Federated Learning Using Multi-institutional Data for Generalizable Detection of Abnormal Chest X-rays

Abstract. There has been a remarkable progress in designing deep learning for chest x-ray diagnosis. However, generalizability of such models still remains a major concern. This is one of the primary reasons hindering the clinical applicability of such models. We propose a strategy to improve generalizability for the task of chest x-ray diagnosis using federated learning. Our model makes use of chest x-ray image data from multiple different sources. We train individual models using each of such datasets for the detection of abnormal chest x-ray images. Subsequently, we note that larger datasets may contain more variety of abnormalities in larger numbers. This may be helpful for a better detection of abnormal chest x-rays. Based on this, we propose a method for aggregating the individual models based on the size of of the training data used. Our model shows superior performance and better generalizability compared to other aggregation strategies alongside existing federated learning approaches. We will make the code publicly available once the review process is complete.

Keywords: Federated learning, Chest X-ray, Generalizable.

1 Introduction

Deep learning models have achieved great success for the automated analysis of chest x-rays [16]. However, many such models lack generalizability, i.e., a model trained in one dataset often performs poorly in a different dataset. The probable source of performance drop lies in the image acquisition procedure and the class distribution as presented in the individual datasets. In this regard, collaborative machine learning through uniting data from all institutions/organizations to learn a deep model helps to include more variety of image samples during training. This can improve the generalizability of the trained model. However, this simple approach has the following two limitations: (a) need huge data storing capacity in the centralized server where the model will be trained as well as data transmission hazard, and (b) data privacy is not preserved which is important for dealing with medical data like chest x-ray. To deal with the above-mentioned limitations, federated learning, a privacy preserving machine learning is getting importance.

Federated learning is an advanced deep learning paradigm that allows multiple institutions to develop a deep model utilizing data from all institutions without sharing the raw data. Literature shows that several methodical advancements are proposed for federated learning [17] and a very small fraction of them have been applied on chest x-ray datasets [13, 19]. The available research works focusing on federated learning for chest x-ray diagnosis hardly uses multi-institutional data except [20] which uses 16, 148 cases containing Chest X-Ray and non-image data from 20 institutes across the globe to learn a federated model. It is evident that there is scarcity to utilize multi-institutional large volume datasets to build federated model to advance the automated chest x-ray diagnosis systems.

In this paper, we aim to improve the generalizability of the chest x-ray diagnosis model using federated learning considering multi-institutional data. Our work considers three publicly available chest x-ray datasets (NIH dataset [1], VinBig Dataset [3] and ChestXpert dataset [23]). DenseNet-121 [4] architecture is configured to develop a binary classifier based on the presence of abnormality in the chest x-rays. The experimental result shows that DenseNet-121 trained on one dataset may not performing well on other datasets. Hence, we opt for aggregating the local trained models obtained from three different datasets. The aggregated model is termed as global model. We hypothesize that if the model trained with more data influences the weight of the global model more significantly, the global model may have better generalizability.

We compare the results with the following three different federated aggregation techniques: (i) simple averaging, (ii) weighted averaging based on train losses received from the local models, and (iii) weighted averaging based on validation losses received from the local models. We find that the our global model with the proposed aggregation scheme improves the generalizability. We believe that this is the first attempt where large number of chest x-ray images from three sites are considered for producing federated learning and improving the generalizability of the Chest X-ray classification.

The rest of the paper organized is as follows. In Section 2, we discuss the related works. Section 3 discusses the proposed method for preparing the chest x-ray diagnosis model. The experimental protocol and analysis of experimental results are presented in Section 4. Finally, we conclude the paper in Section 5.

2 Related Work

We divide our literature review in three parts. First part focuses on deep learning for chest x-ray analysis, the second part focuses on federated learning and final part focuses on federated learning for chest x-ray analysis.

