The Potential Benefits of Transfer Learning in AI algorithm development for COVID-19 Classification models using Chest X-ray Images

Xin Tie[1], Ran Zhang[1], Zhihua Qi[2], and Guang-Hong Chen[1,3]

[1] Department of Medical Physics, University of Wisconsin-Madison, Madison, WI
[2] Department of Radiology, Henry Ford Health System, Detroit, MI
[3] Department of Radiology, University of Wisconsin-Madison, Madison, WI

DEPARTMENTS OF Medical Physics & Radiology
UNIVERSITY OF WISCONSIN SCHOOL OF MEDICINE AND PUBLIC HEALTH

Acknowledgements

- The work is partially supported by an NIH grant 3U01EB021183-04S1 and by a grant from Wisconsin Partnership Program

Introduction

- COVID-19 has put healthcare systems under immense pressure, especially in the early pandemic
- Chest X-ray is the first-line imaging test
  - Almost all clinics are equipped with radiography units
- It can be potentially used in patient triage and management

Introduction

- COVID-19 has put healthcare systems under immense pressure, especially in the early pandemic
- Chest X-ray is the first-line imaging test
  - Almost all clinics are equipped with radiography units
- It can be potentially used in patient triage and management
- Machine learning (ML) methods can quickly and tirelessly learn how to differentiate COVID-19 from Non-COVID-19

Challenge: Limited Data

- Limited data
  - Data processing and sharing are time-consuming

- How can we develop ML models with limited data?
  - Transfer learning

Challenge: Limited Data

- Limited data
  - Data processing and sharing are time-consuming
- How can we develop ML models with limited data?
  - Transfer learning
- Two scenarios
  - From previous tasks to the task of interest
  - From large clinical sites to small clinical sites

Research questions

- Can transfer learning from the previous task benefit our current task?
  - Previous task: identify different pulmonary diseases
  - Current task: differentiate COVID-19 from Non-COVID-19

Research questions

- Can transfer learning from the previous task benefit our current task?
- Can transfer learning from large clinical sites benefit small clinical sites?
  - Task: differentiate COVID-19 from Non-COVID-19

Research questions

- Can transfer learning from the previous task benefit our current task?
- Can transfer learning from large clinical sites benefit small clinical sites?

Can transfer learning from the previous task benefit our current task?

- Data:
  - An NIH (CXR) dataset curated in the pre-COVID era: 112,200 CXRs[1]
  - A COVID dataset from the Henry Ford Healthcare System: 27,845 CXRs

NIH CXR dataset
112,200 CXRs

Atelectasis	Cardiomegaly	Effusion	Infiltration	Consolidation	Nodule	Pneumonia
Emphysema	Pneumothorax	Fibrosis	Pleural Thickening	Edema	Mass	Hernia

1. Wang X, Peng Y,… & Summers RM. (2017). IEEE CVPR

Can transfer learning from the previous task benefit our current task?

- Data:
  - An NIH (CXR) dataset curated in the pre-COVID era: 112,200 CXRs[1]
  - A COVID dataset from the Henry Ford Healthcare System: 27,845 CXRs

