A Systematic Benchmarking Analysis of Transfer Learning for Medical Image Analysis

Mohammad Reza Hosseinzadeh Taher[1], Fatemeh Haghighi[1], Ruibin Feng[2], Michael B. Gotway[3], and Jianming Liang[1]

[1]Arizona State University [2]Stanford University [3]Mayo Clinic

Project page: github.com/JLiangLab/BenchmarkTransferLearning

View PDF

Task	Initialization	Metric	Scratch	ImageNet	ChestX-ray14	CheXpert	ImageNet→ ChestX-ray14	ImageNet→ CheXpert
Fourteen thorax diseases classification		AUC	80.31±0.10	81.70±0.15	---	81.99±0.08	---	**82.25±0.18**
Five thorax diseases classification		AUC	86.60±0.17	87.10±0.36	**87.40±0.26**	---	87.09±0.44	---
Tuberculosis detection		AUC	89.03±1.82	95.62±0.63	96.32±0.65	97.07±0.95	**98.47±0.26**	97.33±0.26
Pneumothorax segmentation		Dice	67.54±0.60	67.93±1.45	68.92±0.98	69.30±0.50	**69.52±0.38**	69.36±0.49
Lung segmentation		Dice	97.55±0.36	98.19±0.13	98.18±0.06	98.25±0.04	98.27±0.03	**98.31±0.05**

The best methods are **bolded** while the others are highlighted in blue if they achieve equivalent performance compared with the best one (i.e., $p > 0.05$).
When pre-training and target tasks are the same, transfer learning is not applicable, denoted by "-".

A3 Continual pre-training bridges the domain gap between natural and medical images.

Q3

Can medical image datasets help bridge the domain gap between natural and medical images?

Benchmarking newly-developed pre-training techniques for medical image analysis

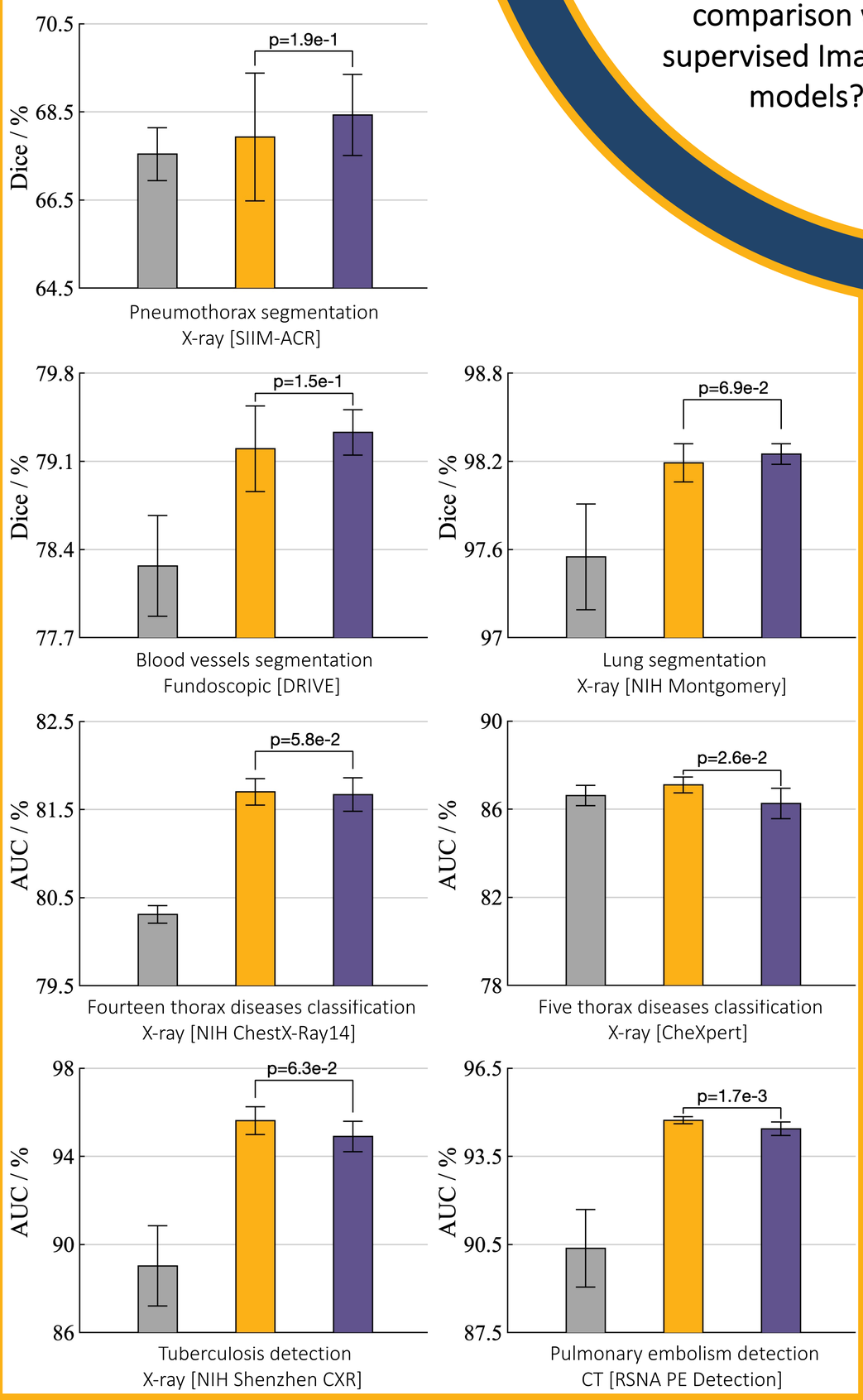
Q1 What advantages can supervised iNat2021 models offer for medical imaging in comparison with supervised ImageNet models?