2.1 Deep Learning for Chest x-ray Analysis

Automated analysis of chest x-rays using deep learning methods have gained an unprecedented success in recent years. The availability of large-scale chest x-ray datasets [1, 23, 24] has facilitated this research. Many such research works have focused on the diagnosis of a specific abnormality from chest x-rays. For the diagnosis of tuberculosis from chest x-rays, a CNN model has been proposed in [25]. For the same application, an ensemble of deep CNN models has been introduced in [26]. In [27], a deep learning-based method has been proposed for an automated detection of malignant pulmonary nodules from chest x-rays. Taylor *et al.* have designed a CNN model for the diagnosis of pneumothorax [28]. See [29, 30] for deep learning methods to detect pneumonia and [31, 32] for the detection of COVID-19 from chest x-rays.

Many researchers have designed methods for the diagnosis of multiple abnormalities from chest x-rays. Rajpurkar *et al.* have used Densely Connected Convolutional Networks [4] for the diagnosis of chest x-rays [34]. In [37], an attention regularized deep learning method has been proposed for thoracic disease classification from chest x-rays. A segmentation-based deep fusion network has been proposed in [38] that performs classification of thoracic diseases from chest x-rays. Using weak-supervision, Ouyang *et al.* have proposed a hierarchical attention based method for chest X-ray abnormality localization and diagnosis [39]. To alleviate the need of large training datasets, several methods have been proposed to perform few-shot learning [40] and zero-shot learning [41] for chest x-ray diagnosis.

2.2 Federated Learning

Federated learning is a privacy-preserving machine learning technique that allows multiple institutions to develop a deep model utilizing data from them without sharing the raw data [17]. Recent research focuses on this field considers three diverse aspects of this technique namely, (i) the learning algorithm and training method [42], (ii) the data privacy protection mechanism [48], and (iii) the participant incentive mechanism [49]. However, we find that development of federated learning algorithm and training method is an active research area. Literature shows that federated averaging (in short FedAvg) [42] is a widely used effective and simpler aggregation strategy. FedAvg assumes that the data in the different clients will be independent and identically distributed and have similar class distribution among the clients. This assumption fails in real scenarios which lead towards developing algorithms to overcome the statistical challenges, communication efficiency/stability issues, device/data heterogeneity issues [43–45]. Federated learning framework for semi-supervised learning [46], continual learning [50], meta-learning [51] are designed.

Based on our study, we find that research study on federated learning algorithm made a noticeable progress. However, the research is confined mostly in algorithm development and performances are evaluated on a synthetically generated multi-source dataset and there is huge scope to apply federated learning approaches in real-life multi-centered dataset.

2.3 Federated Learning for Chest X-ray Diagnosis

Literature shows that several attempt have been taken to build deep convolutional neural network models with federated learning. The available works mostly focuses on simulated environment for multiple client system. The developed algorithms are evaluated either with in-house dataset [19] and using publicly available dataset [13, 22]. To cope up with inter-client different class distribution, partially shared network is used in [13]. In [21], a remarkable research work is done to build a Federaedly learned deep model for predicting the future oxygen requirements of symptomatic patients with COVID-19. Patients' electronic medical records (EMR) and chest x-ray are analyzed for the prediction task. The pre-trained ResNet34 with spatial attention is used to extract 512-dimensional feature vector from the CXR images. The image feature along with the 21 EMR feature serve as input for the Deep & Cross network which is trained with federated learning.

In this paper, we propose a federated learning model that aggregates client models based on the training data. We notice that most of the existing works focus on simulated environment for multiple client system. In contrast, we consider large scale chest x-ray datasets from multiple sources and evaluate the efficacy of our federated learning model to improve generalizability for chest x-ray diagnosis.

3 Method

3.1 Overview

We propose a federated learning framework to facilitate the detection of abnormal chest x-ray images. In situations where datasets can be shared across institutions, all the datasets can be used together for training a machine learning model. However, due to privacy and security concerns it may not be possible to share medical data across institutions. Federated learning is useful in such scenarios. In federated learning, each institute (termed as client in federated learning literature) may train a model using its own data. Then, each individual model is shared to a central server which aggregates the individual models to produce a final model.

We use this framework to design a generalizable model for detecting abnormal chest x-ray images. We use three separate large-scale datasets to train three local (client) models. These datasets are NIH Chest X-ray dataset [1], CheXpert dataset [23], and the VinBigData Chest X-ray dataset [3]. These models are subsequently aggregated by a central server. We use the size of the individual training data to perform the aggregation. Thus, we employ the decentralized training approach of federated learning followed by an aggregation scheme to create a global model. Our method of aggregation helps in improving generalizability. The steps of our method are discussed next.

3.2 Local Model Training

In Federated learning process the individual local models are trained first. We use Densenet-121 [4] as the backbone for the local models because of its successful application in chest x-ray diagnosis [34]. In our design, we a local model corresponding to each dataset. Thus, we have one local model each trained with the NIH, CheXpert, and VinBig datasets.

3.3 Federated Aggregation

The next step in the process is to combine the weights from the local models to generate a global model. Consider a system with k clients. If the local model of client i is $f(w_i)$, the resultant weight after federated aggregation is given by

$$F(w) = \frac{1}{k} \sum_{i=1}^{k} c_i f(w_i), \tag{1}$$

where c_i is the weight assigned to client i and $\sum_{i=1}^{k} c_i = 1$. In case of simple averaging, the weights assigned to each client are equal.

There may be different strategies for assigning weights to individual clients. We may compute weights based on validation losses and training losses of individual models. However the losses do not directly capture the distribution of the data. On the other hand, we consider the following fact. If a local model is trained with a large number of training data, the model may be expected to be exposed to a wider variety of data compared to a local model trained with less number of data. Consequently, the former local model may be trained with a wider variety of abnormalities compared to the later. More significant contribution from such a local model during aggregation is likely to facilitate the overall performance.

Therefore, during aggregation, if we assign weights to local models in proportion to the number of training the data, the resultant model may be expected to capture the distribution of the data more efficiently. Consequently, such an aggregation is likely to produce superior performance across datasets from different institutions. Hence, we aggregate the local models with weights proportional to the number of training data. Thus, considering T_i to be the number of training data for local model i, we write

$$c_i = \frac{T_i}{\sum_{i=1}^{k} T_i}. \tag{2}$$

Subsequently, we use the values of c_i from (2) to aggregate the local models using (1). Upon aggregation, we obtain a global model that is used for performance evaluation using test datasets from different sources. Next we describe the experiments and results.

4 Experiments and Results

4.1 Datasets

Several radiographic findings are annotated in the datasets we use. The NIH Chest X-ray dataset contains 1, 12, 120 images divided classified for 15 target classes, 14 of which are for chest abnormalities and one is No Finding. The CheXpert dataset has 2,23,414 records divided into 14 classes, 11 of which are pathologies and the other three are Fracture, Support Devices, and No Finding, respectively. The VinBigData Chest X–ray dataset contains 15000 images divided into 14 classes, 13 of which are for pathologies and one is No Finding. We conduct our experiments for binary classification, considering two classes. These classes are normal (images with No Finding, Fracture, and/or Support Devices) and abnormal (images with at least one of the other abnormalities). These datasets are either skewed toward the normal or the abnormal class.

A recent research using the NIH Chest X-ray dataset found that training a moderate dataset is more effective than training a suitably big dataset [4]. Hence, we only use 11,624 images from the NIH dataset for our experiments. There are 5,787 normal images and 5,837 abnormal images in this training set. We take a total of 8,574 images for training, 1,706 for validation, and 1,344 for testing following [4]. The model trained using this dataset is named as NIH model.

We use only the frontal images from the CheXpert dataset. This dataset is heavily skewed toward abnormal ones. As a result, we use data augmentation techniques to increase the count of normal images in the training set. We perform data augmentation in five different ways. We rotate the normal images by 5°,10° in both clockwise and anticlockwise rotation (four ways here), and performs horizontal flip. There are now 2,66,761 images in the whole collection, including 1,00,200 normal images and 1,44,500 abnormal ones. For training, we use $2,44,700$ images, $12,000$ for validation, and $10,061$ for testing. The model trained using this dataset is named as CheXpert model.

