MICCAI2022 Singapore

DART

International Workshop on Domain Adaptation & Representation Transfer

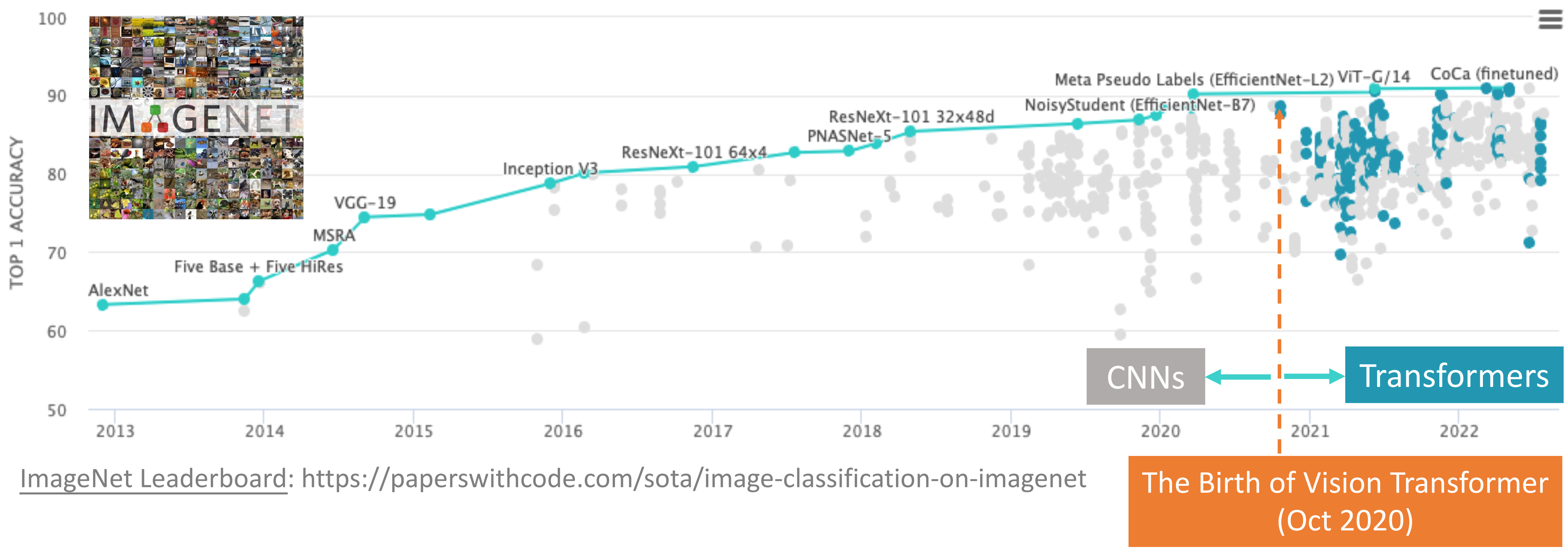
Benchmarking and Boosting Transformers for Medical Image Classification

DongAo Ma[1], Mohammad Reza Hosseinzadeh Taher[1], Jiaxuan Pang[1], Nahid Ul Islam[1], Fatemeh Haghighi[1] , Michael B. Gotway[2], Jianming Liang[1]

