AI Assisted Annotation and Federated Learning: How to Standardize Data Labeling to Boots Training Performance

John Garrett, PhD, Director, University of Wisconsin, Madison
Iman Zare Estakhraji, PhD; Farzana Ali, PhD; Gian Marco Conte, MD, PhD; Prerna Dogra, PhD; Shahriar Faghani, MD; Mona Flores, MD; Reza Forghani, MD, PhD; Brad Generaux; Amilcare Gentili, MD; Kristopher Kersten, PhD; Meghan Lubner, MD; Andrew Missert, PhD; Spencer Workman, MD; Joseph Yacoub, MD; Khaled Younis, PhD; Yuankai Huo, PhD

Background/Problem Being Solved

Federated learning presents a promising approach to leveraging extensive datasets in medical imaging without central data collection, thus maintaining data privacy and reducing ePHI breach risks. This method involves simultaneous model training across multiple sites with only the network weights being shared. The focus of this project is on renal cell carcinoma (RCC) segmentation, where variability in data labeling across sites poses significant challenges. RCC is a pertinent choice due to its high occurrence and potential benign nature, necessitating accurate segmentation for subsequent diagnostic analyses.

Intervention(s)

To address labeling inconsistencies, AI-assisted annotation tools will be employed to standardize the data annotation process across different sites. This project utilizes a hybrid dataset comprising a public dataset (TCGA-KIRC) for training and a comprehensive dataset from UW Madison for testing. The intervention includes manual and AI-assisted annotations using open source tools like ITK Snap and 3D Slicer, followed by federated learning using open source libraries to evaluate model performance and data requirements.

Barriers/Challenges

The primary challenge is the inherent variability in training masks generated by different annotation tools and protocols used at various sites, potentially degrading model performance and increasing data requirements.

Outcome

The project aims to demonstrate whether AI-assisted annotations can enhance the efficiency and consistency of data labeling in federated learning setups, thus potentially improving model performance and reducing the number of samples needed for effective training.

Conclusion/Statement of Impact/Lessons Learned

By integrating AI-assisted annotation within a federated learning framework, this initiative expects to set a benchmark for improved segmentation accuracy and operational efficiency in medical imaging. Successful outcomes will provide crucial

insights into optimizing deep learning tasks across diverse clinical environments, offering a scalable model for future multi-site medical imaging projects. Results and methodologies will be shared publicly to aid further research and development in this domain.

Figure(s)

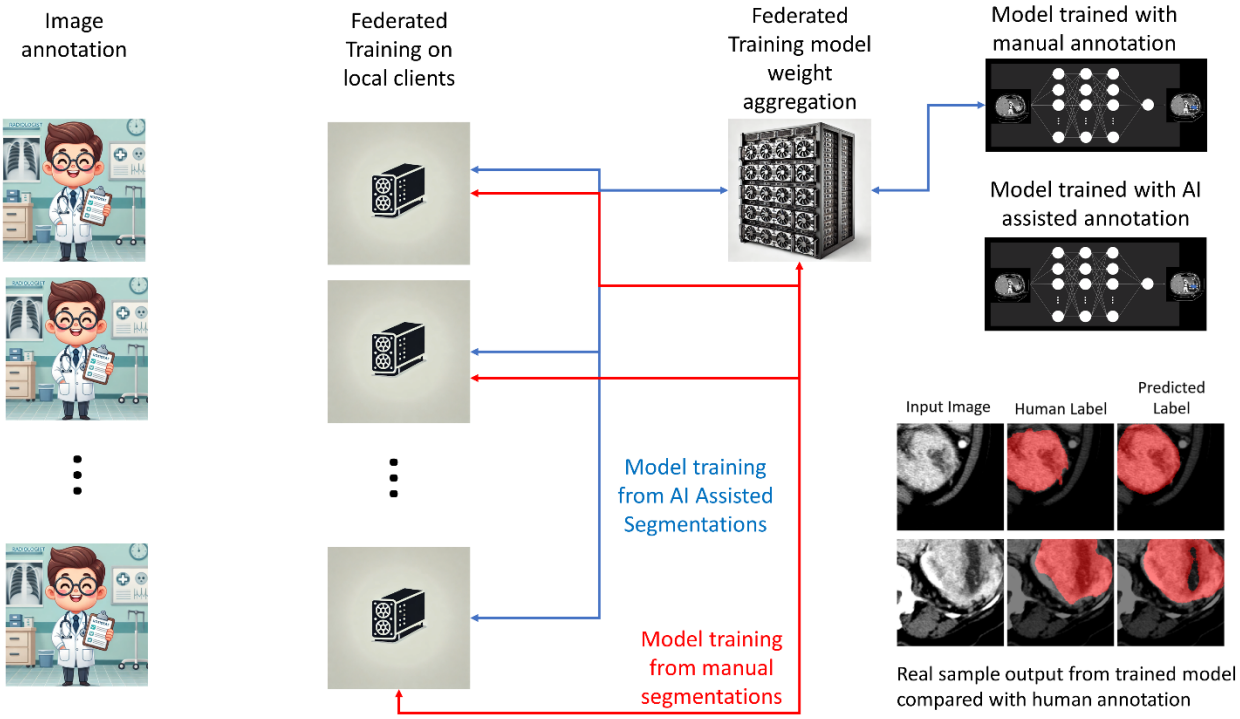

Figure 1. Workflow used in the study and described in the abstract for AI Assisted annotation and federated learning.

Keywords

Artificial Intelligence/Machine Learning; Imaging Research