The VinBigData Chest X-ray dataset has a strong bias in favour of normal images. To eliminate the bias, we increase the abnormal images by rotating 5° clockwise and anticlockwise, respectively and by horizontal flip. There are now 19,788 images in the whole collection, including $8,106$ normal images and $7,182$ abnormal images. We use 15,288 images for training, $2,750$ images for validation, and $1,750$ for testing. The model trained using this dataset is named as VinBig model.

4.2 Implementation

We use Densenet-121 as the backbone for the local (client) models. We use models pre-trained on the ImageNet dataset. We replace the model's final classification layer (1000-way softmax) with a single neuron that outputs the estimated chance that an input image is abnormal through sigmoid activation function. With a batch of 64 images, we train local models using 256×256 input images and 224×224 centre crop. Binary cross entropy loss (BCE) is the loss function employed. With a momentum of 0.9, a decay rate of 0.0001, and a learning rate of 0.001, the stochastic gradient descent (SGD) optimizer is used. Each model goes through 100 epochs of training. The open source pytorch deep learning framework is used to build this network.

4.3 Comparative Analysis of Performance

To look into the efficacy of the proposed method, we compare our method with the method proposed in [9]. The original paper uses simple averaging as their aggregation policy. We also compare our aggregation technique with three different

Table 1: Classification performance in terms of AUC using different models on test data from different datasets. NIH, CX, VB represent the NIH Chest X-ray test data, CheXpert test data, and VinBigData Chest X-ray test data, respectively. NIH, CheXpert, and VinBig are the Densenet-121 model trained with NIH, CheXpert and VinBigData Chest X-ray test dataset, respectively. DN121-SA represents Densenet-121 models from three clients aggregated using simple averaging. DN121-TL and DN121-VL represent Densenet-121 models from three clients aggregated using weights that are inversely proportional to training losses and validation losses, respectively.

Type	Method	AUC NIH	AUC CX	AUC VB
Local	NIH	0.98	0.83	0.76
Local	CheXpert	0.97	0.87	0.82
Local	VinBig	0.92	0.76	0.98
Global	[9]	0.96	0.83	0.62
Global	DN121-SA	0.97	0.83	0.91
Global	DN121-TL	0.95	0.81	0.95
Global	DN121-VL	0.96	0.82	0.94
Global	Proposed	0.97	0.87	0.83

aggregation methods using Densenet-121 backbone for local models. These methods are simple averaging (abbreviated as DN121-SA), weighted averaging with weights inversely proportional to training loss (abbreviated as DN121-TL) and weights inversely proportional to validation loss (abbreviated as DN121-VL). We also compare the performance of the proposed method with local models (Densenet-121 trained on either NIH, or CheXpert, or VinBig dataset). Sample image results using the proposed method are presented in Fig. 1.

Performance of Local Models All of the local models are trained for 100 epochs with the same parameters. The results of the local models are presented in Table 1, Table2 and Fig 2. Notice that all of the local models outperform their peers on their individual test data. The performance of the NIH model on the NIH test data is superior to that of the VinBig and CheXpert models.

On CheXpert test data, the CheXpert model performs the best, whereas the other local models do not. It can be observed that none of the local models performs consistently across datasets from different sources. Similar to the other local models, the VinBig model performs best on the VinBigData test data. There is a huge variance in performance between the various models. The NIH model shows the least performance than the other two local models. Thus, local models have poor generalizability.

Performance of Global Models During weighted federated averaging, the weights (c_i) of each local model are equal for simple averaging (model DN121-

Table 2: Classification performance in terms of F1 score and accuracy of different models on test data from different datasets. NIH, CX, VB represent the NIH Chest X-ray test data, CheXpert test data, and VinBigData Chest X-ray test data, respectively.