[1] Arizona State University [2] Mayo Clinic

Transformers

- Refreshing ImageNet Leaderboard
- Dominating Computer Vision

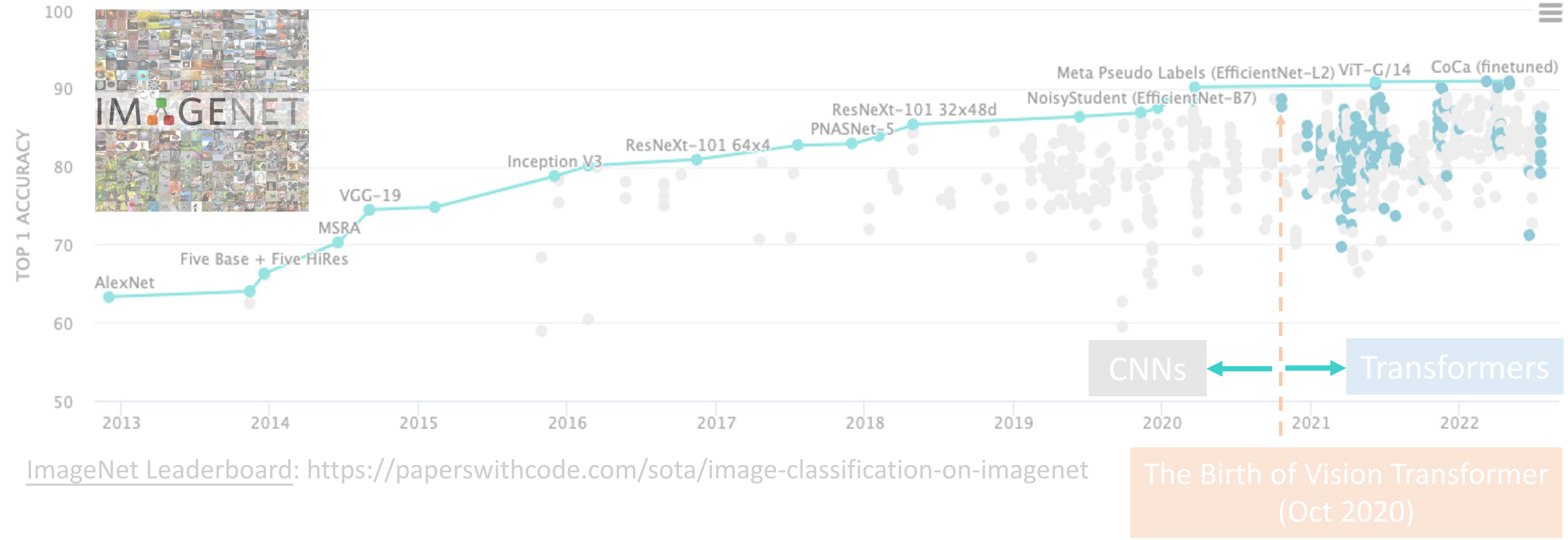


ImageNet Leaderboard: https://paperswithcode.com/sota/image-classification-on-imagenet

Transformers

- Refreshing ImageNet Leaderboard
- Dominating Computer Vision

How do visual transformers perform on medical images?



ImageNet Leaderboard: https://paperswithcode.com/sota/image-classification-on-imagenet