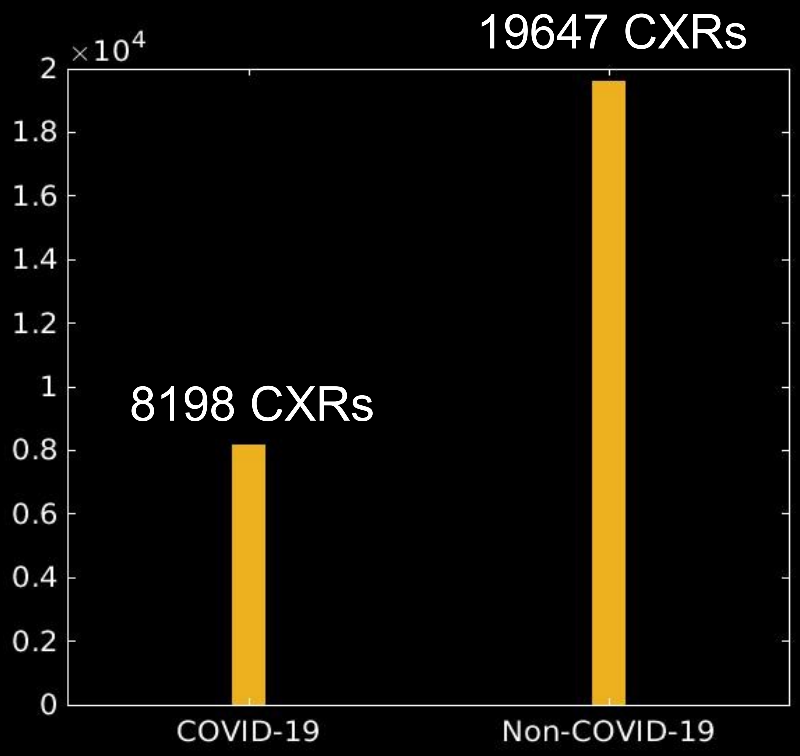
NIH CXR dataset
112,200 CXRs

1. Wang X, Peng Y,… & Summers RM. (2017). IEEE CVPR

Can transfer learning from the previous task benefit our current task?

- Data:
  - An NIH (CXR) dataset curated in the pre-COVID era: 112,200 CXRs
  - A COVID dataset from the Henry Ford Healthcare System: 27,845 CXRs

NIH CXR dataset
112,200 CXRs



HF COVID-19 dataset
27,845 CXRs
COVID-19: 8,198 CXRs
Non-COVID-19: 19,647 CXRs

Can transfer learning from the previous task benefit our current task?

- Data:
  - An NIH (CXR) dataset curated in the pre-COVID era: 112,200 CXRs
  - A COVID dataset from the Henry Ford Healthcare System: 27,845 CXRs

Training and validation set	Internal test set
from four hospitals in the Henry Ford Healthcare system	from four hospitals in the Henry Ford Healthcare system
5725+ / 14059- CXRs	1000+ / 1000- CXRs

External test set
from other hospitals / clinics in the Henry Ford Healthcare system

