From Competition to Collaboration

Making Toy Datasets on Kaggle Clinically Useful for Chest X-Ray Diagnosis Using Federated Learning

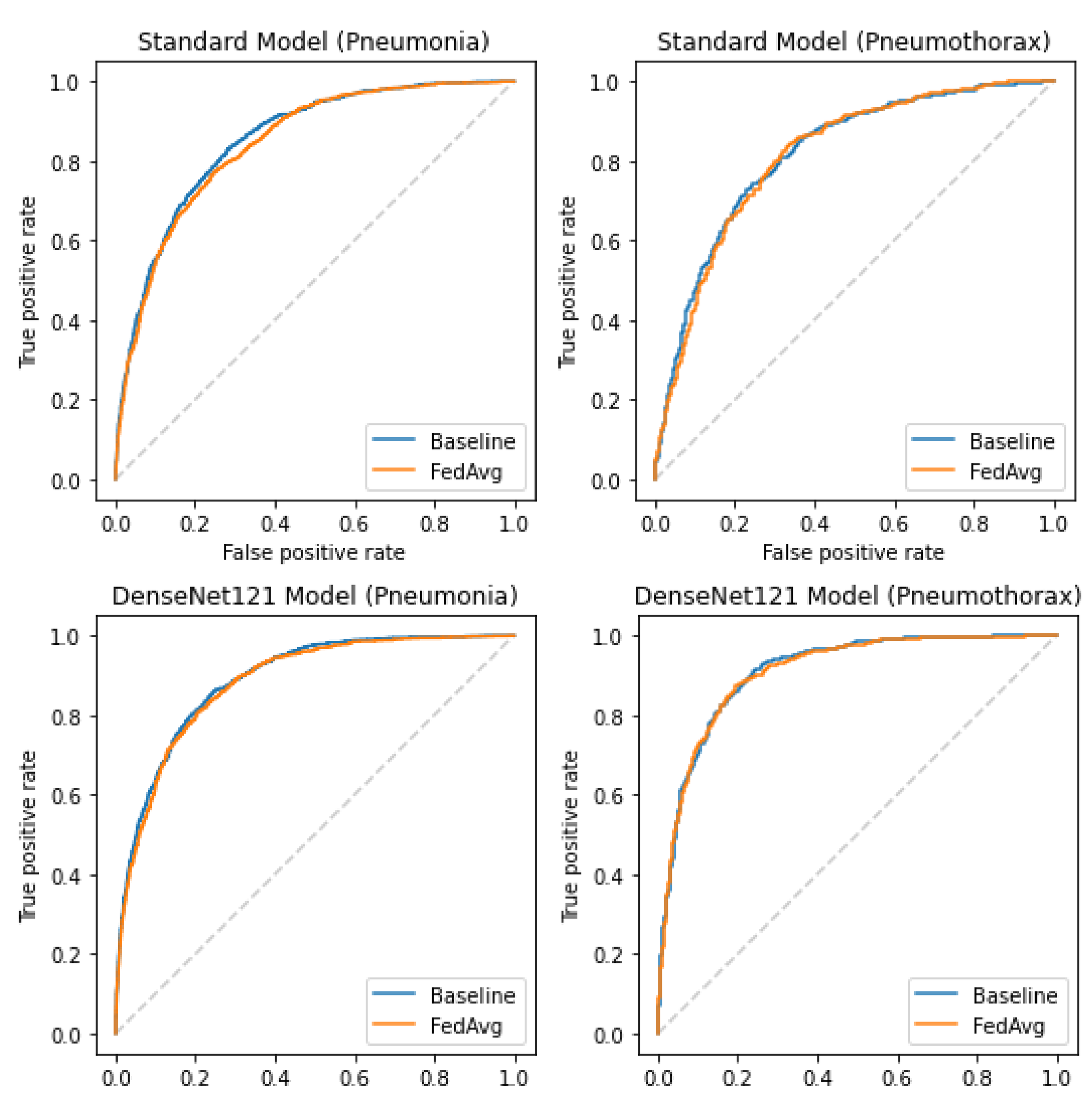
Pranav Kulkarni, Adway Kanhere,
Paul H. Yi, Vishwa S. Parekh

