Automation, Surgery Support and Intuitive 3D visualization to optimize workflow in image guided therapy SysTems

DELIVERABLE D2.1

Requirements for data storage solution for federated learning (for selected use cases)

Project number:	ITEA 20044
Document version no.:	v 1.1
Edited by:	Nishat Raihana
Date:	2022.04.24

ITEA Roadmap challenge:
Smart Health

HISTORY

Document version #	Date	Remarks
V0.1	2022.02.14	Starting version, template
V0.2	2022.05.10	Compilation of first input by partners + review
V1.0	2022.05.31	Final version

Deliverable review procedure:

- **2 weeks before due date**: deliverable owner sends deliverable –approved by WP leader– to Project Manager
- **Upfront** PM assigns a co-reviewer from the PMT group to cross check the deliverable
- **1 week before due date**: co-reviewer provides input to deliverable owner
- **Due date:** deliverable owner sends the final version of the deliverable to PM and co-reviewer

TABLE OF CONTENTS

1 Abbreviations

AI	Artificial intelligence
CNN	Convolutional neural network
CT	Computed tomography
CTV	Clinical target volume
DICOM	Digital imaging and communications in medicine
EHR	Electronic healthcare records
FL	Federated learning
GDPR	General data protection regulation
GTV	Gross tumour volume
MDL	Medical data lake
MR	Magnetic resonance
MRI	Magnetic resonance imaging
PACS	Picture archiving and communication system

2 Executive summary

This document introduces federated learning (FL) within medical imaging and presents requirements for training deep networks for automatic analysis of medical images (e.g., image segmentation, image classification). Throughout the document, the concept of FL will be introduced along with a brief description of different FL models, and how to implement FL. The main concept of a data lake will also be discussed. This document thoroughly discusses the GDPR issues concerning patients' data while doing R&D on federated learning. The document also presents how FL can be used for different use cases in ASSIST.

3 Introduction of medical data lake

A data lake stores current and historical data from one or more systems in its raw form, which allows business analysts and data scientists to easily analyze the data. The advantage of using a data lake is easy data storage which simplifies raw data. A schema is applied afterward to make working with the data easy for business analysts, application developers, and data scientists. The Medical Data Lake (MDL) is a secure and scalable distributed service for medical images, structured and unstructured data. A database stores the current data required to power an application. A data lake stores current and historical data for one or more systems in its raw form to analyze the data. The basic difference between data lake and a database is that a data lake is a superset of a database, where data lake has additional set of API for the stored data as well as acting as a data storage. Inovia's MDL system solution is a microservice based solution that not only aims to store medical images but also provides the intelligence and API to segment images and machine learning to retrain the module that been used for image segmentation. Additionally, it serves as an execution platform for analytics, and is optimised for easy training and deployment of AI. During this, it is possible to use open-source machine learning algorithms/models, as well as the ability to develop proprietary algorithms/models. Moreover, the MDL includes the option for data anonymization (for GDPR compliance), to allow broader analyses. The aim of ASSIST is to develop a medical data lake for storing the data and use the data for training different models and implement it in different hospitals. For FL a data lake provides better flexibility than saving the data in a data warehouse or database for training.

3.1 Anonymization

Research with health data is concerned with the General Data Protection Regulation (GDPR), which aims to ensure patients' privacy. A few anonymization techniques will be introduced in this section in compliance with GDPR guidelines.

Radiotherapy based cancer treatment requires medical images of patients. Medical images can for example be obtained in DICOM or NIFTI format. DICOM images contain a lot of information that needs to be removed for anonymization. In every DICOM file, the information is embedded in the header, and this information is organized into four levels of hierarchy — patient, study, series, and instance.

- "Patient" is the person receiving the exam
- "Study" is the imaging procedure being performed, at a specific date and time, in the hospital
- "Series" — Each study consists of multiple series. A series may represent the patient being physically scanned numerous times in one study (typical for MRI), or it may be virtual, where the patient is scanned once, and that data is reconstructed in different ways (typical for CT)
- "Instance" — every slice of a 3D image is treated as a separate instance. "instance" is synonymous with the DICOM file itself in this context.

To illustrate this hierarchy, Figure 1 shows a few DICOM files from the publicly available pancreatic cancer dataset from The Cancer Imaging Archive (TCIA).

Figure 1. Example of a DICOM dataset organized by the “Patient”, “Study”, “Series”, and “Instance” levels. For a 3D volume, each DICOM file normally represents one slice of the volume.

Table 1 shows a printout of the header using PyDicom (a python package which allows reading and writing of DICOM files). Here most of the non-relevant information has been removed, and some “fake” patient data has been added.

(7fe0, 0010)	Pixel Data	Array of 524288 elements

Table 1: Printout of a portion of a DICOM header displaying Patient, Series, Study, and Instance UIDs and text descriptions.

3.2 Data flow

Inovia’s MDL will receive DICOM images from hospitals via the API-Gateway Module and save it using the ASSIST - Proxy-Service module. Figure 2 shows the process of data flowing from hospital to Inovia’s MDL.

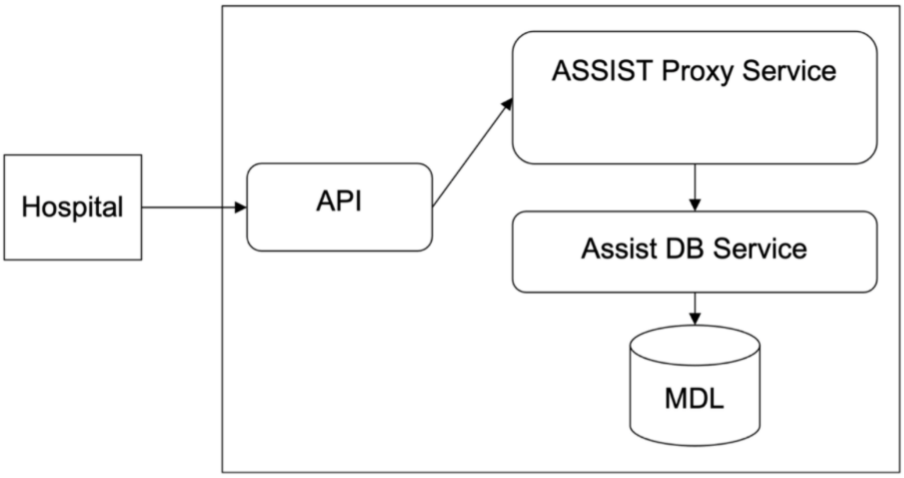

Figure 2. An overview of how medical images can flow from a hospital to the MDL.

As shown in Figure 2, the ASSIST Proxy-Service is a microservice responsible for handling patient data. A microservice is a small independent application that performs a highly focused service, the counter model to a huge monolithic application that serves many different kinds of requests, Typically a microservice is lightweight and can be started in microseconds. For that reason, they are a good choice for handling scalability issues, if the load increases on the service(s) just start more services.
In this microservice Inovia has implemented a Data Anonymizer module which replaces, empties, or removes patient's, physician's, and any other information from DICOM files.

4 Introduction to Federated Learning

This section will introduce federated learning, interested readers are referred to recently published papers about FL in health care for more information (Rieke et al., 2020; Antunes et al., 2022; Kairouz et al., 2021; Xu et al., 2021).

4.1 Why federated learning?

One could argue that "deep-learning" is the state-of-the-art for machine learning. Architectures such as deep CNNs or transformers typically outperform other ML methods on several established benchmarks. But these networks have two drawbacks; training is costly both in time and computation, and a huge amount of training data is required. No element is more essential in machine learning than high quality training data, and this is especially true for deep learning. The work involved in acquiring, labelling, and preparing training data is daunting. Quantity and quality are both important. To collect a large training set is especially difficult in medical imaging, as researchers and companies then need to follow more regulations compared to other types of data.

