Federated Learning for Privacy-Preserving Prediction of Disease Risk in Healthcare Networks

Digital healthcare has brought immense opportunities for leveraging patient data to enhance medical decision-making, disease prevention and prediction, and personalized treatments. However, the sensitive nature of medical data and stringent privacy regulations, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), pose significant challenges to data sharing across healthcare institutions. To address these concerns, Federated Learning (FL) offers a novel solution by enabling distributed model training without sharing sensitive patient data, thus preserving privacy and regulatory compliance.

This project proposes developing an FL framework for predicting disease risks, such as diabetes or cardiovascular conditions, using synthetic healthcare, real or mimic datasets. The framework will simulate real-world scenarios where multiple healthcare institutions collaboratively train a global model while keeping patient data decentralized. This approach ensures that data remains local to each institution, protecting privacy.

The project involves designing and implementing an FL system using the Flower framework. Privacy-preserving techniques, such as differential privacy and secure aggregation, will be integrated to enhance data security. Performance will be evaluated against centralized learning approaches, focusing on accuracy, communication efficiency, and privacy guarantees. The system's scalability and robustness in handling diverse data distributions will also be analyzed.

Aim

By addressing the technical, ethical, and practical challenges of deploying FL in healthcare, this project aims to contribute to the growing field of privacy-preserving AI. The outcome will demonstrate the feasibility of collaborative innovation in healthcare while maintaining the confidentiality of sensitive patient information.

This project explores the development of a **federated learning (FL) framework** for predicting disease risks, leveraging decentralized patient data to enhance privacy and comply with regulations like GDPR and HIPAA. FL addresses critical challenges in healthcare data sharing by enabling collaborative model training without sharing raw data.

The project will involve:

- Designing and implementing an FL system based on the FLOWER framework using synthetic healthcare, real or mimic datasets.
- Integrating privacy-preserving techniques, such as differential privacy and secure aggregation.
- Evaluating the system's performance in terms of accuracy, efficiency, and scalability compared to traditional centralized approaches.

The expected outcomes include a functional FL framework for disease prediction, a comprehensive performance evaluation, and a detailed report. This work demonstrates the potential of FL to revolutionize healthcare through privacy-preserving AI, equipping students with practical skills to address future challenges in secure, data-driven medicine.

Supervisors:
Sadi Alawadi (sadi.alawadi@bth.se)
Prashant Goswami (prashant.goswami@bth.se)