Domain

4

Update the selectors and experts for only a few epochs



Transferring to Stanford-CXR

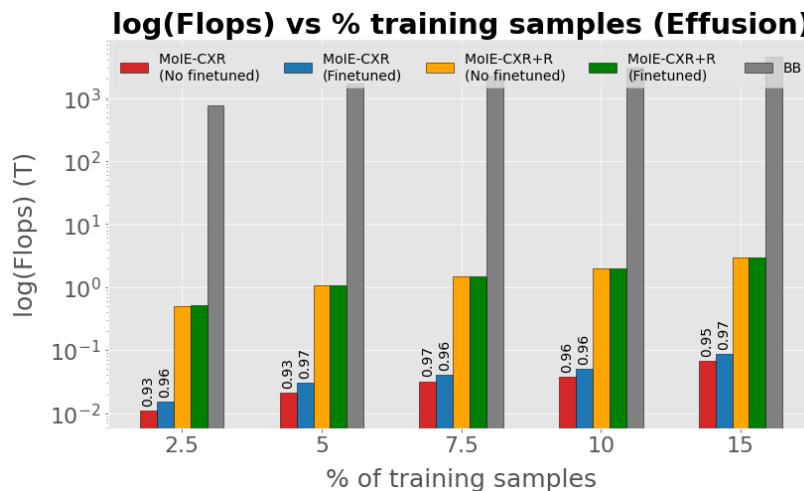
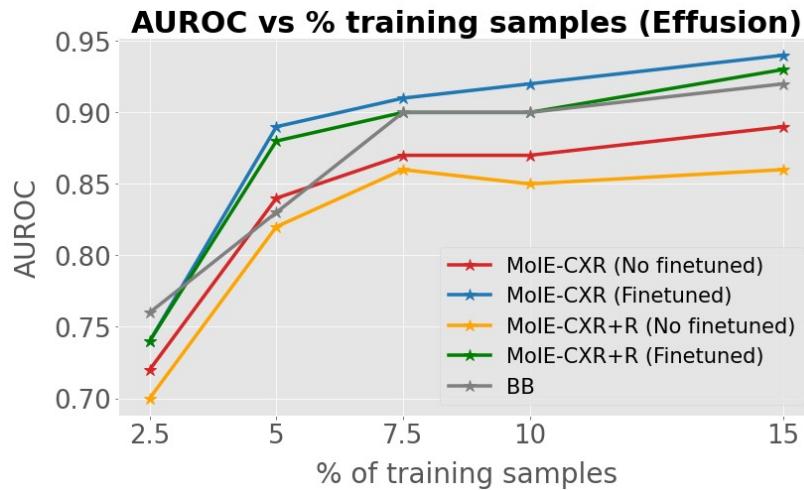


MoIE-CXR (No finetuned): only selectors are fine-tuned.

MoIE-CXR (Finetuned): selectors and experts are fine-tuned.

MoIE and MoIE-CXR + R consists of mixture of experts excluding and including the final residual respectively.