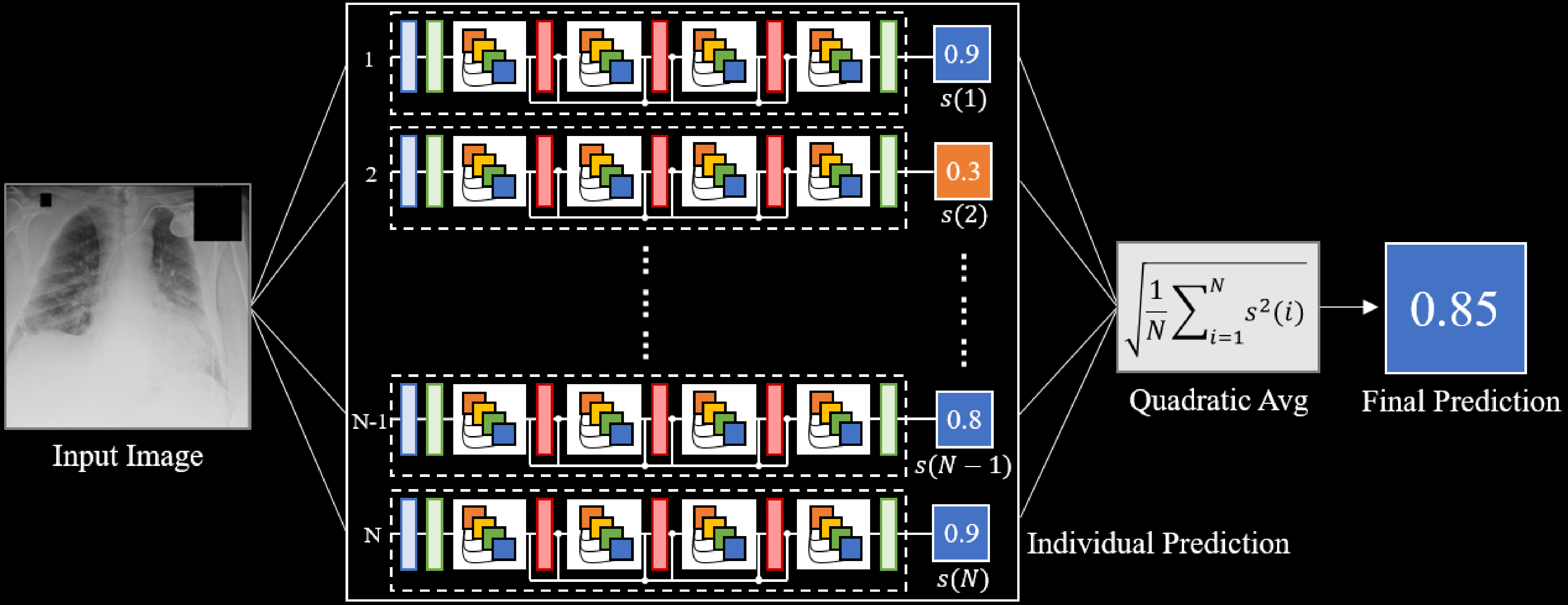
1473 + / 4588- CXRs

Can transfer learning from the previous task benefit our current task?

- Develop a COVID-19 classification model for the Henry Ford (HF) dataset.
  - With transfer learning: The model was pretrained in the NIH CXR dataset and then fine-tuned on the HF COVID-19 dataset
  - Without transfer learning: The model was directly trained on the HF COVID-19 dataset

Can transfer learning from the previous task benefit our current task?

- Model:
  - CV19-Net[1]



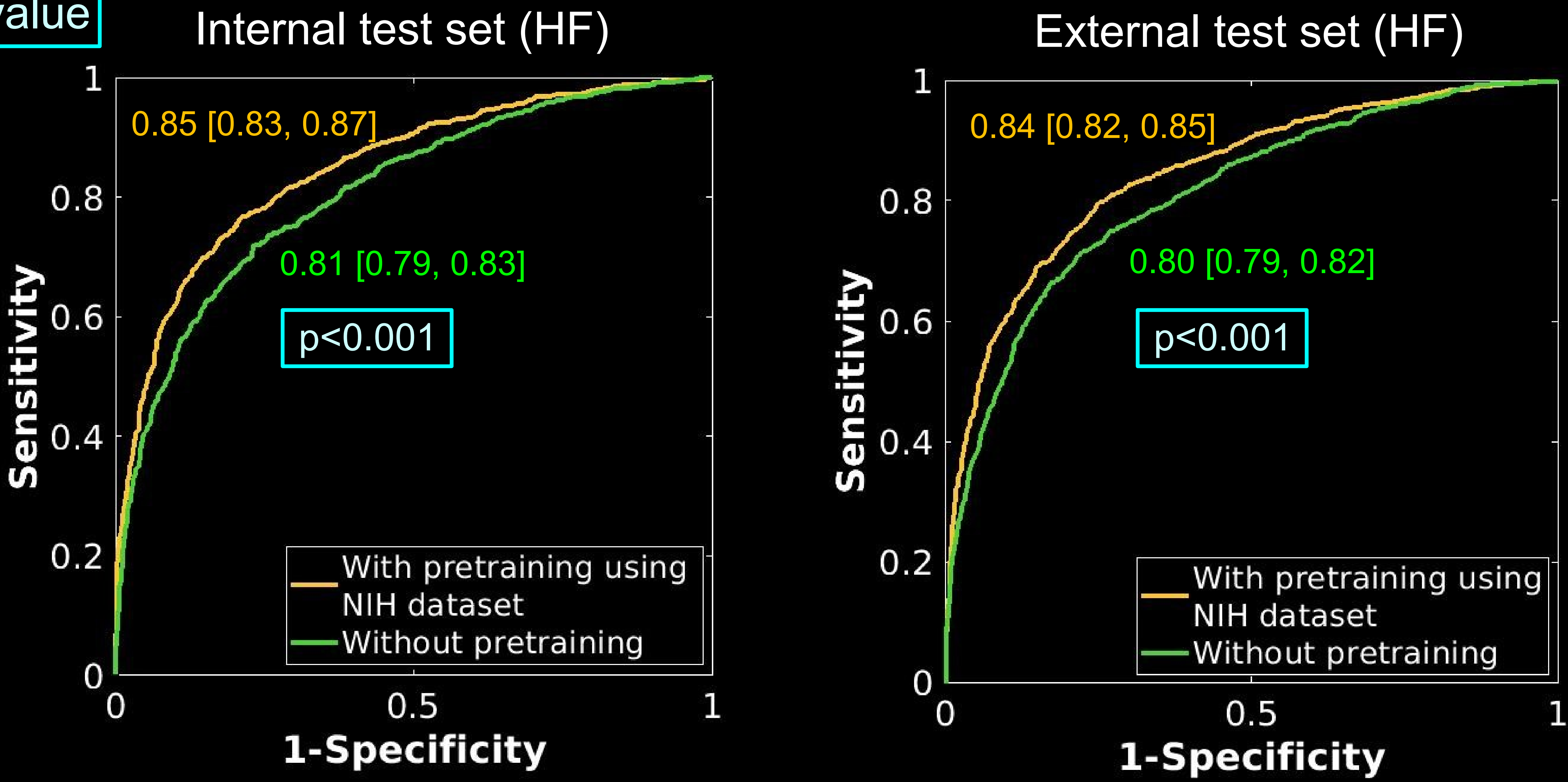
1. Zhang, R., Tie, X., Qi, Z., Bevins, N. B., Zhang, C., Griner, D., ... & Chen, G. H. (2021). Radiology, 298(2), E88-E97.

