Funded by
the European Union

D2.1 SoA of FL-tailored AI models, PET methods and tools

Lead Author: Shaila Calvo (GRAD)
With contributions from: Carmen García (GRAD), Adrián Vázquez (GRAD), Carlota Cañamero (GRAD), Jan Ramon (INRIA) and Twaha Mukammel (TVS)
Reviewers: Zeev Pritzker (AVO), Anna Nogué (QBIM), Gemma Urbanos (QBIM), Alejandro Vergara (QBIM)

Deliverable nature	**Report (R)**
Dissemination level	Public (PU)
Delivery date	30-04-2024
Version	2.0
Total number of pages	46
Keywords	Federated Learning (FL), csPCa, Privacy-enhancing Technologies (PETs), Trusted Execution Environments (TEEs), Secure Multi-party Computation (SMPC), aggregation methods

EXECUTIVE SUMMARY

This document reports the research results corresponding to Task 2.1 “Research FL-tailored AI models, aggregation methods and PET tools” of the FLUTE “Federated Learning and mUlti-party computation Techniques for prostatE cancer” project. In line with what it is expected from this task, the information gathered in this deliverable can be divided into three fundamental aspects:

- **An analysis of Artificial Intelligence (AI) models** from two complementary perspectives. On the one side, the application of AI to the healthcare field, focusing on the clinical use case targeted by this project: clinically significant prostate cancer (csPCa) prediction, with the aim of driving the development of FLUTE’s predictive models based on this research, together with the clinical support provided by the project’s clinical partners. On the other hand, the application of federated ML models, or Federated Learning (FL), ranging from its high impact in clinical field to identifying the state-of-the-art central aggregation methods to be used in the federated setting of the FLUTE project.

- **State-of-the-art study of Privacy-enhancing Technologies (PETs)** that can potentially be used for ensuring privacy in the federated FLUTE environment. In this regard, two approaches will be explored: software methods and hardware methods. In the former case, Secure Multi-party Computation (SMPC) has been explored from a scalability point of view. In the latter, the study was focused on the advantages of using Trusted Execution Environments (TEEs) in FL.

- **A description of the software tools to be used in FLUTE project**. On one hand, the use of specific tools for FL settings, with a particular focus on PySyft technology, will be defined. On the other hand, the software tools that will be provided by the partner QBIM to extract information from Magnetic Resonance Images (MRI) to be used for csPC predictive models training, will also be described.

The research conducted in this task builds upon the work developed and already reported in the TRUMPET project, which has been expanded in this deliverable.

DOCUMENT INFORMATION

Grant agreement No.	**101095382**	**Acronym**	**FLUTE**
Full title	**Federate Learning and mUlti-party computation Techniques for prostatE cancer**		
Call	HORIZON-HLTH-2022-IND-13-02		
Project URL	https://cordis.europa.eu/project/id/101095382		
EU project officer	Mrs Serena Battaglia		

Deliverable	**Number**	**D2.1**	**Title**	**SoA of FL-tailored AI models, PET methods and tools**
Work package	Number	WP2	Title	Scalable privacy enhanced Federated Learning and AI
Task	Number	T2.1	Title	Research FL-tailored AI models, aggregation methods and PET tools

Date of delivery	**Contractual**	**M12**	**Actual**	**M12**
Status	2.0 ☒ Final version			
Nature	☒ R ☐ DEM ☐ DMP ☐ DEC ☐ ETHICS ☐ OTHER			
Dissemination level	☒ Public ☐ Sensitive			

Authors (partners)	**GRAD, INRIA, TVS**
Responsible author	Shaila Calvo Almeida
	scalvo@gradiant.org

Summary (for dissemination)	This deliverable contains an analysis of AI models and central aggregation methods for the FL setting, as well as state of the art of the PET methods and SW tools for FL scenarios.
Keywords	Federated Learning (FL), csPCa, Privacy-enhancing Technologies (PETs), Trusted Execution Environments (TEEs), Secure Multi-party Computation (SMPC), aggregation methods

VERSION LONG			
Issue Date	**Rev. No.**	**Author**	**Change**

31/10/2023	0.1	Shaila Calvo, Carmen García, Adrián Vázquez (GRAD); Jan Ramon (INRIA); Twaha Mukammel (TVS)	First draft including document's structure, work distribution between partners and introductions to each of the technologies.
31/01/2024	0.2	Shaila Calvo, Carmen García, Adrián Vázquez (GRAD); Twaha Mukammel (TVS); Jan Ramon (INRIA)	Advancing the state of the art of each of the technologies. Re-structuring of some sections.
31/03/2024	0.3	Shaila Calvo, Carmen García, Carlota Cañamero, Adrián Vázquez (GRAD); Twaha Mukammel (TVS); Jan Ramon (INRIA)	Research final version
10/04/2024	1.0	Shaila Calvo (GRAD)	Document assembly and final complete version
12/04/2024	1.1	Zeev Pritzker (AVO)	First revision
16/04/2024	1.2	Anna Nogué, Gemma Urbanos, Alejandro Vergara (QBIM)	Second revision
29/04/2024	2.0	Shaila Calvo (GRAD)	Inclusion of revisions. Final document.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ADC: Apparent Diffusion Coefficient
AI: Artificial Intelligence
AL: Active Learning
ANN: Artificial Neural Network
bpMRI: bi-parametric MRI
CART: Classification and Regression Tree
csPCa: clinically significant Prostate Cancer
CT: Computed Tomography
DCE: Dynamic Contrast-Enhanced
DL: Deep Learning
DNNs: Deep Neural Networks
DRE: Digital Rectal Examination
DWI: Diffusion-weighted Imaging
EES: Extravascular Extracellular Space
EHR: Electronic Health Records
FDA: Food and Drug Administration
FL: Federated Learning
FOSTER: Federated Out-of-Distribution Synthesizer
GDPR: General Data Protection Regulation
HSMs: Hardware Security Modules
LR: Logistic Regression
LV: Lesion Volume
ML: Machine Learning
mpMRI: multi-parametric MRI
MRI: Magnetic Resonance Image
MRTB: MRI-targeted biopsy
ncsPCa: non-significant csPCa
OOD: Out-of-Distribution
PET: Privacy-enhancing technologies
PSA: Prostate-specific antigen
PSAD: PSA density
PV: Prostate Volume
QIB: Quantitative Imaging Biomarker
RF: Random Forest
SGD: Stochastic Gradient Descent
SMPC: Secure Multi-party Computation
sSVM: sparse Support Vector Machine
SV: Shapley Value
T2WI: T2-weighted Image
TEE: Trusted execution environment
TPM: Trusted Platform Module
TML: Traditional ML
WP: Work Package

1. Introduction

In the last decade, Artificial Intelligence (AI) has undergone exponential growth, revolutionising the ways in which we interact with technology and reshaping decision-making across diverse sectors. Currently, we can say that AI is part of everyday life, making headlines in the news daily due to its remarkable advances. Everyone has heard about AI and how it is changing our day-to-day lives. Among the most talked-about technologies we can find, for example, self-driving cars, recommendation systems used by streaming platforms like Netflix or HBO Max, or fraud detection systems. In addition, it could be said that in recent years, AI technologies have generated great interest among the general population with the emergence of tools such as ChatGPT[1] or image generation tools on demand, like DALL-E[2].

