FEDERATED LEARNING IN MEDICAL IMAGING, DECENTRALIZATION OF TRAINING

Erfan Darzidehkalani[1,2] **Mohammad Ghasemi-Rad**[3] **P.M.A. van Ooijen**[1,2]

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

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

[3]Department of Interventional Radiology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

{e.darzidehkalani,p.m.a.van.ooijen}@umcg.nl

ABSTRACT

With recent developments in medical imaging facilities, extensive medical imaging data is produced every day. This increasing amount of data provides an opportunity for researchers to develop data-driven methods and deliver better healthcare. However, data-driven models require a large amount of data to be adequately trained. Furthermore, there is always a limited amount of data available in each data center. Hence, deep learning models trained on local data centers might not reach their total performance capacity. One solution could be to accumulate all data from different centers into one center. However, data privacy regulations do not allow medical institutions easily combine their data and this becomes increasingly difficult when institutions from multiple countries are involved. Another solution is to use privacy-preserving algorithms, which can make use of all the data available in multiple centers while keeping the sensitive data private. Federated learning (FL) is such a mechanism that enables deploying large-scale machine learning models trained on different data centers without sharing sensitive data. In FL, instead of transferring data, a general model is trained on local datasets and transferred between data centers. FL has been identified as a promising field of research, with extensive possible uses in medical research and practice. This paper introduces FL, with a comprehensive look into its concepts and recent research trends in medical imaging.

Keywords Federated learning · Privacy-preserving machine learning · Medical imaging

1 Introduction

Deep learning has shown great promise in the field of radiology. It has been used extensively in various medical imaging domains and has already helped clinicians and radiologists in numerous ways. The area of radiology has dramatically benefited from deep learning research. It has been shown that deep learning can improve the existing models of tumor detection, from early processing stages such as image enhancement in Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), noise reduction, lesion detection and segmentation, disease monitoring. All of these areas have shown great promise for the use of artificial intelligence in clinical settings.

Deep neural networks are made up of many layers with billions of parameters, and they train to learn a complex, high-dimensional mapping from raw input data to desired labels.[1] The main issue with training deep neural networks in real-world medical practice is that a massive amount of diverse data is needed. A neural network trained on a single dataset from a single institute may be easily overfitted, resulting in a strong bias towards that institute and poor

generalization. Furthermore, latent patterns in one client's imaging data may influence the performance of a neural network in ways that have nothing to do with the actual biological way in the image. For example, datasets containing only one modality or images registered on a specific atlas may bias deep learning models towards that modality or atlas, capturing irrelevant data as significant predictors. The quality of data of a single institution depends on a variety of factors such as number of patients, type or number of imaging machines available, and number of experts available at that institution. Apparently, not all the healthcare facilities have vast amounts of diverse imaging data, and deep learning models are thus usually trained on limited datasets. Many times, it is also impossible for a single instution t One potential solution to this data shortage is to obtain imaging datasets from different clients. This method has the potential to increase both the amount and diversity of data collected. The most frequent method for establishing such a collaboration is to centralize vast and diverse datasets from multiple institutions and train a deep neural network on an accumulated dataset situated in a central hub, as can be seen in Figure 1. However, this technique is fraught with difficulties; strict national or regional privacy rules, such as GDPR in Europe or HIPAA in the United States, preclude institutions from easily sharing their patients' data. Other impediments may arise due to the multiple stakeholders, including hospitals, patients, researchers, medical physicians, and industrial corporations, each pursuing their own interests. The significant amount of time and effort (and hence money) that an institution spends to collect and clean data makes it hesitant to share it with other institutions.

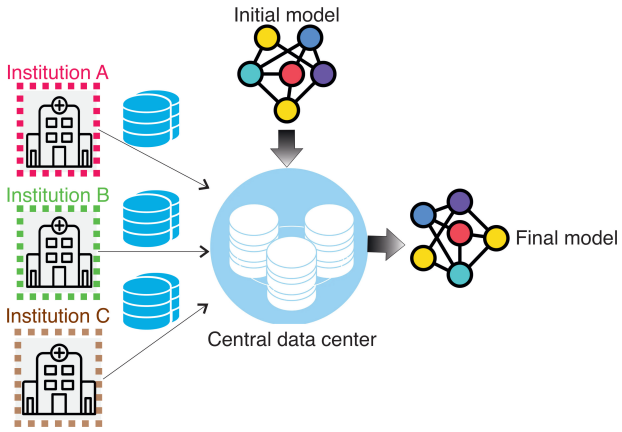

Figure 1: Centralized data sharing (CDS)

Recent advancements in privacy-preserving AI algorithms play an essential role in solving this . They enable researchers and institutions to train their networks on diverse imaging data from multiple institutions while ensuring that data will be kept locally, thus avoiding many of the issues concerned with building and maintaining an extensive central database.A general methodology in deep learning is decentralized or distributed learning. Distributed learning can be defined as a group of algorithms in which multiple clients do part of the computation or data storage tasks. The data distribution allows numerous clients to participate in the learning process and enables higher performance with a larger input data size. It generally involves multiple nodes and clients doing partial computation each on then own local database. Distributed learning is done for a variety of reasons, including performance boost and large scale computation. Federated learning is a version of distributed learning tailored for tasks where data privacy is essential so that researchers can preserve privacy while performing distributed learning. This feature enables healthcare centers to train deep learning models without compromising the privacy of their local data.