Can transfer learning from the previous task benefit our current task?

- Training details
  - The initial learning rate of Adam optimizer was 1e-4
  - The batch size was selected to be 40
  - Early stopping and model ensemble were used to avoid overfitting and reduce variance of predictions
  - No layers were frozen during the training
- Evaluation metric
  - Area under the ROC curve (AUC) and its 95% CI

Result

p-value



Summary

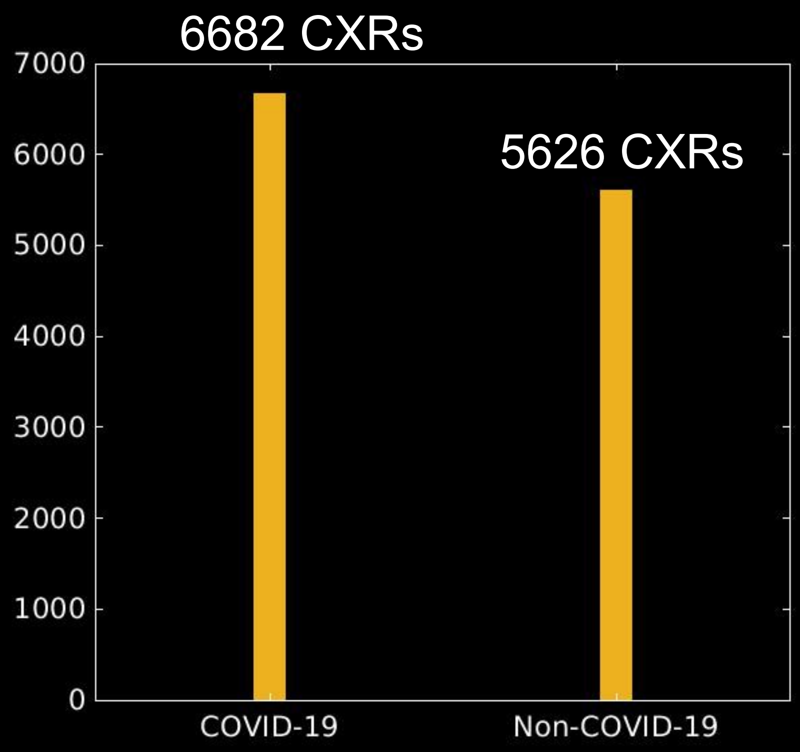
- **Question**: Can transfer learning from the previous task benefit our current task?
- **Answer**: Yes. Pretraining using a public CXR dataset significantly improved the performance (p-value<0.001) of the HF COVID-19 classification model.

