CREATIS
LIRIS

Whole-brain radiomics for clustered federated personalization in brain tumor segmentation

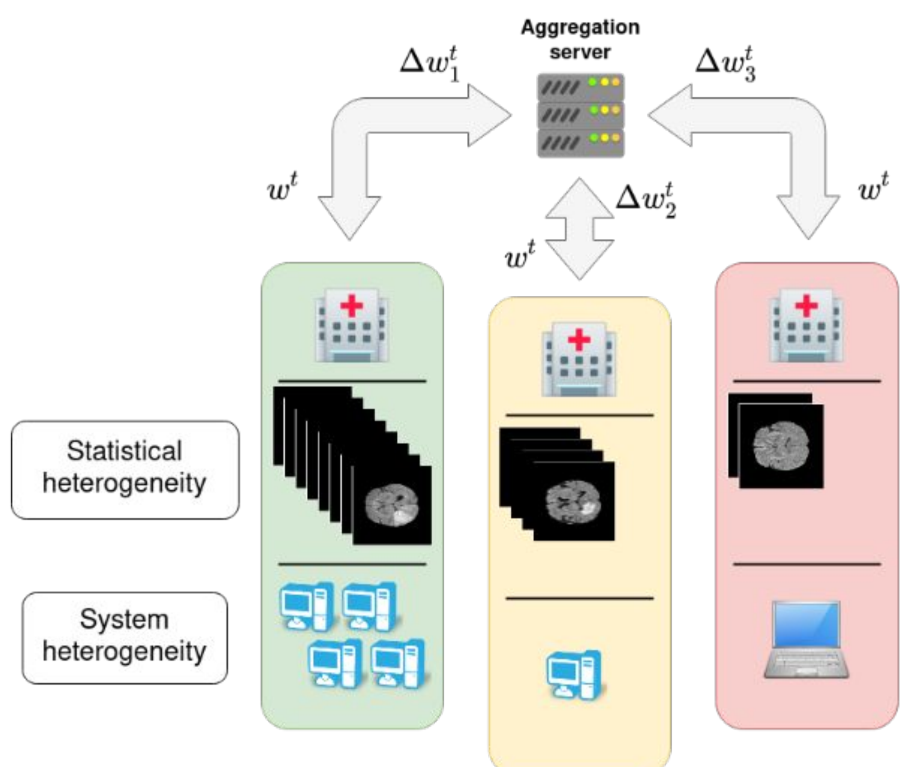
Matthis Manthe[1-2], *Stefan Duffner*[2], *Carole Lartizien*[1]

[1]**CREATIS**, *CNRS (UMR 5220), INSERM (U1294), INSA Lyon, Université de Lyon,*
[2]**LIRIS**, *CNRS (UMR 5205), INSA Lyon, Université Claude Bernard Lyon 1, Université Lumière Lyon 2, Ecole Centrale Lyon*
matthis.manthe@insa-lyon.fr

UNIVERSITÉ DE LYON
INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Motivation

