Ferdelance: addressing data sharing issues using federated machine learning

D. Malpetti[1], C. Bonesana[1], L. Azzimonti[1], S. Mitrović[1],
L. Khoury Aerts[2], A. Pham[3], J.A. Bielicki[2]

[1] Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland
[2] Universitäts-Kinderspital beider Basel, Basel, Switzerland
[3] University of Basel, Basel, Switzerland

Spearhead
Innovating against AMR

Ferdelance

Introduction

In the biomedical domain, the integration of diverse data sources, such as data located at different hospitals, presents challenges due to privacy and security requirements. Frequently, regulations on both a national and international scale hinder the ability of hospitals to exchange data with one another. However, leveraging these data sources can yield substantial benefits, as a larger and more diverse dataset generally leads to more robust results in machine learning models.

This scenario directly applies to **SPEARHEAD**, a flagship project funded by Innosuisse. The consortium aims at performing a comprehensive study of antimicrobial resistance, a phenomenon where microorganisms become resistnat to antimicrobial treatments, in patients suffering with urinary tract infections. Among the aims of the project is the creation of a machine learning model designed to support medical practitioners in prescribing antimicrobials optimally. This model necessitates training via four distinct data sources, without the possibility of relocating the data from their original locations.

Federated machine learning (FL), a technique introduced by Mc Mahan et al. in 2017 [1], emerges as a solution to fulfill these specific requirements.

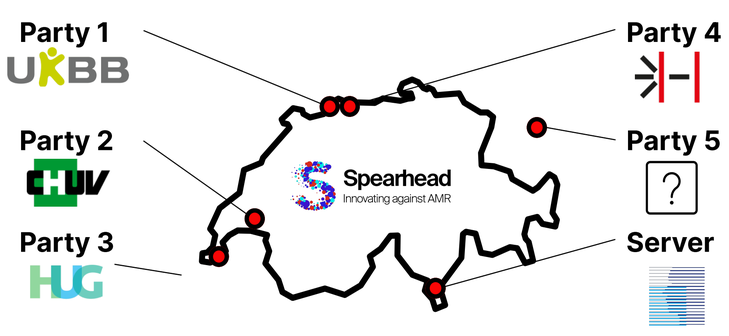

Figure 1
SPEARHEAD's distributed database, spread across four different Swiss hospitals. In addition to data sources, a server is located in Lugano, at IDSIA. In the future, the consortium could be joined by one or more data sources located outside of Switzerland.

Methods

FL allows multiple institutions to collaboratively train a common model **without sharing their data,** but only sharing fully anonymous model parameters. FL is typically coordinated by a server that builds a **global** model out of **local** contributions received from data holders (clients).

FL was first proposed in 2017 by McMahan et al [1]. It ensures greater data privacy and security as it eliminates the need for central collection or storage of data, which is a major concern for several organizations. As a result, FL is experiencing an increase in popularity in the healthcare sector [2-3].

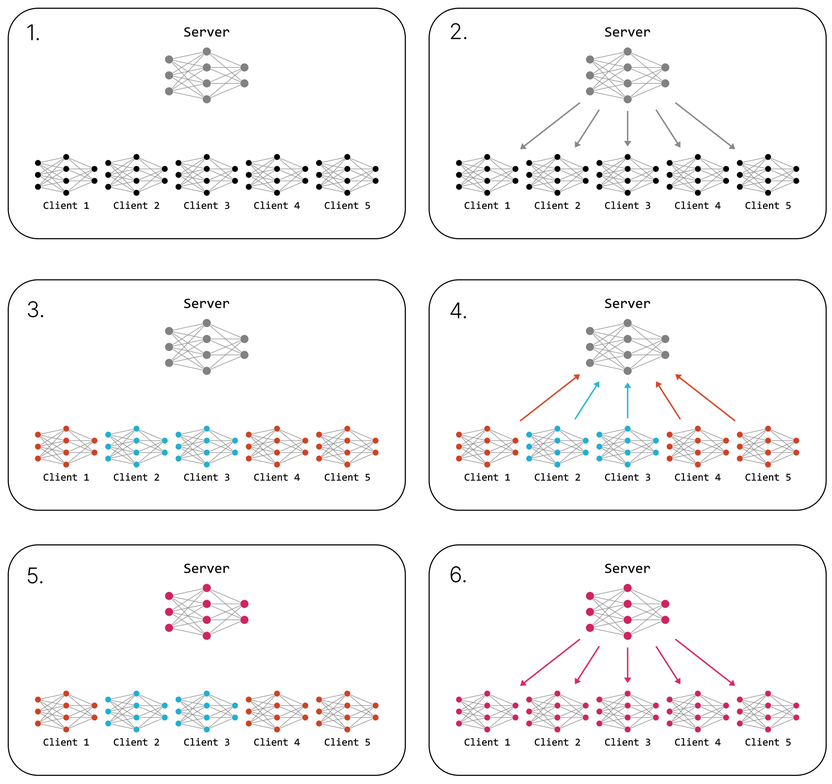

Figure 2
Frames providing a schematic representation of a simple federated learning training scheme, for a setting with a server and five clients (holding data locally).
1. Starting configuration.
2. Server initiates communication with clients, in some cases sending an initial model.
3. Clients train local models.
4. Clients send local models to server.
5. Server combines local models to build a global model.
6. Server sends global models to clients.