As of today, AI is present in all sectors of society, including manufacturing, education, banking or medicine. However, in the realm of AI, advances in innovation face a challenge due to limited access to high-quality datasets that enable the development of truly useful and efficient algorithms. Unfortunately, this is greatly evident in one of the sectors where AI can have great potential: the healthcare sector. Clinical data are typically scattered across several locations, including hospitals, clinics, and medical devices, so that to obtain large training datasets for the algorithms, it is often necessary to aggregate all this information in one node. The consolidation of clinical data from diverse sources raises concerns about privacy and security. The health data, being highly sensitive, is subject to stringent privacy regulations, like the General Data Protection Regulation (GDPR) in the European Union or hospital's Ethical Committees. Due to these concerns, it is necessary to search for alternatives to train more efficient algorithms, making critical the technologies that address privacy and security issues. Seeking to be one of these alternatives, Federated Learning (FL) emerges as a relevant solution.

Under the research framework of the **FLUTE** project, a scalable privacy enhanced FL environment will be designed, deployed and tested, which will host the training of Deep Neural Networks (DNNs) based algorithms for clinically significant prostate cancer (csPCa) prediction. This document fits the framework of **WP2 "Scalable privacy-enhanced Federated Learning and AI"** and collects a study of the state-of-the-art of the different technologies that will be researched and developed for this purpose.

As shown in Figure 1, the work carried out in WP2 will follow the specifications defined in WP1 based on the needs and requirements of users and stakeholders (T1.1 and T1.2) and will comply with the legal and ethical guidelines and standards defined in WP6. Technological outcomes of WP2 will be integrated, similarly to those in WP3, in the FLUTE platform developed in WP4. Furthermore, WP2 work will be used as input for WP9, where WP2 results will be examined and reused in terms of AI and the developed PETs will be used for secure training. Finally, these technologies will be validated in WP5.

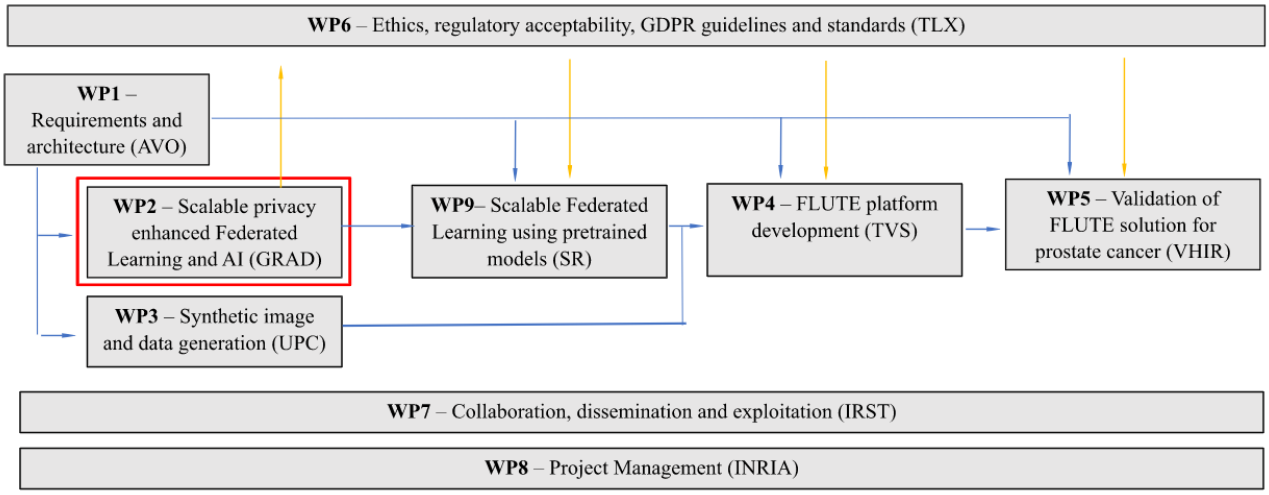

Figure 1. Relationship of WP2 with the different FLUTE WPs

This document is directly related to T2.1 "Research FL-tailored AI models, aggregation methods and PET tools" and its main objective is to report an analysis of AI models and different aggregation methods for a FL environment, as well as the state of the art in PET (Privacy-enhancing Technologies) methods and software tools for FL environments:

- **Section 2** aims to analyse the state-of-the-art of FL with a special focus on the medical field, and more specifically on the same use case as FLUTE: csPCa prediction. It will start by analysing the relevance of AI in the clinical field and providing a detailed analysis of several AI-based research on csPCa prediction. Furthermore, it will show how FL emerges as a fundamental tool in the clinical domain. Different studies using FL solutions in the clinical setting will be presented, ending with some examples related to csPCa prediction. Finally, different techniques to enhance AI model performance in a federated environment will be discussed and also different available aggregation methods for FL environments will be described.
- **Section 3** will be focused on the study of PET methods that ensure information privacy and also the scalability of results in a FL environment. Specifically, the study conducted here focuses on two types of PET technologies, one of them software-based and the other, hardware-based: Secure Multi-party Computation (SMPC) and TEEs (Trusted Execution Environments).
- **Section 4** will discuss the different frameworks, tools and software that will be used in the FLUTE. It will focus, on the one hand, on the tool that will be used in the FLUTE's FL environment: Pysyft. On the other hand, two software tools will be presented, that will be used within the framework of WP2 to process MRI images and extract Quantitative Imaging Biomarkers (QIBs) for use as input in the csPCa prediction algorithms: the QP-Care® and the QP-Prostate.

It is worth mentioning that this document builds on the work already carried out and reported in the **TRUMPET**[3] **"TRUstworthy Multi-site Privacy Enhancing Technologies"** project. Therefore, with the aim of avoiding redundant information, some sections of this document may include references to this project.

1.1. Study outcomes and WP continuity

From the analysis of the various technologies in this study, the best techniques will be selected for use in the development of subsequent tasks within the WP. As shown in Figure 2, this study will yield insights into how the csPCa prediction algorithms should be approached, as well as the techniques to enhance their performance and convergence within the scope of T2.2 "Design and development of AI models for prostate cancer diagnosis and aggregation methods for FL scenario". Furthermore, it will provide the necessary foundation to start working on software-based and hardware-based PETs, the responsibility of T2.3 "Design, development and assessment of software-based PETs for FL settings" and T2.4 "Design, development and assessment of hardware-based PETs for FL settings".

In this WP, validation of the PET methods will also be performed in task T2.5 "Validation and benchmarking of combined PET methods for privacy-armored FL", where various combinations of hardware and software-based PET techniques will be studied to find the optimal combination for the FL environment. The validation of the csPCa prediction model will be carried out in WP5 "Validation of FLUTE solution for prostate cancer prediction".

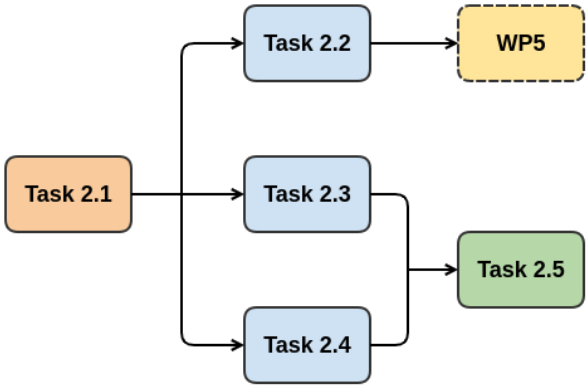

Figure 2. Relationship of FLUTE WP2 tasks and with WP5

2. State-of-the-art on FL-tailored AI models for prostate cancer prediction

Artificial Intelligence (AI) is at the forefront of computer science, dedicated to the development of systems capable of emulating and performing tasks that traditionally require human intervention and intelligence. These systems can learn from data, recognise patterns, make decisions and adapt to new situations; abilities that make AI a very useful tool that is increasingly used across various sectors of society.

