Federated Patient-Level Prediction

Byungjin Choi, Dong Yun Lee, Chungsoo Kim, Jimyung Park, Rae Woong Park

Background

Recently, several studies have attempted to predict patients' clinical outcomes by analyzing electronic medical records based on machine learning. However, existing machine learning can learn only from centralized data, so the use of multi-institutional data is limited.

Recently, federated learning has been proposed. Federated learning is a method of developing a multi-institutional model by sharing only weights without sharing data(1,2)

For federated learning, frameworks such as nvFLARE have been proposed and based on this framework, multinational, multi-institutional federated learning to predict the prognosis of COVID-19 has been conducted.(3)

However, in the above study, there was no common pipeline for model input extraction, so researchers at individual institutions had to manually extract input features such as chest x-ray data and vital signs. This process has limitations in that it lowers the study's reproducibility and increases the study's difficulty and cost.

Using the CDM (common data model), it is possible to extract clinical data across multiple institutions using the same method and code. In particular, the OHDSI community has simplified the cohort definition and feature extraction process using GUI (graphic user interface) tools such as ATLAS. Through a common data model and supportive OHDSI analytic tools, several clinical studies have been conducted.

Accordingly, by integrating OHDSI tools and federated learning frameworks, we aim to create a framework that makes clinical data extraction and model federated learning into one process. We call the above framework FPLP (federated Patient-level prediction).

Methods

Researchers can define custom PLP (Patient-level prediction) using ATLAS and create code for DLM (deep learning model) with PyTorch in the same way as in the existing analysis. After that, using FPLP API, a researcher can upload the custom PLP and DLM code to the server and give the command to initiate FPLP.

Using cohort configurations in the custom PLP package, each client generates the target cohort and outcome cohort, and the label is defined with the time-at-risk. And the clinical features are extracted through OHDSI's FeatureExtraction package. As a result, DLM input features and target outcomes were extracted.

After that, federated learning is conducted based on the DLM code. In the process of federated learning, the server sends the global model weight to individual clients. Local training is performed on each client, and the server collects client weight values and aggregate weights to update the global model weight. This process is repeated for a predefined number of rounds by the nvFLARE package.

In the test phase, the global model created through FL and all local models trained only with individual client data is validated on the test dataset of each client. This process is called cross-site evaluation.

