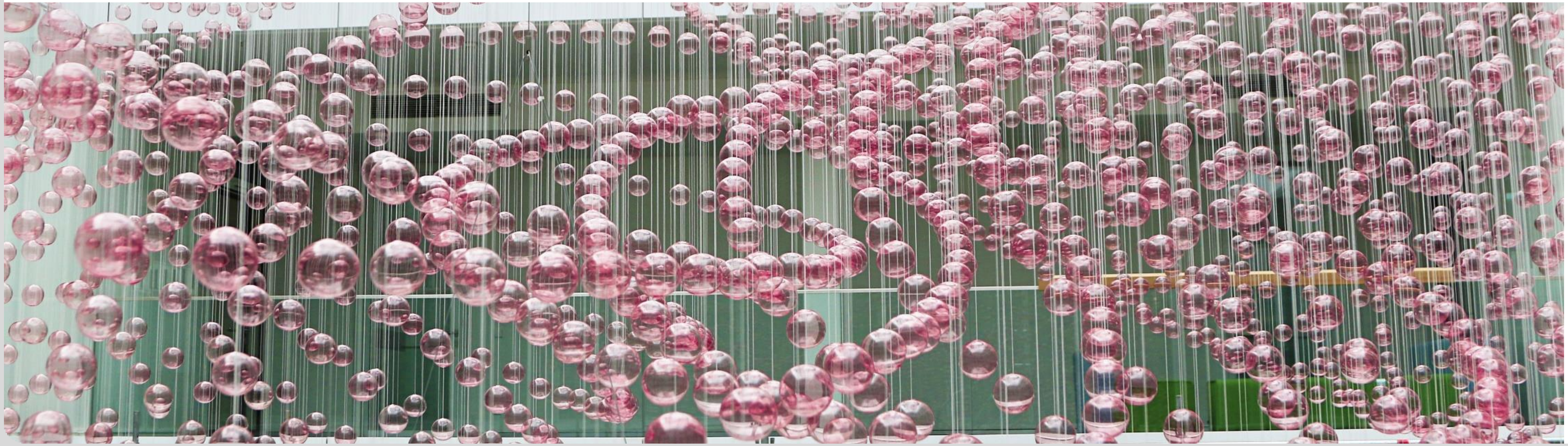
Federated Learning on Medical Image Data: Closing Gaps to the Real-World

Master’s Thesis Presentation by Marius Kempf
Supervisors: Klaus Kades (DKFZ), Maximilian Zenk (DKFZ), Niklas Kühl

KARLSRUHE SERVICE RESEARCH INSTITUTE (KSRI)
INSTITUTE OF INFORMATION SYSTEMS AND MARKETING (IISM)

GERMAN CANCER RESEARCH CENTER (DKFZ)
Research Group: MEDICAL IMAGE COMPUTING (MIC)



Motivation: Medical (image) data is rare, highly sensitive and precious – accordingly, it is the main bottle-neck for deep learning applications.

DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

Karlsruhe Institute of Technology

From Machine Learning to Deep Learning

- Machine Learning learns from samples using various types of models (i.e. Neural Networks)
- Deep Learning is a part of ML methods using Convolutional Neural Networks (CNNs)
- Typical tasks with image data: Object Detection, Object Segmentation, …