AI offers innovative and efficient solutions to a diverse array of challenges, ranging from automating tedious tasks to optimising complex processes. Its capacity to analyse extensive datasets and generate useful insights has transformed industries such as healthcare, manufacturing, finance and many other, offering tangible benefits in terms of efficiency, accuracy and quality of life.

Machine Learning (ML) is one of the most widely employed AI technologies. It is a sub-field of AI that enables machines to learn patterns and perform tasks without being explicitly programmed, providing a unique capability to analyse data and offer automated predictions or decisions. ML is a broad field that encompasses several types of AI technologies which offer a wide range of possibilities depending on the type of data available for training and the task to be addressed. Among the most well-known ML techniques are technologies such as regression models, decision trees, clustering technologies and Deep Learning (DL) algorithms. Chauhan et al.[4] defines DL as branch of ML designed to replicate the function of the human cerebral cortex. DL algorithms are representations of Deep Neural Networks (DNNs), which are characterised by numerous hidden layers. The neurons in these networks operate in a distributed manner, collaboratively learning from the input to produce a desired output. This approach has proven to be particularly effective in the interpretation of unstructured data such as images, sound and text, leading to significant advances in fields such as computer vision, natural language processing and speech recognition. The broad capabilities have made these techniques widely used in many sectors of the society, from traffic control or detecting imperfections on production lines, to interpreting medical images.

Federated Learning (FL)[5] [6] is a field of AI that extends the principles of ML by allowing models to be trained in a decentralised way at the different data locations, rather than the traditional approach of centralising data to train models. This allows to avoid the need for direct transfer of raw data, avoiding potential privacy and security issues.

This section will examine AI models from two main perspectives. Firstly, it will introduce the significant impact of AI in the medical field, delving into solutions that address the same use case as the FLUTE project: predicting csPCa. Next, it will discuss the relevance of FL tools in the clinical setting, concluding with some examples of FL application in csPCa prediction. Finally, various methods to enhance FL algorithms' performance will be analysed, as well as the different aggregation methods for FL studied in the state-of-the-art.

2.1. AI-powered csPCa prediction models

One highly useful application of AI is clinical practice. According to a study by S. Secinaro et al.[7], since the 1950s, doctors have attempted to improve diagnosis using computer-assisted programs[8] [9]. Since then, interest and advances in AI medical applications have increased significantly due to the major technological advancements and the vast amount of data available for collection and utilization.

AI algorithms have shown tremendous efficiency in analysing large amounts of data of diverse nature and finding patterns sometimes hidden to the human eye. In the clinical domain, these technologies are proving to be of great assistance in analysing data from various sources, such as electronic health records (EHR) of patients, medical images such as X-rays or magnetic resonance imaging (MRI), real-time patient data, hospital management data, and even, more recently, genomic data of individuals. A large number of studies and research focusing on the application of AI in different health fields have been published. These range from detecting possible patterns of depression or anxiety from patient data or automatically identifying tumours in medical imaging, to identifying risk factors for certain diseases, discovering new relevant biomarkers, or even optimising clinical healthcare resources. The possibilities are as numerous as ailments, types of data, and technological opportunities.

The application of intelligent algorithms and ML systems is revolutionising the ways in which diseases are diagnosed, treated, and managed today. In addition to providing valuable support to clinical staff in decision-making, these tools provide tremendous support to patients in terms of monitoring and empowerment. As part of the FLUTE study, the efficacy of AI in detecting diseases will be demonstrated, with a specific focus on csPCa prediction.

According to the WHO, cancer is one of the leading causes of death globally, responsible for nearly 10M deaths per year (statistics from the year 2020). The same study reported approximately 1.41M cases of prostate cancer around the world, placing this disease among the top five types of cancer with the highest incidence[10]. Thus, prostate cancer stands as one of the prevailing malignancies among the global male population.

When making decisions regarding prostate cancer treatment, the ability to differentiate tumours of clinical significance is paramount and aims to prevent both unnecessary treatment and instances of underdiagnosis. csPCa involves the presence of cancer cells in the prostate with the potential to grow and spread. Uncontrolled proliferation of cancer cells can lead to the formation of a tumour, and if these cells manage to leave the prostate and enter the bloodstream or the lymphatic system, they can travel to other parts of the body, creating metastasis. In contrast, non-clinically significant prostate cancer implies the presence of cancer cells in the prostate, but the cells do not pose the same risk of aggressive growth and spread. For this reason, early detection and accurate assessment of the grade and stage of prostate cancer are crucial to determining the risk and aggressiveness of the disease, and to guiding treatment decisions. Conventional detection techniques, including

prostate-specific antigen (PSA) testing and biopsy, while useful, are invasive and prone to false positives or negatives. And this is where new technologies have emerged as a promising tool to enhance accuracy and efficiency in csPCa prediction.

Many researchers have focused their studies in recent years on trying to find alternative ways to identify or predict cases of csPCa in an efficient and less invasive way. According to a study conducted by D.J. Van Booven et al.[11], the use of AI has been shown to be beneficial in assisting with pathological classification to assess stratification and treatment of prostate cancer. Furthermore, AI shows promise in automating the evaluation of prostate cancer characterization and severity based on clinical data and image-based tasks.

2.1.1. Clinical variable-based models

Focusing on the use of clinical data for the detection of this type of cancer, we encounter several interesting studies that are based mainly on the use of variables such as serum PSA or digital rectal examination (DRE). Identifying elevated levels of PSA is a common clinical approach for diagnosing prostate cancer. However, elevated PSA levels can also occur in various benign prostatic conditions. Approximately 75% of men who undergo a prostate biopsy do not have cancer even if they have elevated PSA levels. The overdiagnosis leads to unnecessary overtreatment of prostate cancer with undesirable side effects, such as incontinence, erectile dysfunction, infections, and pain[12]. For this reason, many studies have assessed the utility of PSA and other clinical characteristics in the context of AI for detecting prostate cancer and/or its progression. Such research carried out by Morote J. et al. at Vall d'Hebron Institute (VHIR): *"The Barcelona Predictive Model"*[13]. This model, based on a binary logistic stepwise regression, has been able to predict csPCa with a good accuracy (AUROC 0.89) on a cohort of 1,486 patients from the Barcelona's area, using some clinical variables as input: age at the time of biopsy, Ca family history, initial or repeated biopsy, serum PSA, DRE, prostatic volume and PI-RADS category. Another related article by M. Stojadinovic et al.[14] described a Classification and Regression Tree (CART) model that could be used to identify patients with csPCa based on clinical data such as age, PSA, DRE, prostate volume, and PSA density (PSAD). This model resulted in an 0.833 area under the receiver operating characteristic curve and determined that PSA density was the most decisive variable, showing that the decision tree provided a net benefit compared to a logistic regression model.