Introduction

- CXR datasets hosted on Kaggle, though useful from a data science competition standpoint, have limited utility in clinical use due to their narrow focus on diagnosing one specific disease.
- Therefore, a way to harmonize these toy datasets could revolutionize how small, narrowly-focused datasets can be used for development of clinically-relevant deep learning models.
- We propose CheXViz, a FL framework for training a 'global' meta-deep learning model on spatially distributed datasets with non-iid annotations.
- In other words, **we aim to demonstrate how CheXViz can be used to make toy datasets from Kaggle clinically useful.**

Methods

- Train a global CheXViz model to classify cases of pneumonia and pneumonia using distributed, non-iid toy CXR datasets from RSNA Pneumonia Detection and SIIM-ACR Pneumothorax Segmentation competitions on Kaggle.
- Two different model architectures: A naïve 3-layer CNN ('standard') and an ImageNet pretrained DenseNet121 using transfer learning [1].
- Utilize FedAvg model weight aggregation strategy for FL [2].
- Compare global model performance with 'baseline' models trained on both tasks separately for each model architecture.
- Use bootstrapping and paired t-test to compare AUROC.
- Statistical significance defined as $p < 0.05$.



ROC curves obtained from baseline and CheXViz models evaluated across both the datasets.

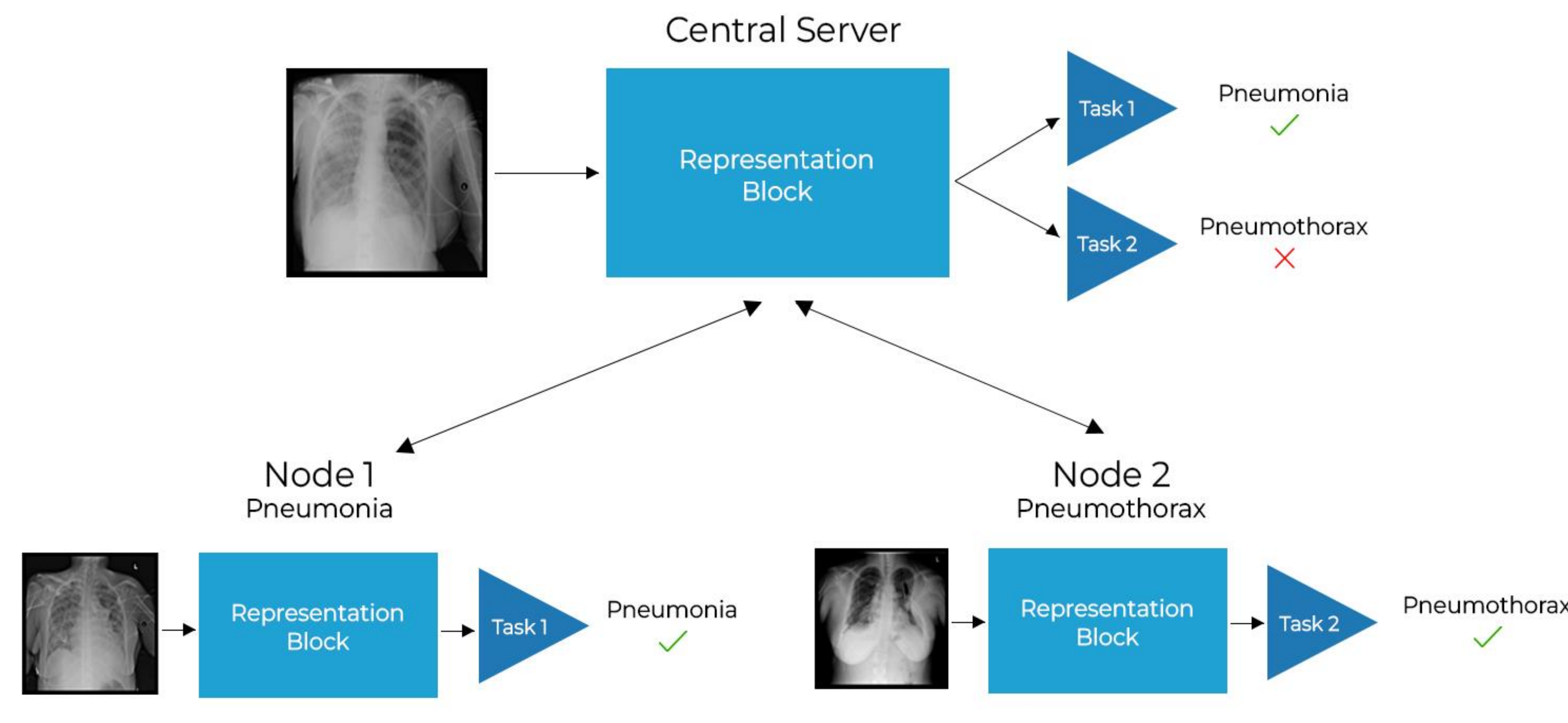


Illustration of the CheXViz framework

Federated learning (FL) is a ML paradigm that approaches problems from a multi-domain and multi-task perspective. By using a decentralized and distributed approach consisting of a central server and nodes, a global meta-model can be trained to generalized distributed tasks with non-iid labels.

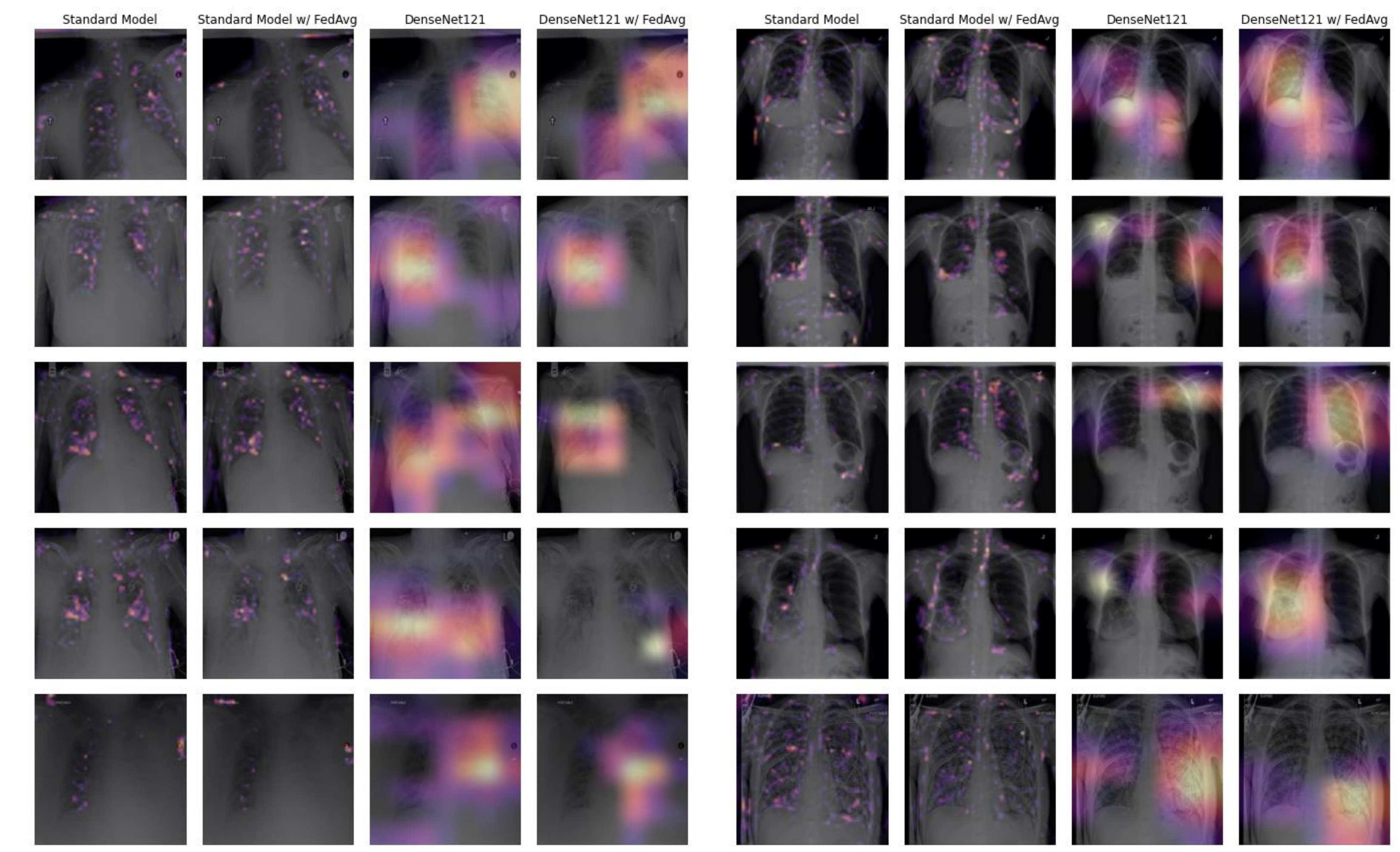
CheXViz is a FL framework is initialized as a deep neural network consisting of a representation block and a task-specific block. During training, only weights corresponding to the representation block are aggregated and redistributed by the central server, thus preserving task-related information for each node