Medical data is sensitive and need to be anonymized before inclusion into any training set. GDPR regulations restrict this further, and the terms of agreement may prohibit sharing of the data. Different hospitals, regions and countries may have different rules (see section 8). In short, the creation of large medical image data sets is hard.

Federated learning seems to be the obvious remedy to the data collection problem. The hospitals/clinics become nodes/clients in an asynchronous training network instead of being simple contributors of raw data. Model updates are shared instead of sharing data. This way, the images still become part of the training set, but the data is never shared between nodes, see Figure 3. See Figure 4 for a comparison of FL and centralized training.

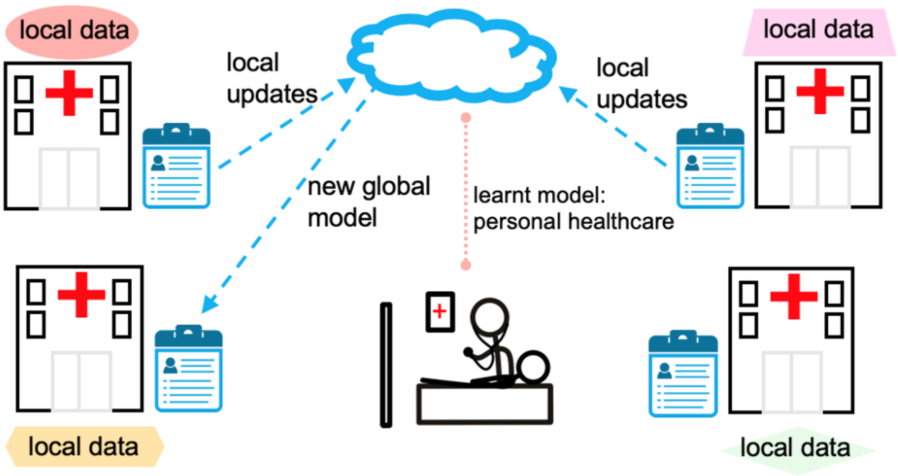

Figure 3. The main idea in FL is to not store all data in a single large, centralized database or data lake, but to instead store for example image data locally at each hospital. Instead of sending medical images and other medical data between the hospitals, the hospitals send updates, or parameters, of deep learning models. This

process is then iterated to convergence. Instead of having one large supercomputer, it is with federated learning sufficient if each hospital has a smaller computer.

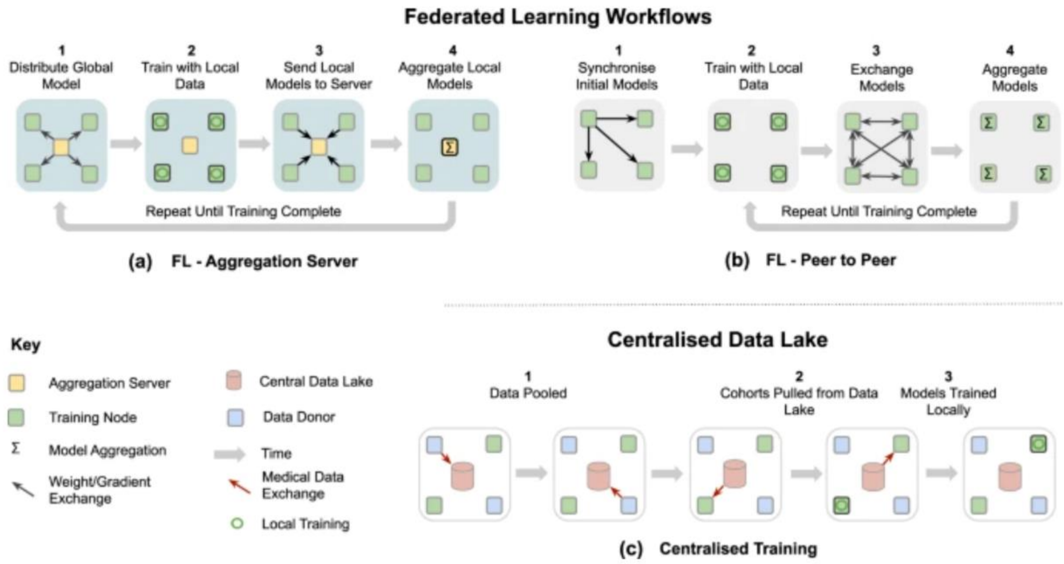

Figure 4. Image and figure text from (Rieke et al., 2020). A comparison of federated learning workflows and centralized training. a) FL aggregation server—the typical FL workflow in which a federation of training nodes receive the global model, resubmit their partially trained models to a central server intermittently for aggregation and then continue training on the consensus model that the server returns. b) FL peer to peer—alternative formulation of FL in which each training node exchanges its partially trained models with some or all of its peers and each does its own aggregation. c) Centralised training—the general non-FL training workflow in which data acquiring sites donate their data to a central Data Lake from which they and others are able to extract data for local, independent training.

4.2 Horizontal and vertical federated learning

In FL we distinguish between two main cases depending on how the distributed data is partitioned. In horizontal FL each participant has the same features but different examples (e.g. all have the same type of images but from different patients) and in the vertical case, clients can hold different features from the same examples / subjects. An example of vertical learning is a bank and an insurance company having different data on the same customer. In this project we are mainly concerned with the horizontal use-case.

4.3 Cross-silo vs cross-device FL

It is also common to distinguish between the cross-silo case and the cross-device case. The separation between the two is not precisely defined, but it is related to the scale of clients. In the cross-silo case we assume a moderate number of client sites, but for each client we assume that is has access to significant hardware and storage resources, and a reasonably stable internet connection. The use-cases in ASSIST all fall in this category. In the cross-device case the FL architecture needs to target millions of connected devices (IoT), and the network and hardware heterogeneity can be expected to be a major concern.

4.4 The difference between FL and distributed machine learning

While the computation in FL shares many common patters with normal (data parallel) distributed machine learning, the infrastructure setting with a significant degree of system heterogeneity, and the communication constraints introduced by weak and sometimes unreliable network connections (the internet) is a challenge. Most importantly, in FL we have no control over the data partitioned amongst clients, i.e. computation needs to be robust to non-identically distributed partitions and significant unbalance in the size of local datasets. A large body of research is currently developed in the wider community addressing these challenges specific to FL, for an overview see the review (Kairouz et al, 2019).

4.5 Algorithms for federated learning

The most common horizontal FL algorithm for deep learning models is federated averaging (FedAvg). It starts with an initial model state that is sent amongst the clients, each client trains the model in a certain number of iterations (or epochs) and is then sent to a central server where the model weights are averaged to an updated model state, the updated state is usually validated on some designated validation data set before it is sent back to the clients for e new round of training. The procedure continues until the model validation converges or reaches a desired score. The algorithm is vulnerable to non-IID data distribution among clients and one way to diminish the issue is to shorten the number of local iterations between the averaging steps but with an increased communication cost. A large body of work is currently pursued in the community to improve details of the basic FedAvg pattern to e.g. improve communication efficiency and robustness to corrupted data on clients. In particular, modifications of the aggregation function to penalize large deviations in averaged gradients between global rounds is a common strategy.

