Performance comparison between federated and centralized learning with a deep learning model on Hoechst stained images

Damien ALOUGES[1,2], Georg WÖLFLEIN[3], In Hwa UM[4], David HARRISON[4,5], Ognjen ARANDJELOVIĆ[3], Christophe BATTAIL[1] and Stéphane GAZUT[2]

[1] Université Grenoble Alpes, IRIG, Laboratoire Biosciences et Bioingénierie pour la Santé, UA 13 INSERM-CEA-UGA, 38000 Grenoble, France
[2] University of Paris-Saclay, CEA, List ; F-91120, Palaiseau, France
[3] School of Computer Science, University of St Andrews, North Haugh, St Andrews KY16 9SX, Scotland, UK
[4] School of Medicine, University of St Andrews, North Haugh, St Andrews KY16 9TF, Scotland, UK
[5] Division of Laboratory Medicine, Lothian NHS University Hospitals, Edinburgh EH16 6SA, Scotland, UK

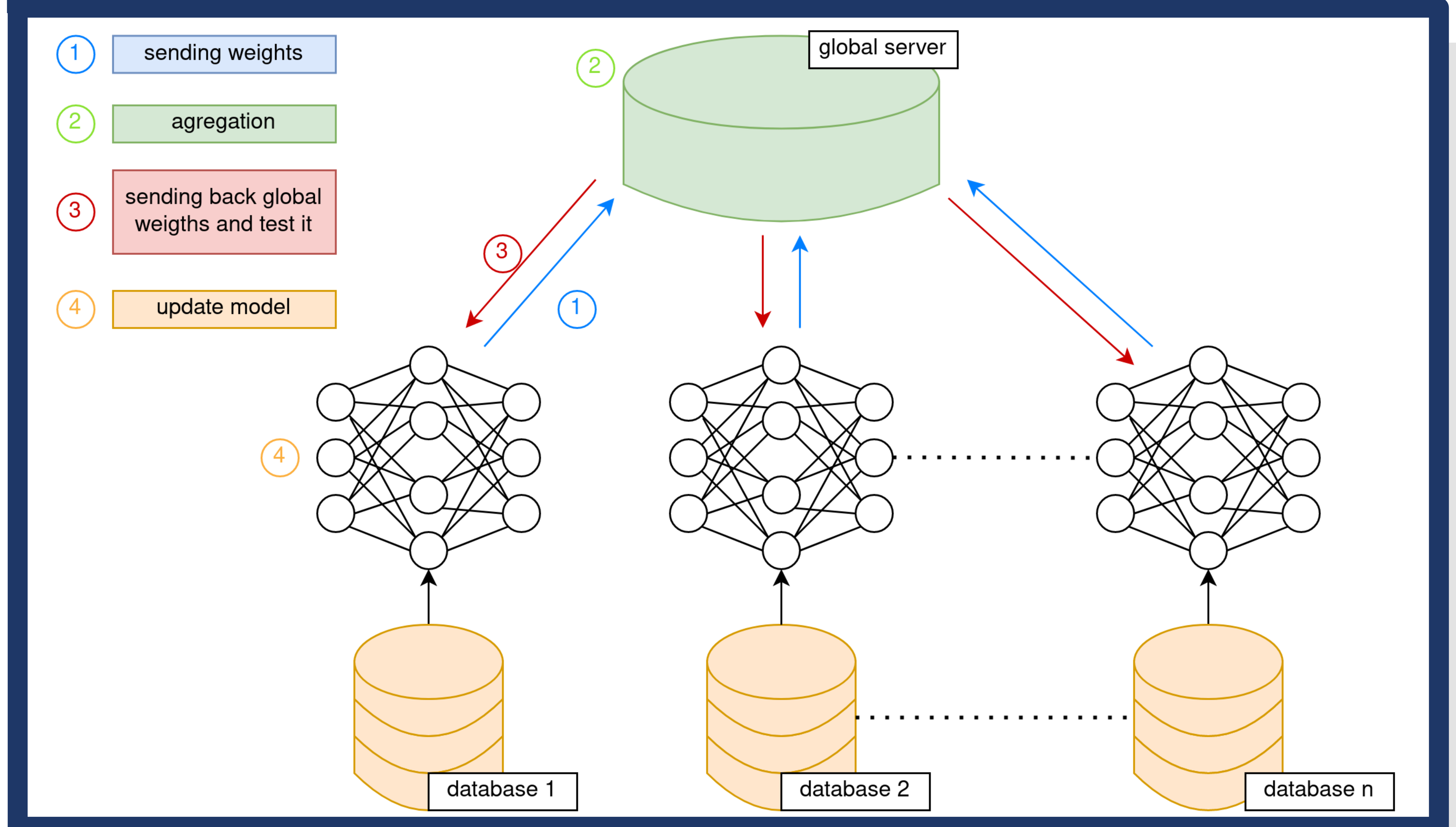
Abstract

Medical data is not fully exploited by Machine Learning (ML) techniques because **the privacy concerns restrict the sharing of sensitive information** and consequently the use of centralized ML schemes. Usually, ML models trained on local data are failing to reach their full potential owing to low statistical power.