Results

- The CheXViz framework trained 'meta' models demonstrated excellent performance compared to the baseline models for the diagnostic classification of pneumonia and pneumothorax abnormalities.
- We further visualized the Grad-CAM outputs for evaluating the explainability and generalizability of the models [3]. Our preliminary analysis suggests that the heatmaps from CheXViz models demonstrate higher and focused activations within the lungs, compared to the baseline models.

Model Metrics

Task	Model	Loss	Sensitivity	Specificity	AUPR	AUROC	p-value
Pneumonia	Standard	0.38	82.57	71.97	0.63	0.85	-
	Standard w/ FL	0.39	78.01	74.41	0.61	0.84	0.10
	DenseNet121	0.34	84.90	76.76	0.71	0.89	-
	DenseNet121 w/ FL	0.35	80.08	79.79	0.70	0.88	0.19
Pneumothorax	Standard	0.41	74.13	75.89	0.54	0.82	-
	Standard w/ FL	0.42	80.22	69.87	0.52	0.81	0.71
	DenseNet121	0.31	91.30	76.31	0.73	0.91	-
	DenseNet121 w/ FL	0.31	84.57	83.10	0.73	0.91	0.76



(a) Pneumonia (b) Pneumothorax

Grad-CAM Visualization of True Positives

Discussion

- Given the challenges in curating expert-level annotations for diseases, it is understandable why Kaggle-hosted competitions have focused largely on single diseases [4, 5].
- Although Kaggle CXR datasets and data science competitions have made an indelible impact on data science and AI for healthcare, they are still a far cry from being clinically useful datasets.
- Our findings demonstrate that CheXViz can be used to create global 'meta' models to make toy datasets from Kaggle clinically useful, a large step forward towards bridging the gap from bench to bedside.
- It is our hope that our work can be a first step towards moving Kaggle CXR datasets from competition to collaboration and transform these toy datasets into clinically useful models.

References

[1] G. Haung, et al. Densely connected convolutional networks. In *Proceedings of the IEEE conference of computer vision and pattern recognition*, 2017. doi: https://doi.org/10.48550/arXiv.1608.06993

[2] B. McMahan, et al. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. 20–22 Apr 2017.

[3] R. R. Selvaraju, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 2017. doi: https://doi.org/10.1007/s11263-019-01228-7

[4] G. Shih, et al. Augmenting the national institutes of health chest radiograph dataset with expert annotations of possible pneumonia. *Radiology: Artificial intelligence*, 1(1), 2019.

[5] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.