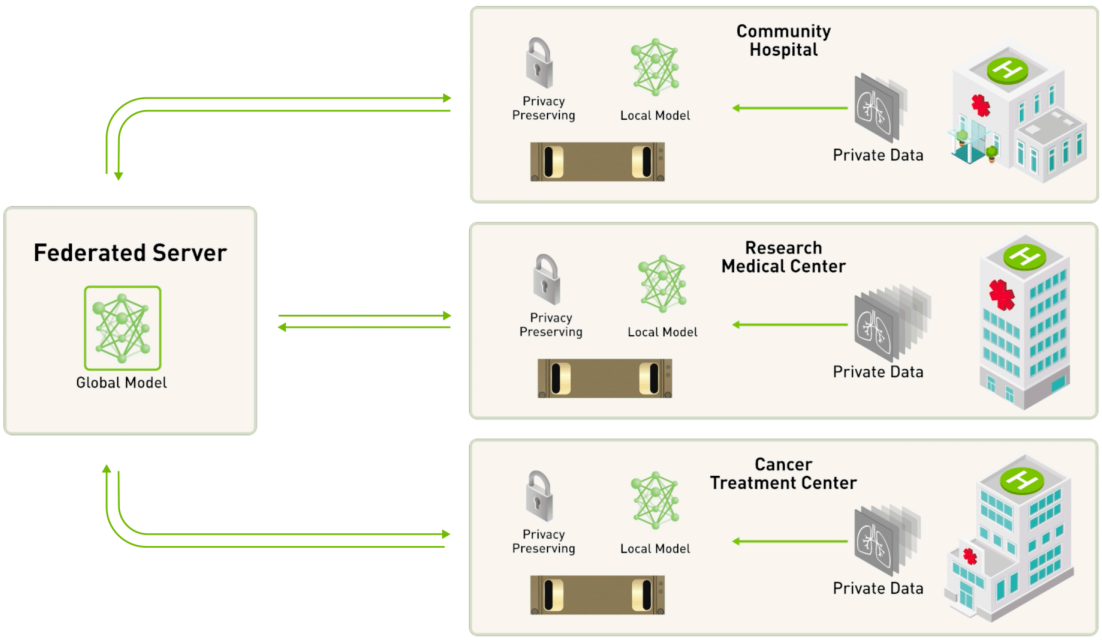
Computational Methods for Biomedical Image Analysis

Federated Learning

Introduction



Dataset is always a problem for machine learning and deep learning research, and the privacy of the data is also a important issue for every machine learning researcher. In biomedical image research, the source of the data mostly comes from hospital, yet the data is privacy. Therefore, Google published a new training idea, called Federated Learning.

Federated learning is a technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging them. This approach stands in contrast to traditional centralized machine learning techniques where all the local datasets are uploaded to one server, as well as to more classical decentralized approaches which often assume that local data samples are identically distributed.

Our project wants to verify that federated learning can still maintain high accuracy of the task, even though classical machine learning technique will still perform better, yet testing federated learning will be very important for biomedical image analysis field.

Dataset

In our project, we select the data from the **link**, which is liver dataset, as part of "Medical Segmentation Decathlon Challenge 2018". It consists of 131 labelled data and 70 unlabelled data. The labelled data was partitioned, based on our own split, into 104 training images and 27 validation images for this training task.

Approach

We are quite fortunate that we get the resource from Professor Chou, and know the application framework, CLARA, which is a healthcare application framework for AI-powered imaging, genomics, and the development and deployment of smart sensors. It includes full-stack GPU-accelerated libraries, SDKs, and reference applications for developers, data scientists, and researchers to create real-time, secure, and scalable solutions.

In our work, we will test tow different machine learning model, traditional and federated learning one. First, we will choose some well-known traditional machine learning model, such as SVM, Random Forest and so forth. Second, we will put the same dataset to the model (clara_ct_seg_liver_and_tumor_amp)

And the end, we will compare the predicting result for both of the training result. And discuss some important issues.

Evaluation

We will evaluate the training result by Confusion Matrix, and observe the accuracy and precision.