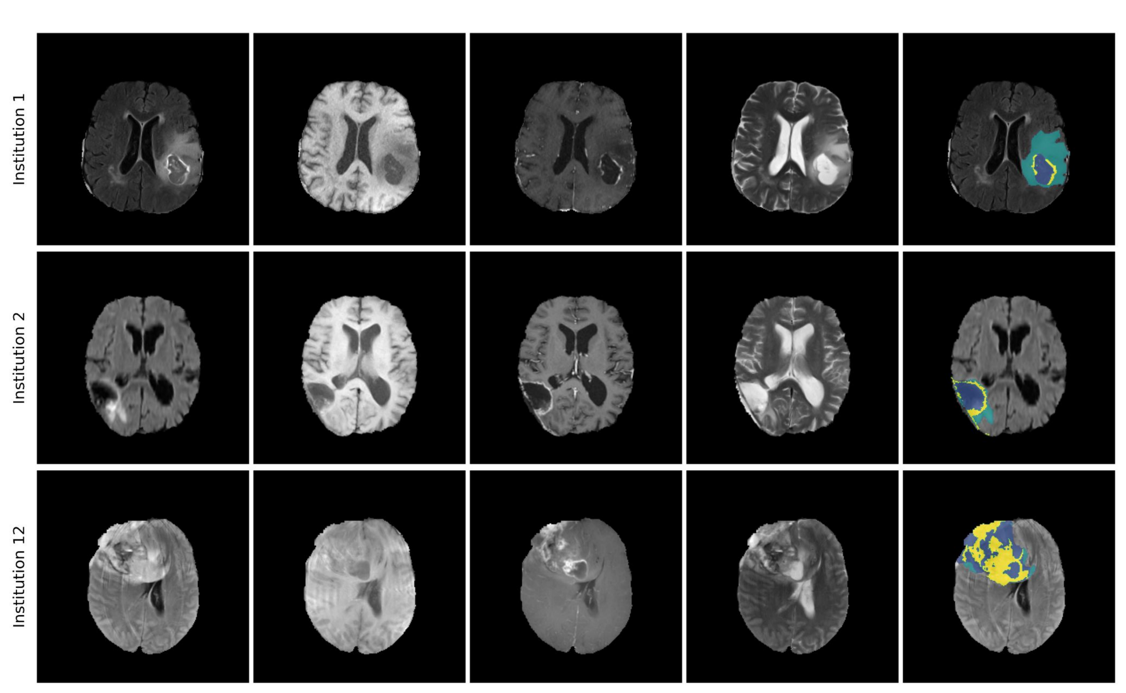
- Federated learning (FL) enables to efficiently train deep neural models on large decentralized datasets without sharing raw data between health institutions,



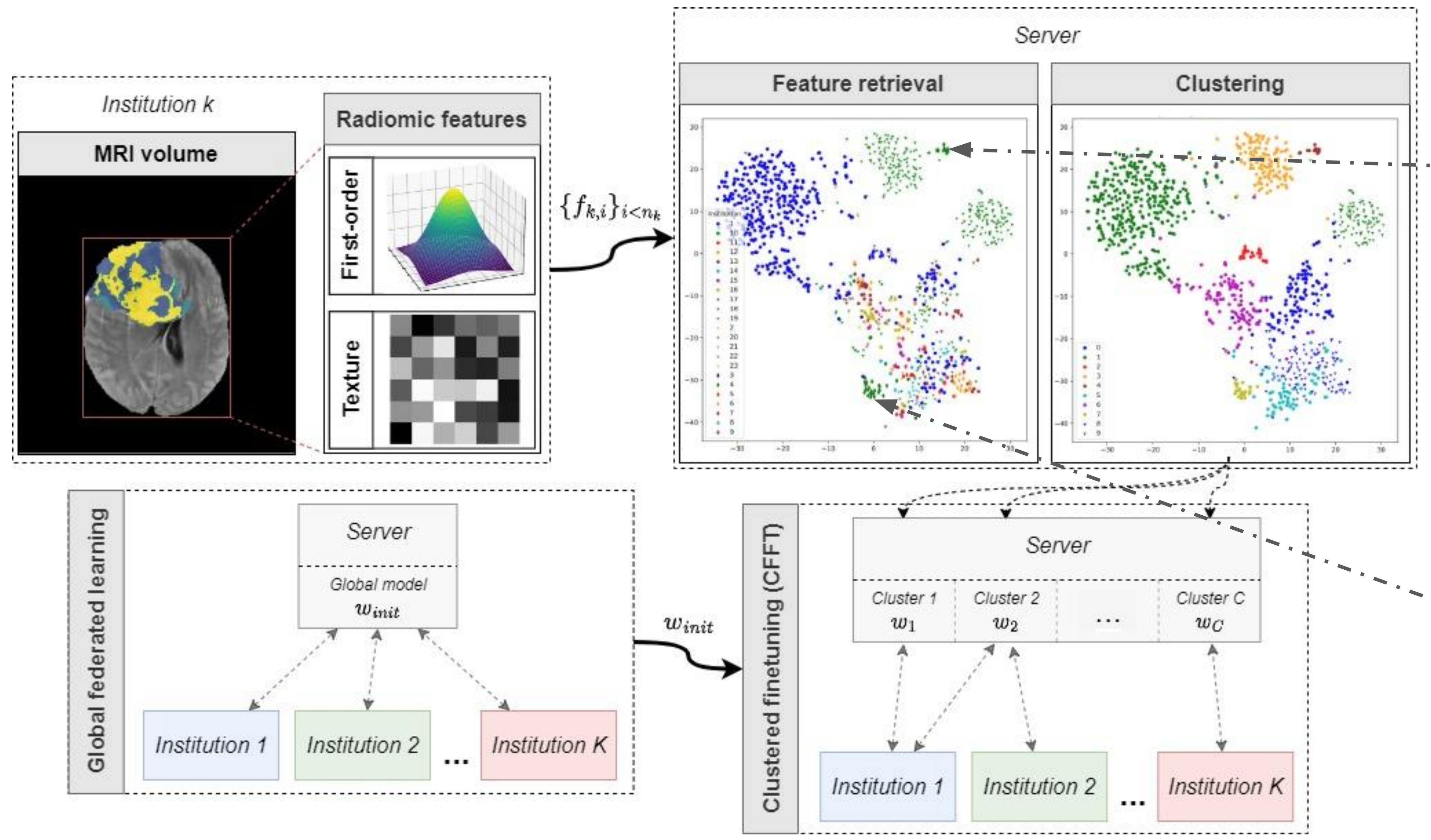
- It underperforms with statistically heterogeneous distributed data,
- Personalization methods mostly assume local datasets homogeneity, hardly true in realistic cases,
- Clustered FL is particularly interesting in case of limited amount of data per institution.