2 FL algorithms

A deep learning model is a form of an algorithm having artificial neural networks. It utilizes high volumes of data to learn the algorithm and find meaningful pattern in the datasets. Artificial neural networks have generally millions of parameters called model weights. Training a model is the process of teaching the algorithm to perform a task (like detection, classification or segmentation in the imaging domain). This training process is done by exposing the model to a specific dataset. FL is a learning method in which multiple participants train (or update) a local model on their data without actually sending data to the central node. A global model is updated according to the updated models received from participants. This way of training allows researchers to ensure the privacy of models and distributes the heavy computing process. FL is also efficient in communication since generally, only model weights will be communicated in this setting, and models have much smaller size than actual data.

In this regard, it tackles the infrastructural barriers of moving large volumes of data from one institution to another. Various ways to harmonize global and local model updates result in multiple versions of FL. Generally, federated networks require multiple clients who hold the data and perform the local training and a central trusted server, which manages the whole process.

Each client trains a model it gets from the central server on its local data. To get the model, the client sends a request to the cloud server, informing that it's ready to start the local training session. Then the request is processed, and the latest global model is sent back to the client. Next, the training session starts using the received model and local data. After the local training session is finished, the model is returned and the center accumulates the received updates. Finally, the global model is updated by the server based on the received model and notifies the client that one training round is successfully completed. A schema of these steps can be found in the Figure 2.

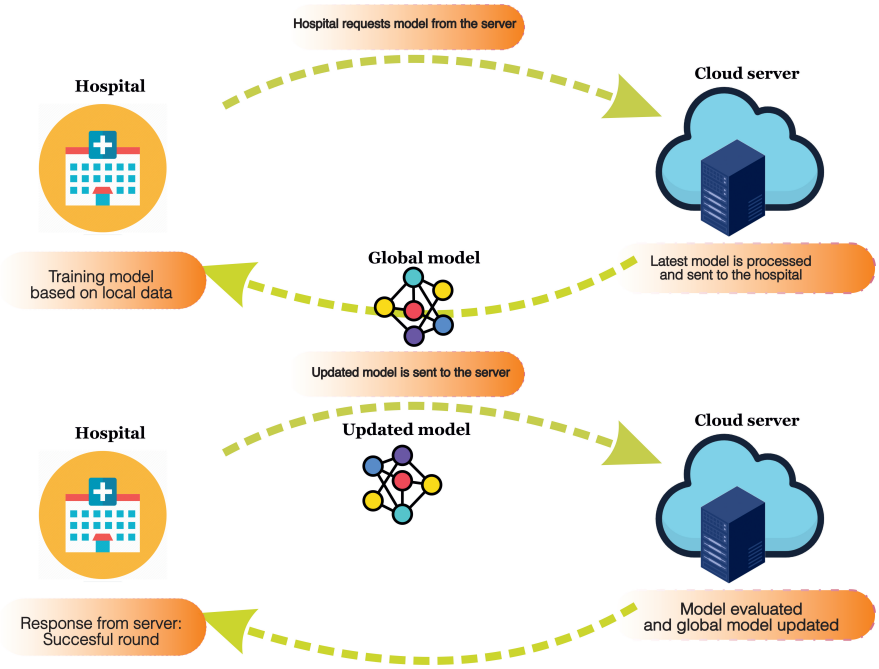

Figure 2: Communication between client and server, exchanging the model

There exist several FL algorithms, and this paper discusses the most important of them. McMahan et al. [2] proposed a federated averaging method (FedAvG) to minimize parameter change. The algorithm is straightforward: a subset of the clients is selected each round. Training is distributed among multiple clients. Each client will compute an updated model on their own local dataset. All model instances on the clients should start with the same random initialization to achieve convergence. Finally, a central server gathers the updates of the prospective clients.
Another approach is averaging the outputs of the local models trained on the clients individually (ensemble single client models). In this algorithm, neither the models nor the data will be shared among clients in the training cycle. All the clients will be assigned a similar model with random initial values. Each client will train its model. Their outputs for the same task will be averaged in the deployment phase, resulting in an accumulated knowledge from multiple models.

A third algorithm is single weight transfer (SWT). In this algorithm, a deep learning model is trained at a single client up to a particular time and then transferred to the next client. There are numerous options to decide when to finish a local training and pass its model to the next client. Standard criteria are the number of epochs per client and validation loss or accuracy depending on the problem. For example, Chang et al. [3] chose to reach the plateau of validation loss as a sign of moving to the next client. Cyclic weight transfer (CWT) is another algorithm in which a model is trained at each client for a predetermined number of epochs, then transferred to the next client. In this algorithm, the model visits each client than once.