Need for (medical image) data

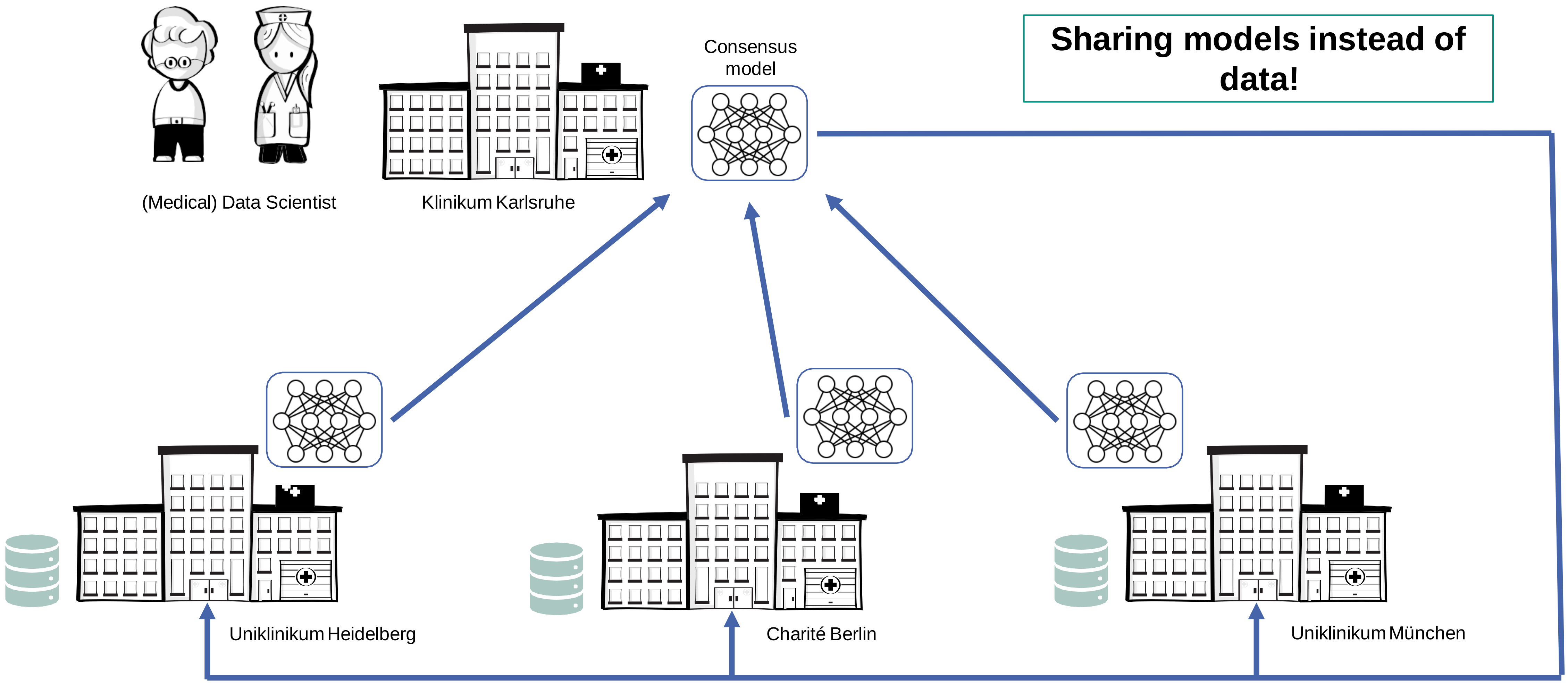
- Huge amounts of (training) samples needed
- Medical data is created and stored everyday, but…
  - Labelling medical data requires experts (in particular, segmentation-labels are expensive)
  - Medical (image) data is highly private and therefore sensitive and strictly regulated

Motivation: When applying Federated Learning (FL), models are exchanged between participants – locally available data stays local.

Sharing models instead of data!

McMahan, HB.,Moore, E. et al. «Communication-efficient learning of deep networks from decentralized data» 2017

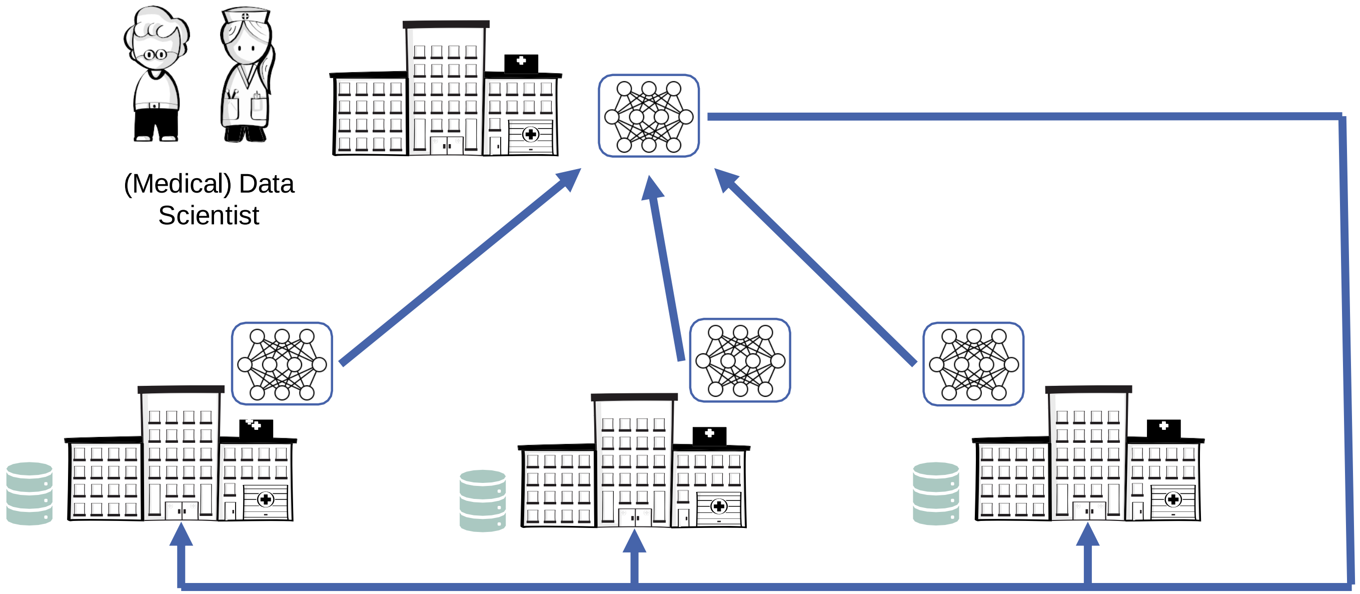
Motivation: Federated Learning (FL) allows the “decoupling of model training from the need for direct access to the raw training data.”



Motivation: Federated Learning (FL) allows the "decoupling of model training from the need for direct access to the raw training data."