At the end of the training all clients possess the global model.

Results

We are developing **Ferdelance**, a versatile framework for training federated machine learning models. **It is open-source, modular, and particularly suitable for research purposes.** It is designed with privacy, security, and traceability in mind, while allowing the data holders to keep full control on their data.

Ferdelance can work with many kind of federated algorithms, thanks to the capability of working both in a **centralized topology**, where a central node has the role of aggregating models together, and in a **decentralized topology**, most suited for iterative algorithms.

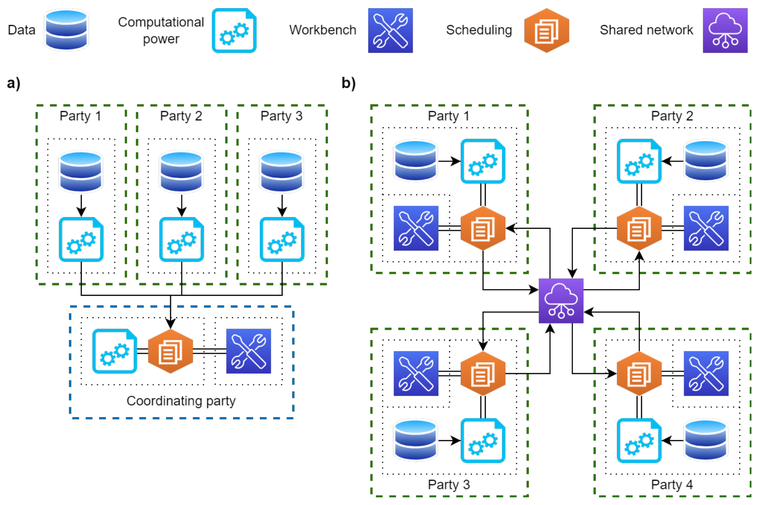

Figure 3
Ferdelance permits to train models both in presence of an aggregator server (centralized topology) and in a peer-to-peer context (decentralized topology). All the parties possessing data need to hold an amount of computational power too. One workbench and one scheduling are necessary for the framework to function. Decentralized topology requires the presence of a shared network too.

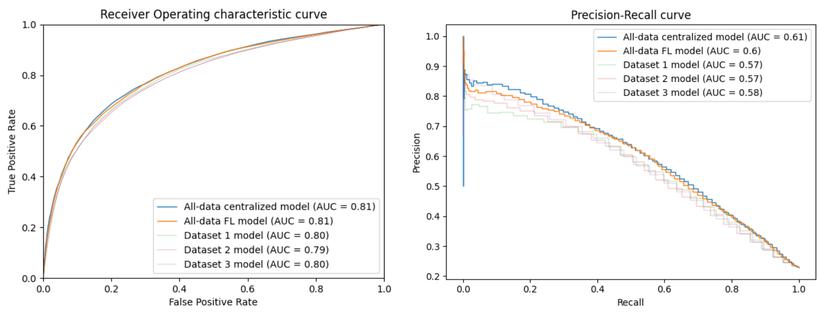

Figure 4
We tested Ferdelance using publicly available data [5]. We mimicked a federated scenario by partitioning the dataset among three distinct parties. A model trained using the entire dataset simultaneously and one employing federated learning show comparable performances.

Conclusions

Ferdelance permits the training of machine learning models within a federated environment, leveraging the complete potential of data sources situated across different sites, without the need for data migration between sites. Key characteristics of the framework are its versatility, the modular structure, and its design, created with privacy, security, and traceability in mind.

The framework was developed as part of the SPEARHEAD project, which will serve as the primary use case for demonstrating the its efficacy. While the present testing phase was performed using publicly available data, access to data from hospitals is set to be provided in the near future, allowing for a comprehensive evaluation within an authentic real-world scenario.

References

[1] Communication-efficient learning of deep networks from decentralized data, McMahan et al., *AISTATS Proceedings*, 2017

[2] The future of digital health with federated learning, Rieke et al., *npj Digital Medicine*, 2020

[3] Federated learning for predicting clinical outcomes in patients with COVID-19, Dayan et al., *Nature Medicine*, 2021

[4] Ferdelance: a Framework for Secure Federated Machine Learning, Bonesana et al., *Unpublished: in preparation*, 2023

[5] A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection, Kanjilal et al., *Science Transl Med*, 2020

Funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Project partners

Dalle Molle Institute for Artificial Intelligence IDSIA · SUPSI · University of Basel · SUPSI · Hôpitaux Universitaires Genève · Universitätsspital Basel · UKBB kompetent & menschlich · CHUV Centre hospitalier universitaire vaudois · Resistell™ · WAAT · Swiss Re · NOVARTIS