The functionality of models and tasks in an FL scenario differs depending on the FL algorithm. Algorithms that transfer models are more versatile and adaptable than other algorithms. Deep learning models' performance in a federated environment can also vary from model to model. Models' adaptability can determine the overall performance of an FL network. For example, research has shown that some deep neural network components (such as batch normalization

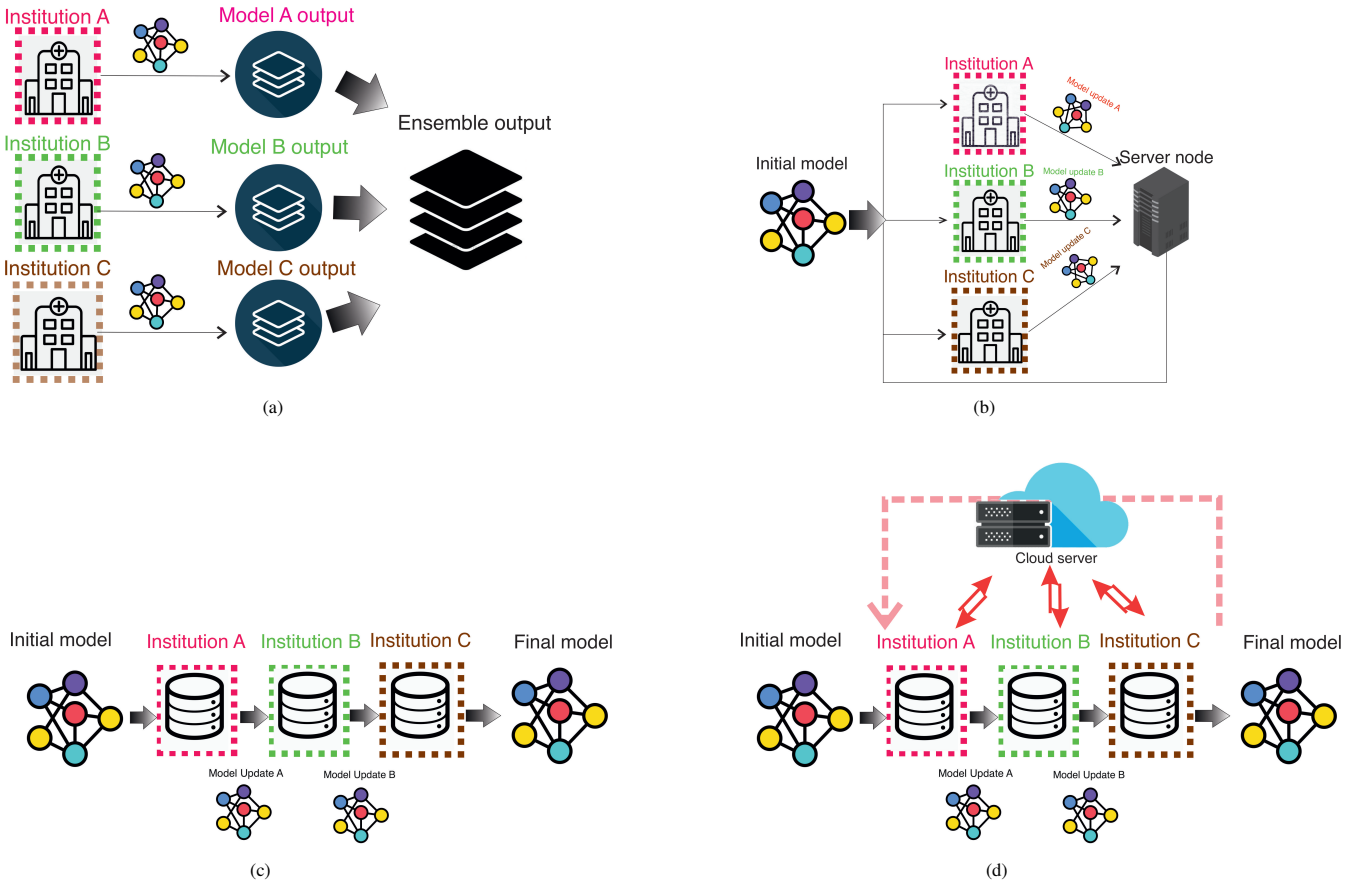

Figure 3: Schema of different decentralized learning methods, (a) Ensemble methods, clients train local models on their own dataset,model outputs of different clients are averaged. (b) FedAvg, an initial model is sent to the clients, each train the model on their own data and the resulting local models are averaged in a central server. (c) SWT, an initial model is sequentially passed through clients and visit each clients once. Final model is the model trained on the latest client. (d) CWT, similar to SWT, however, model is passed through institutions multiple times.

layers) cause performance issues and are harder to adjust in a federated setup. On the other hand, components like convolutional layers could be easily averaged, averaging their results in a proper global model. As a result, deep learning models that have more suitable components are a better choice for FL. Reseach is going on to develop specific models that perform better in a federated environment[4].