Type	**Method**	**F1-score** NIH	CX	VB	**Accuracy** NIH	CX	VB
Local	NIH	0.93	0.89	0.64	0.93	0.81	0.68
Local	CheXpert	0.90	0.90	0.70	0.90	0.84	0.75
Local	VinBig	0.80	0.87	0.88	0.83	0.79	0.90
Global	[9]	0.89	0.88	0.49	0.88	0.80	0.61
Global	DN121-SA	0.90	0.89	0.70	0.91	0.82	0.79
Global	DN121-TL	0.83	0.88	0.65	0.85	0.81	0.78
Global	DN121-VL	0.88	0.88	0.69	0.89	0.81	0.80
Global	Proposed	0.90	0.90	0.71	0.90	0.84	0.76

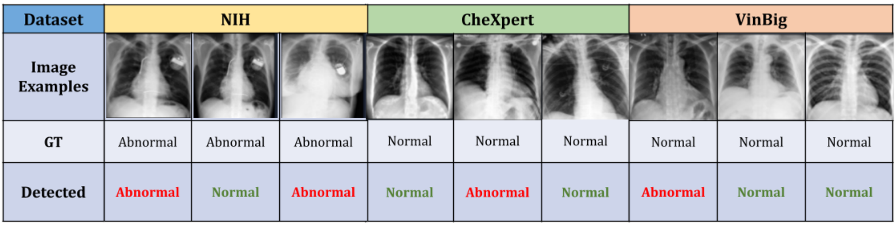

Fig. 1: Sample images from the different datasets with their predicted labels and the ground truth (GT) labels.

SA). Because the CheXpert model uses big train data, its weight is greater than the weights of the other two models in model Proposed. The weights of the VinBig model are higher than the weights of the other two models for models DN121-TL and DN121-VL. This is due to the VinBig model's lower train and validation loss when compared to the other two models. From Table 1, Table2 and Fig 2, it can be observed that the proposed model performs most consistently across datasets from different sources. Thus, we find the usefulness of the proposed federated learning scheme towards designing a generalizable method for detecting abnormal chest x-ray images.

5 Conclusion

In this paper we employ federated learning technique to incorporate the variance of multiple datasets into our model's performance while maintaining privacy. We attempt to include the diversity of diverse datasets and create multiple global models by aggregating local models, which yield very good results. We

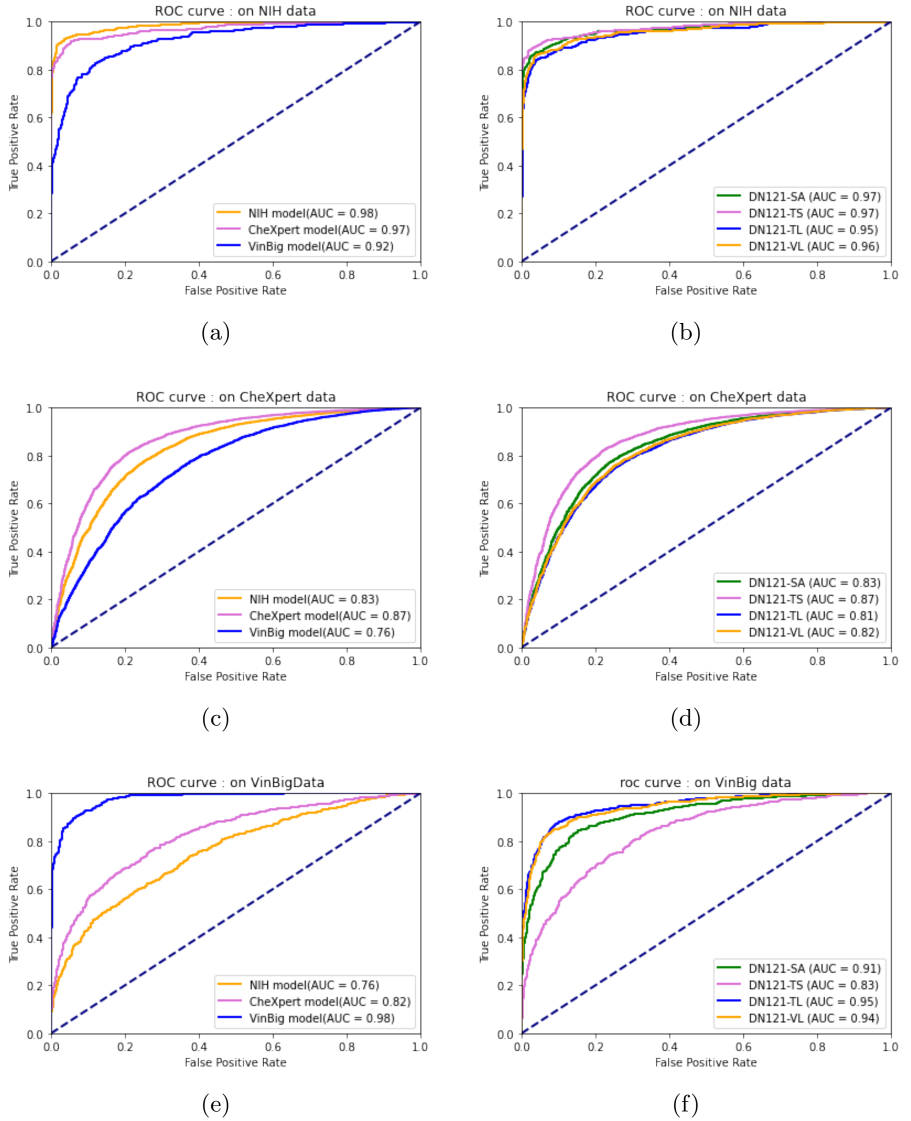

Fig. 2: ROC curves of different local and global models. Fig. 2 (a), 2 (c), 2 (e) represents the ROC curves when NIH model, CheXpert model, VinBig model are tested on NIH Chest X-ray test data, CheXpert test data, and VinBigData test data, respectively. Fig. 2 (b), 2 (d), 2 (f) represents the ROC curves when different global models are tested on NIH Chest X-ray test data, CheXpert test data, and VinBigData test data, respectively. The evaluation metric is AUC.

can observe that the global model, which is aggregated with weights in the ratio of training dataset sizes of the individual local models, produced good results among the offered models.

6 Acknowledgement

The authors are thankful to the NIH Clinical Center for making the chest x-ray dataset publicly available.

References

1. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. & Summers, R. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 2097-2106 (2017)
2. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K. & Others Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *Proceedings Of The AAAI Conference On Artificial Intelligence*. **33**, 590-597 (2019)
3. Nguyen, H., Lam, K., Le, L., Pham, H., Tran, D., Nguyen, D., Le, D., Pham, C., Tong, H., Dinh, D., Do, C., Doan, L., Nguyen, C., Nguyen, B., Nguyen, Q., Hoang, A., Phan, H., Nguyen, A., Ho, P., Ngo, D., Nguyen, N., Nguyen, N., Dao, M. & Vu, V. VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. (2020)
4. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Densely connected convolutional networks. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 4700-4708 (2017)
5. Krizhevsky, A., Sutskever, I. & Hinton, G. Imagenet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*. **25** (2012)
6. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*. pp. 770-778 (2016)
7. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 2818-2826 (2016)
8. Tang, Y., Tang, Y., Peng, Y., Yan, K., Bagheri, M., Redd, B., Brandon, C., Lu, Z., Han, M., Xiao, J. & Others Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digital Medicine*. **3**, 1-8 (2020)
9. Khan, S. & Alam, M. A Federated Learning Approach to Pneumonia Detection. *2021 International Conference On Engineering And Emerging Technologies (ICEET)*. pp. 1-6 (2021)
10. McMahan, B., Moore, E., Ramage, D., Hampson, S. & Arcas, B. Communication-efficient learning of deep networks from decentralized data. *Artificial Intelligence And Statistics*. pp. 1273-1282 (2017)
11. Çallı, E., Sogancioglu, E., Ginneken, B., Leeuwen, K. & Murphy, K. Deep learning for chest X-ray analysis: A survey. *Medical Image Analysis*. **72** pp. 102125 (2021)

12. Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T. & Yu, H. undefined. *Federated Learning*. (2019)
13. Chakravarty, A., Kar, A., Sethuraman, R. & Sheet, D. Federated Learning for Site Aware Chest Radiograph Screening. *2021 IEEE 18th International Symposium On Biomedical Imaging (ISBI)*. pp. 1077-1081 (2021)
14. Feki, I., Ammar, S., Kessentini, Y. & Muhammad, K. Federated learning for COVID-19 screening from Chest X-ray images. *Applied Soft Computing*. **106** pp. 107330 (2021)
15. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*. (2014)
16. Çallı, E., Sogancioglu, E., Van Ginneken, B., Van Leeuwen, K. & Murphy, K. Deep learning for chest X-ray analysis: A survey. *Medical Image Analysis*. **72** pp. 102125 (2021), https://www.sciencedirect.com/science/article/pii/S1361841521001717
17. Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T. & Yu, H. undefined. *Federated Learning*. (2019)
18. Alpers, J., Reimert, D., Rötzer, M., Gerlach, T., Gutberlet, M., Wacker, F., Hensen, B. & Hansen, C. 2.5 D Thermometry Maps for MRI-guided Tumor Ablation. *International Conference On Medical Image Computing And Computer-Assisted Intervention*. pp. 311-320 (2021)
19. Feki, I., Ammar, S., Kessentini, Y. & Muhammad, K. Federated learning for COVID-19 screening from Chest X-ray images. *Applied Soft Computing*. **106** pp. 107330 (2021)
20. Al., I. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*. **27**, 1735-1743 (2021,9), https://doi.org/10.1038/s41591-021-01506-3
21. Dayan, I., *et al.*. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*. **27**, 1735-1743 (2021,9), https://doi.org/10.1038/s41591-021-01506-3
22. Banerjee, S., Misra, R., Prasad, M., Elmroth, E. & Bhuyan, M. Multi-diseases Classification from Chest-X-ray: A Federated Deep Learning Approach. *AI 2020: Advances In Artificial Intelligence*. pp. 3-15 (2020)
23. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K. & Others Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *ArXiv Preprint ArXiv:1901.07031*. (2019)
24. Johnson, A., Pollard, T., Berkowitz, S., Greenbaum, N., Lungren, M., Deng, C., Mark, R. & Horng, S. MIMIC-CXR: A large publicly available database of labeled chest radiographs. *ArXiv Preprint ArXiv:1901.07042*. (2019)
25. Hwang, S., Kim, H., Jeong, J. & Kim, H. A novel approach for tuberculosis screening based on deep convolutional neural networks. *Medical Imaging 2016: Computer-aided Diagnosis*. **9785** pp. 750-757 (2016)
26. Lakhani, P. & Sundaram, B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. **284**, 574-582 (2017)
27. Nam, J., Park, S., Hwang, E., Lee, J., Jin, K., Lim, K., Vu, T., Sohn, J., Hwang, S., Goo, J. & Others Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology*. **290**, 218-228 (2018)
28. Taylor, A., Mielke, C. & Mongan, J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. *PLoS Medicine*. **15**, e1002697 (2018)