Research questions

- Can transfer learning from the previous task benefit our current task?
- Can transfer learning from large clinical sites benefit small clinical sites?

Can transfer learning from large clinical sites benefit small clinical sites?

- Data:
  - A COVID dataset curated and made public in Spain (BIMCV): 12,308 CXRs[1]



BIMCV COVID-19 dataset
12,308 CXRs
COVID-19: 6,682 CXRs
Non-COVID-19: 5,626 CXRs

1. Vayá, Maria de la Iglesia, et al. arXiv:2006.01174 (2020).

Can transfer learning from large clinical sites benefit small clinical sites?

- Develop COVID-19 classification models for various sizes of simulated clinical sites in the Spain dataset.

Spain COVID-19 dataset
12,308 CXRs

Internal test set
2,201 CXRs
COVID-19: 1,087 CXRs
Non-COVID-19: 1,114 CXRs

Can transfer learning from large clinical sites benefit small clinical sites?

- Develop COVID-19 classification models for various sizes of simulated clinical sites in the Spain dataset.

Spain COVID-19 dataset
12,308 CXRs

COVID-19: 50 patients, 163 CXRs
Non-COVID-19: 50 patients, 132 CXRs

...

COVID-19: 200 patients, 576 CXRs
Non-COVID-19: 200 patients, 471 CXRs

...