Transferring to Stanford-CXR

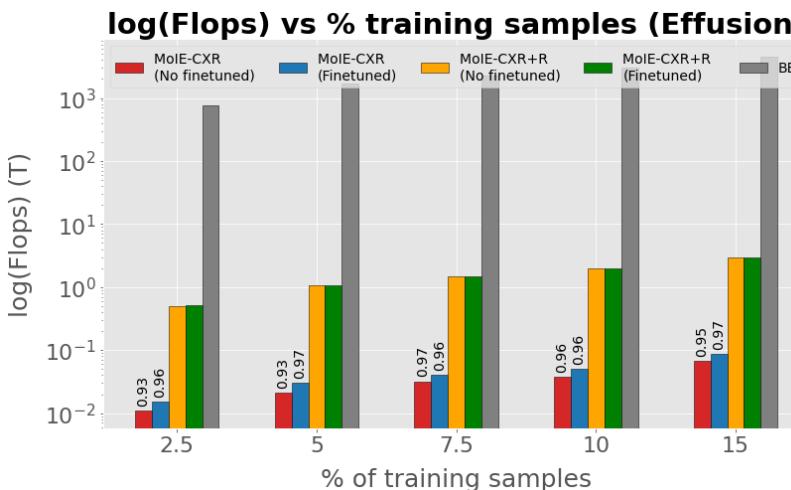
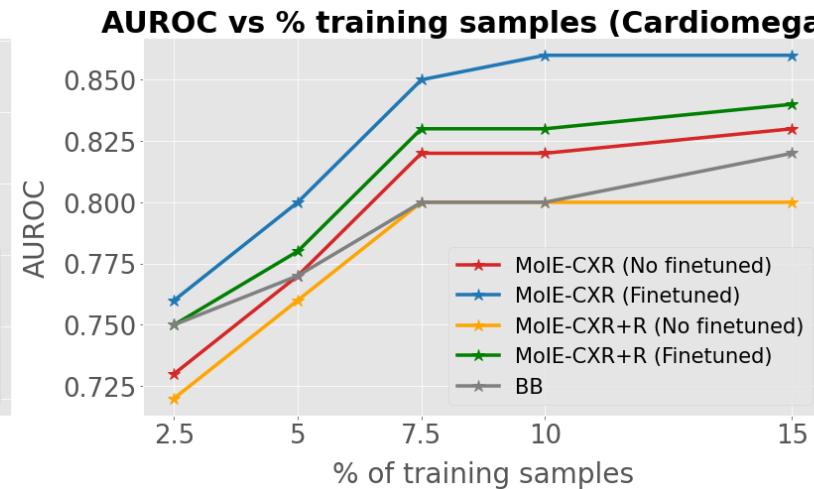
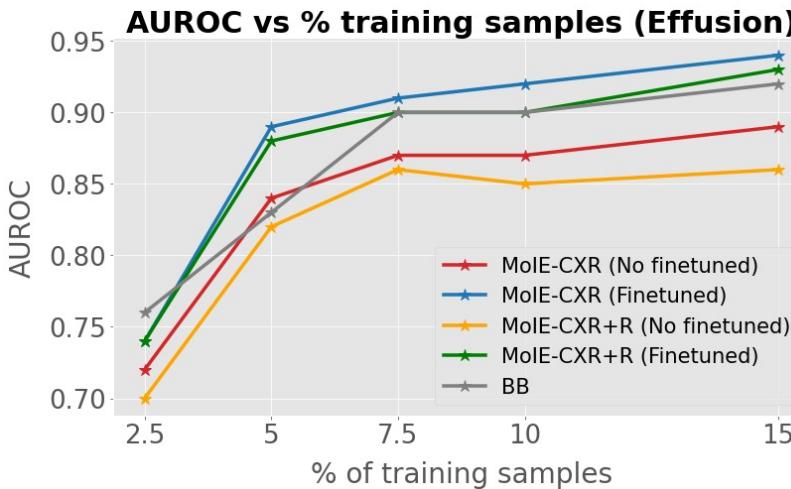


MolE-CXR (No finetuned): only selectors are fine-tuned.

MolE-CXR (Finetuned): selectors and experts are fine-tuned.

MolE and MolE-CXR + R consists of mixture of experts excluding and including the final residual respectively.