Brain tumor segmentation

- **Dataset :** Federated Brain Tumor Segmentation 2022 (**FeTS2022**[10]),
- **Input :** 1251 3D *T1*, *T1ce*, *T2*, *FLAIR* brain MRIs.
- **Task :** Multi-label brain tumor segmentation : Enhancing tumor (***ET***), Tumor core (***TC***) and Whole tumor (***WT***),
- **Partitioning :** original 23 clinical institutions (*~61% with less than 15 samples*),
- **Model :** 3D U-Net *(~1.4M parameters)*.



Clustered Federated Finetuning (CFFT)



- Each institution extracts ***intensity*** (median, energy, …) and ***texture*** (GLCM, GLRLM, …) metrics from each of its volumes : **93 features** per modality and volume.
- The server performs standard clustering on the features of every institution : **feature normalization ⇨ PCA ⇨ GMM**
- After several rounds of standard FL (such as FedAvg), the server performs federated trainings in parallel on the C clusters of decentralized data, based on the following **sample-level clustered federated objective with C clusters**

$$\{w_c^*\}_{c=1}^C = \underset{\{w_c\}_{c=1}^C \in \mathcal{W}^C}{argmin} \sum_{k=1}^K \sum_{i=1}^{n_k} l(w_{Cluster(\hat{f}_{k,i})}, x_{k,i}, y_{k,i})$$

Segmentation performance

Dice

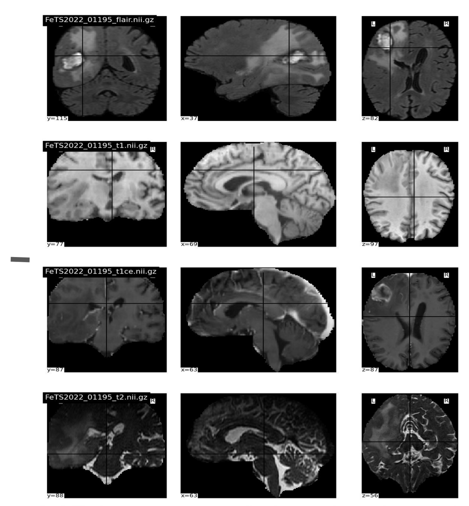
Training algorithm	Average	TC	WT	ET
Centralized	0.8912 ± 0.1201	0.8896 ± 0.1878	0.9188 ± 0.0859	0.8651 ± 0.1956
FedAvg	0.8803 ± 0.1414	0.8722 ± 0.2251	0.9099 ± 0.0906	0.8588 ± 0.2005
Local finetuning	0.8879 ± 0.1202	0.8876 ± 0.1863	0.9132 ± 0.0891	0.8629 ± 0.1960
CFFT_{ideal}	0.8887 ± 0.1239	0.8867 ± 0.1891	0.9139 ± 0.0892	0.8654 ± 0.1960
CFFT	0.8874 ± 0.1254	0.8799 ± 0.2056	0.9120 ± 0.0907	0.8704 ± 0.1852