Other researchers have been analysing the use of techniques based on Artificial Neural Networks (ANNs) for the detection of such ailments for over 20 years. B. Djavan et al.[15] conducted a study aimed at prospectively developing two ANNs for early detection of prostate cancer in men with total PSA levels of 2.5 to 4 ng/mL and 4 to 10 ng/mL. The area under the receiver operating characteristic curve (ROC AUC) was 0.876 and 0.913 for the ANN models of 2.5 to 4 ng/mL and 4 to 10 ng/mL, respectively. Another notable example is the study conducted by Finne et al.[16], which aimed to assess whether an ANN (multilayer perceptron, MLP) and Logistic Regression (LR) could provide an additional reduction of false-positive PSA results compared to the proportion of free PSA in prostate

cancer screening. The model was built using data on total PSA, free PSA ratio, DRE, and prostate volume. The results showed that with a sensitivity level of 95%, a 19% reduction of false-positive PSA results could be achieved using the free PSA ratio, compared to 24% with the logistic regression model and 33% with an ANN.

Recent studies also focus on the advanced use of DL techniques to aid in the diagnosis of csPCa. F. Gentile et al.[17] developed a DL model based on ANN with an input layer, an output layer, and 7 hidden layers. The output layer returns a value between 5 to 10, representing the hypothesised Gleason score of the cancer. The clinical variables inserted into the model are total PSA, free PSA, p2PSA, PSA density, and age. Their model achieved sensitivity values as high as 86% and a specificity of 89%.

As seen from this literature analysis, numerous studies have been conducted for the diagnosis of csPCa using only clinical variables such as PSA or DRE, along with other information such as patient age. All this research, ranging from simple AI models to more complex techniques, have shown high performance in detecting this ailment, being a great support for clinicians to effectively diagnose positive cases and avoid over-diagnosis, thus reducing unnecessary testing.

2.1.2. Magnetic Resonance Imaging (MRI)-based models

A very relevant technique for detecting csPCa, in addition to evaluating values such as PSA or DRE, is the assessment of MRI results. The two known modalities of this type of technique, multiparametric MRI (mpMRI) and biparametric MRI (bpMRI), are powerful tools in the assessment of prostate cancer, as they provide a detailed and accurate image of the prostate gland and the surrounding tissues. mpMRI employs three imaging sequences — typically T2-weighted, diffusion-weighted (DWI), and dynamic contrast-enhanced (DCE) images— and demonstrates high sensitivity and specificity in detecting csPCa. Current guidelines advocate its use before biopsy. However, the accuracy of DCE is currently being debated. bpMRI, using only T2 and DWI, has been suggested as a feasible alternative[18].

In the classic diagnostic workflow of csPCa detection using MRI information, the clinician is responsible for analysing the image and associating the severity based on their observation. As explained by H. Lu et al.[19], most clinical diagnoses follow a consensus reporting standard with the adoption of Prostate Imaging Reporting and Data System (PI-RADS)[20], which provides qualitative guidelines for clinical assessment. However, there is variability in the interpretations of scans among radiologists, which can be attributed in part to the steep learning curve required for scan interpretation[21]. Therefore, in recent years, there has also been a lot of interest in the study of Quantitative Imaging Biomarkers (QIBs). Many studies have shown that these specific characteristics that can be extracted from MRI scans can help predict csPCa and improve the performance of PI-RADS[22]. A QIB refers to an objective and numerical measure used to assess the presence or progression of a disease, health status, or treatment response, as opposed to a

qualitative biomarker, which relies on subjective observations. When referring to images, particularly MRI images, a quantitative biomarker involves precise numerical measurements obtained from MRI images to quantify specific tissue characteristics under examination. In the research on the diagnosis of csPca through MRI image analysis, several interesting QIBs have been identified. Some of these are mentioned below:

- **Apparent Diffusion Coefficient (ADC)** measures the mobility of water in tissues. Areas with csPCa tend to have lower ADC values due to restricted cellular movement. A low ADC can indicate the presence of aggressive tumours.

- **Diffusion Fraction (f)** evaluates how ADC varies with different echo times. Specifically, it can provide information about the mobility of water in tissues, which is useful for characterizing prostate cancer.

- **Ktrans**. The transfer constant between the vascular and the extravascular extracellular space (EES). This measure assesses vascular permeability, that is, how easily molecules can pass from the bloodstream to the extravascular extracellular space. It is extracted from modelling the contrast perfusion in a DCE sequence, and it is related to tumor vascularisation.

- **Kep**. This measure quantifies the washout, which is the process opposite to Ktrans. While Ktrans evaluates the entry of molecules from the vascular to the extravascular space, Kep assesses the rate at which those molecules are cleared from the extravascular space and return to the bloodstream. It is extracted from modelling the contrast perfusion in a DCE sequence, and it is related to tumor vascularisation.

- **Ve**. The fraction of volume of EES. This measure indicates what proportion of the extravascular space is occupied by the contrast agent or the molecules of interest. It is a measure of the distribution of these molecules in the extravascular extracellular space relative to the total volume of that space. It is extracted from modelling the contrast perfusion in a DCE sequence, and it is related to tumor vascularisation.

- **Radiomics,** which refers to high-throughput quantitative imaging features (also known as texture characteristics) in MRI images, such as tumour tissue heterogeneity, that can provide information about the aggressiveness of prostate cancer.

- **Tumour-to-Gland Volume Ratio (TGV)**. This measure compares tumor volume to the total volume of the prostate gland and can help to assess the relative extent of the tumor in the prostate gland.

Among the most prominent metrics are the aforementioned ADC and DCE. Some researchers have focused on demonstrating that these types of metrics can be highly relevant in distinguishing cases of prostate cancer[23]. Recently, radiomic features have also proven to be very useful in assessing the severity of the disease[24]. ÁS Iglesias. et al.[25] published a study aiming to identify potential imaging

biomarker profiles (perfusion/diffusion + radiomic features) extracted from MRI that could discriminate patients according to their risk or the occurrence of biochemical recurrence (BCR) 10 years after diagnosis, as well as to evaluate their predictive value with or without clinical data. In this study, which involved the FLUTE project partner QBIM, three types of characteristics were extracted: texture/radiomic features and quantitative parameters (diffusion and perfusion features) from T2-w and DWI and DCE sequences respectively. In their results, they observed that prostate region-wise imaging biomarker profiles, mainly composed of radiomic features, allowed for discriminating between risk groups and patients with BCR. Overall, the image biomarker profiles retained good predictive capability (AUC values exceeding 0.725 in most cases), which improved overall when some clinical variables were included.

Seeing the utility of this type of data, some researchers have ventured into developing AI models capable of predicting cases of csPCa based on them. DJ Winkel et al.[26] published a study that investigated whether supervised ML classifiers could predict csPCa from a set of quantitative image features and compared these results with established PI-RADS v2 assessment scores. To do this, they took perfusion maps, ADC, and absolute T2-signal intensities as input. Specifically, they trained four AI models for this task: Gradient Boosting Machines (GBM), Neural Networks (NNet), Random Forest (RF), and Support Vector Machines (SVM). All ML models outperformed PI-RADS v2 assessment scores in the prediction of csPCa (RF, GBM, NNet, and SVM vs. PI-RADS: AUC 0.899, 0.864, 0.884, and 0.874 vs. 0.595). These results clearly indicate that QIBs contain relevant information for csPCa prediction from image features and that AI can lead to a significant advancement and aid in its analysis.

