# Federated Learning in the Clinical Environment

## FMML (Federated Medical Machine Learning)

Gabriel Coral, Joshua Hait, Kyle Isaak, and Adam Watkins

CMPT 340 - G. Hamarneh
Simon Fraser University, Canada

**Abstract.** In recent years, the desire to keep one's data private and secure has become highly sought after in our society. One aspect of this is the privacy of medical data. While patient data can be used by Artificial Intelligence (AI) and Machine Learning (ML) models to better understand, diagnose, and treat diseases, stricter access to patient data adds an additional layer of complexity to using these techniques.

Federated Learning (FL) is a technique that allows patient data to remain private, while still using that data to build and improve ML models. This project explores and implements a proof of concept of FL, which could then be adapted to a clinical environment to build ML models to help improve patient care. Specifically we used MRI images of brains with 4 different classifications of tumors. We used Tensorflow framework to create our ML models and we used Flower framework for the FL methods. Our FL model accuracy results were within 10% of our centralized model results, but there is still room for improvement. Our results show that there is definitely value in exploring FL methods, as we have room for improvement and other research applied to the current pandemic shows the possible efficacy of FL in a clinical environment.

**Keywords:** Federated Learning · Patient Privacy · Model Aggregation · Machine Learning in Medicine

## 1 Introduction

### 1.1 Problem Description

The goal of this project is to create a ML model that can be trained on patient data while keeping that data private. Ideally, the model would be sent to a medical centre that stores data for training. Once it has finished being trained, the model can be sent back to a remote server for aggregation with other models.

In this project, the data used consists of several thousand MRI images of human brains. Some of these brains are tumour free while the others have one of three types of tumors: glioma, meningioma, and pituitary tumors. With the use of FL, our model is able to be trained and aggregated across multiple clients simultaneously without the MRI images ever leaving the local machine.

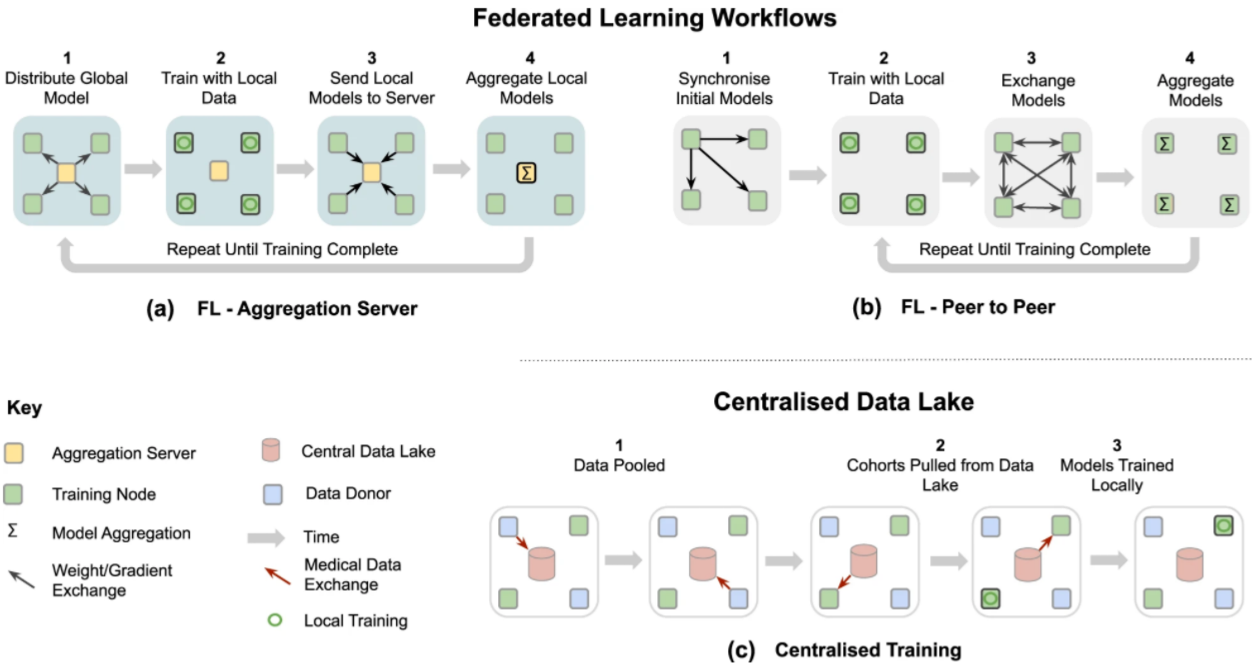
### 1.2 Project Motivation

In clinical environments, data privacy is a huge concern. Typical ML models require direct access to patient data to train models, which is not ideal as patients prefer their personal health data to remain confidential with their doctors, rather than that data being sent around to different research labs. However, the benefits of training ML models on real data is extremely valuable to future patient health, and the restricted access to patient data makes it difficult to train ML models easily. Therefore, there is a need for a way to train ML models on patient data, while retaining patient privacy, confidentiality, and anonymity.