Q2 How generalizable are the self-supervised ImageNet models to medical imaging in comparison with supervised ImageNet models?

A1 Pre-trained models on fine-grained data are better suited for segmentation tasks, while pre-trained models on coarse-grained data prevail on classification tasks.

- Random initialization
- ImageNet
- iNat2021*

**The most recent large-scale fine-grained datasets*

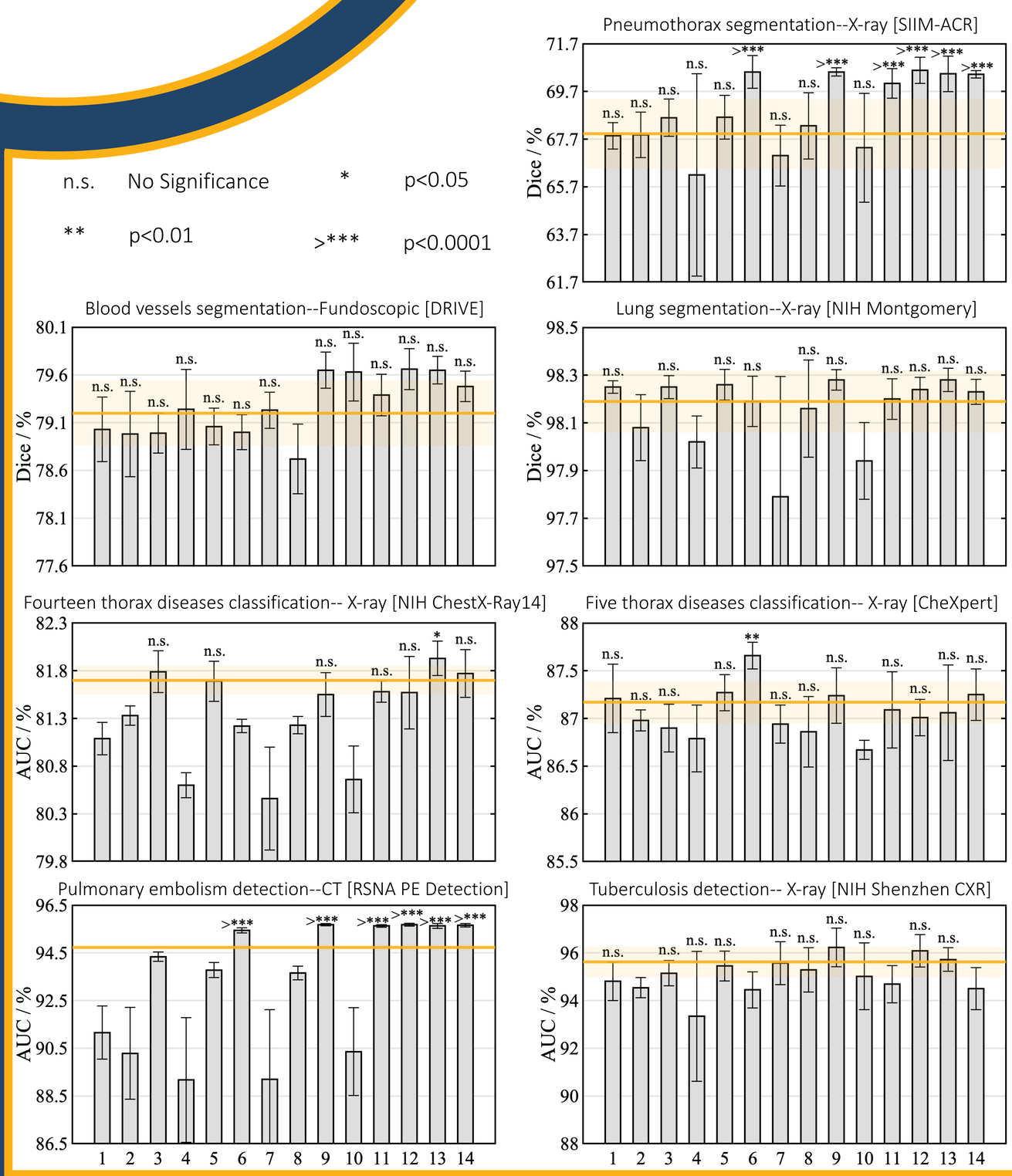


A2 Self-supervised ImageNet models learn holistic features more effectively than supervised ImageNet models.

1: InsDis	8: SimCLR-v2
2: MoCo-v1	9: Sela-v2
3: PCL-v1	10: InfoMin
4: PIRL	11: BYOL
5: PCL-v2	12: DeepCluster-v2
6: SimCLR-v1	13: SwAV
7: MoCo-v2	14: Barlow Twins

Baseline (supervised)

n.s.	No Significance	*	p<0.05
**	p<0.01	>***	p<0.0001



This research has been supported partially by ASU and Mayo Clinic through a Seed Grant and an Innovation Grant, and partially by the NIH under Award Number R01HL128785.