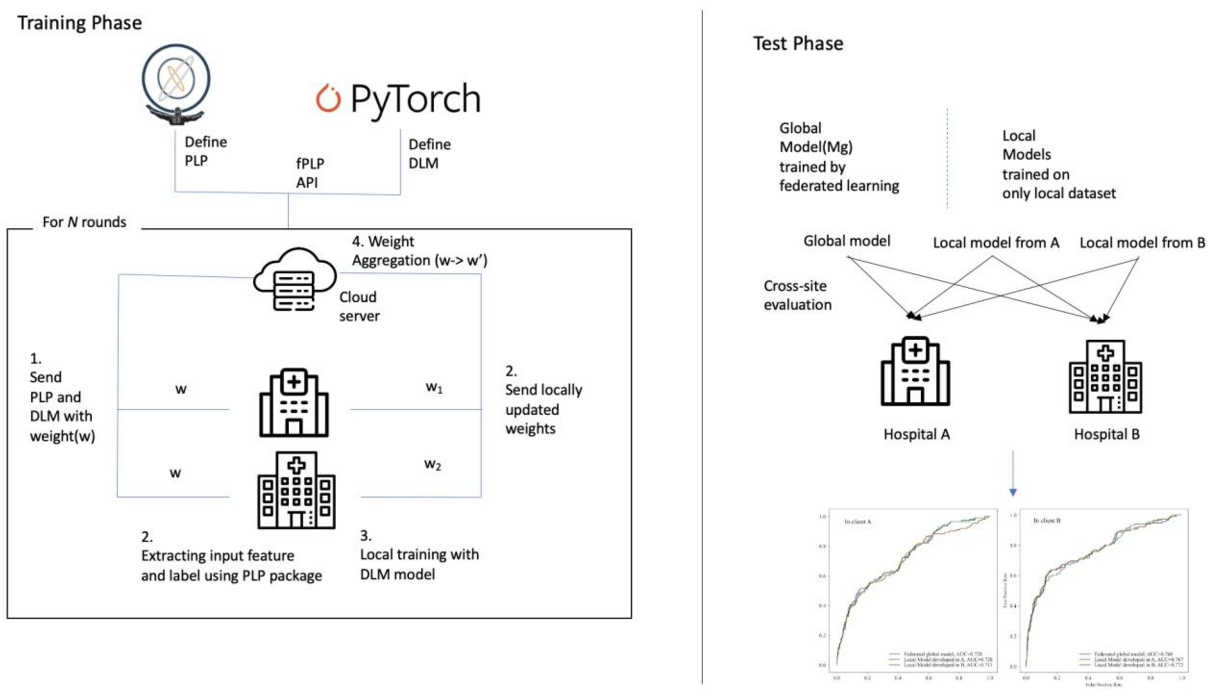

Figure 1. Overview of federated patient-level prediction

For the Proof-of-concept study, we randomly divided the electronic health records of the Ajou University Medical Center in South Korea into two simulated distributed clients based on patients (Client A, B). Then, we trained the prediction DLM through FPLP. An AWS (amazon web services) instance is the Central Server for FPLP. For security, the client connected only outbound connection to the port of AWS and communicated using gRPC (google remote procedure call), and intermediate communication is encrypted.

Through FPLP, we created a prediction model to predict AKI (acute kidney injury) within three days after PCI (percutaneous coronary intervention). We define target cohort, outcome cohort, and time-at-risk and set demographic variables, diagnoses, and medications as model input features through ATLAS. A deep & cross network algorithm was adopted for DLM. After training, we calculated the AUC (area under the receiver operating characteristic curves) as the performance metric.

Results

In each simulated client, the target cohort included 15,296 and 15,062 patients, respectively, and about 351 cases and 335 cases of AKI occurred. Input features and labels were extracted from the target cohort. Based on federated learning, the global model was developed. And only with the data of clients A and B, local model A and local model B were produced.

In the test dataset of client A, the global model showed AUC 0.728, local model A showed AUC 0.728, and local model B showed a performance of 0.711. And in the test dataset of client B, the global model showed AUC 0.768, local model A showed AUC 0.767, and local model B showed a performance of 0.772. Overall, the best performance was established when the validation was performed on the same client as the training client. The local model made in another place showed lower performance, and the global model showed intermediate performance.

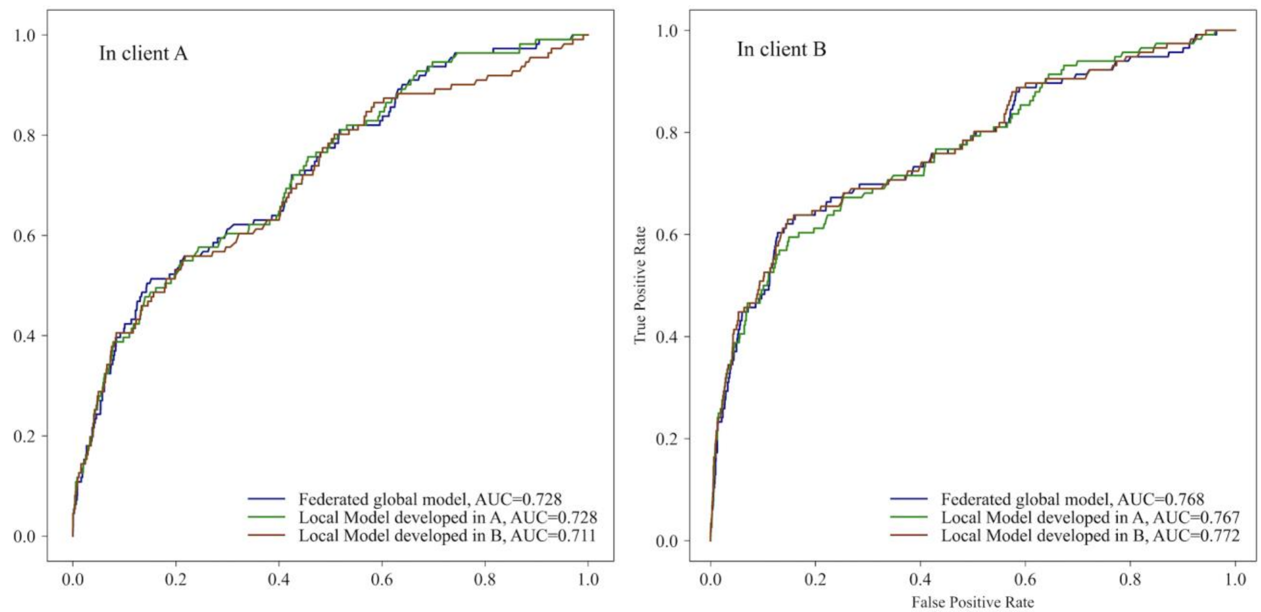

Figure 2. Performance of models

Conclusion

We integrated a common data model and federated learning process into one framework. Through this framework, the researcher can extract data and create distributed, multi-institutional prediction models with a unified code.

Acknowledgment

This work was supported by the Bio Industrial Strategic Technology Development Program (20003883, 20005021), funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea), and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR16C0001).

References/Citations

1. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3(1):1–7.
2. Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:161005492. 2016;
3. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021;27(10):1735–43.