### 1.3 Background Information

In the medical setting, FL is a method that can be used to improve data privacy while training ML models across multiple centers. The basic idea of federated learning is to distribute the model to nodes (or clients) that will train the models individually, send the trained models back to a central server, and aggregate the models together to produce a central model. FL achieves data privacy by not directly sharing sensitive personal information such as MRI images to the central server, but by sharing the resulting trained models from the clients. The model training utilizes ML techniques to provide accurate models. The method we will be going through in this paper will be similar to the aggregation server method of FL as seen in Figure 1.

**Fig. 1.** Example types of federated learning in a. and b. A typical centralized ML training method in c.
Source: Rieke, N., Hancox, J., et al



## 1.4 Related works / Literature survey

Other works have also looked into using federated learning in a practical environment by simulating model aggregation between institutions to assist with diagnosis of glioma patients []. The focus of the model training in the paper looks at an imaging segmentation method that uses U-Net Topology which has a wide range of imaging applications, but was used to identify regions of the brain in an MRI scan. Overall, when compared to a non-federated learning method, there is a significant improvement in model accuracy as more data for training would be aggregated, which previously was not possible due to strict limitations on data privacy and patient confidentiality.

## 1.5 Map of Report

The rest of the report is divided as follows: Section 2 describes the dataset we will be using to implement a FL example. Section 3 explains the ML model we created and FL framework we used. Section 4 presents the results of our FL experiment, as well as a comparison to regular ML methods. Section 5 discusses the accomplishments and a summary of what was learned throughout the project. Section 6 shows the contributions to the project by each team member. Section 7 presents our conclusions and discussion about the project and topic of FL. Section 8 explains how we or others could continue this project. Section 9 covers a list of references used during the project. Section 10 gives acknowledgement to sources that have guided us in our project.

## 2 Materials

For our dataset, we used a set of MRI images of the brain. The dataset features 3264 images, with four classes of images: 1) Brains with Glioma Tumors; 2) Brains with Meningioma Tumors; 3) Brains with Pituitary Tumors; 4) Brains without Tumors. The task for our FL/ML model will be to classify the type of tumor found in an MRI scan of the brain. The dataset we are using are already split into training and testing directories of images. The training directory is used by our FL clients to train on, while the testing directory was used by the FL server to evaluate the aggregated model. Note, as we are with four clients, the training data is separated in four separate equally sized partitions.

The dataset can be downloaded here: `https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri`

## 3 Methods

The FL Framework we have chosen to work with is called Flower (`https://flower.dev/`). Flower operates in a client-server architecture, where the centralized model on the server is acquired by training individual models on a number of clients, and aggregating the results. For our project, we chose to divide our training dataset into four partitions, and create four clients, each with one partition of the training dataset. These four clients each simulate a distinct 'clinic' sending weights, based off of their own data, to a central server; helping to train a model to predict classify brain tumors. Flower trains aggregates model in rounds. During each round, the server selects and samples from a fraction of the clients (2-3 in our project), to improve the central model. After every round, the server evaluates the central model on the testing dataset.

For the ML model used by our FL clients and on its own, our research led us to using a simple neural network to perform our image classification [6]. Our neural network has a flatten layer, followed by two dense layers. This ML model provided a good balance of simplicity, speed, and accuracy.

We will be comparing FL with four clients, each with a partition of the training dataset to a ML model that has access to the full training dataset. This will show how FL performs while keeping patient data private, versus a ML model that does not keep patient data private.