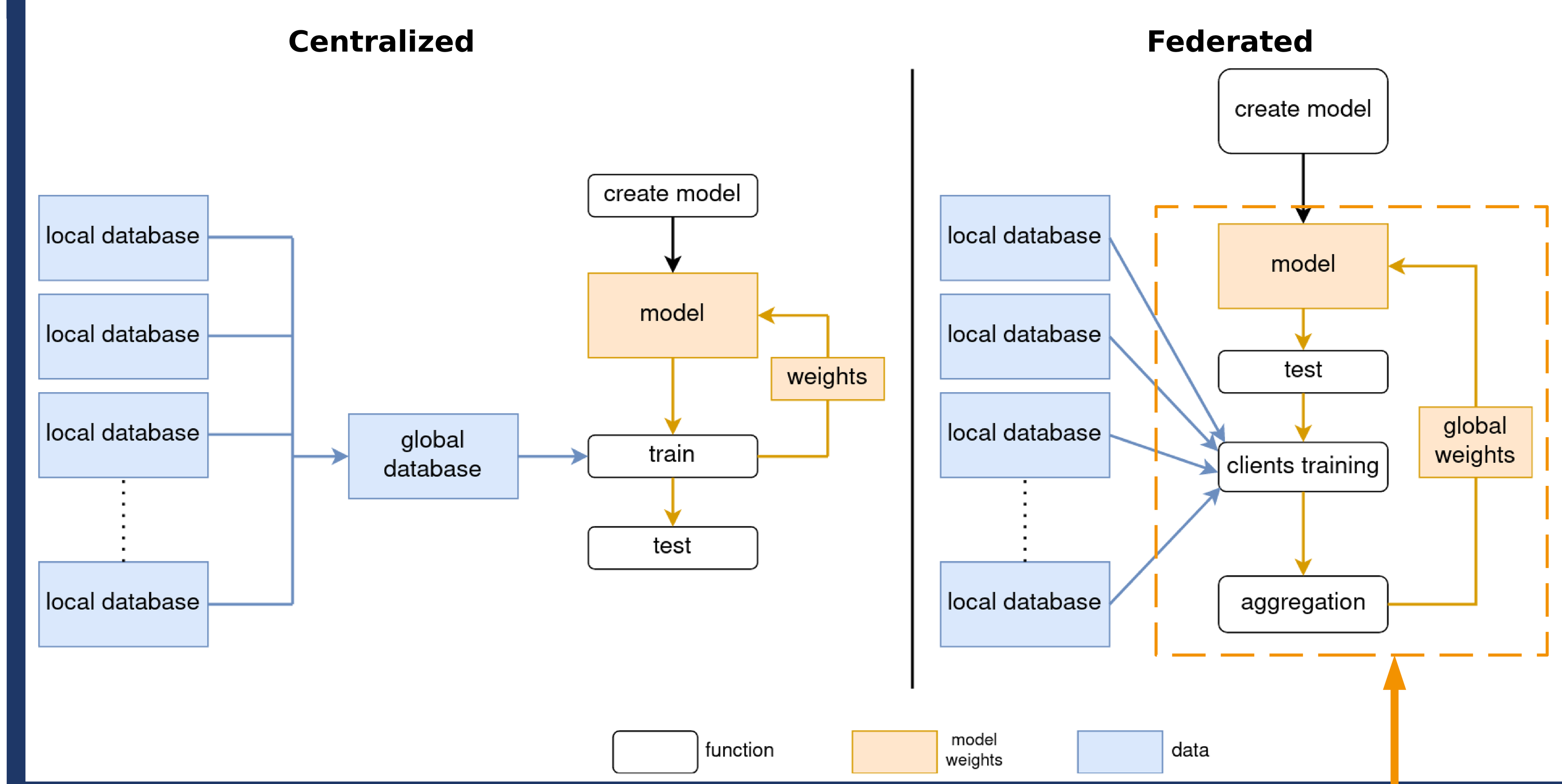
Federated Learning (FL) solves critical issues in the healthcare domain such as data privacy and enables **multiple contributors to build a common and robust ML** model by sharing local learning parameters without sharing data. FL approaches are mainly evaluated in the literature using benchmarks [1] and the trade-off between accuracy and privacy still has to be more studied in realistic clinical contexts.