4.6 Challenges and considerations for local data in relation to ASSIST

Federated learning has its own challenges. Every node (clinic) needs to have the capacity to do local training on their local data. If the model is trained with supervised learning all local images used in the training need to be labelled correctly and consistently to ensure the quality of the training data. The nodes with the least processing capacity in the federation set the constraint on how complex the model can be. It goes without saying that the communication between the nodes need to be protected from intrusion. Big architectural changes in the model may make nodes obsolete until they retrieve the updated model architecture. Running the model training at a local node need to be user friendly and as automatic as possible.

There are also special constraints for federated learning. The part of the data available locally will probably have a local bias and not fulfil the criteria of independent and identically distributed data for the full set of data. Model updates sent from the nodes will not be synchronized. A node could even become adversarial, sending updates that damages the model performance.

A hybrid approach would be to freeze the first layers of the network and having the nodes simply inputting the training images to the frozen part of the network and sending the network output as model updates. The rest of the training is then done centralised. The computational load on the node is low, but the amount of data to transfer will be high, training cannot adjust the frozen layers, and the centralized training site needs a lot of capacity.

Experiments, where a dataset has been partitioned homogeneously and trained with federated averaging (FedAvg) settings, have proven to converge to the same accuracy as a model trained centralized on the complete dataset. In most real cases, there are no guarantees that the data distributions between the clients are homogeneous. Typical distributed optimization problems are mentioned in (McMahan et al., 2017). In cases where FL is applied, it is not possible to compare model scores with a centralized trained model, instead, it is important to show that a federated model outperforms models trained locally on each of the clients' databases.

5 FL frameworks, network communication and privacy

This section will provide an overview of different existing frameworks for FL, and mention pros and cons of each framework. The section will also describe the main requirements when it comes to communication between the nodes. Federated learning frameworks are distributed computing frameworks that need to be designed with scalability and resilience in mind. Implementation might range from centralized client-server architectures to fully decentralized peer-to-peer systems. FEDn, which is the framework developed by Scaleout, implements a hierarchical, or tiered architecture where deployments might range from client-server to highly distributed with multiple aggregations servers.

5.1 Notable open-source FL frameworks

NVIDIA FLARE : Nvidia runtime environment for Federated learning (https://nvidia.github.io/NVFlare)

Pros:

- Configuration based code
- Privacy-Preserving Algorithms: Differential privacy
- Tools for data management
- Different aggregation algorithms benchmarked with different settings
- Active development
- Solid reputation in the open-source community
- Command line deployment of new applications

Cons:

- Basic admin management tool

Flower : Friendly Federated Learning Framework: (https://flower.dev/)

Pros:

- Usability: Easy to use
- ML Framework Agnostic
- Scalability: Tested with 1000s of client simultaneously
- Large open source community

Cons:

- Less benchmarking details
- No Differential privacy

PySyft + Grid : (https://github.com/OpenMined/PySyft)

Pros:

- Decouples private data from model training, using Federated Learning, Differential Privacy, and Encrypted Computation
- One of the first open-source FL frameworks
- Large open source community

- Differential privacy support

Cons:

- Latest released version is not working and the framework is going to be substantially refactored
- Not production-grade

Scaleout FEDn : (https://github.com/scaleoutsystems/fedn) (Ekmefjord et al., 2021)

Pros:

- Hierarchical architecture for efficient model aggregation
- Scalability: Tested with 1000s of clients simultaneously
- Open-source framework for federated machine learning
- Production-grade
- Integrated user interface to get insights about the training and aggregation processes
- Machine learning framework agnostic
- Deployment can be done both using Docker containers and Kubernetes
- Single platform for both cross-device and cross-silo settings
- Active development
- Allows both alliance-based and custom compute packages for model training

Cons:

- No differential privacy / secure aggregation
- Relatively complex setup due to natively distributed architecture

5.2 Requirements for network communication

Depending on the details of the implementation of the messaging, the clients may or may not need open ingress ports to participate in the federation. FEDn is designed in a way to avoid open ingress ports on clients. In all frameworks however, the Central/Aggregator node needs dedicated ports open for communication with the clients.

The clients send an update to the Aggregator when they have completed the specified number of local iterations (one or a few batch updates up to several full epochs, depending on settings). There are no special requirements for the network connecting the nodes. However, there must be enough bandwidth to communicate with the central node. The required bandwidth depends on how large the deep learning model is, e.g. if it contains 100,000 or 100,000,000 trainable parameters, and how often the nodes send updates to the Aggregator. Benchmark case studies for both large CNNs (natural language processing) and smaller LSTM models (IoT use case) for FEDn can be found in (Ekmefjord et al, 2021).

5.3 Requirements for privacy

A general challenge in federated learning is to guarantee that sensitive information is not communicated between the different nodes, and this is especially important for medical data. Differential privacy is a mathematical principle which consists of adding a well-defined amount of noise to the weights in the deep learning model to secure the privacy while preserving most of the general information from the dataset. The drawback of this operation is that the performance of the aggregated model is reduced. As a concept, differential privacy has a wider scope than federated learning and is mainly used for privacy preserving analytics / database queries. It is sometimes used as an output privacy enhancing technology in FL where controlled noise is added to the weight updates before sending and aggregating them at the server. It is not clear to what extent this is important for a given application, and there is a risk that it will negatively impact convergence of training. Still, differential privacy is a viable way to protect against inference attacks (independent of whether the training is federated or not).

5.4 Data preparation

It is of importance that the data stored at all clients are pre-processed as similarly as possible. Challenges with medical data can be that the MR and CT scanners generate images with different appearance (see section 6.1.2). Another problem is that the size and resolution of the images can differ between the clients. These issues need to be addressed before a federation can be started.

6 FL in different ASSIST use cases

6.1 Brain tumours

Brain tumours compose about 2% of the cancer incidences, affect some 300,000 subjects globally each year (Leece et al., 2017), with a low survival rate and a high morbidity for the patients. Though not being the most prevalent cancer type, brain tumours are prone to complicated and challenging treatment procedures that are often a combination of surgery, radiotherapy and chemotherapy, where treatment planning and follow up of the treatment is highly dependent on radiology images. The best treatment for a specific patient depends on if there is one tumour or many small metastases, and the size and location of each tumour or metastasis. Furthermore, the size of the tumour is required to calculate how much radiation to apply to kill the cancer cells. MRI is normally used to obtain this information, and to plan the treatment, as MRI provides very good contrast between soft tissue types (and different MR sequences provide slightly different information / contrast). It is also necessary to segment important risk organs (e.g. the optic nerve) which should not be damaged by the radiation, see Figure 5. The treatment plan, i.e. how much radiation to apply to different parts of the brain, can be generated manually, through mathematical optimization or through machine learning.

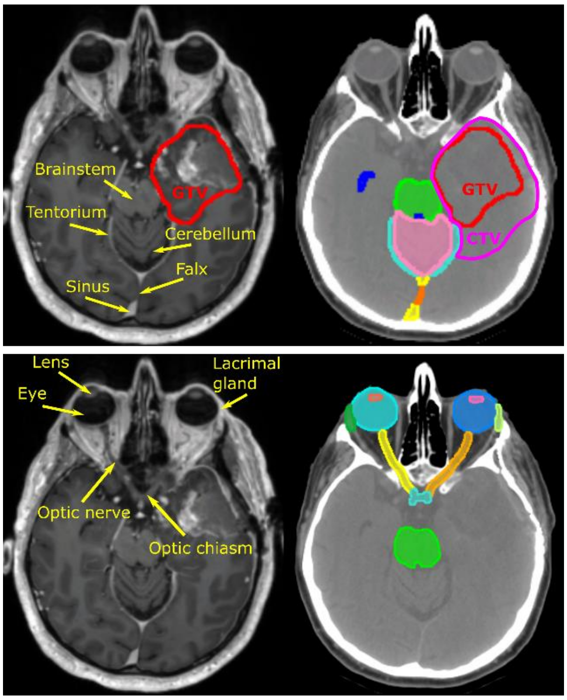

Figure 5. Illustration of brain tumour (red, to be killed by radiation) and risk organs (yellow, which should receive as little radiation as possible). GTV = gross tumour volume, CTV = clinical target volume (CTV). Deep learning can reduce the treatment planning time substantially, by performing automatic segmentation of tumour(s) and risk organs (instead of doing manual time-consuming segmentations). Image from an open dataset in the cancer imaging archive (see references).