## 4 Results

We were able to create an ML model that identified whether MRI scans contained tumours with an accuracy of roughly 70% (see Fig. 2 for results) when training a model with the complete dataset. Training the model on a quarter of the dataset, simulating the data availability in each clinic, we achieved just a 50% classification accuracy. However using FL to aggregate models between clinics, we were able to train a model that could predict with an accuracy of roughly 60% (see Fig. 3 for results). Our accuracy was high enough for the FL model to be considered for practical improvement in a medical setting. Figure 4 shows the results of a ML model being trained on a single partition of the data with an accuracy of roughly 50%, which approximately shows what one of the FL client's training would look like. These results will be further discussed in Section 7.

## 5 Accomplishments

### 5.1 Summary of Learning

We learned ML image analysis techniques through using tensorflow to prepare an ML model capable of FL aggregation. Specifically we learned how to prepare and optimize image data to be used for training and testing

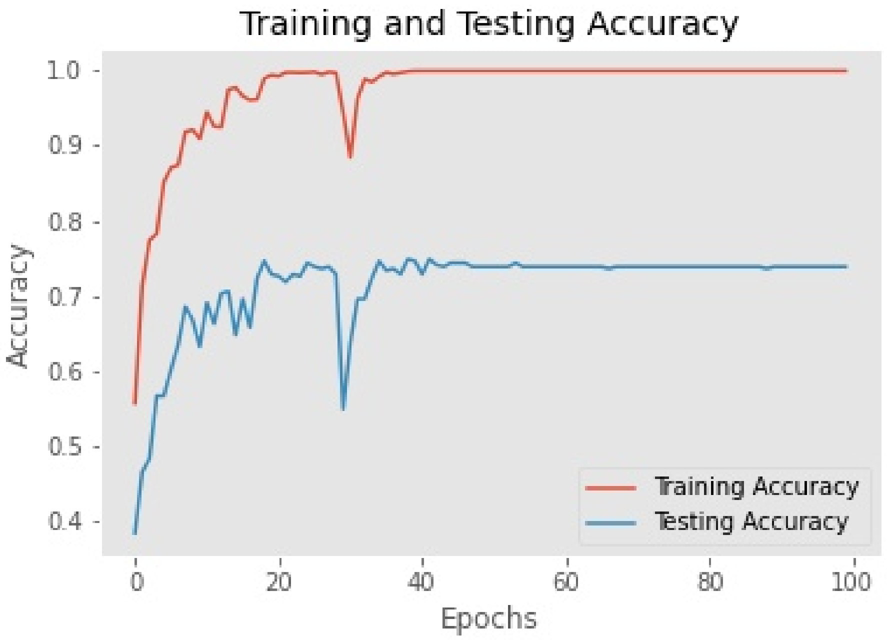


**Fig. 2.** Results of ML model training and testing accuracy

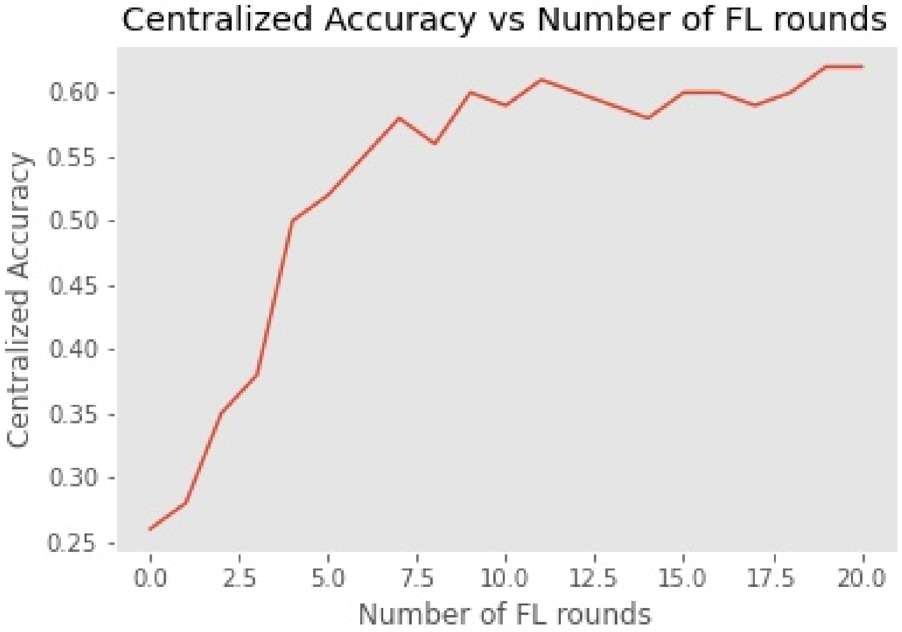


**Fig. 3.** Results of FL model centralized accuracy

using ML. We also gained experience in ways to evaluate trained models accuracy through testing and learned methods to improving model accuracy. We learned how implement FL strategies using the flower framework and some ways to improve trained model accuracy.

