Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Distributed Intelligence: Federated Learning in Cardiovascular Data



Project Manager
Satya Aditya Kanuri

Principal Investigator
Prof. Dr. Christoph Hoog Antink

Project Term
2024 - 2024

Clusters
Lichtenberg II Cluster Darmstadt

Institute
Artificial Intelligent Systems in Medicine

University
Technische Universität Darmstadt

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

Introduction

In recent years, Cardiovascular diseases (CVDs) and other cardiac abnormalities have been on the rise, becoming a leading cause of death worldwide. The rapid development of Internet-of-Things (IoT) technology has opened up new possibilities in healthcare. Devices such as smartwatches and fitness bracelets now feature medical-grade ECG capabilities. The advancement of artificial intelligence (AI), particularly in machine learning (ML) and deep learning, has led to remarkable innovations in various medical fields including radiology , pathology , genomics, injury risk assessment, and disease prognosis. However, the use of health data is heavily regulated by stringent legal requirements. Additionally, the decentralized nature of medical data collection is significant. Hospitals and medical universities collect patient data for routine care and research, extending beyond primary use. Despite this, sharing data between institutions, known as secondary use, is rare due to legal constraints, posing risks to data holders. Balancing privacy protection with data availability is essential for optimal AI application outcomes. In 2015, Google researchers McMahan et al. introduced federated learning (FL) to improve typing predictions in their Android operating system without centralizing data. The core principle of FL is to avoid aggregating data from various clients (referred to as nodes) into a central infrastructure. Instead, data remains in the nodes' secure local environments, and only models are exchanged. This approach eliminates the risks associated with secure data transfer and storage. Despite its potential, the practical application of FL in healthcare, particularly for ECG data, is still in its early phase. The unique characteristics of heterogeneous ECG data, including variations in sampling frequency, patient demographics, and recording devices, pose significant challenges in implementing effective FL strategies. Additionally, the scalability of FL and its performance in different scenarios, such as varying the number of participating clients, needs thorough investigation. This research aims to advance the understanding and practical application of Federated Learning in healthcare by implementing established FL strategies tailored to the unique characteristics of heterogeneous ECG data. Specifically, the study will create and evaluate several FL scenarios to assess their impact on metrics like accuracy, F1 score, Precision and Recall. By exploring how these scenarios perform under different conditions, including varying scaling methods and the number of clients, this research seeks to provide insights into the optimal configurations for FL in real-world healthcare settings.

Methods

For the experiments, several FL scenarios are created and simulated based on the data distribution (IID/Non-IID) and how the data is scaled/normalized(Central/Distributed Scaling) constituted to four scenarios and also a scenario where the datasets of hospitals are kept separate, which makes in a total of five scenarios. The Flower framework is utilized to simulate the training process with 4, 6, and 8 clients, ensuring that the data distribution and model training are conducted in a manner

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN TECHNISCHE UNIVERSITÄT DARMSTADT UNIKASSEL VERSITÄT Philipps Universität Marburg

that reflects real-world federated learning conditions. For all the scenarios discussed below, key performance metrics such as precision, recall, and F1 score are recorded after every communication round to evaluate the effectiveness of the distributed and central model. These metrics provide a comprehensive view of the model's performance in terms of ability to correctly identify and classify arrhythmia types.

Results

For all the scenarios described above, all scenarios were simulated for 15 communication rounds and metrics like Precision, Recall and F1-score were recorded for both the central and distributed model. The following sections presents the metrics for all the scenarios in the form of a box plot. Initially, box plots were generated for each metric (Precision, Recall, F1) by grouping all the scenarios based on the number of clients for each strategy. This was done separately for both the central and distributed models. Next, the box plots for each metric were created by grouping all the strategies together and presenting them for each client count. This was also done for both the central and distributed models. These two approaches of presenting box plots enhances the clarity and depth of the analysis by focusing on the interaction between strategies, client counts, and model types.

Discussion

The results and analysis of the three communication strategies(FedAvg, FedOpt, and FedProx) demonstrates their overall effectiveness in federated learning scenarios. The both sections of the results clearly reveal a performance decline when transitioning from IID (independent and identically distributed) to non-IID data distributions, as well as from centralized to distributed scaling across all tested scenarios and also the impact of client count from 4-6 clients. For IID scenarios, whether in centralized or distributed settings, all three strategies exhibited nearly comparable performance across varying client counts, indicating their robustness when data is uniformly distributed. However, the results diverged in the non-IID scenarios. While FedAvg maintained good performance with a small client count (specifically 4 clients), its effectiveness diminished as the number of clients increased. On the other hand, FedOpt and FedProx performed on par with FedAvg up to 4 clients in the non-IID setting. Notably, as client counts increased to 6, FedOpt and FedProx outperformed FedAvg, suggesting that these strategies are better suited for handling higher client counts in non-IID environments. For separate hospitals scenario, since the experiment is conducted based on a 3 client system, similar performance was observed between the 3 strategies with FedAvg performing slightly better than the other two strategies and also with lower variance of metrics between the communication rounds.

Last Update: 2025-03-03 15:41

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg