Distributed System Architecture for Decentralized Learning in Telemedicine

Master Project, Summer 2020

Machine learning approaches are gaining momentum in the medical domain, especially to inform and drive clinical processes through decision support systems (CDSS). With the development of large-scale telemedical systems providing care to thousands of patients at home, machine learning technology has the potential to improve the efficiency of daily health data assessments in the telemedical center.

As part of the Telemed5000 research project, we investigate the applicability of machine learning for telemedical care in heart failure (HF), which is among the main causes for unplanned hospitalizations in Germany. Each patient provides daily measurements of their weight and blood pressure as well as a 2-minute ECG recording and a self-assessment score, all of which are afterwards assessed by clinical personnel.

As the recorded patient data is subject to high data privacy concerns, the machine learning system must be designed with data minimization and patient anonymity in mind. On the other hand, the trained ML model should ideally be public as to benefit all affected HF patients regardless of whether or not they are currently enrolled in a telemedical care program.

Recent developments in the field of federated learning indicate the feasibility of implementing machine learning systems in which user data remains private. Moreover, in decentralized learning, no central aggregation server is needed. However, regarding the real-world application of such architectures in the medical domain many questions remain, especially regarding the dependability requirements for medical care systems.

This master project focuses on architecture design and evaluation of decentralized learning from a distributed systems perspective. High-level project goals include:

- Designing a distributed architecture to ensure the fault tolerance and dependability of the system, including the collaboratively trained model
- Creating a prototypical implementation for use in a lab environment, incorporating peer-to-peer networking and distributed ledger technologies
- Extending a distributed application testbed evaluate the interactions of thousands of participating nodes in a simulated environment

We expect previous experience in systems and network development, including system-level programming languages. Previous experience in the areas of machine learning architectures, distributed ledgers or fault-injection techniques is welcome.

Contact: Robert.Schmid@hpi.de (C-1.13), Jossekin.Beilharz@hpi.de, Andreas.Polze@hpi.de