Federated Learning

- Sharing models instead of data
- Patients' **data remains local & private**
- Avoidance of extensive legal agreements
- "Access" to much more samples which can be used for model training
    - Better performance
    - Higher generalizability of models



McMahan, HB.,Moore, E. et al. «Communication-efficient learning of deep networks from decentralized data» 2017
Rieke, N., Hancox, J., Li, W. et al. «The future of digital health wit federated learning» 2020

Related Work: Federated Learning has been shown to be working – however, it is used little for real-world applications with medical image data.

Reference	FL Setting	FL Solution	Participants	Compute Plan	Algorithm	Task
Xu et al. (2020)	real-world	custom solution	4	FedAvg	3D-Densenet	Classification
Balachandar et al. (2020)	simulated	-	4	Sequential	GoogleLeNet	Classification
Wang et al. (2020)	real-world	NVIDIA Clara Train	2	FedAvg	C2FNAS	Segmentation
Remedios et al. (2020)	real-world	custom solution	2	Sequential	U-Net	Segmentation
Remedios et al. (2019)	real-world	custom solution	2	Sequential	CNN (Incept.)	Segmentation
Chang et al. (2018)	simulated	-	4	Sequential & Ensembling	ResNet34	Classification
Kaissis et al. (2021)	synthetic	PriMIA (based on PySyft)	3	FedAvg	ResNet18	Classification
Dou et al. (2021)	synthetic	custom solution	3	FedAvg	CNN	Segmentation
Roth et al. (2020)	real-world	NVIDIA Clara Train	7	FedAvg	DenseNet-121	Classification
Feki et al. (2021)	simulated	-	4	FedAvg	VGG16 & ResNet50	Classification
Sarma et al. (2021)	real-world	NVIDIA Clara Train	3	FedAvg	3D AH Net	Segmentation
Sheller et al. (2020)	simulated	-	10	FedAvg & Sequential	U-Net	Segmentation
Baheti et al. (2020)	simulated	-	3	FedAvg	V-Net	Classification
Yang et al. (2021)	synthetic	NVIDIA Clara Train	3	FedAvg	3D U-Net	Segmentation
Sheller et al. (2019)	simulated	-	4, 6, 8, 32	FedAvg	U-Net	Segmentation
Li et al. (2019b)	synthetic	NVIDIA Clara Train	13	FedAvg	CNN	Segmentation
Andreux et al. (2020)	simulated	-	2, 5	FedAvg	CNN	Segmentation
Yan et al. (2020)	simulated	-	2, 4, 8	FedAvg	CNN	Classification
Lee et al. (2021)	real-world	PySyft	6	FedAvg	multiple	Classification
Flores et al. (2021)*	real-world	NVIDIA Clara Train	20	FedAvg	ResNet34	Classification

Digital Service Innovation
Karlsruhe Service Research Institute
Institute of Information Systems and Marketing

Related Work: Federated Learning has been shown to be working – however, it is used little for real-world applications with medical image data.

- Total: 20 articles
- Real-world: 8 articles
- Applied FL solutions:
  - NVIDIA Clara (6)
  - PySyft (2)
  - Custom solution (4)

Technical Solutions are required to bring FL into actual medical institutions!

Reference	FL Setting	FL Solution	Participants	Compute Plan	Algorithm	Task
Xu et al. (2020)	real-world	custom solution	4	FedAvg	3D-Densenet	Classification
Balachandar et al. (2020)	simulated	-	4	Sequential	GoogleLeNet	Classification
Wang et al. (2020)	real-world	NVIDIA Clara Train	2	FedAvg	C2FNAS	Segmentation
Remedios et al. (2020)	real-world	custom solution	2	Sequential	U-Net	Segmentation
Remedios et al. (2019)	real-world	custom solution	2	Sequential	CNN (Incept.)	Segmentation
Chang et al. (2018)	simulated	-	4	Sequential & Ensembling	ResNet34	Classification
Kaissis et al. (2021)	synthetic	PriMIA (based on PySyft)	3	FedAvg	ResNet18	Classification
Dou et al. (2021)	synthetic	custom solution	3	FedAvg	CNN	Segmentation
Roth et al. (2020)	real-world	NVIDIA Clara Train	7	FedAvg	DenseNet-121	Classification
Feki et al. (2021)	simulated	-	4	FedAvg	VGG16 & ResNet50	Classification
Sarma et al. (2021)	real-world	NVIDIA Clara Train	3	FedAvg	3D AH Net	Segmentation
Sheller et al. (2020)	simulated	-	10	FedAvg & Sequential	U-Net	Segmentation
Baheti et al. (2020)	simulated	-	3	FedAvg	V-Net	Classification
Yang et al. (2021)	synthetic	NVIDIA Clara Train	3	FedAvg	3D U-Net	Segmentation
Sheller et al. (2019)	simulated	-	4, 6, 8, 32	FedAvg	U-Net	Segmentation
Li et al. (2019b)	synthetic	NVIDIA Clara Train	13	FedAvg	CNN	Segmentation
Andreux et al. (2020)	simulated	-	2, 5	FedAvg	CNN	Segmentation
Yan et al. (2020)	simulated	-	2, 4, 8	FedAvg	CNN	Classification
Lee et al. (2021)	real-world	PySyft	6	FedAvg	multiple	Classification
Flores et al. (2021)*	real-world	NVIDIA Clara Train	20	FedAvg	ResNet34	Classification

Search: „distributed learning" OR „federated learning" on EMBASE and PubMed, until May 31, 2021

Methods: The application of FL in a medical environment comes with specific requirements that should be met by the technical FL solutions.

DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

Karlsruhe Institute of Technology

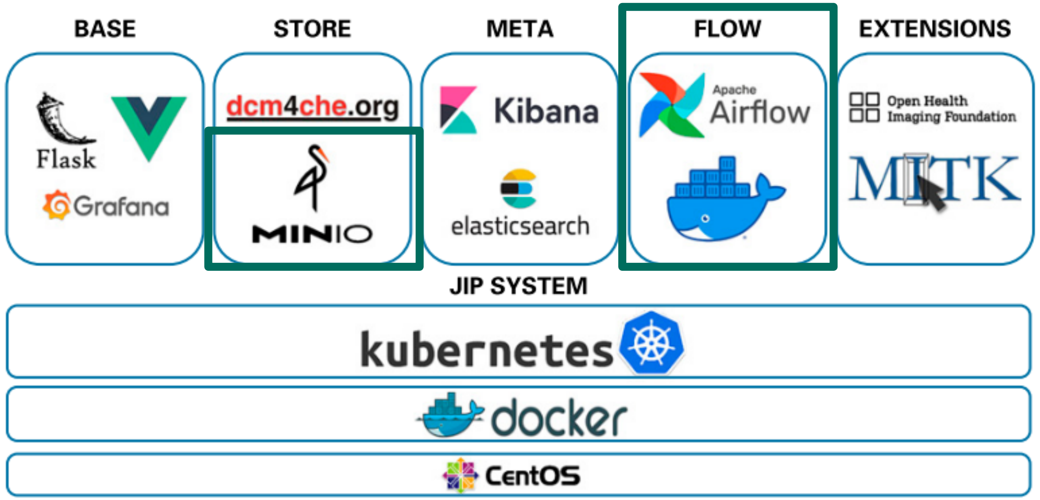
Technical Federated Learning Solutions

- Selection of the solutions to be investigated:
  - Used or compared in previous scientific projects or reviews
  - Significant number of GitHub stars & aactive community with current contributions
- 6 FL solutions were selected and are compared (NVIDIA Clara, TFF, FedML, FATE, PaddleFL, PySyft)

Requirements

- Derived from related work in the field of FL with medical images
- Interviews & talks with three domain experts of the field of
  - Medical Image Computation
  - Federated Learning

Methods: JIP Federated is based on the Joint Imaging Platform which comes with powerful open-source technologies such as Apache Airflow.



Joint Imaging Platform

- Platform solution for cross-institutional collaboration in the field od medical image analysis
- Technological backbone is developed with open-source tools

Apache Airflow

- Scheduling & monitoring of workflows
- Accessible via its powerful API
- Workflows configured as Directed Acyclic Graphs (DAGs) consisting of multiple operators
- Each operator performs exactly one task of the workflow
- Exemplary workflow (DAG):

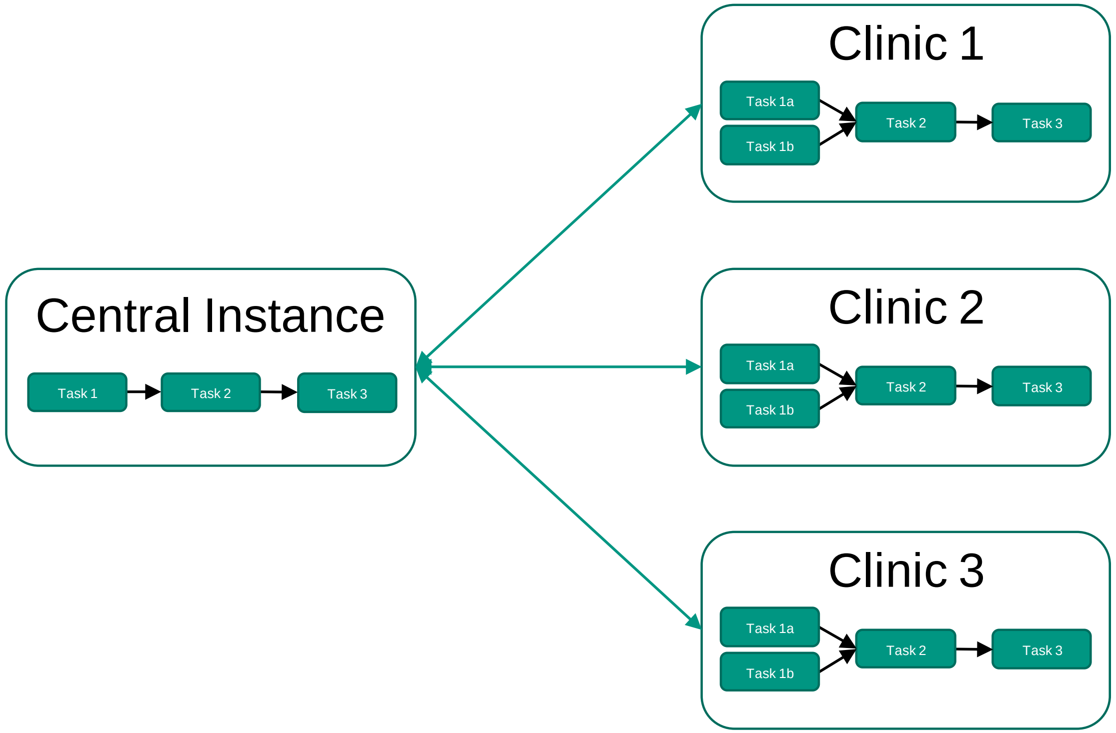
Scherer, J. et al «Joint Imaging Platform for Federated Clinical Data Analytics» 2020