6.1.1 Federated brain tumour segmentation

To segment tumour(s) and risk organ(s) is currently often performed manually or semi-automatically by a neuro radiologist, medical physicist or radiation oncologist. Manual segmentations can be very time consuming, e.g. 10 – 60 minutes per patient, especially for many metastases and risk organs. To train a segmentation network like U-Net (Ronneberger et al., 2015) to perform automatic segmentation, in 10 – 120 seconds, requires annotated brain tumour images (see Figure 5). In the BraTS challenge (Menze et al., 2014) the number of training subjects is about 400, but to scan and annotate images from 400 subjects is a lot of work for a small hospital. Through federated learning it is sufficient if each hospital provides a smaller number of annotated images, see Figure 6.

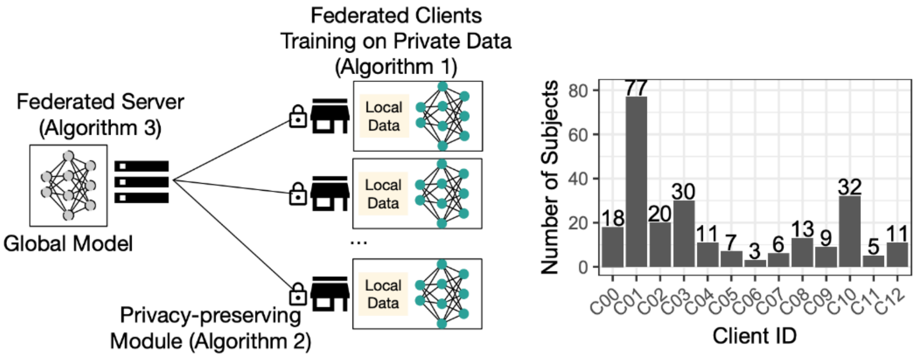

Figure 6. Left: illustration of a privacy-preserving federated learning system used to train a brain tumour segmentation network (Li et al., 2019). Right: distribution of 242 training subjects into 13 federated clients. A challenge in FL is having different numbers of subjects (images) in each node, as each node then will complete a training epoch very quickly or more slowly.

A general challenge in federated learning is to obtain metrics (e.g. Dice score) that are as close as possible to non-federated learning (data-centralised training). This is especially true when privacy-preserving algorithms are used, which for example add noise to the learned weights. Figure 7 (from Li et al., 2019) shows the obtained performance when using FL, compared to non-FL, for different settings of the FL training. The obtained performance is sometimes lower compared to data-centralised training. Several other researchers have also used FL for brain tumour segmentation (Tedeschini et al., 2022; Yi et al., 2020).

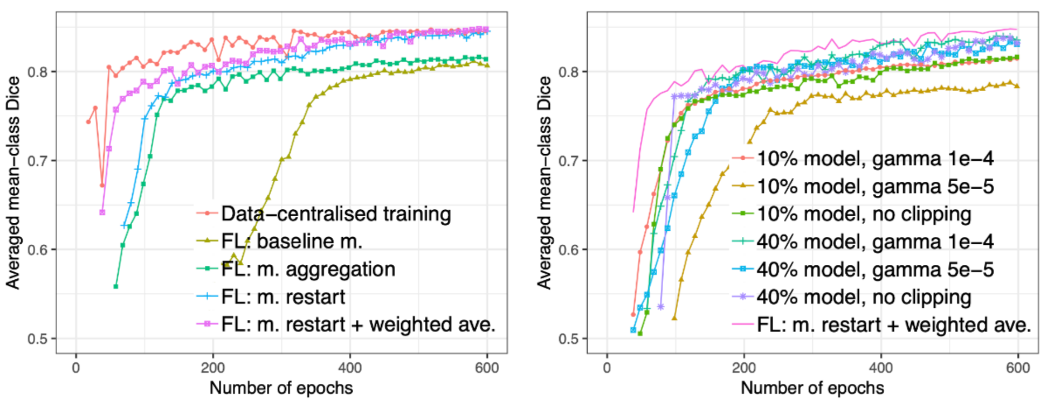

Figure 7. Comparison of brain tumour segmentation performance (Dice score) on the test set with (left): FL vs. non-FL training, and (right): partial model sharing (Li et al., 2019). A general challenge is to with FL obtain similar performance as non-FL training, this is especially true for privacy-preserving algorithms.

6.1.2 Domain shift problem

A common challenge is that MR images collected at different hospitals / scanners have different characteristics (used MR sequence, image intensity, image resolution, noise), commonly known as domain shift in medical imaging. A segmentation network trained on images from MR scanner / hospital A will therefore in general not perform as well for images from MR scanner / hospital B. Deep learning models such as CycleGAN (Zhu et al., 2017) can be used for image harmonization (Bashyam et al., 2022), i.e. to make MR images from different MR scanners look more similar, but it is not obvious how to train such models in a federated setting (as each node only stores images from one specific MR scanner). Furthermore, the images collected at one hospital may originate from patients with a mean age of 40 years, while the mean age is 55 years for patients at another hospital. Another challenge is that the annotations of tumours and risk organs may be done a bit differently at each hospital, as neuro radiologists do not always agree where the tumour border is.

6.1.3 Preliminary FL results for the BraTS dataset

Scaleout has in the ASSIST project performed some initial experiments on the open brain tumour dataset for radiation therapy (BraTS) where the data is split so that each partition includes complete sets of subjects. This causes a small non-IID effect on the partitions compared to splitting the images at random. The results of the experiments showed that the federated setting converged to the same score as the centralized model and that none of the local trained models (i.e., models trained on one partition) performed as well. This is the behaviour we expect from prior studies on different datasets and with different models, see e.g. (Ekmefjord et al, 2021) for additional examples and references.

6.1.4 Federated treatment planning

When the image data has been segmented, the next step in the radiation oncology workflow is to create a treatment plan that can be sent to the delivery system for treatment delivery. The treatment plan creation is typically a tedious task where the treatment planner spends many hours to create a treatment plan fulfilling the clinical goals of the treatment such as sufficient dose to the target volume while avoiding excessive dose to sensitive structures in the vicinity of the target. Recently, machine learning technology has been used to automate the treatment plan generation by predicting a dose distribution based on patient geometry and treatment protocol (McIntosh et al, 2021). The predicted dose is then used as input to an optimization problem to generate a deliverable treatment plan without the need of manual input. The model training typically requires data transfer of medical image data and treatment data for the selected treatment protocol which is often a time-consuming and cumbersome process. Federated learning could help in the model development process by enabling data access to multiple clinics without the need of data transfer to develop planning models based on data from multiple clinics. Such data may also be used for multi-node testing to ensure the trained model works on a variety of patient data from different clinics and countries.

6.2 Prostate enlargement

6.2.1 Clinical Background and Type of Data

Prostate enlargement, also called Benign prostatic hyperplasia (BPH), is a noncancerous increase in size of the prostate gland. (BPH), proliferation of the glandular and stromal tissue in the transition zone of the prostate, results in lower urinary tract symptoms (LUTS) and bladder outlet obstruction, see Figure 8.