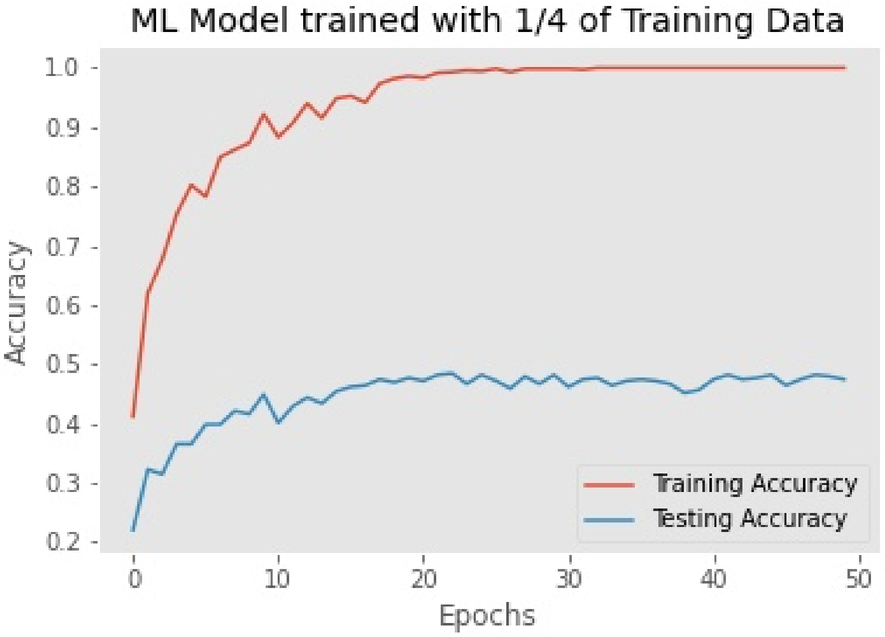


**Fig. 4.** Results of ML model trained on one client's partition

### 5.2 Obstacles that were Overcome

One of the biggest accomplishments of this project was successfully using the Flower framework to succesully implement FL with a custom tensorflow model. At the beginning of our project, most members on our team had never heard of the concept of FL, and had fairly minimal experience with ML. After completing this project, we all have a better understanding of both ML and FL, as well as the significance they both have to the medical community.

Another obstacle we had to overcome was how to use the data we acquired. The dataset used had a decent number of images (3264), but to demonstrate FL, we needed to divided that dataset up for our 'clients', as well as save a good portion of it for the 'server' to be used for testing.

An additional obstacle was finding a dataset to use. There are hundreds of available medical datasets online, but many of those datasets are extremely large (hundreds of gigabytes), or in hard to use formats. We settled on the dataset we did due to its relatively large size, and the simple file format (jpg images, which we converted to DICOM to simulate a medical environment) made the dataset easy to work with. The proper selection of our dataset was critical, as an easy to use dataset allowed us to focus our efforts more on the ML models and FL framework.

### 5.3 Obstacles that weren't Overcome

Originally, we had planned to create the FL framework ourselves. This was not pursued due to the complexity of the problem and the time constraints we had. We ended up using Flower (`https://flower.dev/`) as our FL framework to help us achieve our results.

Another thing we planned to do was to create an interface for clients to add more data to the model. Again, this was not pursued due to complexity and time constraints, as well as a shift in our project's goal, from designing our own FL system to using an existing framework to show a proof of concept.