JIP Federated as part of KAAPANA on GitHub: https://github.com/kaapana/kaapana

Methods: JIP Federated consists of two DAGs (workflows) serving as workflow on a central instance and on each participating JIP instance.

JIP Federated

- One central scheduling DAG on the central
  - Organises overall procedure
  - Aggregates model updates
- Equal DAGs on each participating institution
  - Performs model training on locally available data
  - Loads/sends model from/to central instance
- All but two experiment-specific components (operators) are fully reusable



Methods: An experiment using JIP Federated for federated brain tumor segmentation serves as use-case to show the applicability of the developed solution.

Experiment Setting

- Central instance & 3 JIP instances (each hosting the same amount of data samples)
- 3D U-Net Architecture
  - 500 federated rounds (one epoch per round)
  - Adam optimizer (reset before each new federated round) – Federated Averaging
- Comparison of training behavior using centralized/pooled image data

Image Data - BraTS

- MRI brain scans containing diagnoses gliomas
- Randomly split into 3 three subsets for training & testing

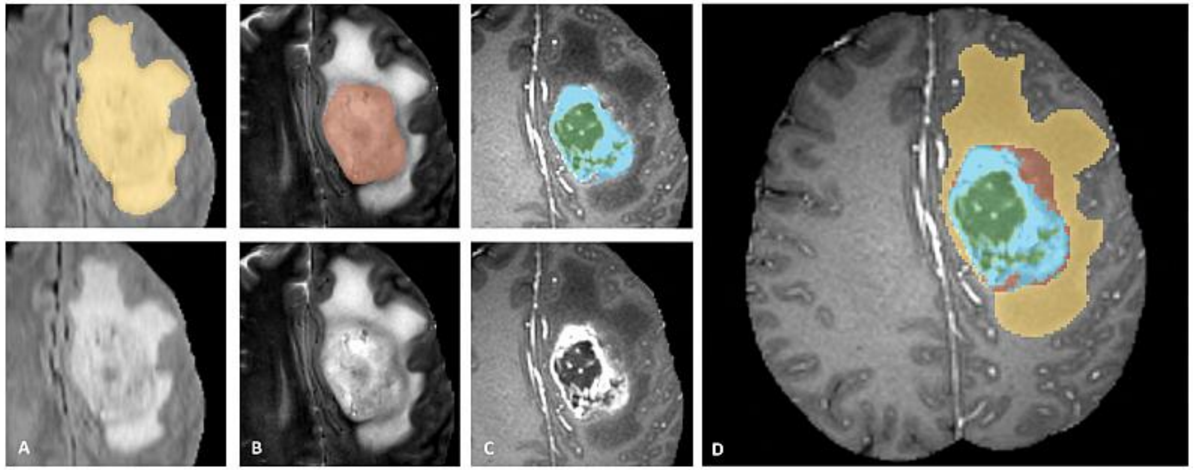


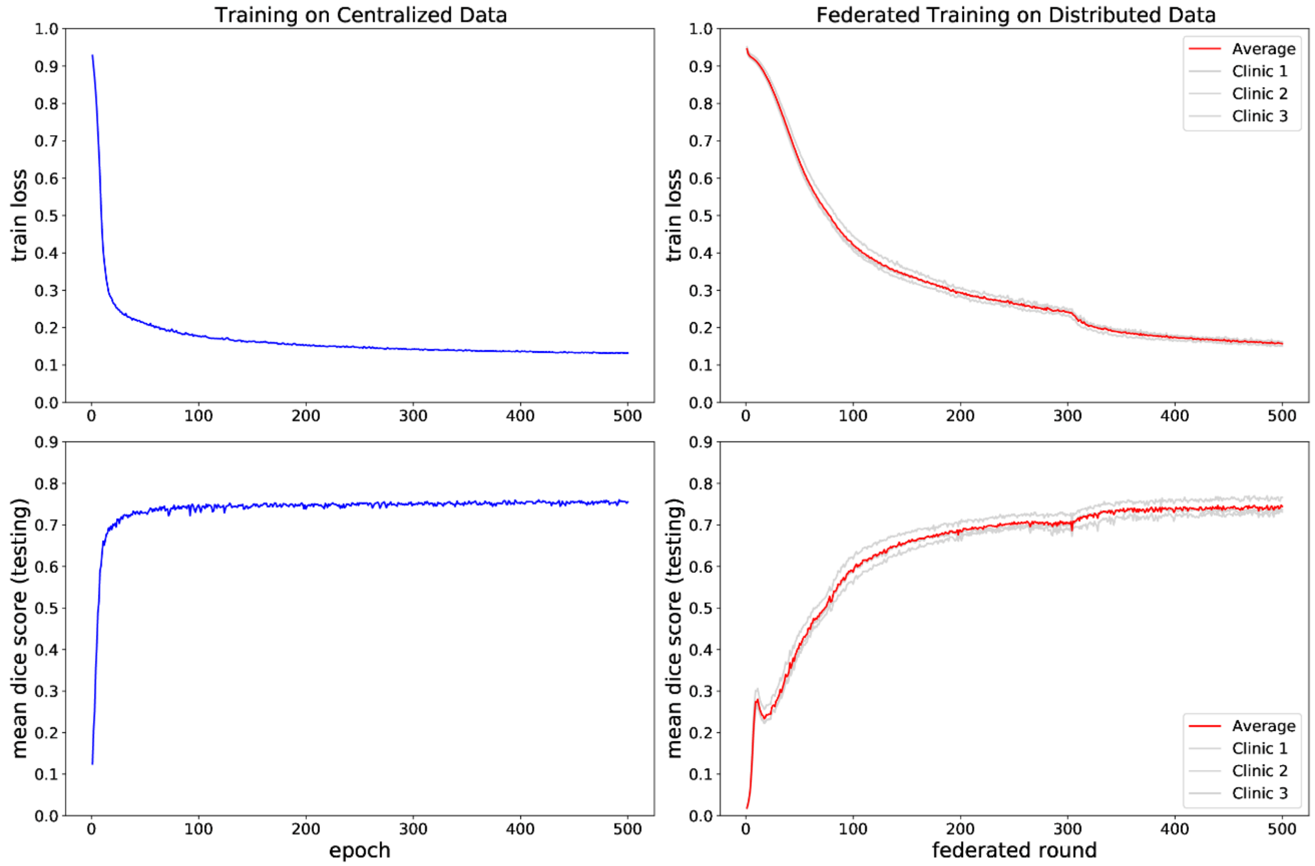
Figure from Menze, B. et al. «The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)» 2015

