Digital Health - Connected Healthcare
Prof. Dr. Bert Arnrich
Bjarne Pfitzner

Master's Thesis Proposal

Predict Post-Operative Complications with Privacy-Preserving Federated Learning

Many surgeries are risky procedures with high chances of developing complications afterwards. Identifying these complications as early as possible is vital to provide the best care for patients. Machine learning models have shown much promise in these kinds of prediction tasks, however, they require large enough datasets to train on. Ideally, data from multiple hospitals would be aggregated to train a model with a diverse dataset, but regulations such as the **GDPR** prevent this sharing of sensitive data.

Federated learning (see Fig. 2) is a promising approach which entails exchanging models and training them directly where the data is, instead of sending out private data. This allows training on a large, distributed dataset without the need to ever access any protected data.

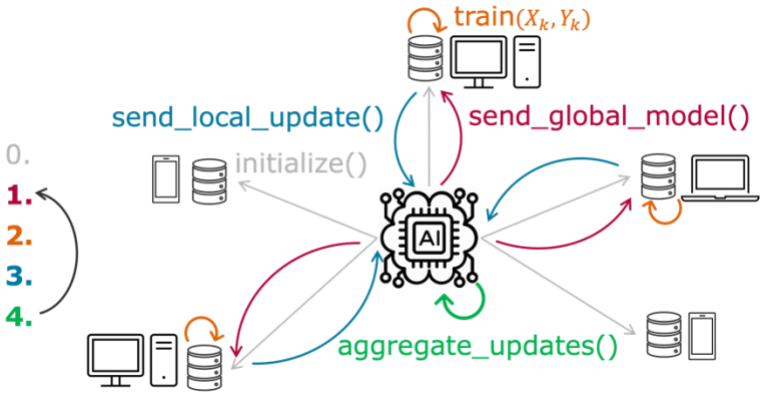

Figure 1: Federated Learning Architecture [1]

In addition to the native privacy benefits of federated learning, recent works have used **differential privacy**, a framework to formally quantify the privacy risk. In its essence, it relies on the addition of noise into the training process to hide the impact of any individual data record on the trained model.

The Charité – Universitätsmedizin Berlin consists of four distinct campuses in addition to multiple teaching hospitals and collaborative research centres. Many surgeries are performed at multiple locations, meaning that patient data is scattered across the different sites. You will receive access to **multimodal, real-world medical data** from campuses Mitte, Virchow and Benjamin Franklin and investigate the benefits of using federated learning instead of training multiple distinct models. The evaluation of differential privacy is an essential component of this thesis. Based on the insights gained, the next steps can be determined. One possible area of research could be generative models for the synthesis of even more data to improve future models. Another would be an investigation of federated multi-task learning, treating different types of surgeries as a separate but related task and evaluating the benefits of this approach.

Your Responsibilities

- Get acquainted with recent developments in the areas of federated learning and differential privacy.
- Implement and evaluate federated learning algorithms for the real-world dataset.
- Analyse the impact of differential privacy.
- Identify and investigate and additional sub-are of federated learning (e.g. generative models, multi-task learning)

Your Profile

- (At least) Basic knowledge of Deep Learning (and potentially classical machine learning)
- Good mathematical foundation
- Good programming skills (preferably Python) and experience in a Deep Learning framework (e.g. TensorFlow, PyTorch)

If you are interested in combining machine learning and data privacy for healthcare, please contact:

Bjarne Pfitzner
bjarne.pfitzner@hpi.de
+49 (0) 331 5509-1374
G-2.1.12, Campus 3

[1] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated learning of deep networks using model averaging", CoRR, vol. abs/1602.05629, 2016. arXiv: 1602.05629. [Online]. Available: http://arxiv.org/abs/1602.05629.