Privacy-preserving training of deep neural networks in large scale infrastructures

E. Darzidehkalani[1,2,*], P.M.A. van Ooijen[1,2]

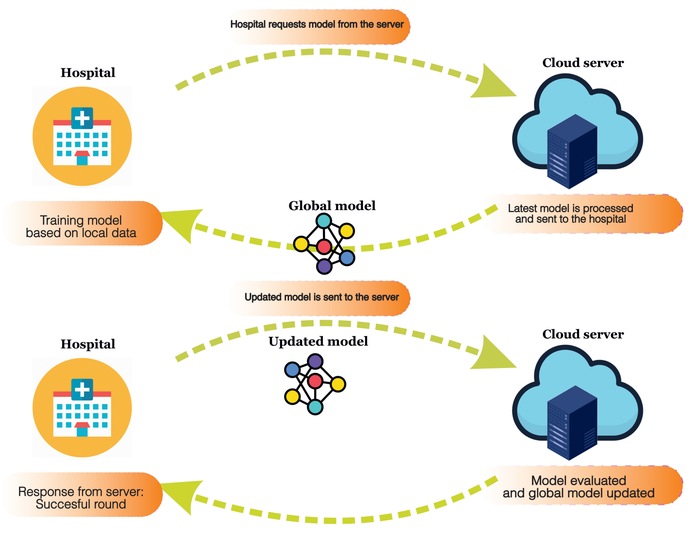
[1]Department of Radiation Oncology, University Medical Center Groningen, University of Groningen

[2]Machine Learning Lab, Data Science Center in Health (DASH),University Medical Center Groningen, University of Groningen, Hanzeplein 1, the Netherlands

*Contact: e.darzidehkalani@umcg.nl

Introduction

Aggregation of medical imaging data can help building accurate deep learning models. However, it is not always feasible due to strict privacy regulations. Federated learning (FL) is an emerging technology that enables researchers to build large scale networks and exchange trained models without compromising patients' private data. In this work we investigate the feasbility of federated sequential model exchange in brain tumor segmeetnation

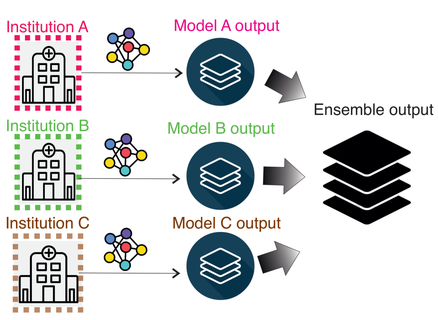


Communication between client and server, exchanging the model

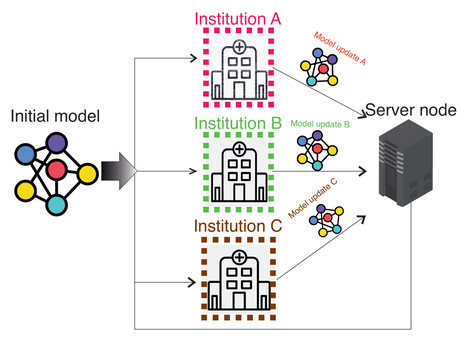
FL is a developing and growing technology that offers secure data access to institutions. It fosters global cooperation and will redefine the paradigms of AI in radiology in the near future.

FL algorithms

The most important algorithms used in the medical imaging domain are FedAvg[7], Ensemble methods, Single Weight Transfer (SWT), and Cyclic weight transfer (CWT).
In Ensemble models, training is done locally outputs of local models are averaged. In FedAvg, local models are trained in each hospital and then the models are averaged by a central server each round.

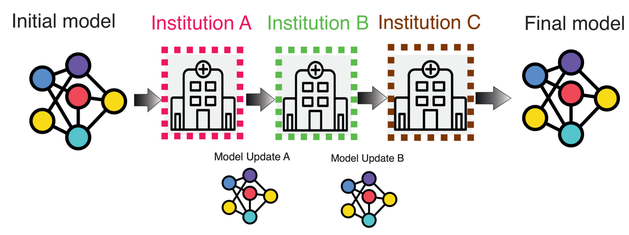


Ensebmle methods

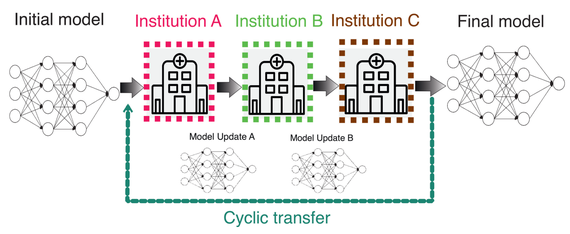


FedAvg

In SWT, models are passed through institutions only once, and the global model is updated after passing each client. CWT is like SWT, only the model is passed through institutions in a cyclic manner and more than one time.



SWT



CWT

Applications in radiology

FL has shown great promise in several areas of radiology. FL has been successfully deployed in COVID-19 research [4] , Lung nodule detection[1], retinotherapy[3], mammography[3], breast cancer detection[2], MR image reconstruction [8], brain tumor segmentation[6], brain tumor type classification, and patient similarity analysis[5]. With preserving patients private information and without revealing sensitive data.

Challenges and considerations

Data-related issues, such as heterogeneous data profiles and low-quality clients, affect FL network performance. Potential solutions are FAIR data collection, data standardization, and bias-reducing algorithms.
Security and privacy are also other important issues. Patient Re-identification, sensitive data retrieval, and adversarial attacks are the most important threats to an FL network. Countermeasures like model encryption, differential privacy (DP), and data perturbation are popular measures to protect private data.

Discussion and conclusion

FL and AI are growing fields and are expected to gain more trust from medical experts and open their way to more medical centers.
More data types, including medical texts/letters and genomics data, can be handled in FL networks in the future. Technologies like Natural Language Processing (NLP) are vital to extracting information from other data types in addition to the imaging data. A large pool of institutions with various data types opens the way to use real-time big data technologies in FL networks.

References

[1] Pragati Baheti et al. "Federated Learning on Distributed Medical Records for Detection of Lung Nodules." In: *VISIGRAPP (4: VISAPP)*. 2020, pp. 445–451.

[2] Brett K Beaulieu-Jones et al. "Privacy-preserving distributed deep learning for clinical data". In: *arXiv preprint arXiv:1812.01484* (2018).

[3] Ken Chang et al. "Distributed deep learning networks among institutions for medical imaging". In: *Journal of the American Medical Informatics Association* 25.8 (2018), pp. 945–954.

[4] Mona Flores et al. "Federated Learning used for predicting outcomes in SARS-COV-2 patients". In: *Research Square* (2021).

[5] Junghye Lee et al. "Privacy-preserving patient similarity learning in a federated environment: development and analysis". In: *JMIR medical informatics* 6.2 (2018), e7744.

[6] Wenqi Li et al. "Privacy-preserving federated brain tumour segmentation". In: *International workshop on machine learning in medical imaging*. Springer. 2019, pp. 133–141.

[7] Brendan McMahan et al. "Communication-efficient learning of deep networks from decentralized data". In: *Artificial intelligence and statistics*. PMLR. 2017, pp. 1273–1282.

[8] Tran Minh Quan, Thanh Nguyen-Duc, and Won-Ki Jeong. "Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss". In: *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1488–1497. DOI: 10.1109/TMI.2018.2820120.