Results: Existing solutions do not yet support all requirements for their application in a medical environment. JIP Federated still lacks additional privacy mechanisms.

- 9 requirements were derived
- NVIDIA Clara & JIP Federated provide extensive platform solutions
- NVIDIA Clara Train does not provide an open code base
- JIP Federated lacks of additional privacy mechanisms

Requirement		TFF	FedML	FATE	PaddleFL	PySyft	NVIDIA Clara Train	JIP Federated
Open-source		✓	✓	✓	✓	✓	✗	✓
Offline installation		✗	✗	✗	✗	✗	✓	✓
Type of solution	Platform Solution						✓	✓
	Programming Framework	✓	✓	✓	✓	✓		
GPU support	Single GPU	✓	✓	-	✓	✗	✓	✓
	Multiple GPUs	✓	✓	-	✓	✗	✓	✓
Data handling	Non-DICOM	✓	✓	✓	✓	✓	✓	✓
	DICOM	✗	✗	✗	✗	(✓)	✓	✓
	PACS Connectivity	✗	✗	✗	✗	✗	✓	✓
Privacy mechanisms	Differential Privacy	(✓)	✗	✗	✓	✓	✓	✗
	Homomorphic Encryption	✗	✗	✓	✗	✓	✓	✗
	Secure Multi-Party Computation	✗	✓	✓	✓	✓	✗	✗
DL frameworks	PyTorch	✗	✓	✓	✗	✓	✓	✓
	TensorFlow	✓	✗	✓	✗	✓	✓	✓
	PaddlePaddle	✗	✗	✗	✓	✗	-	✓
Compute plans	Aggregation	✓	✓	✓	✓	✓	✓	✓
	Sequential	-	-	✓	✗	✓	✓	✓
Information	Documentation, Tutorials, & Examples	✓	✓	✓	✓	✗	✓	✓
	Community	✓	✓	✓	✗	✓	✓	✓

Results: While convergence is slower, FL with JIP Federated achieves comparable performance. JIP Federated can be a solution for real-world FL.