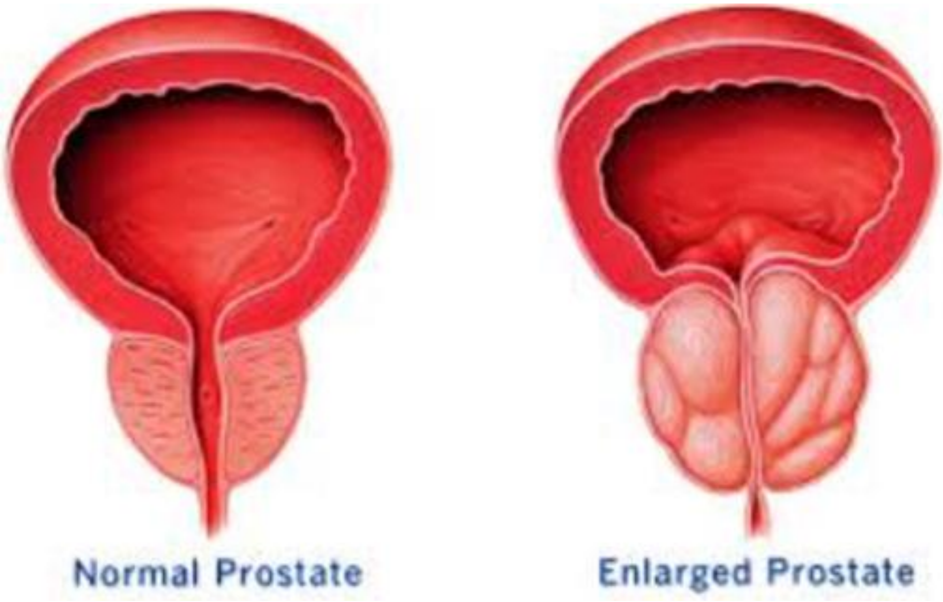

Figure 8. Normal vs enlarged prostate (see references).

The prevalence increases with age and 25% of men older than 70 years old have moderate to severe LUTS that effect their quality of life (QoL). About 105 million men are affected by BPH globally. It typically begins after the age of 40. Half of males age 50 and over are affected. After the age of 80 about 90% of males are affected.
A wide variety of medical and surgical options are available for the management of BPH with LUTS. In patients with moderate to severe LUTS refractory to medical management more invasive treatments are considered. Transurethral resection of the prostate (TURP) and open prostatectomy (OP) are the gold standard treatment methods for prostate glands of 30-80 cm3 and ≥ 80 cm3 respectively. However, these procedures have considerable morbidity rates including retrograde ejaculation, erectile dysfunction, urethral stricture, urinary retention, transfusion requirement and incontinence. Also, in patients with existing comorbidities, increasing age and large prostate volume the complication rates are higher and hence the eligibility for surgical therapies are limited.

Medical and surgical options are available; however, these procedures have considerable morbidity rates. Prostate artery embolization (PAE) has emerged as a minimal invasive treatment method which has a lower risk of urinary incontinence and sexual side effects. Clinical and laboratory data and radiological imaging are the primary sources in patient preparation.

6.2.2 Benefits & challenges of FL compared to conventional training

A major unsolved challenge with AI in medicine is the ability to generalize results of a model trained on data from a single institution. Creating large, centralized collections of images from different hospitals raises some problems about patient privacy. “Federated Learning” is a good alternative where the algorithm is brought to the data (instead of the reverse), consequently it permits institutional collaboration without sharing any data. Infrastructure for this consists of a central server sharing parameters of AI models trained individually at different sites, with each site sharing the model weights (but not patient data) during training. Consequently, FL combines the benefit of using different data sources with the need of sensitivity for the privacy of the medical data.

On the other side, FL in medicine comes up with some challenges, the major challenges for this concept can be listed as below. The all challenges listed here also applies to our Prostate Enlargement use case.

Expensive Communication: Communication is a critical bottleneck in federated networks, which, coupled with privacy concerns over sending raw data, necessitates that data generated on each device remain local. Indeed, federated networks may potentially be comprised of a massive number of devices, and communication in the network can be slower than local computation. In order to fit a model to data generated by the devices in the federated network, it is therefore necessary to develop communication-efficient methods that iteratively send small messages or model updates as part of the training process, as opposed to sending the entire dataset over the network.

Systems Heterogeneity: The storage, computational, and communication capabilities of each device in federated networks may differ due to variability in hardware (CPU, memory), network connectivity (3G, 4G, 5G, Wi-Fi), and power (battery level). Additionally, the network size and systems-related constraints on each device typically result in only a small fraction of the devices being active at once. Each device may also be unreliable, and it is not uncommon for an active device to drop out at a given iteration due to connectivity or energy constraints. In order to cope with this challenges, federated learning methods to be developed should anticipate a low amount of participation, tolerate heterogeneous hardware, and be robust to dropped devices in the network.

Statistical Heterogeneity: Devices frequently generate and collect data in a non-identically distributed manner across the network, moreover, the number of data points across devices may vary significantly, and there may be an underlying structure present that captures the relationship amongst devices and their associated distributions. This data generation paradigm violates frequently-used independent and identically distributed assumptions in distributed optimization, increases the likelihood of stragglers, and may add complexity in terms of modelling, analysis, and evaluation. Indeed, although the canonical federated learning problem aims to learn a single global model, there exist other alternatives such as simultaneously learning distinct local models via multi-task learning frameworks. There is also a close connection in this regard between leading approaches for federated learning and meta-learning. Both the multi-task and meta-learning perspectives enable personalized or device-specific modelling, which is often a more natural approach to handle the statistical heterogeneity of the data.

Privacy Concerns: Finally, privacy is often a major concern in federated learning applications. Federated learning makes a step towards protecting data generated on each device by sharing model updates, e.g., gradient information, instead of the raw data. However, communicating model updates throughout the training process can nonetheless reveal sensitive information. While recent methods aim to enhance the privacy of federated learning using tools such as secure multiparty computation or differential privacy, these approaches often provide privacy at the cost of reduced model performance or system efficiency. Understanding and balancing these trade-offs, both theoretically and empirically is a considerable challenge in realizing private federated learning systems.

6.2.3 Requirements for hardware & network

Unlike running federal learning algorithms through consumer devices, healthcare institutions have relatively powerful computational resources and reliable, higher-throughput networks enabling training of larger models with many more local training steps, and sharing more model information between nodes. These unique characteristics of FL in healthcare also bring challenges such as ensuring data integrity when communicating by use of redundant nodes, designing secure encryption methods to prevent data leakage, or designing appropriate node schedulers to make best-use of the distributed computational devices and reduce idle time.

The administration of such a federation can be realised in different ways. In some situations, which require the most stringent data privacy between parties, training may operate via some sort of “honest broker” system, in which a trusted third party acts as the intermediary and facilitates access to data. This setup requires an independent entity controlling the overall system, which may not always be desirable, since it could involve additional cost and procedural viscosity. However, it has the advantage that the precise internal mechanisms can be abstracted away from the clients, making the system more agile and simpler to update.

In a peer-to-peer system each site interacts directly with some or all of the other participants. In other words, there is no gatekeeper function, all protocols must be agreed up-front, which requires significant agreement efforts, and changes must be made in a synchronised fashion by all parties to avoid problems. Additionally, in a trustless-based architecture the platform operator may be cryptographically locked into being honest by means of a secure protocol, but this may introduce significant computational overheads.

In our study, peer-to-peer administration method will be used in order to reduce additional effort for the development of the honest broker system.

6.2.4 Requirements of trained network

One of the main challenges for the usage of FL is to harmonize their data in order to make them usable for same training infrastructure. In our use case, the main source for data is the X-Ray images obtained from one institution, and the synthetic data derived from this data by using traditional and contemporary techniques. Consequently, there will not be any requirement for the harmonization of data.

On the other hand, in order to prevent any problem about the usage of the synthetic data, the data production and enlargement algorithms to be used should be approved. Any problem in this phase may lead to inaccurate data sampling. In addition to this, the distribution of the segmented data through federal units should also be examined with respect to the source of data (original vs. synthetic)

By using the Auto-ML tool which will be developed for the project, we will perform data preprocessing, feature engineering, hyperparameter optimization and algorithm selection, tasks. With this toolbox, we will also have the opportunity to combine, compete and manage different training models. The performance of the models will also be compared with respect to the usage of FL and non-FL methods.

6.2.5 Similar studies

In our literature survey, we could not find any study that focuses on the solution for our use by using AI either by using federated learning techniques or not. In this section, we want to mention three important studies related with our case. Two of these studies are about federated learning on medical data sets, and the other one was about the prostate segmentation in MR images.

Magnetic Resonance Imaging-based prostate segmentation is an essential task for adaptive radiotherapy and for radiomics studies. In order to prevent the manual delineation which is a time-consuming task, Comelli et. al. (2021) suggested three deep learning approaches aim is to tackle the fully-automated, real-time, and 3D delineation process of the prostate gland on T2-weighted MRI. The first model UNet is used in many biomedical image delineation applications, on the other hand ENet and ERFNet are mainly applied in self-driving cars to compensate for limited hardware availability while still achieving accurate segmentation. They applied these models to a limited set of 85 manual prostate segmentations using the k-fold validation strategy and the Tversky loss function and they compare their results. According to their findings, ENet and UNet are more accurate than ERFNet, with ENet much faster than UNet. Specifically, ENet obtains a dice similarity coefficient of 90.89% and a segmentation time of about 6s using central processing unit (CPU) hardware to simulate real clinical conditions where graphics processing unit (GPU) is not always available. They finally concluded that ENet could be efficiently applied for prostate delineation even in small image training datasets with potential benefit for patient management personalization.

Feki et. al (2021) focused on the hottest topic of the world for the last three years and try to develop a deep learning approach that can assist radiologists to analyze the vast amount of chest X-ray images, which can potentially have a substantial role in streamlining and accelerating the diagnosis of COVID-19. In order to cope with medical data privacy regulations, they present a collaborative federated learning platform which learns without sharing patient data. They investigated several key properties and challenges of federated learning setting including the not independent and non-identically distributed (non-IID) and unbalanced data distributions that naturally arise. They demonstrated that the proposed federated learning framework provided competitive results to that of models trained by sharing data, considering two different model architectures. They concluded that their study will encourage medical institutions to adopt to collaborative process and reap benefits of the rich private data in order to rapidly build a powerful model for COVID-19 screening.

Ma et. al (2022) proposed an assisted diagnosis model for cancer patients based on federated learning. They implied the importance of the studies about the location of cancer recurrence and its influencing factors for the clinical diagnosis and treatment of cancer. In terms of data, the factors influencing cancer recurrence and the special needs of data samples required by federated learning were comprehensively considered. They determined the six first-level impact indicators, and the historical case data of cancer patients were collected. Based on the federated learning framework combined with convolutional neural network, various physical examination indicators of patients were taken as input. The recurrence time and recurrence location of patients were used as output to construct an auxiliary diagnostic model, and linear regression, support vector regression, Bayesian regression, gradient ascending tree and multilayer perceptron neural network algorithm were used as comparison algorithms. CNN's federated prediction model based on improved under the condition of the joint modeling and simulation on the five types of cancer data accuracy reached more than 90%, the accuracy is better than single modeling machine learning tree model and linear model and neural network, the results show that auxiliary diagnosis model based on the study of cancer patients in assisted the doctor in the diagnosis of patients, As well as effectively provide nutritional programs for patients and have application value in prolonging the life of patients, it has certain guiding significance in the field of medical cancer rehabilitation.

6.3 Hepato pancreato biliary oncology

Hepato pancreato biliary oncology deals with malignant or cancerous tumours originating in the liver, pancreas, bile-ducts and gallbladder are some of the leading causes of cancer related deaths world-wide.

Liver cancer is the third leading cause of cancer death world-wide and pancreatic cancer is the fourth leading cause of cancer death in men and women and is projected to be the second leading cause within a decade. Early detection and complete removal of the tumour while saving as much as possible healthy tissue is important for the survival outcome and improved quality of life of the patient.

Pancreatic cancer is amongst the leading causes of cancer death in the developed world (Pancreatic cancer research fund, 2022), with a 5-year survival rate of 5% (American cancer society, 2010). This is largely because patients are asymptomatic until late in course of pancreatic cancer or have nonspecific symptoms, delaying treatment.

For the early diagnosis of HPB related cancers, CT and MR scans are the primary source of information and therefore automatic segmentation of tumours in CT or MR scans is of vital importance. As with most (medical) segmentation challenges nowadays deep-learning AI techniques offer the best results. CT scans are the most common method to diagnose pancreatic cancer because they have a clear contrast of the pancreas, with pancreatic tumours being visible as hypointense regions, as shown in Figure 9.

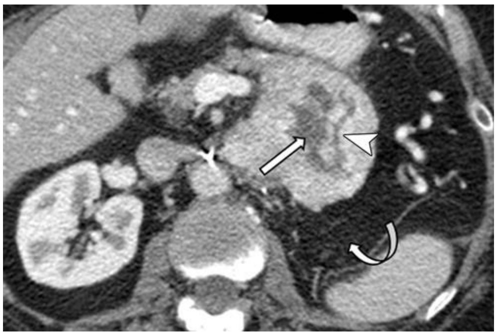

Figure 9. An example of a pancreatic tumour in a CT volume (Low et al., 2011).

Patient data privacy regulations such as GDPR hamper the sharing of medical data, both between hospitals and between hospitals and commercial parties. Although clinical sites can develop algorithms using their own patient database, this often provides a limited set of data, originating from a single scanner, thereby hampering generalizability of the algorithms on other datasets. Commercial parties, such as companies developing medical image analysis software, typically do not have large datasets at their disposal, and rely on publicly available datasets (for example from public segmentation challenges).

Federated learning allows these parties to jointly develop a single model, combining all available data, whilst not requiring any private data transfer. FL can therefore result in far larger datasets for algorithm development than single institutes can muster on their own. In addition, inexperienced centres can leverage on the knowledge of experienced partners.

Especially in the case of pancreatic cancer only a relatively small number of cases are treated at clinical centres yearly, thereby limiting the amount of data available for a single center. If sufficient centres collaborate, federated learning has the potential to thoroughly increase the number of cases available for training, therefore improving model performance, which is the first step towards improving treatment outcome. One of the issues is consolidating the data from different clinical care providers and ensuring consistency in labelling, before jointly training a model.

6.3.1 Requirements for hardware & network

One of the main pitfalls of federated learning is that during the aggregation step, data (model gradients) should still be communicated outside of the participating center. This may not be allowed by strict firewalls that are in use (e.g. at participating hospitals). The specific rules of the network pose a restriction on the FL framework that can be used, and determines which mode of communication should be used.

Participating centers should have access to their own (or cloud based) hardware, allowing for the neural network training for a few generations before aggregation is performed.

