Asynchronous Federated Learning for Personalized Healthcare

Lucas Pacheco [1] Torsten Braun [1]

[1]Communication and Distributed Systems, Institute of Computer Science, University of Bern

Motivation

- Personalized healthcare is essential for tailored treatments and improved patient outcomes
- Increasing health data and machine learning techniques enable disease prediction, treatment optimization, and improved patient care
- Data privacy and security concerns arise due to sharing sensitive health data among institutions
- Federated learning is a promising approach to maintain data privacy, but it faces several challenges

Objectives

- Develop a novel privacy-preserving federated learning algorithm tailored for personalized healthcare applications
- Integrate advancements in machine learning and computer networking to address data privacy and communication overhead challenges
- Evaluate the algorithm's effectiveness and scalability using real-world health datasets

Literature review

- Personalized healthcare: importance in recent years for tailored treatments and interventions
- Rapid advancements: genomic sequencing, medical imaging, and electronic health records
- Machine learning: successful in utilizing health data for disease prediction, treatment optimization, and improved patient care [4]
- Data privacy and security challenges: increasing with volume and complexity of health data
- Sharing health data: raises privacy concerns due to potential data breaches and misuse [2]
- Federated learning: promising approach to address privacy concerns [3]
- Federated learning process: institutions train local models, share updates, and aggregate for global model without exchanging raw patient data
- Benefits: accurate and generalizable models while maintaining data privacy
- Challenges: efficient, privacy-preserving algorithms for heterogeneous, distributed healthcare data [1]
- Existing algorithms: communication overhead, inadequate data privacy protection during model aggregation

Research gaps

- Existing federated learning algorithms may not provide adequate data privacy protection during model aggregation, leaving sensitive health data potentially exposed.
- Current approaches often have high communication overhead, limiting their efficiency and practicality in real-world healthcare settings.
- Heterogeneous health data and varying client capabilities require tailored solutions, which are not yet fully addressed by existing algorithms.

Contributions

- Privacy-Preserving Asynchronous Federated Learning Algorithm (PPAFL) for collaboration among healthcare institutions while preserving data privacy
- Adaptive learning strategy for accommodating heterogeneous health data and varying client capabilities
- Comprehensive evaluation methodology, including benchmarking against existing techniques, to demonstrate the algorithm's effectiveness and scalability

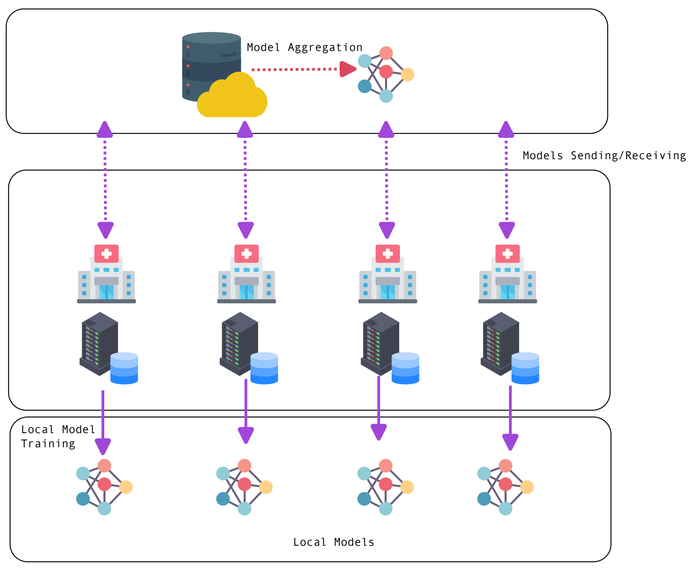

Figure 1. Caption

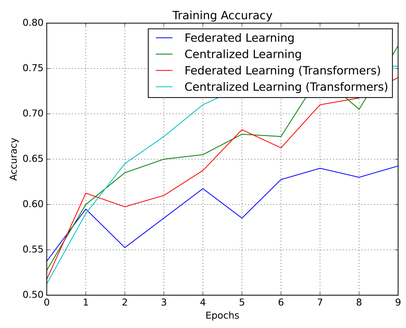

Figure 2. Training Accuracy for the different models experimented.