Mean dice score: whole tumor (WT), enhancing tumor, tumor core (TC)

- Convergence is slower in FL setting
- Similar performance after 500 rounds
  - Centralized data: 0.7554
  - FL: 0.7447 (98.58 %)
- Shapes show similar behavior across participants

The segmentation experiments shows and proves the functionality of JIP Federated

Conclusion:
JIP Federated closes some gaps and can be the basis for future

DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

Karlsruhe Institute of Technology

Summary

- FL can enable science to „access“ more data
- FL in the medical environment needs technical solutions that fulfill specific requirements
- JIP Federated was successfully demonstrated by conducting a segmentation experiment

Limitations

- Scope of derived requirements
- Privacy mechanisms (i.e. Differential Privacy, Secure multi-party computation)
- Synthetic setting used for segmentation experiment

Contributions

- Overview of FL solutions in the healthcare environment, by comparing existing solutions based on requirements derived from the medical image domain
- Open-source solution JIP Federated to conduct real-world FL on medical images in a clinical environment

Thesis: “Federated Learning on Medical Image Data: Closing Gaps to the Real-World”

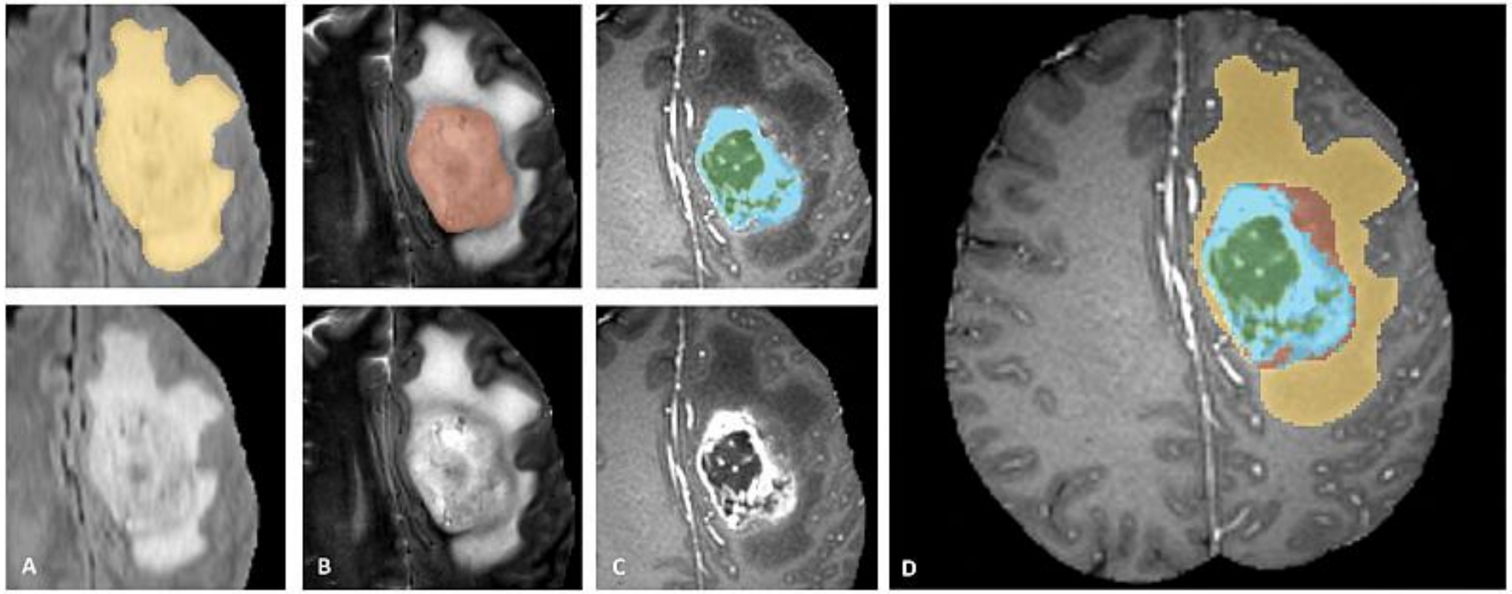
By Marius Kempf

THANK YOU !

BACKUP

BraTS – Glioma segmentation details

- A: Whole Tumor (WT)
- B: Tumor Core (TC)
- C: Tumor Core distinguished into
  - Blue: Enhancing Tumor
  - Green: cystic/necrotic components