COVID-19: 2717 patients, 5595 CXRs
Non-COVID-19: 3063 patients, 4512 CXRs

Small clinical site

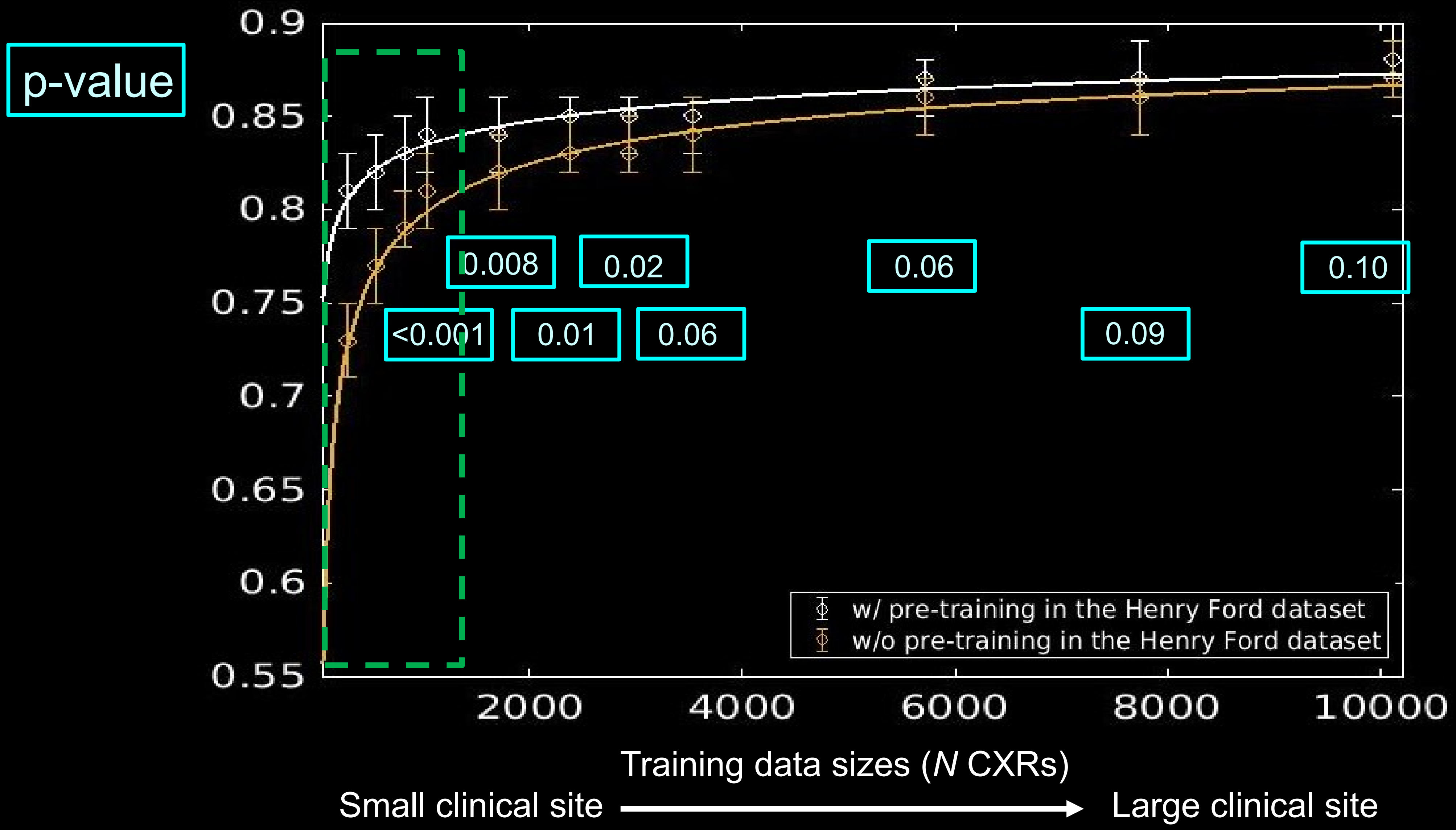
Large clinical site

Internal test set
2,201 CXRs
COVID-19: 1,087 CXRs
Non-COVID-19: 1,114 CXRs

Can transfer learning from large clinical sites benefit small clinical sites?

- Develop COVID-19 classification models for various sizes of simulated clinical sites in the Spain dataset.
  - With transfer learning from large clinical sites: The model pretrained in the HF dataset was fine-tuned on the simulated clinical sites in the Spain dataset.
  - Without transfer learning from large clinical sites: The model was only pretrained in the NIH dataset and then directly fine-tuned on the simulated clinical sites in the Spain dataset.

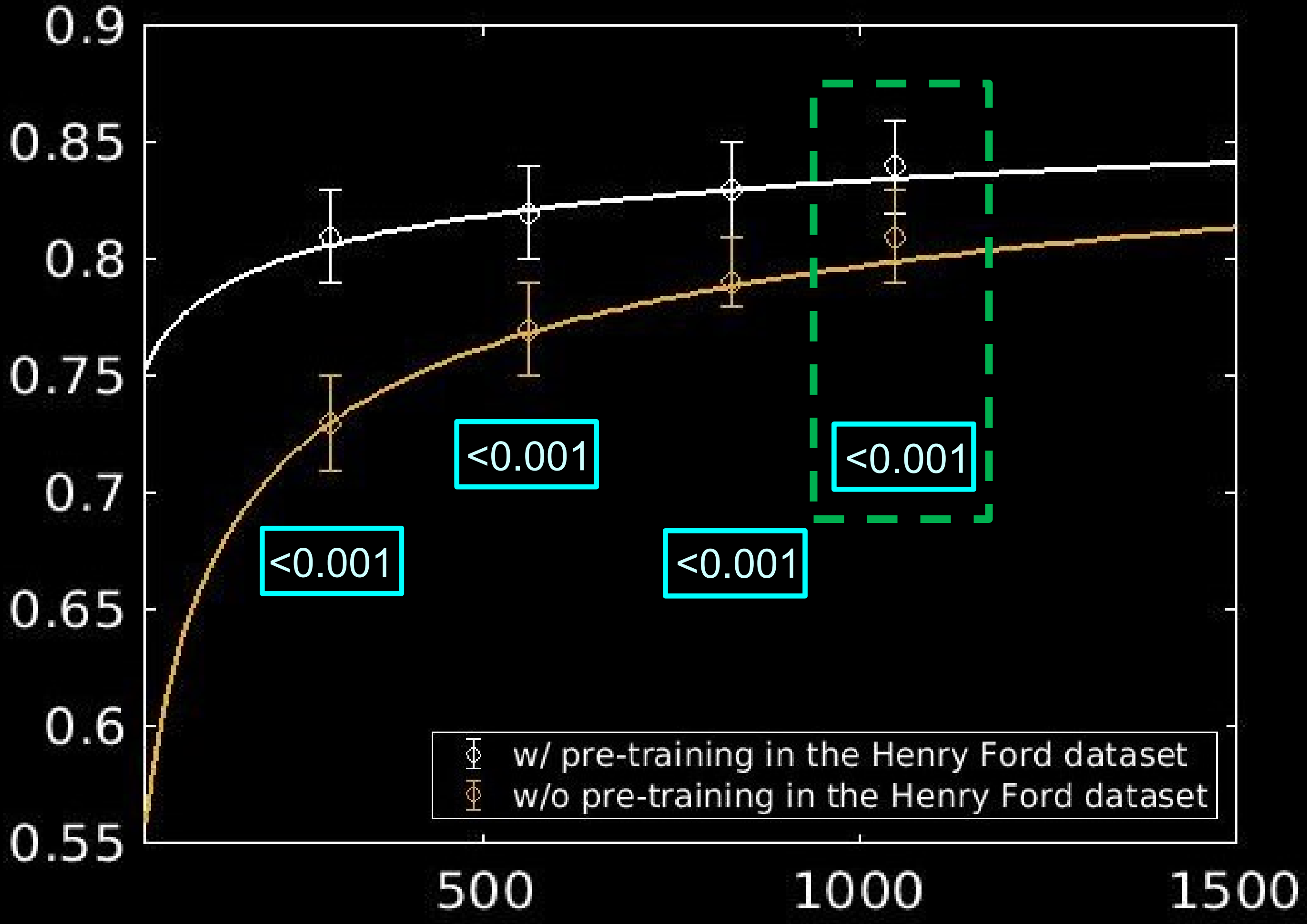
Result



p-value
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.008
0.02
0.06
0.10
<0.001
0.01
0.06
0.09
w/ pre-training in the Henry Ford dataset
w/o pre-training in the Henry Ford dataset
2000
4000
6000
8000
10000
Training data sizes (N CXRs)
Small clinical site
Large clinical site

Result

p-value



0.84 [0.82, 0.85]

0.81 [0.80, 0.83]

Training data sizes (*N* CXRs)

Small clinical sites

Summary

- **Question**: Can transfer learning from large clinical sites benefit small clinical sites?

- **Answer**: Yes. In the simulated small clinical sites (#CXR<3000), the performance gain was statistically significant (p-value<0.05) by fine-tuning a well-trained COVID model from a large clinical system (HF dataset).

Conclusion

- Transfer learning from a public NIH chest x-ray dataset or a large COVID dataset from another site improves the performance of the classification model under the two different scenarios.

xtie@wisc.edu

Thank You

University of Wisconsin-Madison

Can transfer learning from large clinical sites benefit small clinical sites?

- Model:
  - CV19-Net
- Training details
  - The initial learning rate of Adam optimizer was 2e-5
  - The batch size was selected to be 40
  - Early stopping and model ensemble
- Evaluation metric
  - Area under the ROC curve (AUC) and its 95% CI