Benchmarking transformers

Two most popular architectures

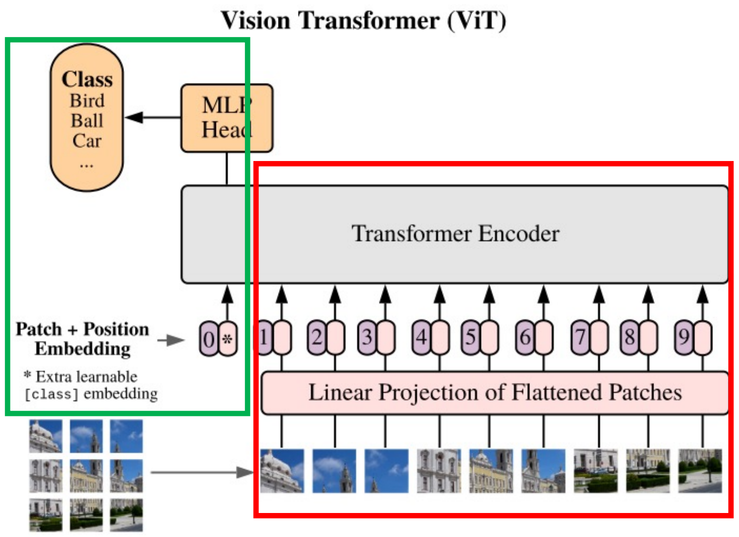
- **Vision Transformer (ViT)**

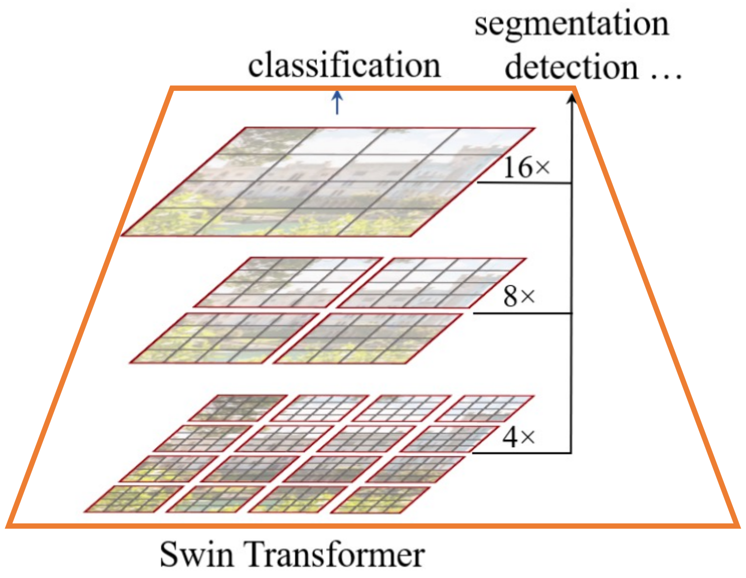
  1. Model global context of an image

  2. Capture patch-wise intra-image relations

- **Swin Transformer (Swin)**

  3. Builds hierarchical feature maps

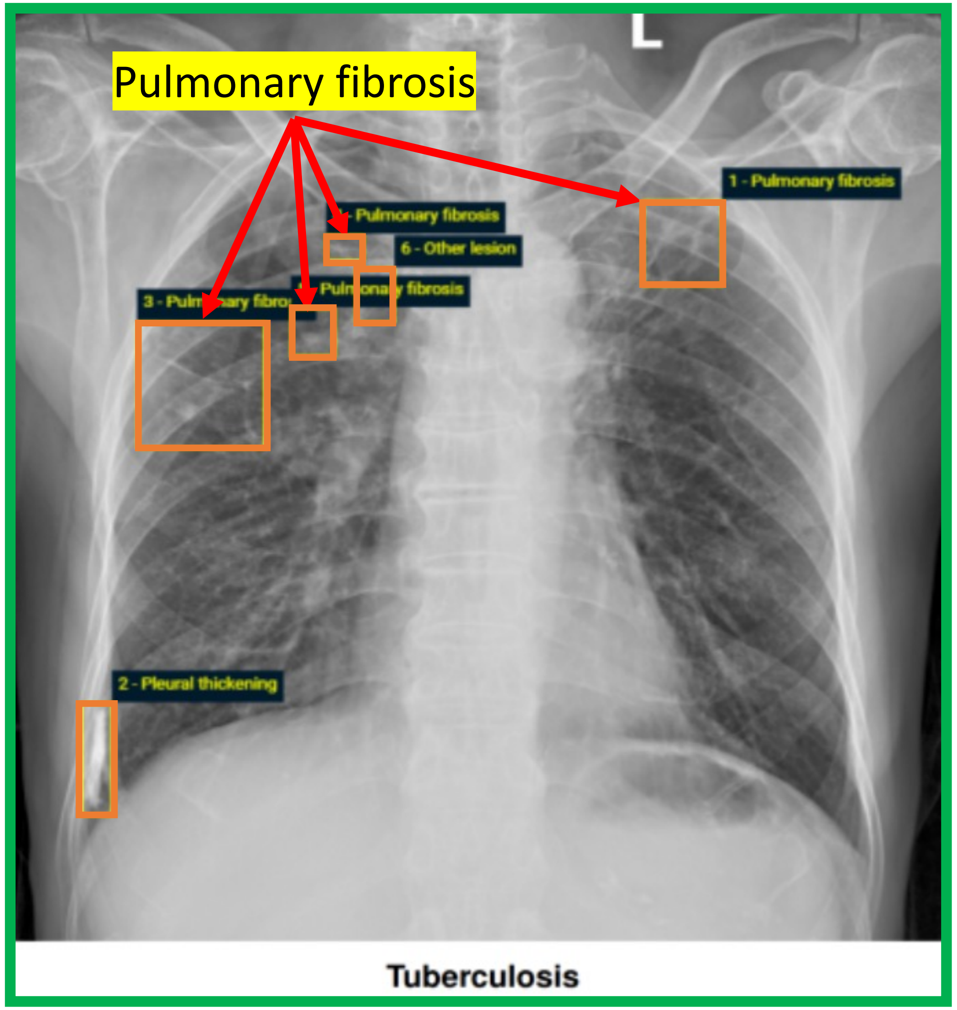




Benchmarking transformers

Two most popular architectures

- Vision Transformer (ViT)
  1. Model global context of an image
     - **Global context of a body region**
  2. Capture patch-wise intra-image relations
     - **Co-occurrence of pathologies**
- Swin Transformer (Swin)
  3. Builds hierarchical feature maps
     - **Lesions at different scales**



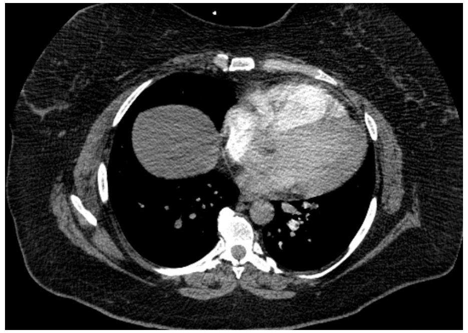
VinDr-CXR: https://vindr.ai/datasets/cxr