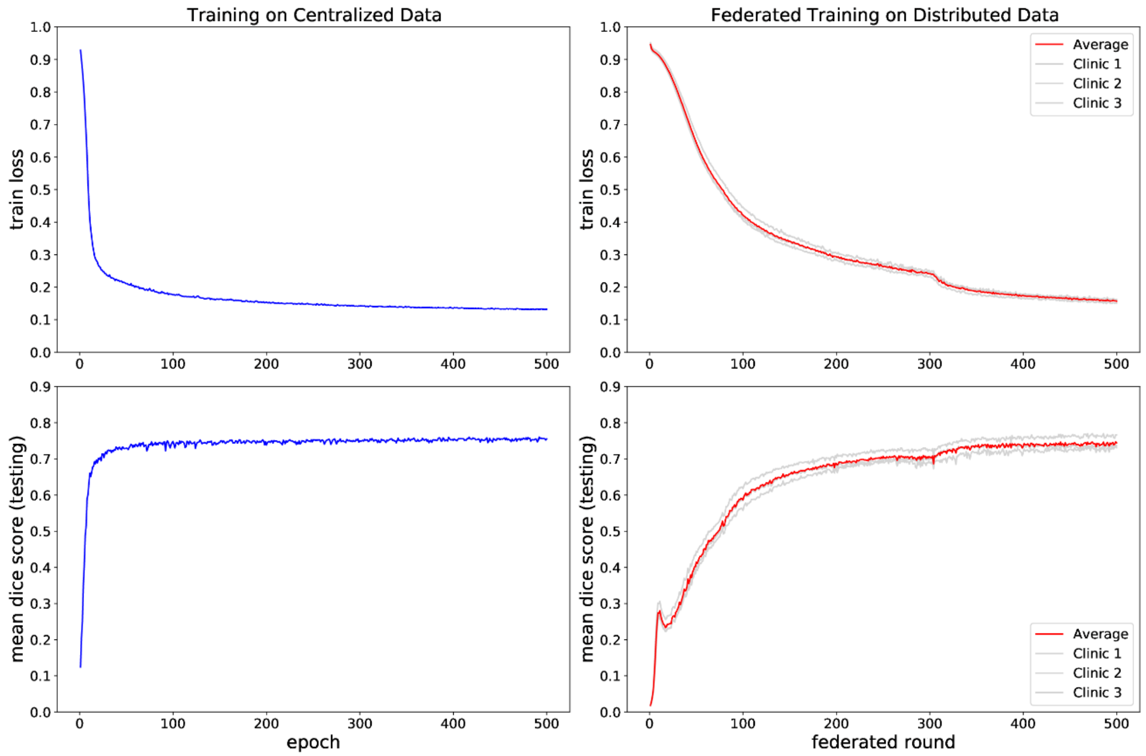


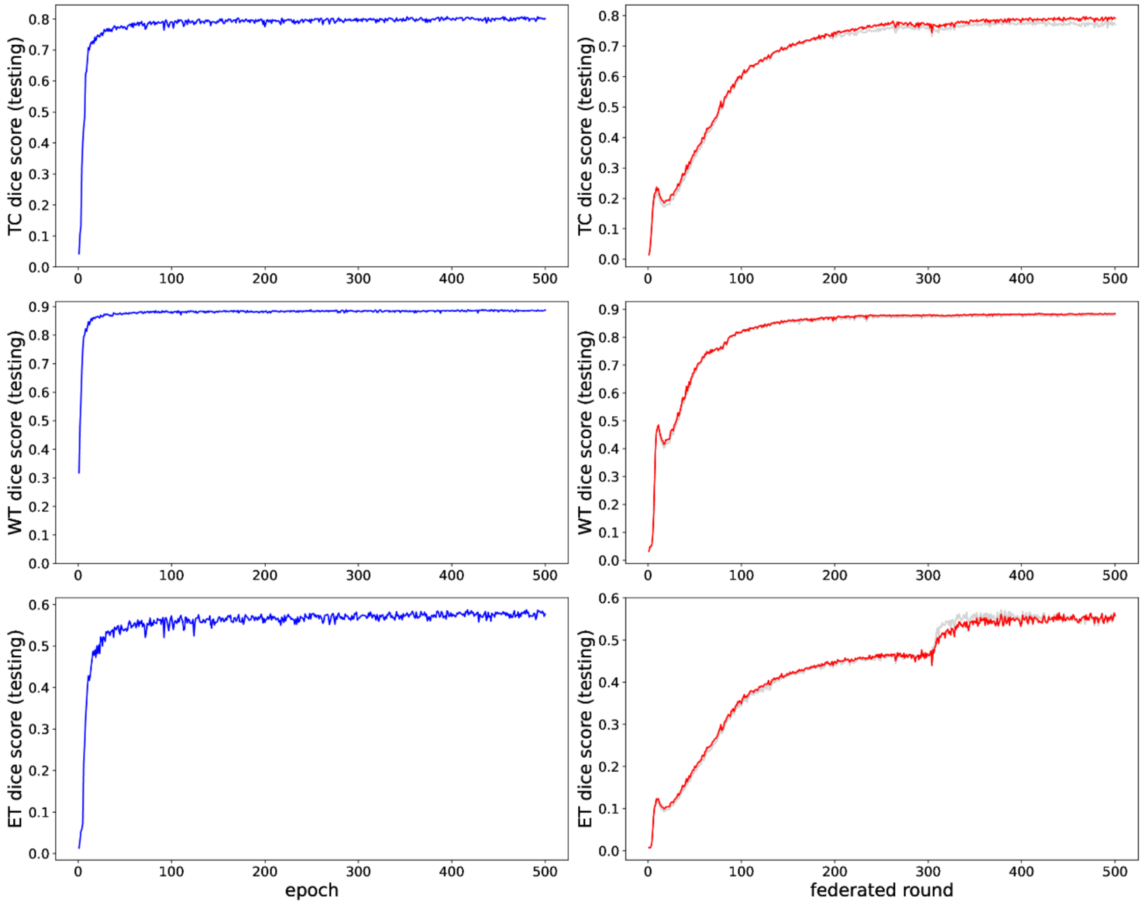
A, B, and C also show different image modalities used (FLAIR, T2, T1c)

Data Distribution

- …
- …

BraTS Experiment – Training Behavior





ET: Enhancing Tumor
WT: Whole Tumor
TC: Tumor Core

Implemented JIP Federated DAGs