In this work, we evaluate this trade-off for a **CD3/CD8 cells labeling model from Hoechst stained images**. Wölflein et al. [2] developed a **deep learning GAN model** that labels CD3 and CD8 cells from kidney cancer tissue slides stained with Hoechst. The GAN model was **trained on 475,000 patches** (256x256 pixels) from 8 whole slide images. We modified the training to simulate a FL approach by **distributing the learning across several clients** and **aggregating the parameters** to create the overall model. We present the performance comparison between FL and centralized learning.

Federated Learning



Experimental design



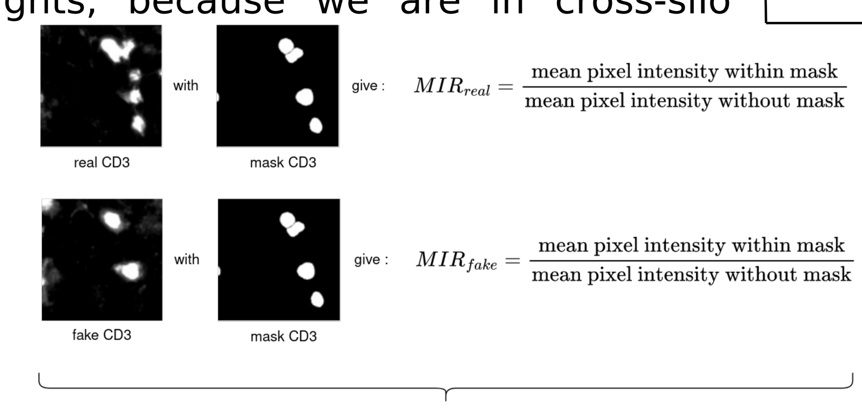
Material and Methods

The HoechstGAN is inspired of **Pix2pix GAN** model [3], It takes **Hoechst* image as input** and create **CD3* and CD8* marked images as output** . The gain of this model is that Hoechst staining is less expensive in time and money than CD3 and CD8 immunohistochemistry experiments. Thus, it could create a huge time and funds saving.

The model was **trained on** 8 whole slide image (WSI) cut in **475 334 patches (256x256)** and **tested on** 2 WSI cut in **152 185 patches**

To simulate the federation we separated the database by whole slide images evenly distributed between the clients. The **aggregation** function is an **average** of the models weights, because we are in cross-silo configuration with trusted clients.

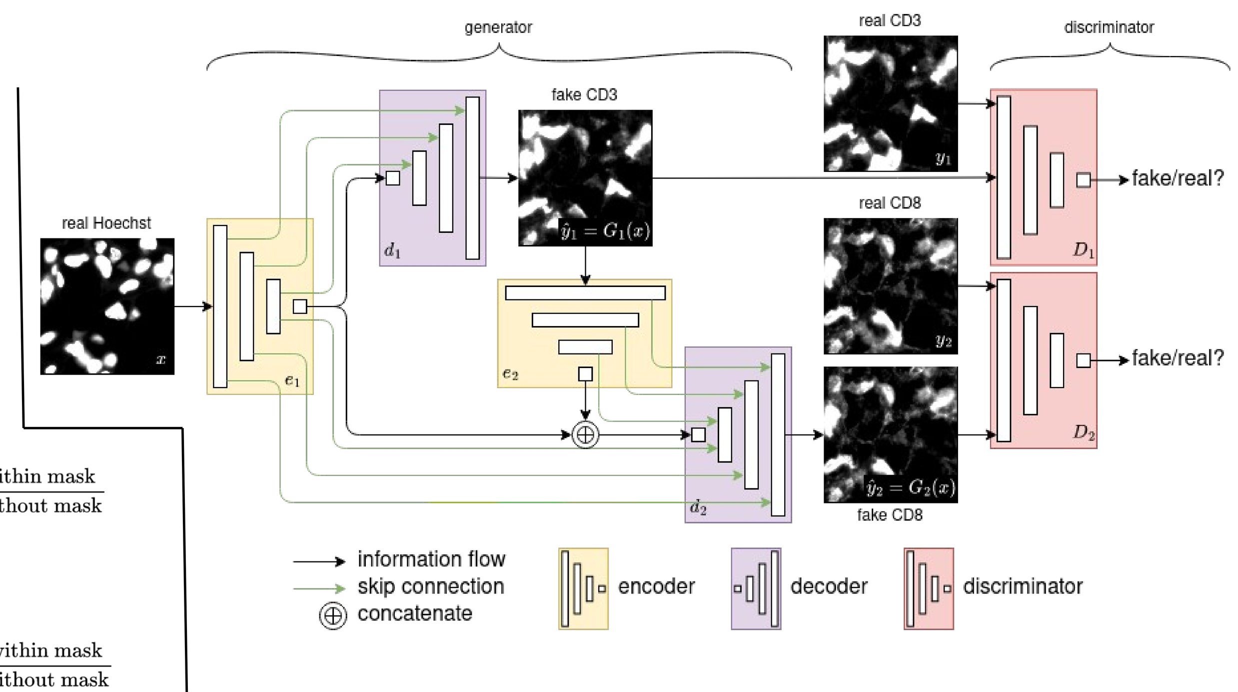
The **metric** used to evaluate the model is the **Masked Intensity Ratio (MIR)**, a custom metric created by G.Wölflein to evaluate the HoechstGAN. It is a metric that **compare the signal to noise ratio** of the **fake staining** and the **real one**, of the same patch.