Transformers have good properties for medical images

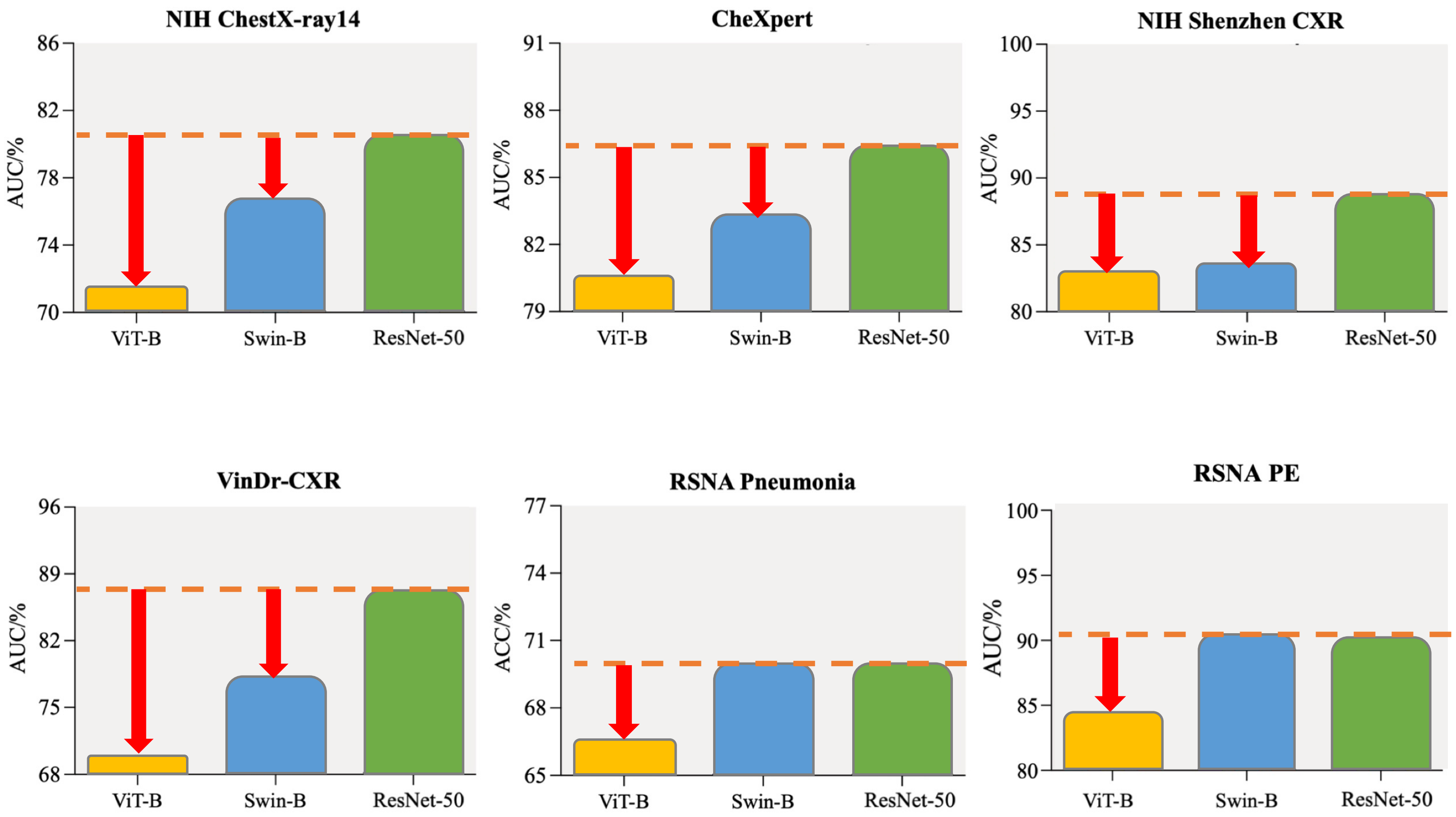
Benchmarking transformers

Target Tasks

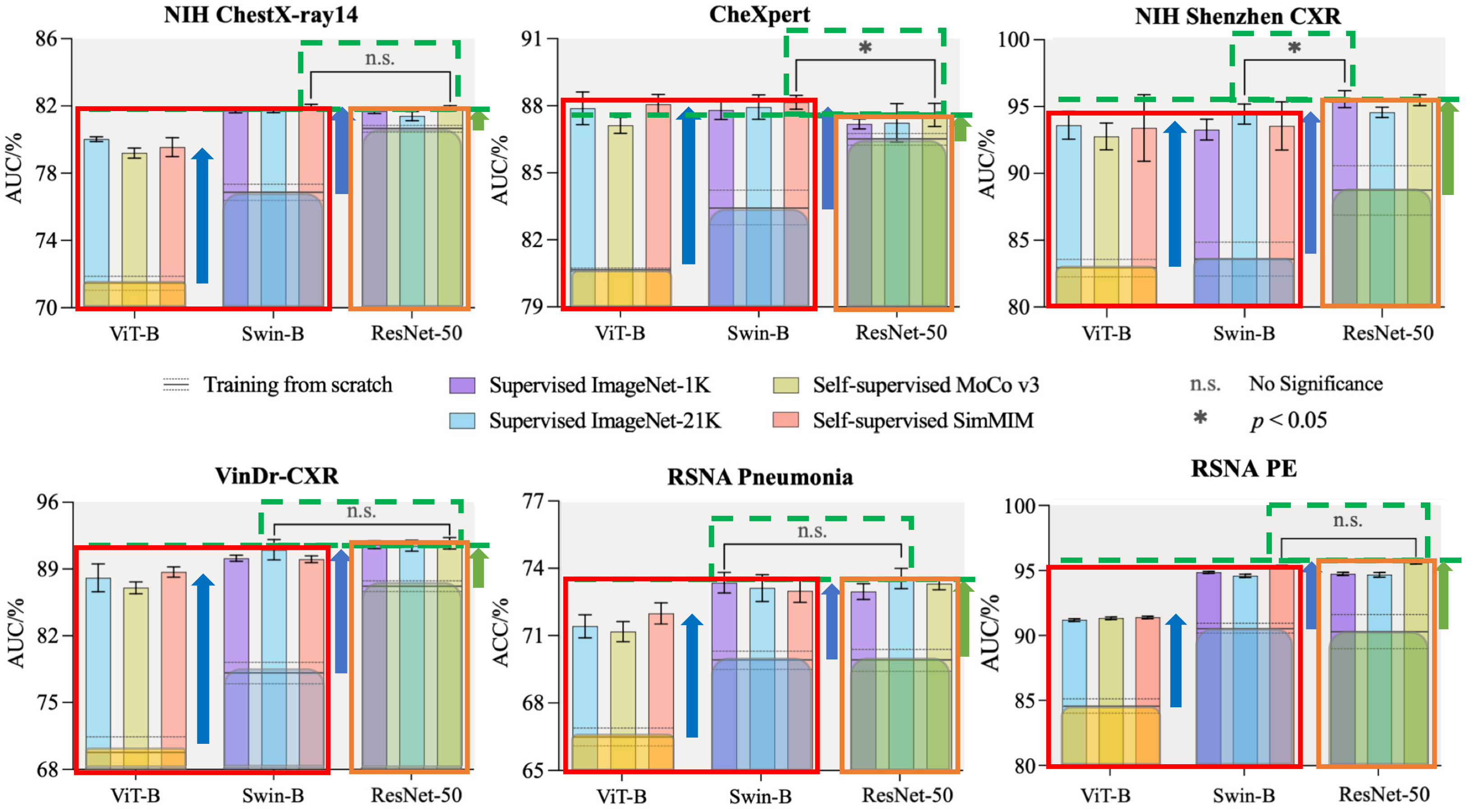
1. **NIH ChestX-ray14**: Fourteen thorax diseases classification (X-ray)
2. **CheXpert**: Five thorax diseases classification (X-ray)
3. **VinDr-CXR**: Six thorax diseases classification (X-ray)
4. **NIH Shenzhen CXR**: Tuberculosis classification (X-ray)
5. **RSNA Pneumonia**: Pneumonia and lung opacity classification (X-ray)
6. **RSNA PE**: Pulmonary Embolism slide-level classification (CT)



Result I: Transformers significantly underperform CNNs when training from scratch



Result II: Transformers can offer similar results as CNNs with ImageNet pre-training

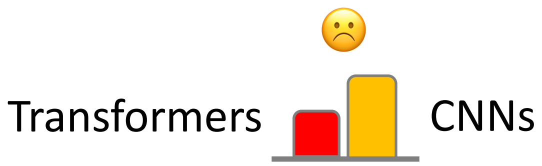


6 pre-trained Transformer models **3** pre-trained CNN models

Question 1

How do transformers perform on medical images?