95% Hausdorff distance

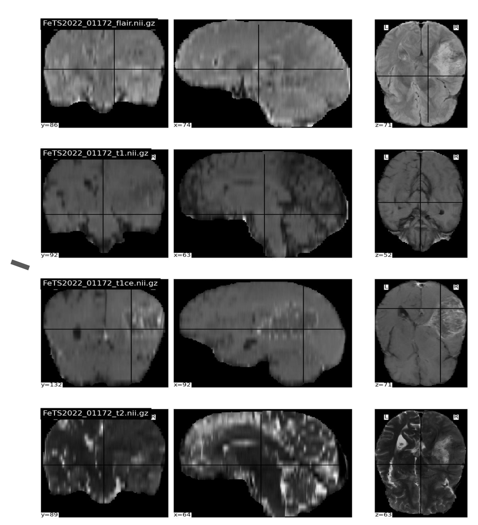
Training algorithm	Average	TC	WT	ET
Centralized	5.1348 ± 6.2212	4.4407 ± 7.2052	7.0318 ± 11.0671	3.7770 ± 8.5314
FedAvg	5.8854 ± 7.7368	4.7745 ± 8.0444	8.8814 ± 15.9499	3.7091 ± 7.6837
Local finetuning	5.9334 ± 7.9817	4.8587 ± 9.3901	8.9284 ± 16.3270	3.5861 ± 7.5650
CFFT_{ideal}	5.5915 ± 7.2588	4.6436 ± 7.3390	8.3688 ± 15.3738	3.5383 ± 7.5576
CFFT	5.8749 ± 7.6108	4.7359 ± 7.4458	8.8709 ± 16.4791	3.6520 ± 7.6291

CFFT matches the performance of other personalization methods while focusing only on the feature shift.