6.3.2 Requirements of trained network

Centers participating in FL model training should harmonize data requirements before commencing, ensuring that the same model architecture can be used for all data. In addition, results of hyperparameter optimization and implementation of fair data sampling strategies should be implemented consistently between centers. Models trained using FL should be benchmarked against baseline results (obtained without FL) to determine whether the use of additional data improves model performance. It should be noted that although FL exposes the model to an increased amount of data, data inconsistency may harm model performance on data of a specific origin.

To the best of our knowledge, no literature regarding the use of federated learning for developing pancreatic and/or liver tumor segmentation networks is available. For recent developments in pancreatic and liver tumor segmentation we refer to D3.1(State-of-the-Art of AI techniques used for personalized diagnosis).

7 Network and hardware specification and configuration for FL in hospitals

7.1 Hospital networks

To use FL in a clinical setting each computer (node) needs to be located behind the hospital firewall and be connected to the PACS, see Figure 10. The PACS is the heart of the imaging activities at every hospital, as it receives images collected from every imaging device (MR scanners, CT scanners, etc), and radiologists then look at the images through different workstations. A trained AI-model needs to run in the PACS or communicate with the PACS. Large hospitals do in general not have a single firewall, but several external and internal firewalls who stop, segment and inspect the network traffic. To convince legal experts and the local IT at each hospital to configure / open the firewall(s) for FL between hospitals is one of the major hurdles for using FL in a clinical setting. A more technical hurdle is to make sure that different hospitals can communicate with each other, as the network architectures may differ and the hospitals need to agree on a protocol for data transfer and data encryption. Here the ASSIST partners can get some inspiration from what has already been done with federated learning in the United Kingdom (Rieke et al., 2020).

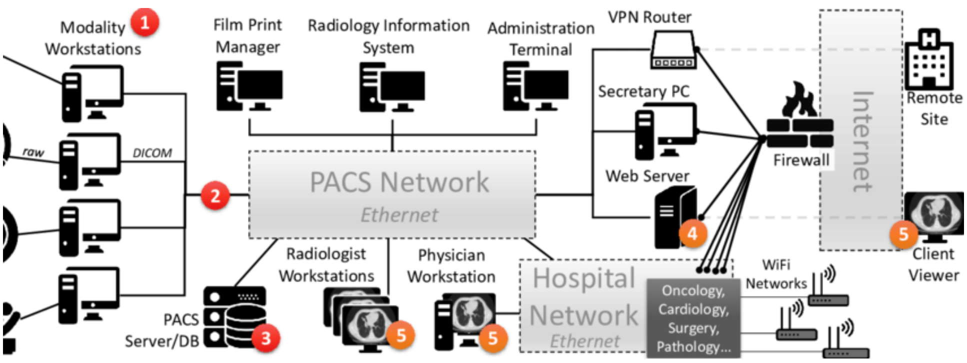

Figure 10. Overview of an IT system at a hospital, focusing on the PACS which is the heart for all imaging activities. The PACS receives images from all imaging modalities (CT, MR, ultrasound, etc) and radiologists can then look at the images through different workstations. One or several firewalls prevent sensitive data from being accessed from outside the hospital. Radiologists working at another hospital can in some cases view images in the PACS by logging in through a VPN.

7.2 Combining image data and other clinical data

It is becoming increasingly common to combine images and clinical data (e.g. sex, age, genetic data, diagnostic history, treatment history) from health care records to further improve deep learning prediction accuracy (e.g. Huang et al., 2020), and this will make federated learning more complicated as each node then needs to access image data as well as clinical data for the same patient. Figure 11 shows how Linköping university hospital (located in region Östergötland, Sweden) in collaboration with CMIV (Center for medical image science and visualization, Linköping university, Sweden) envisions how researchers can get access to both types of data for training AI algorithms. One challenge with combining different types of data is that the anonymization becomes harder.

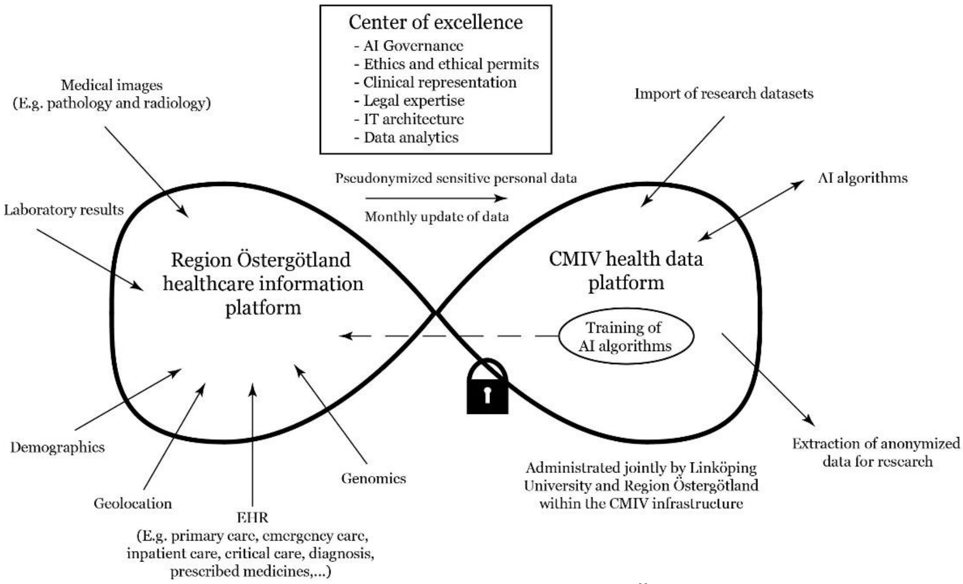

Figure 11. A sketch of the relationship between region Östergötland healthcare information platform (storing clinical data from hospitals in Östergötland for health care) and the CMIV health data platform (focusing on research and training of AI algorithms). Image provided by Håkan Gustafsson at CMIV.

7.3 Required hardware

The required hardware of each FL node depends on the type and amount of data to be used for training the AI model. For clinical data from health care records (e.g. 50 values per patient), it is normally sufficient to use a strong CPU for the training. For image data it is normally required to have one or more graphics cards in each node, as a CNN performs many time-consuming convolutions of high-resolution images in every training iteration. The required specification of the graphics cards depends on if the CNNs work in 2D or in 3D, as 3D CNNs typically require more GPU memory. For 2D CNNs it may be sufficient if the graphics cards have 8 – 11 GB of memory, while 3D CNNs may require 24 – 48 GB of memory.

8 Legal considerations of FL for medical data

As methods for FL already exist, which can even guarantee that sensitive data is not communicated between the nodes (e.g. Li et al., 2019), one of the major challenges for using FL in a clinical setting are the legal aspects. In Sweden this is further complicated by the fact that the 21 hospital regions are allowed to have their own legal interpretation of the Swedish laws (what is allowed in one hospital region is therefore not necessarily allowed in another region, even though the regions are in the same country). To combine data from different hospital regions is in Sweden allowed for research, after ethical approval, but in general not for clinical work. Recently a legal review concluded that the Swedish patient data law needs to be changed to allow for general secondary clinical use of health care data (e.g. to combine data from different hospitals or hospital regions) (Genomic medicine Sweden, 2022).

In Belgium, the federated learning approach is made more difficult due to the large number of hospitals. Furthermore, the hospitals are split in 5 different categories. AZ (general hospital), UZ (university hospital), RZ (regional hospital), PZ (psychiatric hospital) and UPZ (university psychiatric hospital). Different laws and regulations may apply depending on the associated organisation, region or institution (general or university), making collaboration using clinical data challenging. Most commonly, data is used for research purposes.