29. Gabruseva, T., Poplavskiy, D. & Kalinin, A. Deep learning for automatic pneumonia detection. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition Workshops*. pp. 350-351 (2020)
30. Kundu, R., Das, R., Geem, Z., Han, G. & Sarkar, R. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PloS One*. **16**, e0256630 (2021)
31. Tang, S., Wang, C., Nie, J., Kumar, N., Zhang, Y., Xiong, Z. & Barnawi, A. EDL-COVID: ensemble deep learning for COVID-19 case detection from chest x-ray images. *Ieee Transactions On Industrial Informatics*. **17**, 6539-6549 (2021)
32. Ozturk, T., Talo, M., Yildirim, E., Baloglu, U., Yildirim, O. & Acharya, U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers In Biology And Medicine*. **121** pp. 103792 (2020)
33. Yan, C., Yao, J., Li, R., Xu, Z. & Huang, J. Weakly supervised deep learning for thoracic disease classification and localization on chest x-rays. *Proceedings Of The 2018 ACM International Conference On Bioinformatics, Computational Biology, And Health Informatics*. pp. 103-110 (2018)
34. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K. & Others Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *ArXiv Preprint ArXiv:1711.05225*. (2017)
35. Zhou, B., Li, Y. & Wang, J. A weakly supervised adaptive densenet for classifying thoracic diseases and identifying abnormalities. *ArXiv Preprint ArXiv:1807.01257*. (2018)
36. Tang, Y., Wang, X., Harrison, A., Lu, L., Xiao, J. & Summers, R. Attention-Guided Curriculum Learning for Weakly Supervised Classification and Localization of Thoracic Diseases on Chest Radiographs. *Machine Learning In Medical Imaging*. pp. 249-258 (2018)
37. Wang, H., Jia, H., Lu, L. & Xia, Y. Thorax-net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography. *IEEE Journal Of Biomedical And Health Informatics*. **24**, 475-485 (2019)
38. Liu, H., Wang, L., Nan, Y., Jin, F., Wang, Q. & Pu, J. SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images. *Computerized Medical Imaging And Graphics*. **75** pp. 66-73 (2019)
39. Ouyang, X., Karanam, S., Wu, Z., Chen, T., Huo, J., Zhou, X., Wang, Q. & Cheng, J. Learning hierarchical attention for weakly-supervised chest X-ray abnormality localization and diagnosis. *IEEE Transactions On Medical Imaging*. **40**, 2698-2710 (2020)
40. Paul, A., Tang, Y., Shen, T. & Summers, R. Discriminative ensemble learning for few-shot chest x-ray diagnosis. *Medical Image Analysis*. **68** pp. 101911 (2021)
41. Paul, A., Shen, T., Lee, S., Balachandar, N., Peng, Y., Lu, Z. & Summers, R. Generalized zero-shot chest x-ray diagnosis through trait-guided multi-view semantic embedding with self-training. *IEEE Transactions On Medical Imaging*. **40**, 2642-2655 (2021)
42. McMahan, B., Moore, E., Ramage, D., Hampson, S. & Arcas, B. Communication-efficient learning of deep networks from decentralized data. *Artificial Intelligence And Statistics*. pp. 1273-1282 (2017)
43. Xu, J., Glicksberg, B., Su, C., Walker, P., Bian, J. & Wang, F. Federated learning for healthcare informatics. *Journal Of Healthcare Informatics Research*. **5**, 1-19 (2021)

44. Lim, W., Luong, N., Hoang, D., Jiao, Y., Liang, Y., Yang, Q., Niyato, D. & Miao, C. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys Tutorials*. **22**, 2031-2063 (2020)
45. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H., Albarqouni, S., Bakas, S., Galtier, M., Landman, B., Maier-Hein, K. & Others The future of digital health with federated learning. *NPJ Digital Medicine*. **3**, 1-7 (2020)
46. Jin, Y., Wei, X., Liu, Y. & Yang, Q. A survey towards federated semi-supervised learning. (2020)
47. Li, X., Jiang, M., Zhang, X., Kamp, M. & Dou, Q. Fedbn: Federated learning on non-iid features via local batch normalization. *ArXiv Preprint ArXiv:2102.07623*. (2021)
48. Briggs, C., Fan, Z. & Andras, P. A review of privacy-preserving federated learning for the Internet-of-Things. *Federated Learning Systems*. pp. 21-50 (2021)
49. Zhang, K., Shi, Y., Liu, Y. & Yan, Z. Power demand response incentive pricing model. *2019 IEEE International Conference On Big Data (Big Data)*. pp. 2605-2614 (2019)
50. Yoon, J., Jeong, W., Lee, G., Yang, E. & Hwang, S. Federated continual learning with adaptive parameter communication. (2020)
51. Chen, F., Luo, M., Dong, Z., Li, Z. & He, X. Federated meta-learning with fast convergence and efficient communication. *ArXiv Preprint ArXiv:1802.07876*. (2018)