As can be seen, MRI is a key tool for clinicians in diagnosing cases of csPCa and, as expected, AI has also taken a step forward in analysing such images. Thanks to the advancement of AI technologies, the study of DL modelling has enabled the interpretation of medical images, such as the aforementioned bpMRI and mpMRI, or ultrasound, with unprecedent accuracy. Specifically, the use of DL techniques for analysing this type of images in diagnosing csPCa is also a field of study for several researchers. Sun Z. et al.[27] compared the performance of radiologists in detecting MRI-visible csPCa in MRI with and without AI software. This software is based on four proprietary DL-based AI models: (i) MRI sequence classification, (ii) prostate gland segmentation and measurement[28], (iii) prostate zonal anatomy segmentation and (iv) csPCa foci segmentation and measurement. They found that this DL software could help to reduce false positive detections (specificity increased from 57.7% to 71.7%), improve reading times, and increase diagnostic confidence.

In a comparative endeavour, Z.Litao et al.[29] conducted a study on how different DL models based on bpMRI can achieve similar performance to the PI-RADS assigned by clinicians for csPCa diagnosis. They developed four 3D neural network models (ResNet3D[30], DenseNet3D, ShuffleNet3D, and MobileNet3D) with two objectives: classifying between benign and malignant lesions and classifying clinically significant and non-significant cancer. Additionally, they developed an integrated model

that combines PI-RADS and the DL-CS model, abbreviated as PIDL-CS. The performances of the DL and PIDL-CS models were compared with those of PI-RADS. The results, validated in four different hospitals, suggest that the performance of the DL-CS-ResNet model and PI-RADS are comparable, with the difference in ROC curves not being significant. Their proposed DL models can therefore serve as a potential non-invasive auxiliary tool for predicting csPCa. Furthermore, PIDL-CS significantly increased the specificity of csPCa detection compared to PI-RADS assessment by expert radiologists, thus greatly reducing unnecessary biopsies and aiding radiologists to get accurate diagnosis of csPCa. Another notable example is the study conducted by M. Hosseinzadeh et al.[31], aimed at assessing PI-RADS-trained DL algorithm performance and investigating the effect of data size and prior knowledge on detecting csPCa in biopsy-naive patients from bpMRI data. Their results showed a sensitivity for detecting PI-RADS ≥ 4 lesions of 87% and an AUC of 0.88. The sensitivity for detecting Gleason score > 6 lesions was 85%. They also concluded that AI for prostate MRI analysis heavily relies on data size and prior zonal knowledge, and that substantially more than 2,000 training cases are needed to achieve expert-level performance.

Furthermore, in a comparison between the use of mpMRI and bpMRI in the diagnosis of prostate cancer, Xu L. et al.[32] conducted a study in which they concluded that there is no significant difference between the two modalities. However, they point out that DCE MRI helps distinguish between csPCA and PCa in patients with bpMRI score of 3 or higher. Specifically, in cases where bpMRI is 4, the inclusion of DCE MRI facilitates clear identification of tumour aggressiveness and allows for individualized cancer treatment to be developed. Additionally, it emphasizes that sensitivity is higher when combining T2-weighted images (T2WI) with diffusion-weighted images (DWI) compared to evaluating T2WI alone.

However, while the use of DL techniques for this type of task has been previously mentioned, it is true that there are also studies that, using more traditional ML techniques, have demonstrated good results in the analysis of MRI information. In the review carried out by Sushentsev[33] titled *"Comparative performance of fully-automated and semi-automated artificial intelligence methods for the detection of clinically significant prostate cancer on MRI: a systematic review"*, a comprehensive comparison is carried out between the performances of DL models and traditional ML (TML) models. These models were specifically trained to discern between cases of csPCa and non-significant (ncsPCa).

According to this study, in the realm of DL, the recurrent use of Convolutional Neural Networks (CNNs) based on U-Net[34] stands out. Meanwhile, in the context of TML, Random Forest (RF), Logistic Regression (LR), and Support Vector Machines (SVM) are identified as the most employed for this purpose. Table 1 summarizes the best studied algorithms along with their AUC metrics and the inputs used.

Model	AUC	Inputs

Logistic Regression	0.98	T2WI, ADC
CNN (MISN)	PZ: 0.89 TZ: 0.97	ADC, BVAL, DWI0, DWI1, DWI2 Ktrans, T2WI-Cor, T2WI-Sag, T2WI-Tra
CNN (VGG16)	0.89	T2WI, ADC
Random Forest	WP: 0.88 PZ: 0.84 TZ: 0.89	T2WI, ADC, b=1500

Table 1. Performance Comparison of Different Models in clinically significant Prostate Cancer Detection (CNN: Convolutional Neural Network, AUC: Area under the curve, WP: Whole Prostate, PZ: Peripheral Zone, TZ: Transition Zone, MISN: Multi-Input Selection Network)

The results indicate that the best performance was achieved with a LR algorithm, followed by two CNN models and a RF. If the focus is on the type of data that has been used for training, it is observed that in over 82.35% of the cases studied, both in DL and TML, the models' inputs are T2WI images. Specifically, in DL models, 100% of the studies considered in this review use T2WI images, and 80% of them combine them with the ADC. Additionally, in 3 out of the 5 evaluated studies, T2WI images are combined with DWI. On the other hand, in TML architectures, 75% use T2WI images, and all of them incorporate the ADC metric. However, only in 3 out of the 12 evaluated studies are T2WI images, DWI, and the ADC combined. Finally, this comparative study highlights that most of the models were trained using the PROSTATEx public dataset[35]. It concludes by noting that the exclusive use of a single dataset can easily lead to overfitting, as it does not consider the variability in annotations among experts and limits the generalization of results.

2.1.3. Fusion models

It is worth mentioning that while most studies in the literature focus on the use of clinical data such as PSA or DRE, or on the use of features extracted from MRI, there are some studies that have focused on combining both types of data to train prediction algorithms for csPCa. A good example of this is the research carried out by X. Cheng et al.[36], whose objective was to develop and validate a predictive model based on clinical features and mpMRI to reduce unnecessary systematic biopsies in patients without prior biopsy suspected of prostate cancer. They used multivariable logistic regression analysis to determine independent predictors of csPCa on cognitive MRI-targeted biopsy, and they established and evaluated three different models: clinical, MRI, and fusion models. They found that the combined model achieved the best discrimination (AUC 0.88) compared to both the MRI model and the clinical model. Regarding clinical variables, PI-RADS score, index lesion (IL) on

the peripheral zone, age, and PSAD were considered independent predictors and included in the combined model.

Along the same lines, Z. Chen et al.[37] performed a statistical modelling on the use of mpMRI data together with PSA values to establish and validate a new diagnostic model for csPCa called the P.Z.A score. According to their analysis, this model works well enough to increase the detection rate of csPCa.

Furthermore, A. Hiremath et al.[38] conducted a study with the aim of constructing an integrated nomogram (referred to as ClaD) that combined DL-based image predictions, PI-RADS score, and clinical variables to identify csPCa in bpMRI. They trained two DL models, AlexNet[39] and DenseNet[40] for image predictions. The results showed that AlexNet outperformed DenseNet, particularly in its ability to differentiate between both grades of the disease. Furthermore, AlexNet enabled the identification of recurrences in patients who had previously overcome cancer. Among the clinical data they used as input, they specifically included prostate volume (PV), PI-RADS score, PSA, and lesion volume (LV). Their findings underscore the utility and potential of combining MRI images with clinical data to enhance the precision and clinical application of DL models in prostate cancer diagnosis and monitoring.

A study that is also highly interesting is the one carried out by M. Li et al.[41], where they aim to evaluate the potential of clinical-based model, a bpMRI-based radiomics model and a clinical-radiomics combined model for predicting csPCa. To this end, they developed three logistic regression models. The first model is based on radiomic features extracted from the MRI images. The second model takes as input risk factors such as age, total PSA, free PSA and PSAD. Lastly, the combined model integrates these two approaches, achieving a final AUC of 0.98.