Transferring to Stanford-CXR

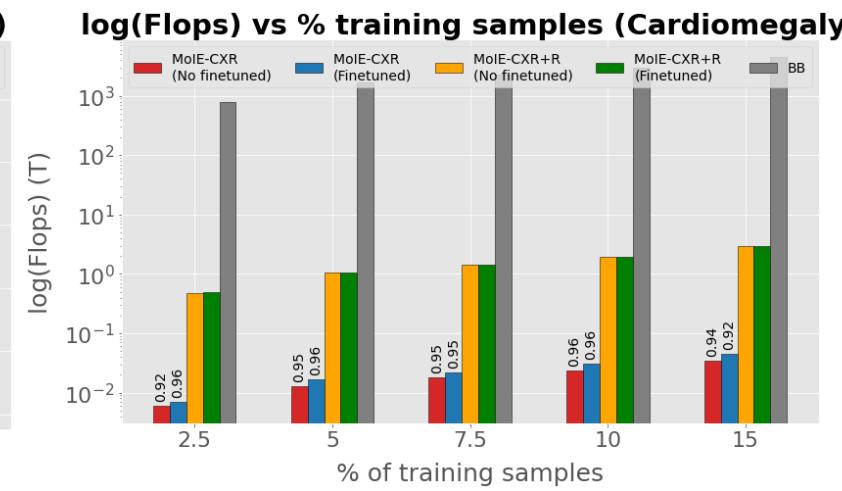
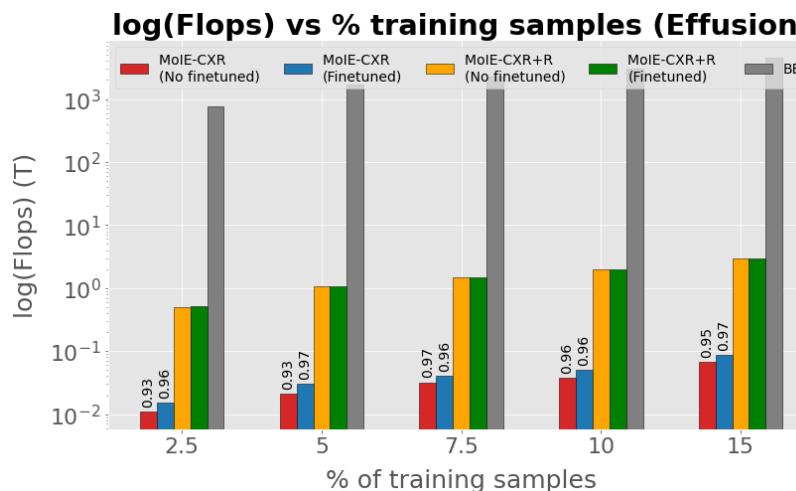
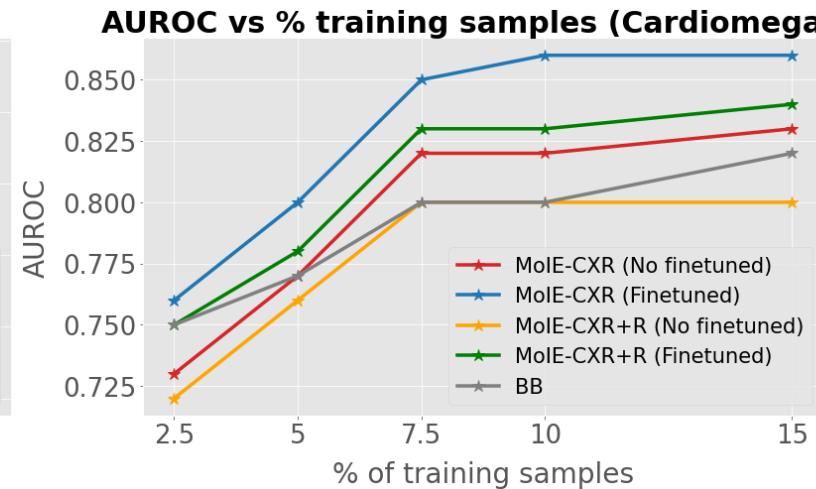
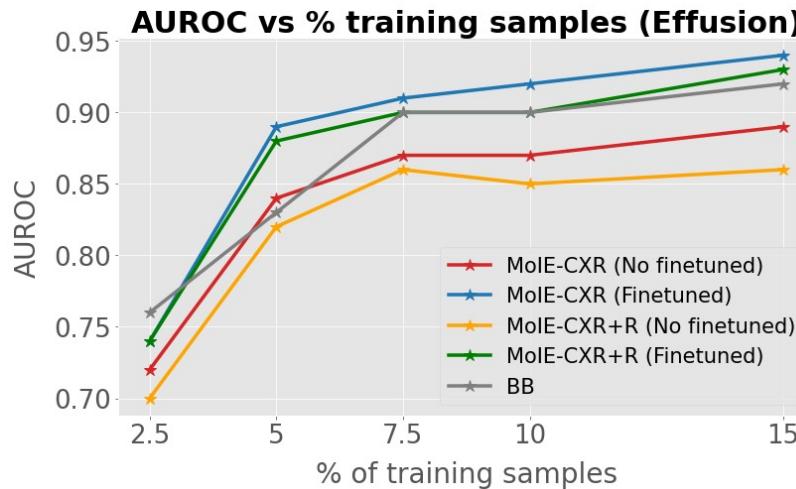