2.1 Comparison of the FL algorithms

We may categorize the algorithms based on what is exchanged between the server and the client to compare federated algorithms. Techniques such as FedAVG, SWT, and CWT, transfer the model between the server and the clients. Approaches like split learning [5] transfer middle layer outputs of a neural network. The middle layer outputs can be regarded as a distorted form of the input data. In other words, as the neural network processes the input data, it undergoes numerous modifications that distort the input. Methods such as ensemble methods share their models' final output and broadcast it to a central server.

The amount of data transferred is relatively tiny in methods in which the model is moved to the central server and is independent of the amount of training data at each site. It is solely determined by the size of the deep learning model. The majority of popular deep learning algorithms are tens of megabytes in size. However, an FL algorithm that transfers a model does not necessarily mean that the overall communication overhead is low. The overall amount of exchanged data also depends on the number of communication rounds between clients and servers. The hyperparameters can determine the number of communication rounds, and communication overhead could be high if there is too much exchange between clients.

Table 1: Comparison of FL methods

Methods	Summary	Transferred data	Communication load	Advantages	Disadvantages	Usecases
FedAvg	In each round, every client trains the global model on local data. Then models are averaged	Model	low	Easily converged	Weak robustness with imbalanced clients distribution	COVID-19 CT scans [6] Lung nodule detection [7]
SWT	model is passed through clients sequentially, visits each client once	Model	Very low	Low communication load	Highly biased towards the latest institution	Diabetic retinotherapy [3], Mammography [3]
CWT	Model is passed through clients sequentially; the sequence is repeated multiple times	Model	low	High performance	Needs many rounds to converge	Breast cancer data[8] EHR [8]
Ensemble methods	All the computations are done locally the outputs are averaged	Output	High	Easy to deploy	High possibility of data leakage, High communication load	Patient health records [9]
CDS	Move data from clients into a central datacenter	Data	-	High performance	No privacy	MR image reconstruction [10] , dermoscopy image synthesis [11]

*Federated Averaging **Single weight transfer

On the other hand, in algorithms that transfer some type of actual data, whether distorted input data (e.g., split learning [5]) or output data (e.g., ensemble models), the size of sent data can vary greatly depending on the data size. However, because medical imaging data is enormous, the amount of communicated information is usually more significant than with methods that transfer the model. CDS also falls into this category, as it requires actual data transfer to a central server. These two groups differ significantly in terms of communication burden as well as privacy level. Because input/output data is not sent in any format, methods which transfer models are more secure since retrieving patient data from deep learning models is difficult.

Another aspect of comparing FL models is that FL algorithms, in which a model is transferred, can consistently be averaged by the central server, regardless of the task they are performing. Deep neural networks performing classification, segmentation, regression, or other tasks could be averaged as long as there is a proper deep learning model for that. All of the mentioned tasks have been demonstrated and proved to work in a federated manner. However, averaging the output from many sources is not always feasible for other federated earning algorithms. For example, if the task is multi-class classification, an ensemble approach cannot simply average the class output of distinct clients. The Ensemble approach is thus limited in the jobs it can tackle.

Several research papers have been published that compare FL implementations. Nilsson et al. [12] compared various FL methods in practice. They demonstrated that FedAvg is the best FL algorithm. Despite having slightly lower performance than CDS, it is practically comparable in their comparative performance analysis to a non-federated architecture. There are numerous variants of the FedAvg algorithm and other FL approaches. However, the original FedAvg method remains one of the top methods in comparison studies. Chang et al.[3]investigated several FL algorithms in the radiology area. According to this study, FedAvg does not impose any bias compared to other algorithms because it considers all clients equally and does not arrange them in any particular order. As shown in Figure 3algorithms such as SWT and CWT, clients are placed in a sequence and trained one after the other. As a result of catastrophic forgetting, the model is more representative of the most recent clients it observed and less of the earlier clients[13]. As a result, there is a bias favoring the most recent institution in models with sequential training. Although CWT can mitigate this effect by running the model through institutions multiple times in a cyclic fashion, bias remains. Table 1 shows the essential characteristics of FL algorithms. There is also a sample of use cases of these algorithms in the medical domain.

3 Applications in radiology

Although FL still needs to be improved before it can be used on a large scale, it has shown promise in a practical medical imaging context in a few implementations in medical images, leading to improvements in patient care. FL can assist underrepresented patients in small clinics where they are a minority and may be overlooked and bring them into a pool with many other similar patients. FL has shown great promise in the research for COVID-19 patients; it was investigated and reported that FL had a clear impact on patient care in a large-scale study on COVID-19 patients across

20 centers on five continents[14]. They used chest X-Ray imaging data in addition to clinical data to determine hospital triage for level of care and oxygen requirement in COVID-19 patients. They demonstrated that the FL model works best for clients with limited datasets. The model performance for these clients is significantly improved compared to when they were trained on their local data, resulting in a change in the patient situation.

Another discovery was that medical centers with unbalanced data had some classes with few samples, resulting in underrepresented categories. These clients saw a significant improvement in prediction for those patient categories, which is especially important because, in COVID-19, patients with severe symptoms are generally in categories with fewer samples than a larger pool of patients with moderate symptoms. However, their care is more critical and requires more attention. In the field of applied FL in radiology, there are numerous projects. As an additional effort to the BraTs challenge, Intel and the University of Pennsylvania launched an extensive effort. This challenge was based on a dataset provided by the University of Pennsylvania's Biomedical Image Analysis section (SBIA)[15].

The Brain Tumor Segmentation dataset from the 2018 BraTs challenge was made available to the public. The dataset consisted of MRI images of glioma patient's brains, gathered from several studies in different institutions. Four radiologists manually annotated the MRI images, categorizing them into various Tumor classes. Tumors were classified into four types. U-Net was the deep learning model used to segment tumors, and the FL network was made up of one master node and numerous clients, each with their data. Two hypothetical clients were developed, and the dataset was assigned to them to evaluate the FL model. To examine different data distribution algorithms, they first divided the data randomly into silos. They also assigned data based on where it was obtained, resulting in non-homogeneous data. After finishing the local training, many clients delivered a model. The central server received updated models from all parties, selected the best models, and returned the aggregated models to the clients. This training strategy allows both the server and the clients to enhance their performance. After receiving the updated model from the central node, clients work on a better model each round. As a result of their experiments, they concluded that in the task of semantic segmentation, federated training could produce MRI segmentation masks that were better or comparable to models trained on-premise.

Sheller et al.[13]proposed a project on brain tumor segmentation using FL and achieved comparable accuracy to CDS. They demonstrated that increasing the number of collaborators improved the FL algorithm's performance and generalizability. Another study suggested a patient similarity analysis to find comparable patterns within different hospitals for possible similar treatments[16]. The goal of this study was the identification of patients with similar profile while protecting their privacy and personal information. They created hash codes to represent patients and a federated environment to control the entire process to achieve this goal. The hashed data had the advantage of being resistant to reverse engineering or adversarial model attacks. They could anticipate five diseases independently, using balanced and unbalanced data to evaluate their proposed algorithm.

They were able to demonstrate how data FL can aid in identifying comparable patients while protecting their privacy. Another effort was made to explore the structural relationship of the brain without revealing any data. The authors used Principal component analysis (PCA) to uncover anatomical relationships between diverse datasets in a federated setup[17]. Federated PCA could extract features from MRI pictures from several medical institutes. Their technique was validated using several databases, including ADNI, PPMI, MIRIAD, and UK Biobank[18].

Another paper that has used federated learning for radiology. Accounting for data variability in multi-institutional distributed deep learning for medical imaging Niranjan Balachandar 1 , Ken Chang 2 , Jayashree Kalpathy-Cramer 2 3 , Daniel L Rubin

4 Challenges and considerations

FL still has a long way to go in radiology. There are numerous challenges both in the theoretical formulation and practical implementation. FL algorithms could be divided into fully decentralized/peer-to-peer methods requiring a trusted central server. Each category comes up with its challenges. Generally speaking, methods with a central server offer more flexibility and better performance, while decentralized methods are more reliable and secure.

The literature is mainly on chest xray, mammogram or retinal images. These are 1 dimensional single pictures. I am not sure how this can address large data from a multi-slices CT or MRI. A single 2.5 mm thickness CTA can have up to 3000 slices. Or an Abdominal MRI can have 30 sequences each containing 30 images.

The main reason that hospitals are interested in bringing in FL algorithms is that privacy and security are their main priorities, and there are strict rules regarding the privacy of patients' data. However, there is still some risk associated with the FL infrastructure [19]. An adversary could reconstruct private data from the local model updates[20]. Hospitals can do additional security measures to prevent adversaries from accessing the exchanged data between the server and clients.

4.1 Data heterogeneity

The FedAvg algorithm authors claim that their proposed method can handle heterogeneous data. However, the decentralized structure of the data makes data processing challenging to verify the completeness and quality of their findings. Further investigations revealed that this claim is not always valid[21]. In almost all cases, heterogenous data deteriorates the accuracy of the FL model. The degree of divergence depends on how heterogenous the data is. Local models are trained on data with different patient profiles, resulting in a global model that could not represent all of the profiles. In some cases, heterogenous data prevents model convergence.

Data homogeneity significantly impacts the version of the federated model to be chosen to train the model. The difference between CDS and FL might vary from similar to CDS better depending on the data. One rule of thumb would be that if data has very different distribution in different data centers, simply averaging each client's data in every round might affect the performance negatively.