Experiments

- COVID-19 Chest X-ray and CT Scans dataset used for evaluation [5]
- PPAFL assessed for learning performance, privacy preservation, communication efficiency, and scalability
- Comparison with centralized and non-federated approaches
- Evaluation of adaptive learning strategy and computer networking techniques for reducing communication overhead and latency

Experimental Results

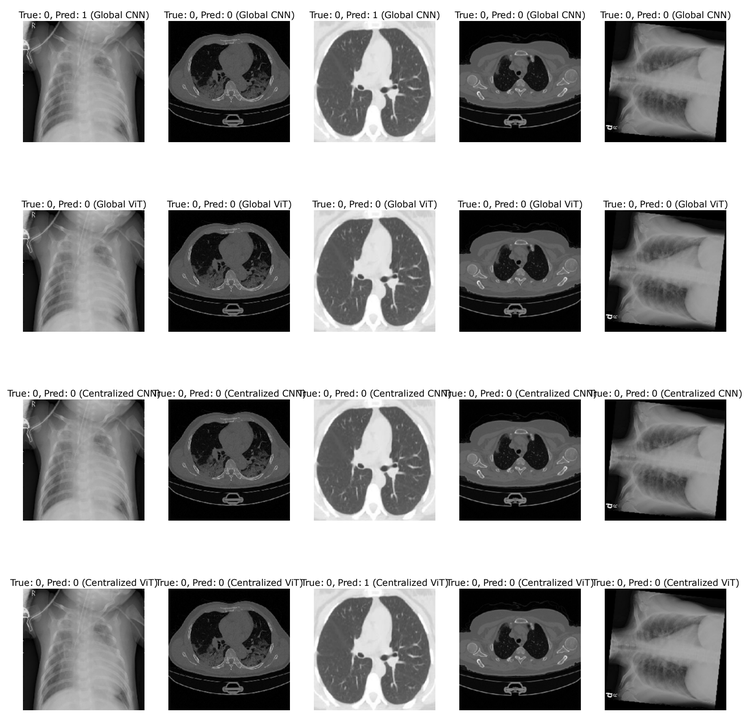

Figure 3. Sample Predictions generated from the model

Conclusions

- PPAFL has significant implications for personalized healthcare applications
- Bridges machine learning and computer networking to enable effective collaboration, data privacy preservation, and accommodation of personalized healthcare requirements
- Potential to revolutionize data-driven decision-making in healthcare, leading to improved patient outcomes and quality of care

Future Works

- Develop more advanced privacy-preserving techniques to further enhance data security during model aggregation in federated learning.
- Investigate new communication-efficient algorithms that reduce the overhead and latency of federated learning in healthcare applications.
- Design adaptive learning strategies that can better accommodate diverse health data and varying client capabilities, leading to more accurate and personalized models.
- Evaluate the proposed federated learning algorithms on a broader range of healthcare applications and datasets to validate their generalizability and effectiveness.
- Explore the integration of additional privacy-preserving techniques, such as homomorphic encryption and secure hardware, to strengthen the overall security of the federated learning framework.

References

[1] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschal idis, and W. Shi.
Federated learning of predictive models from federated electronic health records.
International Journal of Medical Informatics, 112:59–67, 2018.

[2] C. Dwork and A. Roth.
The algorithmic foundations of differential privacy.
Foundations and Trends® in Theoretical Computer Science, 9(3–4):211–407, 2014.

[3] H. B. McMahan, E. Moore, D. Ramage, and S. Hampson.
Communication-efficient learning of deep networks from decentralized data.
In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282, 2017.

[4] Z. Obermeyer and E. J. Emanuel.
Predicting the future - big data, machine learning, and clinical medicine.
New England Journal of Medicine, 375(13):1216–1219, 2016.

[5] E. Soares, P. Angelov, S. Biaso, M. Higa Froes, and D. Kanda Abe.
Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification.
medRxiv, 2020.

{lucas.pacheco, torsten.braun}@unibe.ch