MoIE-CXR (No finetuned): only selectors are fine-tuned.

MoIE-CXR (Finetuned): selectors and experts are fine-tuned.

MoIE and MoIE-CXR + R consists of mixture of experts excluding and including the final residual respectively.

Transferring to Stanford-CXR

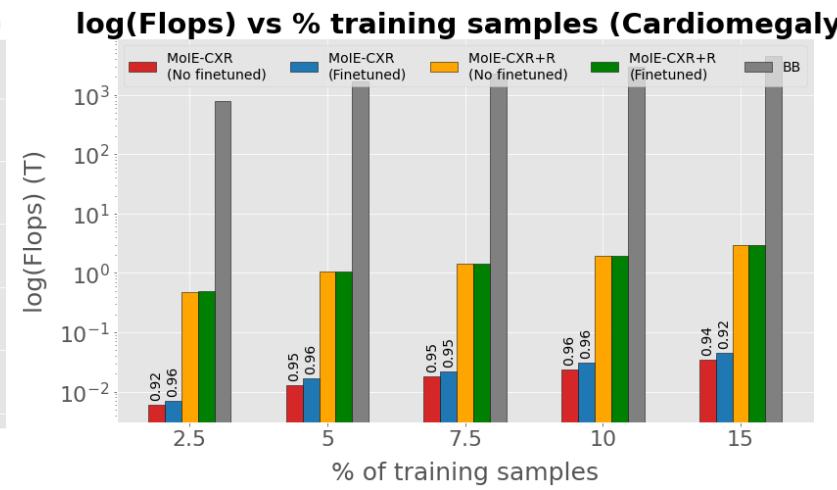
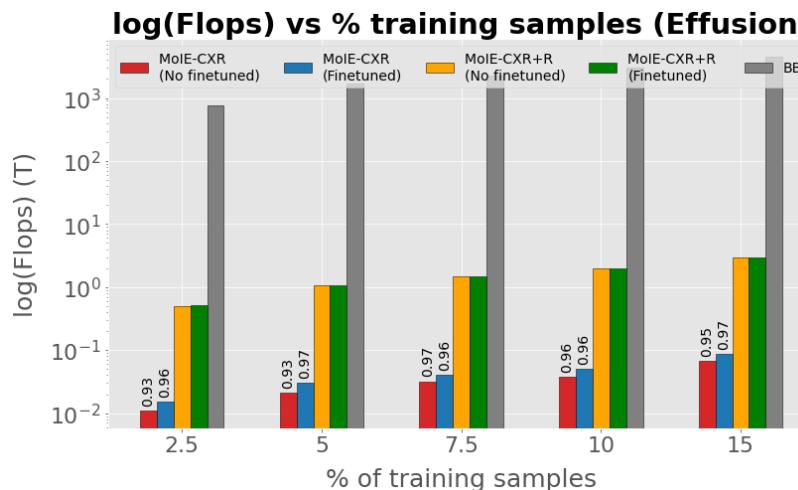
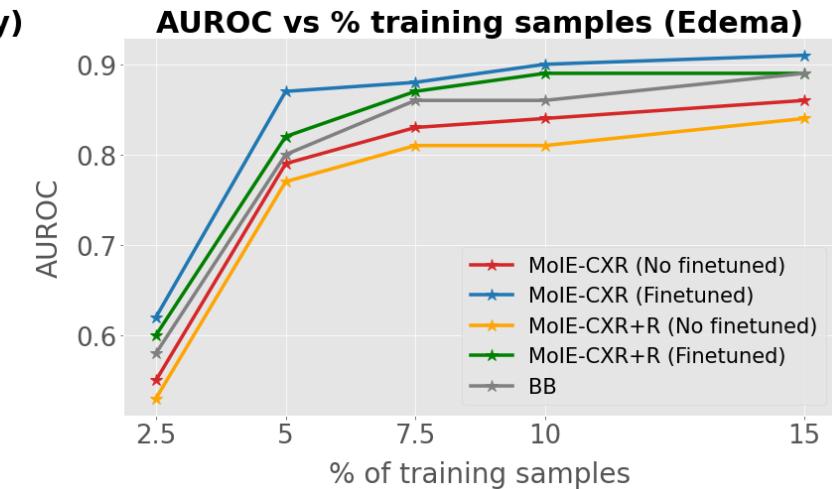
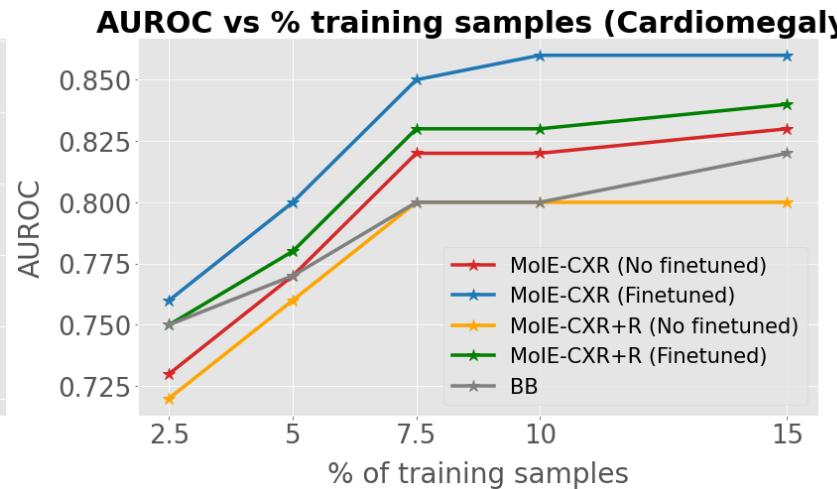
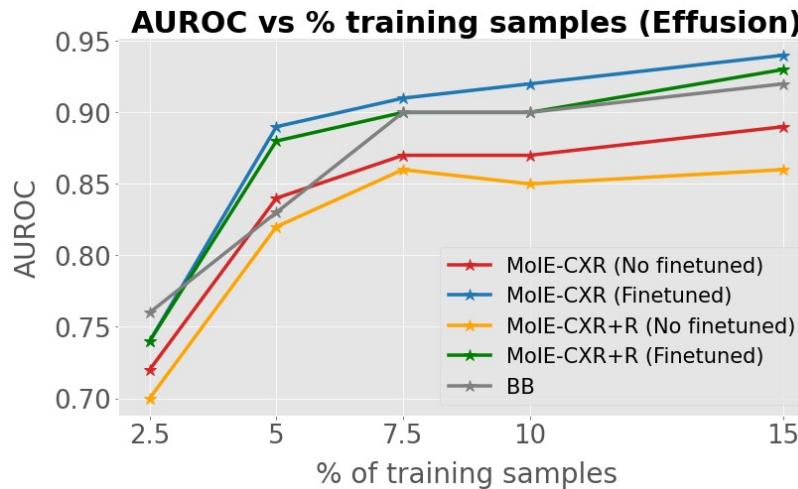


MoIE-CXR (No finetuned): only selectors are fine-tuned.

MoIE-CXR (Finetuned): selectors and experts are fine-tuned.

MoIE and MoIE-CXR + R consists of mixture of experts excluding and including the final residual respectively.