When training from scratch



With ImageNet pre-training

<u>Transformers need pre-training</u> to perform well on medical images

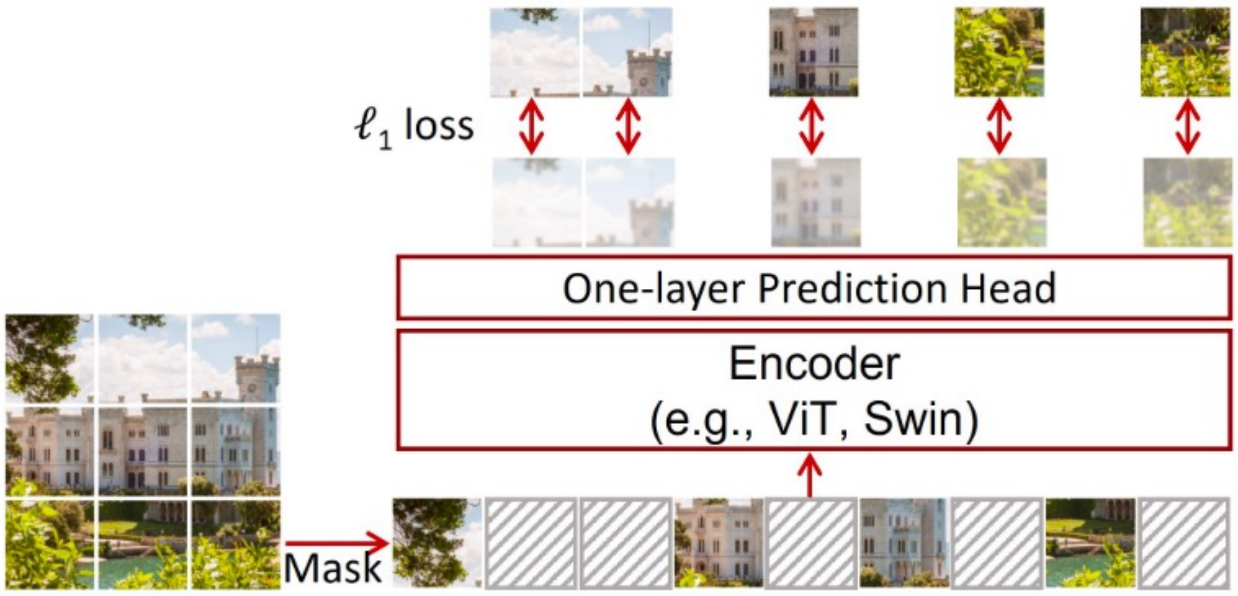
Question 2

Which ImageNet pre-trained transformer performs better on medical image classification?

Supervised or Self-supervised?

SOTA self-supervised learning technique

- **Masked Image Modeling (MIM)** using transformers
- Mask input patches and reconstruct them
  - Develop a holistic understanding of the image
  - Learn fine-grained features via reconstruction

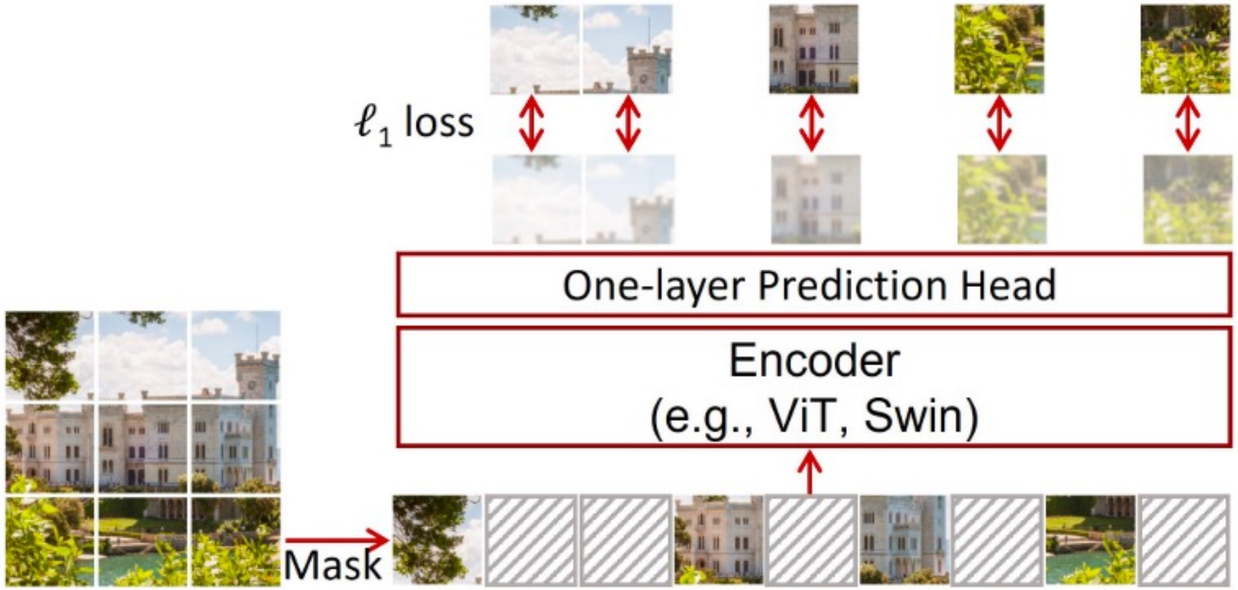


SimMIM (Xie et al., CVPR2022)