The interpretation of GDPR can also differ between countries, and to use FL between countries inside and outside the European union can be even more complicated. In the United Kingdom, federated learning has already been used between different hospitals (Rieke et al., 2020), and a reason for this is that their laws are less restrictive compared to other countries. As the regulations can differ between the countries in the ASSIST project, we provide a small overview in Table 2.

	Sweden	Netherlands	Belgium	Turkey
Data protection regulations	GDPR	GDPR	GDPR	KVKK (Personal Data Protection Law)
Number of hospital regions / organisations / hospitals	21 hospital regions, 85 hospitals	101 regular hospitals, 8 academic medical centers.	164 hospitals, some collaborate under an association.	N/A
Interpretation of national laws	Depends on hospital region.	All hospitals have to comply with the GDPR, no differences.	Every hospital is GDPR compliant, but differences apply per hospital.	Nationwide
Combining data between	Allowed for	For exchange of	This is covered by	It is free for research reasons

hospitals in different regions / organisations	research (after ethical approval) , in general not for clinical work.	data, a data transfer agreement is necessary. Next to the permission of the science committee which also determines whether an approval is needed from the ethics committee or not. Exchanging data for clinical work depends on the care relation of the physician. It is only allowed when it adds to the care of the patient.	collaborative contracts per category/region for providing healthcare services, physicians request and send identifiable data for specific patients amongst these regions in order to provide the best quality of care, anonymized data is still very sensitive, only validated for a specific purpose and hospitals still see even this data as their own data.	with the obligation of anonymization. For other reasons, it depends on the choice of the patient. The citizens who have an account for E-Nabiz system (National EHR system) determine their privacy policy by selection of one of the options presented by the application. They are informed about the scope and results of their selection in detail. For the citizens they don’t have this account the principles mentioned below are applied. · The family practitioner of the person can Access the information without any time limit. · When an appointment with a physician is taken, this physician can access the EHR of the patient limited by the date of the appointment. · The physicians of health care provider organization which was applied by the patient to take medical service can access the EHR of the patient limited with 24

				hours. · When the patient is hospitalized, the physician of the hospital can access the EHR until the discharge of the hospital.
Accessing images at collected / stored at hospital A from hospital B	Manually through tele radiology, or through external review (logging in to PACS at hospital A).	Exchange of data is possible when there is a data transfer agreement and approval of the involved hospitals and if necessary, patient consent. The exchange can be done by e.g. xNAT connections. There is not a specific national exchange platform for this. Direct logging into the PACS is not allowed.	This is covered by collaborative contracts per category/region for providing healthcare services, physicians request and send identifiable data for specific patients amongst these regions in order to provide the best quality of care, anonymized data is still very sensitive, only validated for a specific purpose and hospitals still see even this data as their own data.	By using the TeleRadiology system of the Ministry of Health due to above mentioned conditions. For research goals; the other methods can be used by regarding the obligation of anonymization.

Table 2. An overview of data protection regulations, number of hospitals, interpretation of national laws, and how to combine or access data from different hospitals, for the participating countries in the ASSIST project.

9 Conclusion

In this document we have presented the main concepts and requirements of FL for training deep networks in the domain of medical imaging. Furthermore, we have discussed the benefits and challenges of using FL for several of the use cases in the ASSIST project. The legal regulations in the different countries in ASSIST have also been discussed, as this is an especially important part for FL in the medical domain.

10 References

American Cancer Society. *Cancer Facts & Figures 2010*. Atlanta: American Cancer Society,https://web.archive.org/web/20150114080100/http://www.cancer.org/acs/groups/content/%40nho/documents/document/acspc-024113.pdf

Antunes, R. S., da Costa, C. A., Küderle, A., Yari, I. A., & Eskofier, B. (2022). Federated Learning for Healthcare: Systematic Review and Architecture Proposal. *ACM Transactions on Intelligent Systems and Technology (TIST)*.

Bashyam, V. M., Doshi, J., Erus, G., Srinivasan, D., Abdulkadir, A., Singh, A., ... & iSTAGING and PHENOM consortia. (2022). Deep Generative Medical Image Harmonization for Improving Cross- Site Generalization in Deep Learning Predictors. *Journal of Magnetic Resonance Imaging*, *55*(3), 908-916.

Comelli, A.; Dahiya, N.; Stefano, A.; Vernuccio, F.; Portoghese,M.; Cutaia, G.; Bruno, A.; Salvaggio,G.; Yezzi, A. Deep Learning-BasedMethods for Prostate Segmentation in Magnetic Resonance Imaging. Appl.Sci. 2021, 11, 782. https://doi.org/10.3390/app11020782

Ekmefjord, M., Ait-Mlouk, A., Alawadi, S., Åkesson, M., Stoyanova, D., Spjuth, O., ... & Hellander, A. (2021). Scalable federated machine learning with FEDn. arXiv preprint arXiv:2103.00148.

Feki, I., Ammar, S., Kessentini, Y., & Muhammad, K. (2021). Federated learning for COVID-19 screening from Chest X-ray images. Applied Soft Computing, 106, 107330. doi:10.1016/j.asoc.2021.107330

Genomic medicine Sweden, https://genomicmedicine.se/2022/03/28/lagliga-utrymmet-for-att-dela-genetiska-data-for-vard-av-andra-ar-begransat/, 2022 (in Swedish)

Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ digital medicine*, *3*(1), 1-9.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), 1-210.

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., ... & Feng, A. (2019). Privacy-preserving federated brain tumour segmentation. In *International workshop on machine learning in medical imaging* (pp. 133-141). Springer, Cham.

Leece, R., Xu, J., Ostrom, Q. T., Chen, Y., Kruchko, C., & Barnholtz-Sloan, J. S. (2017). Global incidence of malignant brain and other central nervous system tumors by histology, 2003–2007. *Neuro-oncology*, *19*(11), 1553-1564.

Low, G., Panu, A., Millo, N., & Leen, E. (2011). Multimodality imaging of neoplastic and nonneoplastic solid lesions of the pancreas. *Radiographics*, *31*(4), 993-1015.

Ma Z, Zhang M, Liu J, Yang A, Li H, Wang J, Hua D and Li M (2022) An Assisted Diagnosis Model for Cancer Patients Based on Federated Learning. Front. Oncol. 12:860532.doi: 10.3389/fonc.2022.860532

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.

McIntosh, C., Conroy, L., Tjong, M.C. et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. Nature Medicine 27, 999–1005 (2021)

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, *34*(10), 1993-2024.

Pancreatic cancer research fund, https://www.pcrf.org.uk/about-pancreatic-cancer/facts-figures/, 2022

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, *3*(1), 1-7.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Tedeschini, B. C., Savazzi, S., Stoklasa, R., Barbieri, L., Stathopoulos, I., Nicoli, M., & Serio, L. (2022). Decentralized Federated Learning for Healthcare Networks: A Case Study on Tumor Segmentation. *IEEE Access*.

Yi, L., Zhang, J., Zhang, R., Shi, J., Wang, G., & Liu, X. (2020). SU-net: an efficient encoder-decoder model of federated learning for brain tumor segmentation. In *International Conference on Artificial Neural Networks* (pp. 761-773). Springer, Cham.

Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, *5*(1), 1-19.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).

Brain tumour dataset for radiation therapy, https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=95224486

Prostate enlargement image, https://commons.wikimedia.org/wiki/File:%D8%A7%D9%84%D8%AA%D9%87%D8%A7%D8%A8_%D9%BE%D8%B1%D9%88%D8%B3%D8%AA%D8%A7%D8%AA.jpg