

Cross-Silo Federated Learning for Radiomics

Enhancing Cross-Site Generalization Through Distributed Collaborative Training

Anthony Tucker

Howard University, anthony.tucker@bison.howard.edu

SAURAV KESHARI ARYAL

HCAI Institute, Howard University, saurav.aryal@howard.edu

Radiomics allows for the extraction of features from medical images for predictive modeling, but due to data privacy regulations, training models on multi-institutional datasets is not feasible. We propose a cross-silo federated learning framework that allows hospitals to jointly train radiomics models without sharing patient data.

Our solution utilizes the Flower framework and the radMLBench benchmark (50+ radiomics datasets) to create a multi-hospital federated learning setting. Each hospital is given unique datasets, as seen in real-world settings. We handle feature heterogeneity by using intersection-based feature alignment, making the model compatible across hospitals.

The framework utilizes a four-layer neural network architecture that is trained locally at each hospital. The central server combines the model parameters using Federated Averaging (FedAvg), weighted by the number of samples. We compare the federated learning method with a centralized approach that trains on aggregated data using accuracy as a metric. Although the centralized approach has the advantage of direct access to all the data, the federated approach is able to attain similar results while preserving the privacy of the data, thus proving that collaborative learning can come close to the centralized approach without breaching the privacy constraints. The framework is applied to brain-related disorders (gliomas and glioblastomas) using MRI images, thus illustrating how different institutions can work together to collaboratively enhance the performance of the model while being HIPAA compliant.

CCS CONCEPTS •Computing methodologies •Artificial intelligence •Distributed artificial intelligence • Multi-agent systems

Additional Keywords and Phrases: Federated Learning, Radiomics, Cross-Silo Learning, Privacy-Preserving Machine Learning, Medical Imaging

REFERENCES

- [1] Aydin Demircioglu. n.d. *radMLBench: A dataset collection for benchmarking in radiomics*. Retrieved January 16, 2025 from <https://github.com/aydindemircioglu/radMLBench/tree/main>