Paper Title:
Federated Learning to Safeguard Patients Data: A Medical Image Retrieval Case

Paper Link:
https://www.mdpi.com/2504-2289/7/1/18

1 Summary

1.1 Motivation:

It is difficult to use centralized detection methods with healthcare data. It is necessary to be able to incorporate a larger and richer dataset by only sharing the parameters of the data through the use of FL for better training and ensuring privacy protection. But healthcare centers collect data in various ways which may contain many unstructured and textual data. Thus, a lot of pre-processing is required at times. Training FL based models can also be resource intensive and need compatible GPUs.

1.2 Contribution:

This paper gives a privacy oriented distributed learning solution that does not transmit the actual data back to the main model. The given solution is robust as it does not depend on the quality of data of the respective clients. The scalability was also ensured through how the model performs when there is a variation in the number of clients. By ensuring these, a neural network was designed that can detect COVID-19 by the use of X-ray images from other pathologies.

1.3 Methodology:

The main steps of approach used in this study are: local training, classification, local parameter sending and global model sharing. A supervised classification approach has been used where the used CNN model consists of 17 convolutional 2d layers, one flattening and one linear layer. The dataset, consisting of X-ray images of patients with COVID-19, pneumonia etc had two distributions: an IID and non-IID case.

1.4 Conclusion:

In the IID case, with 200 epochs, 96% accuracy is achieved with 5 clients, 90% for 10 clients and 93% for 15 clients. With non-IID case, accuracy is 90% for 5 clients, 87% for 10 clients and 87% with 15 clients. IID case in general performs slightly better.

2 Limitations:

2.1 First Limitation:

A centralized model has a slightly better accuracy than the proposed model, which is 98% for both IID and non-IID cases.

2.2 Second Limitation:

For non identical and dependent datasets, the results may be slightly less promising then identical and independent datasets.

3 Synthesis:

Even though centralized models may get better accuracy sometimes, it comes at the cost of data privacy. In such cases, Federated learning can provide a better alternative. However, how to keep the accuracy intact while implementing data protection strategies in FL is still a point of discussion.