$$MIR_{real} = \frac{\text{mean pixel intensity within mask}}{\text{mean pixel intensity without mask}}$$

$$MIR_{fake} = \frac{\text{mean pixel intensity within mask}}{\text{mean pixel intensity without mask}}$$

$$MIR_{relative} = \frac{MIR_{fake}}{MIR_{real}}$$



HoechstGAN architecture

* hoechst: chemical staining of cells by binding to DNA
CD3/CD8: protein surface markers of immune cells

Results

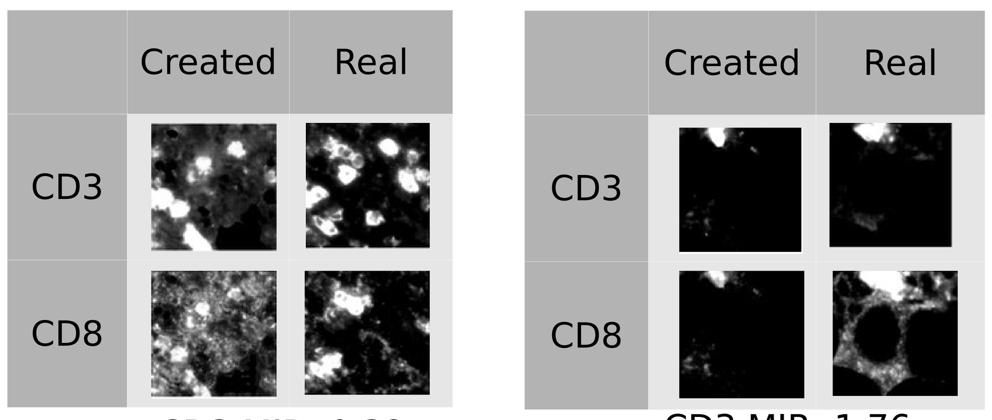
We compared the performance of the HoechstGAN algorithm between centralized and federated modes, for 4 and 8 clients. The performance assessment of CD3 and CD8 cell labeling predictions is based on the MIR metric.

Figure 1: For CD3 or CD8 cells, the **centralized mode provided better predictions than the federated modes**. The difference in performance between centralized and federated was less significant for CD8 cells. **Federated modes using 4 or 8 clients produced similar prediction quality**.

Figure 2: The MIR metric being greatly influenced by the level of noise in the real or created images. We are working on validating the prediction performance using a second metric that is less sensitive to noise.

Predicted cell labeling	Centralized (MIR)	Federated 4 clients (MIR)	Federated 8 clients (MIR)
CD3	1.73	0.7	0.8
CD8	0.84	0.61	0.65

Figure 1: MIR values at the 30th epoch.



CD3 MIR: 0.32
CD8 MIR: 0.24

CD3 MIR: 1.76
CD8 MIR: 6.9

Figure 2: example of a bad and a good prediction obtained in centralized mode

Discussion

The main advantage of federate learning is the **data privacy** because there is no data sharing. However, federated learning is associated with a **longer training time** and shows, in the case study of the HoechstGAN algorithm, **poorer performance** than centralized learning.

In perspective, we will evaluate the predictions obtained by HoechstGAN in centralized and federated mode using a **second metric less sensitive to image noise**. In addition, we will compare the performance of centralized and federated learning approaches on a second use case dedicated to **predicting patients' response to anti-tumor treatments based on clinical and genetic information**.

References

[1] Kairouz, P., et al. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning 14(1–2), 1-210.

[2] Wölflein et al. (2023) Virtual Lymphocyte Staining Using Generative Adversarial Networks. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023. p. 4997-5007.

[3] Isola et al. (2018) Image-to-image Translation with Conditional Adversarial Networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1125-1134.

KATY CH
Horizon 2020 KATY project (grant No 101017453)
Horizon Europe CANVAS project (grant No 101079510)

cea Inserm UGA Université Grenoble Alpes ISMB ECCB 2023 LYON FRANCE