Zhao et al.[22]examined the effect of bias that distributing data can have on final performance of FL algorithms. According to their study, difference in data distribution can have negative effect on the model accuracy up to 55%. Another difficulty is that data heterogeneity might lead to a situation where a best global model might be a poor model for some clients, or a best global model might work quite well on some clients and perform poorly on other clients. Consequently, all participants should agree on the concept of optimum model training in advance of the training. Further technical studies should be carried out to find the optimum technique for updating the central model with heterogeneous data. FedAvg is the standard method for accumulating the data from clients. Still, other distributed optimization methods that can tackle distribution differences are a subject of research.

4.2 Bias

Bias is a prevalent issue in distributed networks. Bias is a state that a neural network is inclined towards the distribution of a client more than other clients. It results in the model performing well on that client by compromising the performance on the other clients. The cause of bias could be the difference in the size or the distribution of clients' data. Also, the FL algorithm itself could be a source of bias.

The global aggregation method, i.e., server algorithm, should be designed to minimize bias . It also should be robust to local variations, as well as perturbations added by security measures. Reducing bias and designing models that capture diversity is possible by calculating the level of bias arising from each client. Then modifying the algorithm to address the difference in the distributions.

However, if the distribution difference is taken into account properly, bias might still emerge later in training. Some features, as well as general data distribution might vary over time. For example, the number of patients with a particular disease in a particular hospital might change for several reasons. This can cause a domain shift: a change in a client's data distribution. There could be more work on data domain shifts and somehow explicitly address the alterations in gender, patient profile, age, and disease among different institutes or one institute. Models could also be further developed to consider economic or racial status into model training and modify a model to handle diversity in images[23].

4.3 Lack of standard data

Standardizing data prevents irrelevant information from being considered meaningful in neural networks. It removes the variability between institutions. Electronic data management is the norm in medical imaging: and medical communications (DICOM) is the globally recognized image data format and the near-global care standard for electronic file storage. However, not all the available data in the medical imaging sector is standardized. Many institutions still lack the infrastructure to handle their imaging data according to current management standards. One factor is the lack of a universal method to organize and manage patient records. Data management is expensive[24]. Not all hospitals have advanced data management facilities. This issue leads to the preselection of hospitals participating in research, which is a source of bias.

Medical data are very diverse because of the diversity of modalities, dimensionalities, and features and because of variables such as variances in the acquisition, medical equipment brand, or local demographics within a specific protocol. There is still no uniform data standardization method. As a result, healtchare federated networks are likely to have clients with disparate data quality and distribution. Methods like FedAvg are generally likely to fail under these circumstances. One way to avoid bias is to harmonize data and make each client data type similar, following a similar preprocessing. This also might require sharing metadata among institutions to find a general method of harmonization in data that suits all the institutions. But this could be tricky given the restrictions of individual institutions. Hence,

one way of further development for FL systems is that the clinicians and computer scientists collaborate to standardize handling data among multiple institutions concerning the privacy restrictions and considerations.

Data collection methods that follow the FAIR principles could also be a significant step forward in data standardization. The FAIR principle consists of Findable, Accessible, Interoperable, and Reusable data collection[25]. FL algorithms that could not use the data from various sites due to the difference in the data type could easily read and analyze data collected in the FAIR manner, which helps add more clients to the network. One example is language protocol differences in sites. Uniform Resource Identifier (URI) could represent the clinical data, enabling automated algorithms to read clinical text queries standardized with FAIR principles[26]. Integrating FAIR data collection and adding it as an initial step of building an FL network could strengthen the FL networks and ease more institutions to join the networks.

4.4 Privacy and Security

Data breach is a major concern and medical data and must be safeguarded in compliance with accepted confidentiality procedures. FL has proven to be effective in protecting patients' privacy and anonymity by keeping data locally. However, there are some privacy-related challenges associated with FL. Despite many attempts to de-identify personal data from DICOM images, patient information could still be re-identified[27] [28]. Recent studies have successfully rebuilt a patient's face from their MRI data. Furthermore, adversaries could steal the data or access the algorithm for non-encrypted networks.

Furthermore, deep learning models still have some sensitive information in the weights they carry. On a decentralized network, it is feasible to reconstitute portions of patients' information only having the local model from one client[29] [30] [31]. Adversaries can decrypt deep learning models and reveal patients' information with a very high accuracy[32]. Malicious parties can distort the deep learning models. False outputs produced by such models can have severe consequences if used in practice. As a result, it should be ensured that models are secure and that adversaries cannot breach models to be employed in the real-world setting[33].

There are specific measures to improve privacy. Particular countermeasures, such as model encryptions, Differential privacy (DP), [34] adversarial defense against malicious clients,[31] and increased communication security, can be done. DP refers to the practice of keeping a dataset's global statistical distribution while minimizing individually identifiable information. DP can be done by adding perturbation to each sample.Adding noise to a dataset to reduce the chance of private data being revealed is based on the argument that one can preserve general data distribution while individual samples are changed by randomly altering a dataset. Adding systematic noise helps machine models to learn the whole distribution of training data while keeping each sample anonymous.