Transferring to Stanford-CXR

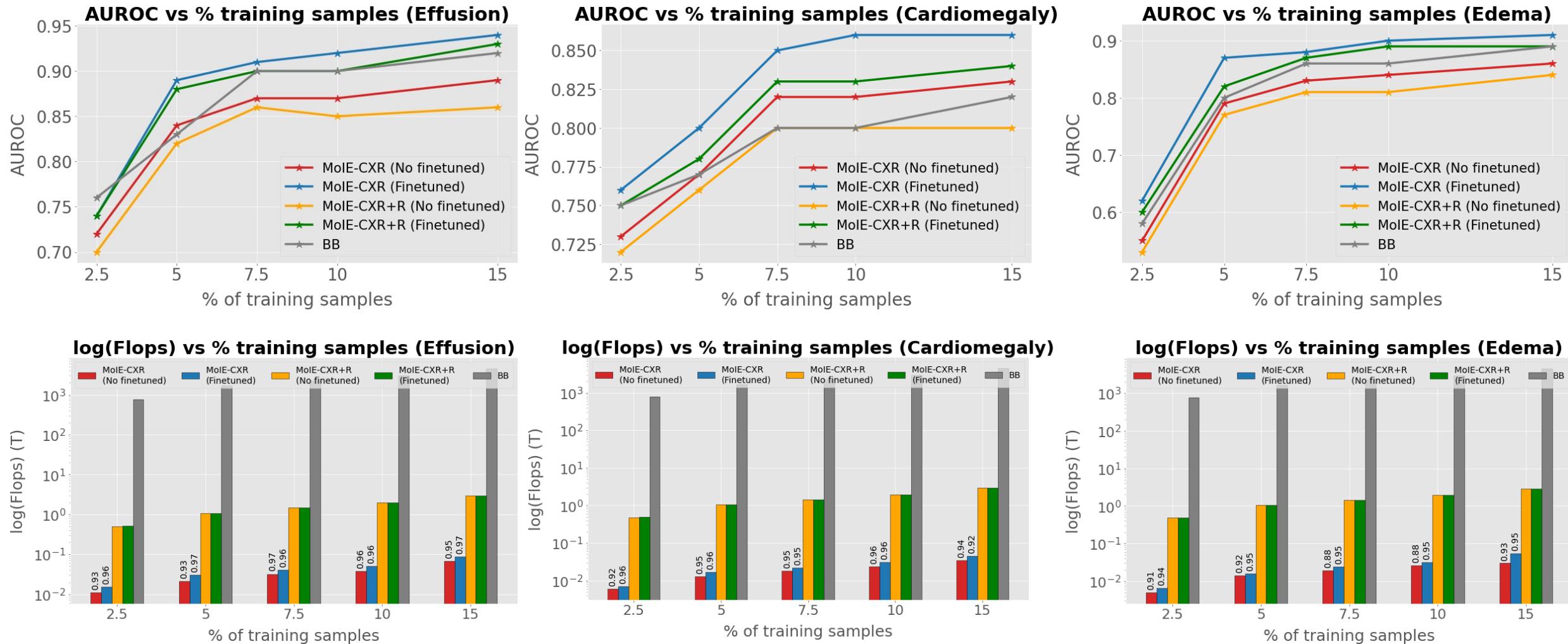


MolE-CXR (No finetuned): only selectors are fine-tuned.

MolE-CXR (Finetuned): selectors and experts are fine-tuned.

MolE and MolE-CXR + R consists of mixture of experts excluding and including the final residual respectively.

Transferring to Stanford-CXR



MolE-CXR (No finetuned): only selectors are fine-tuned.

MolE-CXR (Finetuned): selectors and experts are fine-tuned.

MolE and MolE-CXR + R consists of mixture of experts excluding and including the final residual respectively.



Project website



Shantanu Ghosh¹, Ke Yu², Kayhan Batmanghelich¹

¹BU ECE, ²Pitt ISP

Poster session #3,
P03-037, Tuesday, Oct 10 2023, 9.30-11.00 AM

Thank you