SOTA self-supervised learning technique

- **Masked Image Modeling (MIM)** using transformers
- Mask input patches and reconstruct them
  - Develop a **holistic understanding** of the image
  - Learn **fine-grained features** via reconstruction

Good for medical imaging tasks



SimMIM (Xie et al., CVPR2022)

Result III: Self-supervised learning based on masked image modeling is a preferable pre-training option for medical tasks

✓ Self-supervised **SimMIM** model with the **Swin-B** backbone outperforms fully-supervised baselines

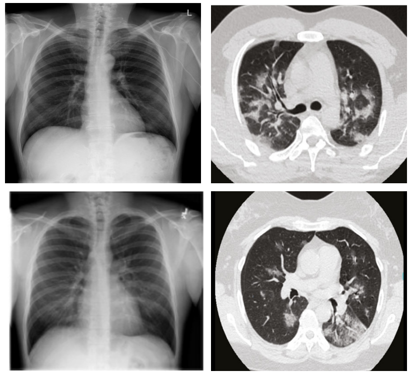
Method \ Task		ChestX-ray14	CheXpert	Shenzhen	VinDr-CXR	RSNA Pneumonia	RSNA PE
Supervised	ViT-B	80.05±0.17	87.88±0.50	93.67±1.03	88.30±1.45	71.50±0.52	91.19±0.11
	Swin-B	81.73±0.14	87.80±0.42	93.35±0.77	**90.35±0.31**	73.44±0.46	94.85±0.07
		*	n.s.	*	n.s.	n.s.	*
SimMIM	ViT-B	79.55±0.56	88.07±0.43	93.47±2.48	88.91±0.55	72.08±0.47	91.39±0.10
	Swin-B	**81.95±0.15**	**88.16±0.31**	**94.12±0.96**	90.24±0.35	**73.66±0.34**	**95.27±0.12**

*The best methods are bolded while the others are highlighted in green if they achieve equivalent performance compared with the best one (i.e., p > 0.05).

Can we further boost the performance of the pre-trained transformers for medical tasks?

Problems

1 Domain Gap



2 Transformers are more data-hungry

Pre-trained on natural images (large-scale dataset)

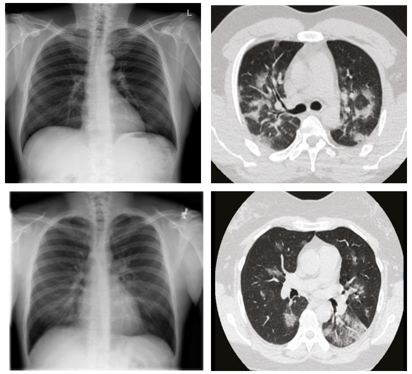
Fine-tuned on medical images (much smaller)

Method \ Task		ChestX-ray14	CheXpert	Shenzhen	VinDr-CXR	RSNA Pneumonia	RSNA PE
Supervised	ViT-B	80.05±0.17	87.88±0.50	93.67±1.03	88.30±1.45	71.50±0.52	91.19±0.11
	Swin-B	81.73±0.14	87.80±0.42	93.35±0.77	**90.35±0.31**	73.44±0.46	94.85±0.07
SimMIM	ViT-B	79.55±0.56	88.07±0.43	93.47±2.48	88.91±0.55	72.08±0.47	91.39±0.10
	Swin-B	**81.95±0.15**	**88.16±0.31**	**94.12±0.96**	90.24±0.35	**73.66±0.34**	**95.27±0.12**

Problems

1

Domain Gap



2

Transformers are more data hungry

Question 3

How to boost transformers' performance for medical image classification?

I. Continue domain-adaptive pre-training with in-domain data

II. Create large-scale in-domain dataset by assembling multiple datasets

X-rays(926K): A large-scale dataset that we assembled

- 926,028 images from 13 different chest X-ray datasets

No.	Source Datasets	Number of Images
1	MIMIC-CXR 2.0.0	377,028
2	CheXpert	223,414
3	PadChest	160,828
4	NIH ChestX-Ray 14	86,524
5	RSNA Pneumonia Detection Challenge	26,684
6	COVID-19 RADIOGRAPHY_DATABASE	21,165
7	VinDR-CXR	15,000
8	Indiana ChestX-ray	7,883
9	Mendeley-V2	5,232
10	COVIDx	1,223
11	Shenzhen Hospital X-ray Set	662
12	JSRT(Japanese Society of Radiological Technology)	247
13	Montgomery County X-ray Set	138
	X-ray(926K)	926,028