However, such countermeasures complicate the training algorithms and can affect training performance. Sometimes much longer training times are required, or accuracy will be dramatically decreased. This can impose an additional cost on the whole network. Hence, it is quintessential to consider whether deploying a countermeasure is necessary. The cost-efficiency of implementing them depends mainly on the level of trust among involved parties and the project scale. If clients do not trust each other, then DP is a necessity. This is because federated clients have regular communication, and critical information can be exchanged in the interactions. So each client's data should be safeguarded from other clients. This shows how important it is to clarify the level of trust among clients. This argument holds true in fully decentralized algorithms, where no central node is involved. And also in algorithms including a central server, where the client/server trust is also essential. Total image anonymization is still a problem. In the absence of encryption, attackers may acquire private information from local datacenters or intercept the communication pathways and rob the passing data.

4.5 System architecture

To allow the local model training hardware (GPU), connections and data centers should be set up in local centers. These bring their challenges, such as high computational power to assure harmony with other clients and high-performance bandwidth and connection between different centers, which is not always feasible in medical centers. Many hospitals still lack computing resources and strong internet connections [35]. Besides, for the whole network to work correctly, redundant computational facilities and data centers should be devised to prevent data loss. If one computational client fails, the network could continue its training, which imposes an additional challenge. Another critical thing to consider while designing and implementing a federated network is the robustness of the network. Federated models should be structured so that adding or removing clients and increasing or decreasing the amount of data in a center does not negatively impact patient data or model privacy.

5 Future of FL research

There are several research trends showing that FL research is growing. The future direction of FL is to integrate it with big data technologies. After establishing FL networks, data could be added to the existing networks in real-time. Allowing the training and inference phase to work in real-time is a potential future direction of FL networks. This can be as streamlining pre-processing, training and, data handling.

It is expected that Federated networks include medical imaging data and work on all other types of medical data. Most of the recent FL implementations make use of imaging data with neural networks specifically designed for image processing. However, other formats of data, especially Electronic health records (EHR), are starting to be added to the current networks and are a contemporary development topic. EHR data are including a wide variety of information from treatment histories to past medication in addition to the medical imaging data EHR data can generally be in text, medical letters, categorical data, quantitative numbers, and binary data[36]. Incorporating this information into the imaging data could help develop better models. For example, making various treatment plans as an input variable to a deep learning model could help radiologists choose between treatment plans. Using EHR data could also help determine the type/stage of disease, as researchers recently used EHR to detect Alzheimer's disease[37].

There is still research to convert EHR data formats to a format usable by deep neural networks. Some progress has already been made using Natural language processing (NLP) to make text records available for deep learning[38]. For this purpose, researchers developed a data standardization framework to extract meaningful features from text data and make them available in the ML pipelines. Medical images combined with genomics data could also be a line of research. Because genomics data are not as prevalent and readily available as imaging data, the data limit problem in genomics is a much bigger issue than medical imaging. Hence, FL can play a pivotal role in bringing genomics data to the medical imaging field. Medical centers can communicate through FL with all types of their data in the future, so the collaboration level is expected to expand.

6 Conclusion

FL is a developing and growing technology that has influenced a variety of aspects across several fields. It offers straightforward and secure data access for institutions. FL utilizes the capacity of several institutions to enhance radiology research while overcoming the limitations of privacy and data sharing laws and regulations. Building a federated environment helps in achieving performance equivalent to the CDS set. It can foster global cooperation among several institutions, therefore redefining the paradigms of artificial intelligence in radiology. It also benefits from continually learning from new data flow, which may compensate for its inferior performance compared to the CDS setting with comparable amounts of data. This article should be helpful for radiologists and data scientists who want to learn about FL ideas and their applications in radiology.

References

[1] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning," *Pattern Recognition*, vol. 58, pp. 121–134, 2016.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[3] K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. L. Rubin, and J. Kalpathy-Cramer, "Distributed deep learning networks among institutions for medical imaging," *Journal of the American Medical Informatics Association*, vol. 25, no. 8, pp. 945–954, 2018.

[4] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "Fedbn: Federated learning on non-iid features via local batch normalization," *arXiv preprint arXiv:2102.07623*, 2021.

[5] M. G. Poirot, P. Vepakomma, K. Chang, J. Kalpathy-Cramer, R. Gupta, and R. Raskar, "Split learning for collaborative deep learning in healthcare," *arXiv preprint arXiv:1912.12115*, 2019.

[6] Q. Dou, T. Y. So, M. Jiang, Q. Liu, V. Vardhanabhuti, G. Kaissis, Z. Li, W. Si, H. H. Lee, K. Yu, *et al.*, "Federated deep learning for detecting covid-19 lung abnormalities in ct: a privacy-preserving multinational validation study," *NPJ digital medicine*, vol. 4, no. 1, pp. 1–11, 2021.