This combined approach of clinical data and features derived from MRI will be pursued in the FLUTE project. We will develop a model that will use advanced DL techniques for csPCa prediction from clinical data such as age, PSA and DRE (starting from the conclusions already obtained by the partner VHIR in the aforementioned “Barcelona predictive model"), and combine it with information derived from MRI, based in particular on the QIBs extracted by the QBIM partner with their proven expertise (see Section 4.2 for more information on the tools to be used for this extraction).

2.2. FL-tailored AI models

2.2.1. Implementation of FL in medicine

The significance of advancements in the field of ML has been demonstrated, as well as how these technologies can provide substantial support in several application domains, such as healthcare, making an in-depth analysis of its use for the detection of csPCa. However, these remarkable advances usually encounter a crucial challenge: data scarcity. Most of the ML models need large amounts of data to be trained, which often reside in different locations. This distribution of data is, in fact, often desirable to improve the generalisation and robustness of models. The problem is that

centralising this data to train algorithms raises fundamental privacy and security concerns, especially when dealing with sensitive information. FL appears to address these concerns.

FL extends the principles of ML by allowing models to be trained in a decentralised way at the different data locations instead of centralising the data to train the models. This allows to avoid the need for direct transfer of raw data. Instead, what is shared are the parameters or weights of the model being trained. The fundamental idea is that models are trained on local nodes, and only the updates to those models are shared. McMahan et al.[42] discuss in their study the clear advantages of FL for privacy compared to data centre training. They say that keeping even an anonymised dataset can jeopardise privacy by linking it to other data. Instead, the information transmitted for FL is the minimum update needed to improve a particular model.

In the contemporary digital era, the convergence of advancements in technology and medicine has resulted in substantial progress in the diagnosis, prognosis, treatment and management of diseases. AI has become a pillar in the transformation of healthcare thanks to its ability to analyse large datasets, identify complex patterns and improve clinical decision-making. A great deal of research focuses on the use of ML algorithms on clinical data, be it raw data, tabular data or medical images. However, in the age of AI, innovation in healthcare is severely limited by the availability and accessibility of high-quality datasets. Researchers often encounter the problem mentioned above: the right data is not available from a single source and training datasets must be composed using data from different organisations. In many cases, patient data are dispersed across several locations such as hospitals, clinics and medical devices. In addition, the use of data from different sources (e.g. from different geographical areas) gives better results in terms of robustness and generalisation of AI algorithms. However, it is in centralising clinical data from different sources that privacy and security concerns arise.

Health data are highly sensitive and subject to strict privacy regulations. As indicated by N. Rieke et al.[43], even if strategies such as data anonymisation might circumvent these limitations, it is widely recognised that simply removing metadata such as patient name or date of birth is often insufficient to preserve privacy. With current innovative technologies it is possible, for example, to reconstruct a patient's face from Computed Tomography (CT) or MRI data. As a result, research is increasingly shifting towards the use of technologies that can ensure maximum security of this type of data. Several methods have been proposed to protect privacy, including de-identification techniques such as differential privacy[44] [45], synthetic data generation, homomorphic encryption, and FL. It is the latter that is the focus of this review.

Initially developed for domains such as mobile and edge device applications[46], FL has recently gained popularity in different healthcare sectors. An example of this is the study carried out by Theodora S. et al.[47] in which they sought to solve a supervised binary classification problem to predict hospitalizations for cardiac events, based on their Electronic Health Records (EHRs) and using a federated algorithm based on a sparse Support Vector Machine (sSVM) architecture. I. Dayan et al.[48] also developed an interesting, federated model called EXAM (electronic medical record (EMR)

chest X-ray AI model), which predicts the future oxygen requirements of symptomatic COVID-19 patients using a combined input of vital signs, laboratory data, and chest X-rays. In developing this model, which was trained in a federated way on data from 20 institutes worldwide, FL facilitated rapid data science collaboration without requiring data sharing and generated a model that generalized across heterogeneous and unharmonized datasets.

In the same study carried out by G. Choi et al. mentioned above, they analyse different research studies related to the application of FL in the medical field published to date. They classified these studies according to the types of data used, the target disease, the use of open datasets, the local FL model and the neural network model. They conclude that this technology was mainly used for training algorithms based on medical images, and the most studied target diseases were COVID-19[49] [50] [51] and cancer. In line with the latter, studies focusing on the use of FL with radiology[52] [53] and pathological[54] [55] images are particularly noteworthy, but other types of data such as the ultrasound images used by H. Lee et al. [56] to train a FL algorithm to predict whether thyroid nodules are benign or malignant. In this study, where different types of DL networks (VGG19[57], ResNet50[58], ResNext50[59], SE-ResNet50 and SE-ResNext50) were trained, it was shown that the performance of FL using decentralised data is comparable to conventional DL using clustered data and is potentially useful for analysing medical images while protecting patients' personal information. This survey concludes that FL in the medical domain appears to be in its early stages at present, with most research using open data and focusing on specific types of data and diseases for performance verification purposes. However, FL medical research is expected to be increasingly applied and to become a vital component of multi-institutional research.

In addition to the above, it is also interesting to mention the review made by A. Chowdhury[60] et al. on the application of FL in the field of oncology and cancer research. This extensive and in-depth review identifies that FL has been explored in many studies on different types of cancer, where the goal is either to compare FL with conventional centralised data analysis approaches in terms of performance or to develop novel methods to solve different challenges. In the most common training scenario, researchers simulate a FL environment by taking an existing dataset and dividing it into subsets using a partitioning scheme, where each subset represents a client in a FL group. This review concludes several relevant points on current research in terms of FL applied to the field of oncology. Among the most interesting is the fact that FL has the potential to become the main learning paradigm for distributed cancer research, but specific obstacles such as the existence of correctly labelled medical data, have slowed its adoption in the clinical setting. In addition, most of the papers they encountered use cancer datasets primarily for benchmarking purposes. There are very few works in FL that address clinically relevant questions. Among the papers they reviewed, many propose new software frameworks, and virtually none follow-up with a clinical trial. Consequently, based on this literature review, FL remains largely absent from the field of clinical oncology. Finally, the compliance and security aspects of healthcare continue to present significant hurdles, and more research is needed into techniques that work with FL to maximise systems security.

If we focus on the specific use case of the FLUTE project, the prediction of csPCa, there are also some recent studies that use an FL-tailored solution. Among them is the work published by A. Rajagopal et al. [61], where they present a flexible FL framework for training, validating, and evaluating customised algorithms for detecting prostate cancer. Specifically, they train a custom DL model for multi-parametric MRI-based detection and classification of prostate cancer based on data from two University hospitals. In this study they observed a positive result, with significant improvements in generalization performance across sites and negligible degradation of performance within the site for both lesion segmentation and csPCa classification. However, they comment that further data and participating institutions may be needed to improve the absolute performance of prostate cancer classification models. This study also mentions previous work focused on applying FL to prostate cancer prediction using MRI data, such as the one published by Sarma et al.[62], where they used multi-centre federated training to improve prostate gland segmentation, an important sub-step in the search for prostate cancer biomarkers. Another study conducted by I. Shiri et al.[63] also focuses on prostate lesion segmentation from MRI images using federated DL algorithms. In this research, a two-stage cascaded U-Net consisting of modified 3D U-Net and Dual Attention 2D U-Net was implemented as the core of DL segmentation. MRI images and a prostate mask obtained from 400 patients with histologically proven prostate cancer through T2-weighted magnetic resonance imaging from eight different centres were used as input in this network. Their FL algorithm outperformed centre-based algorithms, where each centre developed a model using their local dataset.