# 6 Contributions

**Gabriel Coral**

- Searching for resources to reference for medical image analysis using tensorflow
- Searched for literature related to Federated Learning
- Contributed to report

**Joshua Hait**

- Searched for usable datasets, and found our dataset of MRI brain scans
- Implemented and tested Python scripts to load images and transform data into a format usable by our FL/ML models
- Created, trained, and evaluated a simple ML model to be used by and compared with the FL framework
- Assisted with the Tuning and Testing of parameters our FL Clients and Server to maximize the accuracy of the FL model
- Contributed to report

**Kyle Isaak**

- Searching for and planning what dataset to work on
- Creation of report
- Demonstration video

**Adam Watkins**

- Implemented federated learning client and sever using the flower framework for federated learning.
- Wrote scripts to convert our data set into the DICOM format using pydicom.
- Refactored machine leaning models from Jupyter notebooks into federated ready format.
- Contributed to the report

# 7 Conclusions and Discussions

## 7.1 Project Summary

Our project's goal was to create a ML model while keeping the data private, and we achieved that through the aggregation of trained models at a central server through the Flower and Tensorflow frameworks. While the accuracy of our model trained through FL was not up to the standards of a centralized ML model with all the data, when comparing to trained ML models of each individual we took a step towards being able to implement this method in a practical clinical setting.

## 7.2 Discussion

Our project tested three possible scenarios: 1) A FL approach is taken, where data is kept private and an aggregated model is created; 2) A ML approach, where data is not kept private, and a single model is used; and 3) A ML approach, where data is kept private, but clinics are only able to train on their local data. The ML approach with non-private data perform the best, with a 70% testing accuracy, followed by FL with a 60% testing accuracy, and finally ML with private data with 50% accuracy. While we had hoped that FL would perform the best, it was still able to beat out the ML approach with private data. This shows that FL has a practical use in a clinical environment, although it still has improvements to be made.

Other research seems to also have promising use for FL in COVID-19 mortality and diagnostic predictions [3]. The diagnostic predictions in particular used an online AI engine that utilizes FL model produces fairly accurate diagnosis of COVID-19 through CT-scans with a 98% sensitivity. This shows the possible present day uses of FL through its efficacy that could lead to more potential uses in the future.

## 8 Future Work

If we continued working on this, we would pick up one of our original goals to design and implement a FL system, rather than using an existing framework. This would likely require all of our team members to develop a deeper understanding of ML as a whole (which was something beyond what we had time to do for the duration of this project), as well as a deeper knowledge of the mathematics behind ML.

Another future goal would be to adapt our system to be able to work across multiple systems, rather than running our server and client on the same system. This would require good knowledge of networking and client-server architecture.

Additionally, Neural Networks and Convolutional Neural Networks seem to be the go-to models for image classification. Furthering our knowledge of these concepts could help us design a better base ML model, as we were only able to implement a very basic Neural Network.

Taking a step into a more practical setting, as mentioned in the related literature [1], there is the idea of considering data pre-processing to get similar results to our controlled research environment regardless of possibility in differences of image characteristics based on multiple sites having different equipment and procedures. The goal would be to use the pre-processing to make the inputted data to work well with a standardized federated learning method regardless of data or image recording procedures. One example of pre-processing that could improve ML model accuracy would be the U-Net Topology used for image segmentation that was in a previously mentioned paper that also looked at glioma patients [].

## 9 References

## References

1. Sheller, M.J., Edwards, B., Reina, G.A. et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep 10, 12598 (2020). https://doi.org/10.1038/s41598-020-69250-1
2. Rieke, N., Hancox, J., et al. The future of digital health with federated learning. npj Digit. Med. 3, 119 (2020). https://doi.org/10.1038/s41746-020-00323-1
3. Qian, F., Zhang, A. The value of federated learning during and post-COVID-19, International Journal for Quality in Health Care, Volume 33, Issue 1, 2021, https://doi-org.proxy.lib.sfu.ca/10.1093/intqhc/mzab010

. Sheller M.J., Reina G.A., Edwards B., Martin J., Bakas S. (2019) Multi-institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation. In: Crimi A., Bakas S., Kuijf H., Keyvan F., Reyes M., van Walsum T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science, vol 11383. Springer, Cham. https://doi-org.proxy.lib.sfu.ca/10.1007/978-3-030-11723-8-9

4. An Introduction to Biomedical Image Analysis with TensorFlow and DLTK, `https://blog.tensorflow.org/2018/07/an-introduction-to-biomedical-image-analysis-tensorflow-dltk.html/`
5. TensorFlow, `https://www.tensorflow.org/`
6. Image Classification with Tensorflow, `https://www.tensorflow.org/tutorials/keras/classification`
7. Flower, `https://flower.dev/`
8. Dataset from Kaggle, `https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri`

## 10 Acknowledgements

Special thanks to Professor G. Hamarneh, SFU, for refering us to possible FL frameworks to use for this project, which we ultimately chose the Flower framework to use for this project.