3

Datasets from different sources have different labels

III. Adopt <u>self-supervised learning</u> for pre-training

Result IV: Self-supervised domain-adaptive pre-training on a larger-scale in-domain dataset further boosts transformer model's performance

I. Continue domain-adaptive pre-training with in-domain data to bridge the domain gap

Model \ Task	ChestX-ray14	CheXpert	Shenzhen	VinDr-CXR	RSNA Pneumonia
Scratch	77.04±0.34	83.39±0.84	92.52±4.98	78.49±1.00	70.02±0.42
ImageNet	81.95±0.15	88.16±0.31	94.12±0.96	90.24±0.35	73.66±0.34
ChestX-ray14	78.87±0.69	86.75±0.96	93.03±0.48	79.86±1.82	71.99±0.55
X-rays(926K)	82.72±0.17	87.83±0.23	95.21±1.44	90.60±1.95	73.57±0.27
ImageNet→ChestX-ray14	82.45±0.15	87.74±0.31	94.83±0.90	90.33±0.88	73.85±0.72
ImageNet→X-rays(926K)	**83.04±0.15**	**88.37±0.40**	**95.76±1.79**	**91.71±1.04**	**74.09±0.39**

ImageNet vs. ImageNet→X-rays(926K): ChestX-ray14 >***; CheXpert n.s.; Shenzhen *; VinDr-CXR >***; RSNA Pneumonia *

*The best methods are bolded while the others are highlighted in green if they achieve equivalent performance compared with the best one (i.e., p > 0.05).

Result IV: Self-supervised domain-adaptive pre-training on a larger-scale in-domain dataset further boosts transformer model's performance

I. Continue domain-adaptive pre-training with in-domain data to bridge the domain gap

II. Use large-scale in-domain data to satisfy transformer's data hunger

III. Adopt self-supervised learning to overcome heterogeneity of expert labels

Model \ Task	ChestX-ray14	CheXpert	Shenzhen	VinDr-CXR	RSNA Pneumonia
Scratch	77.04±0.34	83.39±0.84	92.52±4.98	78.49±1.00	70.02±0.42
ImageNet	81.95±0.15	88.16±0.31	94.12±0.96	90.24±0.35	73.66±0.34
ChestX-ray14	78.87±0.69	86.75±0.96	93.03±0.48	79.86±1.82	71.99±0.55
X-rays(926K)	82.72±0.17	87.83±0.23	95.21±1.44	90.60±1.95	73.57±0.27
ImageNet→ChestX-ray14	82.45±0.15	87.74±0.31	94.83±0.90	90.33±0.88	73.85±0.72
ImageNet→X-rays(926K)	**83.04±0.15**	**88.37±0.40**	**95.76±1.79**	**91.71±1.04**	**74.09±0.39**

*The best methods are bolded while the others are highlighted in green if they achieve equivalent performance compared with the best one (i.e., p > 0.05).

Benchmarking Transformers

✓ Good initialization (Pre-training) is crucial for transformers

✓ Self-supervised pre-training based on masked image modeling is preferable

Boosting Transformers

✓Self-supervised domain-adaptive pre-training using a larger-scale in-domain dataset can further boost transformer's performance

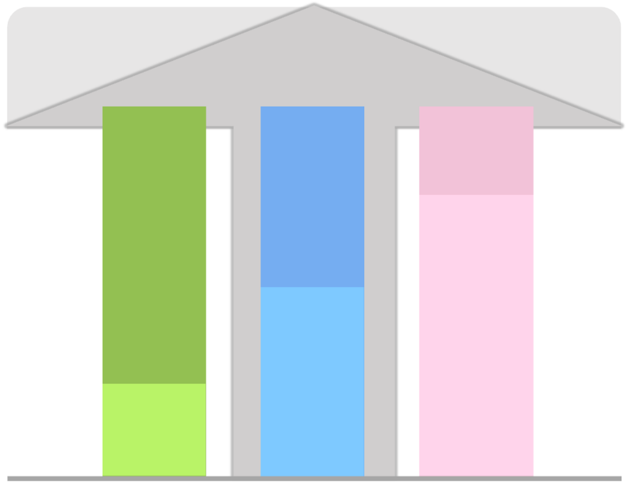
MICCAI2022 Singapore

DART

International Workshop on Domain Adaptation & Representation Transfer

MAYO CLINIC

Benchmarking and Boosting Transformers for Medical Image Classification



Github page: https://github.com/JLiangLab/BenchmarkTransformers