2.2.2. Techniques for improving AI model performance

The landscape of ML has been significantly influenced by the demand for large-scale training data, a requirement for building robust and powerful models. It has been shown that FL, an emerging decentralized ML paradigm, offers a novel solution by enabling collaboration among multiple data owners while prioritizing data privacy. This approach hinges on aggregating models learned on diverse clients to construct a more general model, subsequently distributing it back to the clients for further refinement.

In this section, we review data valuation and model initialization techniques to enhance the performance of AI models and their aggregation, which is one of the main objectives of the FLUTE project.

Data Valuation

In the FLUTE project, data owners are hospitals that exhibit significant **heterogeneity in terms of data** quality, quantity, and distribution. This heterogeneity may stem from variations in instruments, data collection methodologies, among other factors. Such heterogeneity poses a challenge as it can result in suboptimal and less robust models due to variations in statistical distribution of the training data[64].

Data valuation techniques are crucial to address this data heterogeneity in FL. In many instances, clients collect non-identically and independently distributed (non-IID) data, which imply to deal with

Out-of-Distribution (OOD) data. In this context, in [65] propose Federated Out-of-Distribution Synthesizer (FOSTER), which learns a class-conditional generator to synthesize virtual out-of-distribution samples, maintaining data confidentiality and ensuring communication efficiency as required by FL.

In addition, Data Valuation methods play an important role in fairly and efficiently **evaluating the contributions of data owners to the training process** in an FL platform. Some contributions in this area utilize adaptations of the Shapley Value to the FL requirements, to quantify the significance of individual data contributions. The Shapley value (SV)[66] defines a payoff scheme that fulfils many desirable criteria for a fair data valuation. It has often been used to assess training data in centralized learning, however it entails prohibitive communication costs in a FL environment as it requires exhaustive evaluation on each subset of data. In 2020, Wang et al.[67] proposed a federated **Shapley value tailored for FL**, covering empirical study tasks such as noisy label detection, adversarial participant detection, and data summarization across diverse benchmark datasets. Subsequently, in 2022, Fan, Zhenan, et al.[68] introduced the completed federated Shapley value to enhance fairness in the federated SV.

Related to this matter, the integration of **self-supervised** techniques has become increasingly important. Self-supervised learning revolves around training models to understand the inherent structure of the data itself, without relying on externally annotated labels. This shift in the learning paradigm offers significant advantages, particularly in settings where data privacy, distribution and diversity vary across different entities. By pretraining models on local data using self-supervised strategies and then refining them in FL settings, self-supervised learning opens up new avenues for efficient, privacy-preserving and scalable FL. Several novel techniques and strategies have emerged in this field, leveraging methods from the self-supervised learning literature such as non-contrastive self-supervised learning[69] or contrastive self-supervised learning[70]; although fundamentally different, both approaches try to address the challenge of scarce labelled data, exploiting the ability of self-supervised learning techniques to extract valuable information from unlabelled data distributed across a set of clients.

Some self-supervised techniques found in the literature [71] also distinguish between more specialized cases such as **vertical FL** or **horizontal FL**. In the horizontal FL case, the different clients have different samples with a shared feature space, while in the vertical FL case, each client has the same set of samples, but the feature space varies from client to client. The choice between these two paradigms depends on the specific data sharing constraints and the nature of the problem at hand.

In this context, **Active Learning** (AL) techniques have also gained attention in FL. They aim to select the most informative or challenging samples for labelling, thereby improving the model's performance. In [[72]], it is proposed to apply Active Learning (AL) and sampling strategy into a FL framework to reduce the annotation workload in an image classification context. While we do not expect to encounter unlabelled data during the FLUTE project, this provides an approach worth

considering for potential future data owners with unlabelled or semi-labelled data. AL can also be useful for leveraging large, unlabelled datasets for model initialisation.

Model initialization

Model initialization influences convergence, performance, and generalization in the training of ML models. Based on the categorisation made in [73] we can distinguish localized and centralized approaches.

Localized initialization:

1. **Random initialization.** In the literature, most works often use random parameters (weights and biases) to initialize the model. They are populated with values drawn from a chosen **probability distribution**, typically Uniform Distribution X ~ U(a, b) , Truncated Normal Distribution TN(μ, σ2 , a, b) or Normal Distribution X ∼ N (μ , σ2) (typically X ∼ N (0,1)).
2. **Client-Specific Initialization** attempts to leverage data owner-specific data features, encouraging a more tailored and efficient learning process. It could be:
   2.1. **Data-Driven Initialization**. Every data owner performs an analysis of their local data, identifying statistical properties (data distribution, imbalances, etc.), which are later used to set the model's parameters.
   2.2. **Domain-Specific Initialization** utilizes domain-specific knowledge that is relevant to the target task in each client before starting the FL process.
   2.3. **Clustering-Based Initialization** initializes groups of clients according to their similarities, then tunes local models based on client-specific attributes before the FL commences. However, clustering-based initialisation may not be considered during the FLUTE project due to the small number of data owners.

In **centralized initialization,** a central server initiates model training, often after **pretraining a global model** on large datasets and then the model is distributed to all data owners for training and fine-tuning on their local data. Initializing from a pre-trained model reduces training time of training more accurate models (up to 40%) and reduces the impact of both data and system heterogeneity[74], [75]. **Transfer learning** initialization [76] may also be used in scenarios with shareable domains, or when not enough data is available for the target task[77].

2.2.3. Aggregation methods for FL applications

In an FL environment, it is necessary to merge individual models to create a common model. This merging process involves the use of techniques known as "aggregation methods". In the previously mentioned TRUMPET project, research has been conducted on various aggregation methods. However, in this analysis, we will not replicate the research already conducted in this project. To provide a more comprehensive and complete overview of what is covered at this point, relevant information reported in TRUMPET D2.1 is cited in italics.

In the task of federated aggregation, all data owners have some data (a scalar, vector, matrix or more compound object) and the goal is to ensure that a data user receives the aggregate (typically sum or average) of these data.

This is a fundamental operation in FL, both for algorithms which can be decomposed directly into averages and for more complex algorithms such as stochastic gradient descent (SGD) algorithms which frequently need to add up gradients computed by data owners.

If there is a trusted central curator, the data owners can simply send their gradients to that trusted party. If data owners have no trust at all, they can add local differential privacy (LDP) noise to their data, but as LDP typically requires a large amount of noise, this usually is not desirable from a utility point of view.

If data owners cannot trust the aggregator, then sending cleartext data without noise is not advisable, even if the data is constituted only of gradients rather than the originally sensitive information.

There has been a lot of interest in strategies which do not require to fully trust parties, resulting in several approaches for different security and threat models. One line of works employs a trusted shuffler to create anonymity, which can be used to obtain increased privacy. While interesting for theoretical analysis, this approach only replaces the need to trust an aggregator by the need to trust a shuffler, or a shuffling system consisting of several nodes. Another line of work studies secure aggregation methods relying on cryptographic techniques. All these strategies assume an honest-but-curious (also called semi-honest) security model and/or have a significant computational cost. Recently, more efficient approaches have been studied which at the same time are robust under more malicious models, e.g., when one assumes at most a fixed fraction of the parties is malicious.