Clustering analysis

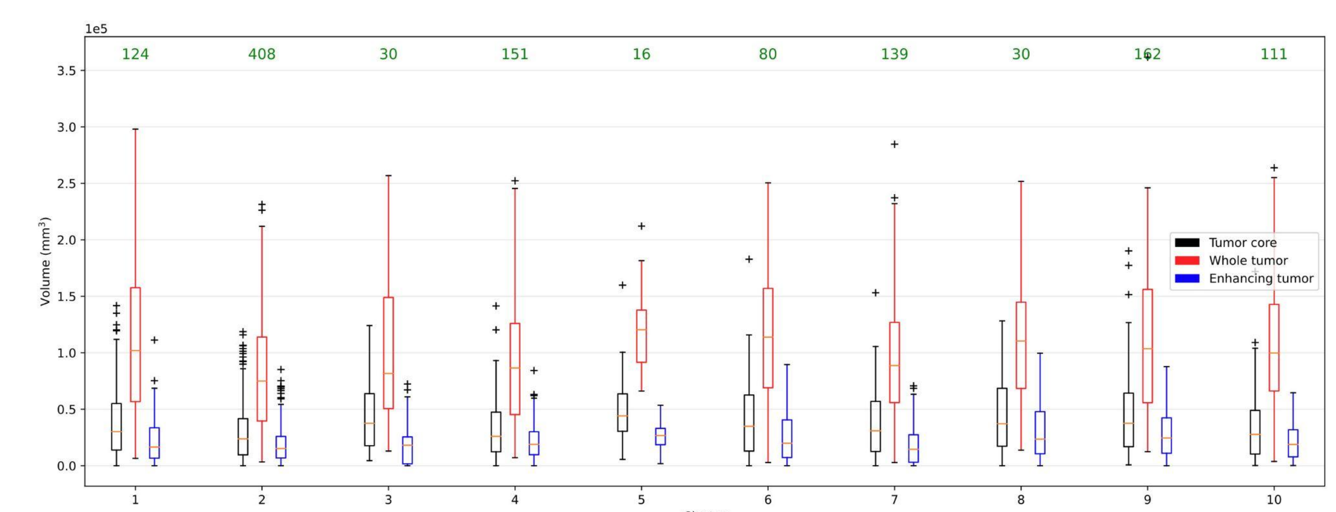


MRI exam of institution 4 assigned to cluster 4



MRI exam of institution 4 assigned to cluster 7

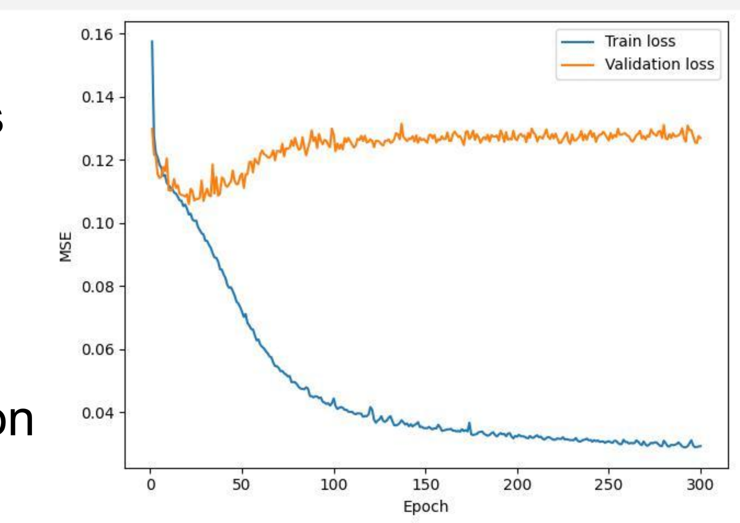
- Volumes from a single institution can have very different appearance.
- They are placed in different clusters, with volumes of other institutions of similar appearance,
- Label distributions are a lot more alike between clusters than between institutions.

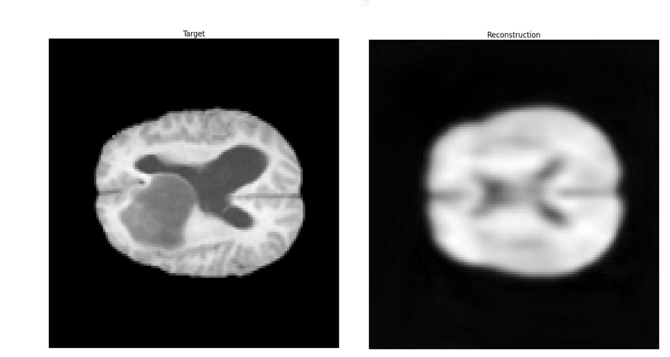


Label distribution per computed cluster

Basic privacy validation

- Sample-level federated clustering requires communicating features for each sample, risking sensitive information leakage.
- **Attack setup** : train a neural network to reconstruct a subsampled volume based on its radiomic features.
- **Results** : Quick overfitting, feature vectors contain only global textural information, tumorous parts are completely withdrawn from the reconstructions.





Conclusion and future work

- The standard personalized federated learning assumption of local datasets homogeneity is too idealistic,
- The isolation of the feature shift in radiomic embeddings and clustered federated finetuning do improve the segmentation performance compared to FedAvg,
- We will explore more complex and rich embeddings to further improve the convergence of federated algorithms on each cluster.

Acknowledgments

This work has been partially supported by the Agence Nationale de la Recherche under grant ANR-20-THIA-0007 (IADoc@UdL). This work was granted access to the HPC resources of IDRIS under the allocation 2022-AD011013327 made by GENCI