[7] P. Baheti, M. Sikka, K. Arya, and R. Rajesh, "Federated learning on distributed medical records for detection of lung nodules." in *VISIGRAPP (4: VISAPP)*, 2020, pp. 445–451.

[8] B. K. Beaulieu-Jones, W. Yuan, S. G. Finlayson, and Z. S. Wu, "Privacy-preserving distributed deep learning for clinical data," *arXiv preprint arXiv:1812.01484*, 2018.

[9] Y. Li, C. Bai, and C. K. Reddy, "A distributed ensemble approach for mining healthcare data under privacy constraints," *Information sciences*, vol. 330, pp. 245–259, 2016.

[10] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1488–1497, 2018.

[11] X. Yi, E. Walia, and P. S. Babyn, "Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by wasserstein distance for dermoscopy image classification," *CoRR*, vol. abs/1804.03700, 2018. [Online]. Available: http://arxiv.org/abs/1804.03700

[12] A. Nilsson, S. Smith, G. Ulm, E. Gustavsson, and M. Jirstrand, "A performance evaluation of federated learning algorithms," in *Proceedings of the Second Workshop on Distributed Infrastructures for Deep Learning*, 2018, pp. 1–8.

[13] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, *et al.*, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[14] M. Flores, I. Dayan, H. Roth, A. Zhong, A. Harouni, A. Gentili, A. Abidin, A. Liu, A. Costa, B. Wood, *et al.*, "Federated learning used for predicting outcomes in sars-cov-2 patients," *Research Square*, 2021.

[15] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.

[16] J. Lee, J. Sun, F. Wang, S. Wang, C.-H. Jun, and X. Jiang, "Privacy-preserving patient similarity learning in a federated environment: development and analysis," *JMIR medical informatics*, vol. 6, no. 2, p. e7744, 2018.

[17] A. Grammenos, R. Mendoza-Smith, J. Crowcroft, and C. Mascolo, "Federated principal component analysis," *arXiv preprint arXiv:1907.08059*, 2019.

[18] S. Silva, B. A. Gutman, E. Romero, P. M. Thompson, A. Altmann, and M. Lorenzi, "Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE, 2019, pp. 270–274.

[19] H. Yin, A. Mallya, A. Vahdat, J. M. Alvarez, J. Kautz, and P. Molchanov, "See through gradients: Image batch recovery via gradinversion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 337–16 346.

[20] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 2512–2520.

[21] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," 2020.

[22] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018.

[23] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results," *Medical Image Analysis*, vol. 65, p. 101765, 2020.

[24] S. J. Wang, B. Middleton, L. A. Prosser, C. G. Bardon, C. D. Spurr, P. J. Carchidi, A. F. Kittler, R. C. Goldszer, D. G. Fairchild, A. J. Sussman, *et al.*, "A cost-benefit analysis of electronic medical records in primary care," *The American journal of medicine*, vol. 114, no. 5, pp. 397–403, 2003.

[25] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[26] L. Masinter, T. Berners-Lee, and R. T. Fielding, "Uniform resource identifier (uri): Generic syntax," *Network Working Group: Fremont, CA, USA*, 2005.

[27] Y. E. Aryanto, M. Oudkerk, and P. Van Ooijen, "Free dicom de-identification tools in clinical research: functioning and safety of patient privacy," *European Radiology*, vol. 25, 06 2015.

[28] E. Monteiro, C. Costa, and J. L. Oliveira, "A machine learning methodology for medical imaging anonymization," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 1381–1384.

[29] N. Carlini, C. Liu, Ú. Erlingsson, J. Kos, and D. Song, "The secret sharer: Evaluating and testing unintended memorization in neural networks," in *28th {USENIX} Security Symposium ({USENIX} Security 19)*, 2019, pp. 267–284.

[30] Y. Zhang, R. Jia, H. Pei, W. Wang, B. Li, and D. Song, "The secret revealer: Generative model-inversion attacks against deep neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 253–261.

[31] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the gan: information leakage from collaborative deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 603–618.

[32] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1322–1333.

[33] R. Tomsett, K. Chan, and S. Chakraborty, "Model poisoning attacks against distributed machine learning systems," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006. International Society for Optics and Photonics, 2019, p. 110061D.

[34] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[35] M. Li, S. Yu, K. Ren, and W. Lou, "Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings," in *International conference on security and privacy in communication systems*. Springer, 2010, pp. 89–106.

[36] T. Seymour, D. Frantsvog, T. Graeber, *et al.*, "Electronic health records (ehr)," *American Journal of Health Sciences (AJHS)*, vol. 3, no. 3, pp. 201–210, 2012.

[37] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of alzheimer's disease stage," *Scientific reports*, vol. 11, no. 1, pp. 1–13, 2021.

[38] C. Dreisbach, T. A. Koleck, P. E. Bourne, and S. Bakken, "A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data," *International journal of medical informatics*, vol. 125, pp. 37–46, 2019.