Hence secure aggregation methods exist which, depending on the threat model, can deliver the aggregation of the data owners' data without revealing the individual terms or intermediate results. Sometimes it is undesirable to reveal this aggregate immediately, as in order to achieve statistical privacy (e.g., differential privacy) one may want to add noise first. One challenge is that none of the parties allowed to know the noise term, as otherwise this party could act as adversary and subtract it from the published result. Fortunately, various strategies exist to generate and add the noise privately.

In FLUTE, our ambition is to scale up MPC-based learning. We will explore multiple strategies to realize this improvement of scalability, most of which will impact both the local computation in the data owner node and the aggregation over the several data owners, including the SMPC used in that context. We will provide some background related to SGD here, and background for other ideas in Section 4.1 on SMPC below.

Compressed gradients. Very recently, the idea has been proposed to compress gradients[78] in the context of SGD. In particular, one can sample part of a gradient, or one can limit every step to only a selection of dimensions in the hope to send over the network shorter gradients while still realizing good convergence, and in particular achieve faster convergence for the same communication cost. Part of our work in FLUTE will further explore this idea.

One can for instance sample uniformly K coordinates of the gradient, and send only these coordinates to the central server, hence reducing the communication cost from d (dimension of the parameters) to K. This compressor (called Rand-K) enjoys nice theoretical properties since it is an unbiased estimate of the gradient. Another choice (among other things) is to consider the K coordinates of the gradients which are the largest in absolute value (Top-K compressor); in this case, each client has extra computations to make, to sort the values of each coordinate of the gradient: this leads to a computation cost of O(d log(K)) at each iteration instead of O(d) in the non-compressed case (note that these extra computation time is little compared to the communication time). The unbiasedness of this compressor makes it more difficult to study; at least in the deterministic case, it can be shown that Top-K compressor outperforms Rand-K compressor [79]

In general, these compressors come with a loss of information, which may prevent the algorithm from converging to a minimum of the objective function, even in some very simple cases [80]. To preserve convergence guarantees, a recent idea has been to incorporate compression errors into subsequent iterations of the algorithm, thus making it possible to obtain convergence guarantees, even for biased compressors like Top-K. This approach, known as "error feedback", is a promising line of research.

All the mentioned modifications of the vanilla SGD are still first order methods. Stochastic optimization methods of order 1 have a slow rate of convergence, of the order of 1/T, where T is the iteration of the algorithm [81] [82]. It can be shown that this rate cannot be improved by any 1st-order method. Thus, quasi-Newton methods (widely used in deterministic optimization) cannot improve substantially on theoretical convergence results.

On the other hand, Newton-type methods, which are of order 2, could achieve better convergence rates. In the deterministic case, a Newton-type algorithm with compression of the hessian was considered in [[83]]. However, calculating the Hessian matrix requires a computation time of d^2, which is not acceptable when considering DNNs. Combining this approach with Top-K or rand-K compressors, an alternative could be to compute the Hessian only on the subspace generated by the K coordinates selected by the compressor.

3. An analysis of PET methods for FL environments

In the previous section, we discussed the use of AI algorithms in the medical field, specifically in detecting csPCa. We also introduced the concept of FL and highlighted its importance in medical AI. Various techniques used in training these technologies were presented, ranging from aggregation methods to techniques for improving model performance. Finally, we discussed available software tools for training FL models. In this section, we will focus on ensuring the security of these environments. We will present a comprehensive overview of the state-of-the-art Privacy Enhancing Technologies (PETs) that will be employed to safeguard medical information. Specifically, we will focus on Secure Multi-Party Computation (SMPC) and Trusted Execution Environments (TEEs). It is worth noting that PETs are a category of secure computing techniques dedicated to protecting the users' personal information. This is achieved through the implementation of data minimization, anonymisation, and encryption strategies[84].

3.1. Secure Multiparty Computation (SMPC)

In its deliverable D2.1, the TRUMPET project provided an overview of SMPC as PET method. For completeness, we here quote this overview in italics.

> *Secure Multiparty Computation (SMPC) protocols allow a group of parties to jointly compute a function while also protecting the privacy of the participants' inputs. It was first introduced by Yao for the two-party case as a solution to the Millionaires Problem, and later generalized to the multi-party scenario by Goldreich, Micali and Wigderson. An important distinctive feature which distinguishes SMPC from other more conventional cryptographic schemes is that SMPC also seeks to protect against adversaries coming from the system itself, hence protecting the privacy of the participants' inputs from each other.*
>
> *The SMPC field covers a wide list of cryptographic techniques ranging from several higher and lower-level tools as Garbled Circuits (GC), Zero-Knowledge proofs, commitments, oblivious transfer and secret sharing, to name just a few. Actually, it is also worth mentioning that strictly speaking, homomorphic encryption techniques can be seen as a subset of SMPC, and many SMPC protocols make use of HE to work properly. Even so, there is a current trend to separate them because SMPC is usually seen as a mechanism to implement concrete policy enforcements among several parties which are interactively performing computation, while HE is usually seen as a mechanism allowing a data owner to securely outsource computation on his/her data. With this in mind, an exhaustive classification of the different SMPC protocols would require to take into account several parameters, as for example, the number of parties involved (mainly two-party or multi-party), the type of corruption (passive, active, covert), the number of corrupted parties (honest majority, dishonest majority), the mobility of the adversary (static, adaptive corruption), etc.*
>
> *Although SMPC originally started as a theoretical curiosity, since the mid 2000's important improvements were made showing that SMPC applications were in fact possible. The first important example of deployment in a real scenario was the implementation of the sugar beet auction in Denmark in 2008. Currently, the most practical general-purpose protocols are based on additive*

secret sharing, and the use of Beaver triples to perform interactive multiplications. They divide computation in two main phases: (1) an offline phase on which some sort of correlated randomness is computed and distributed among the parties, and (2) an online phase on which given some participants' inputs the previous correlated randomness is consumed to perform the intended secure computation.

The most representative example of the previous schemes is the case of the SPDZ protocol together with its different variants MASCOT, Overdrive and SPDZ2k. A common trait of all of them is their low computational cost. On the contrary, the number of interactions, and hence communication cost, grows with the number of parties and circuit depth, which makes the latter its weakest point in real scenarios.

SMPC technology has made a huge progress in the last few years, and there are already many companies offering SMPC-related solutions as Sharemind, Sepior, Zcash, Unbound. This also includes several frameworks and software developments as the SCAPI library, SCALE-MAMBA SMPC-system, MP-SPDZ, Rmind, Jana relational database from Galois, etc. A very detailed list of several software solutions is available on the web of the TPMPC workshop.

Many of the above schemes have been used to implement important ML or FL tasks as the ones required for TRUMPET use cases. Actually, there is also a rapidly growing list of works specifically designed to deal with training and prediction/classification phases. Some of the most recent are ABY, SecureNN, Quotient, Gazelle and Delphi.

In FLUTE, our ambition is to scale up MPC, in particular for using optimization strategies such as stochastic gradient descent (SGD) using MPC we aim to reduce the communication cost by adapting the SGD algorithm and aligning SGD and MPC (see also Section 2.2.3 above)

Next to aggregation-level optimization, we aim to exploit the fact that in ML algorithms often similar tasks need to be repeated many times. For example, in the context of secret sharing a significant cost is the distribution of secret shares over the data owners (every data owner gets one secret share of every secret variable). To avoid this communication cost, one could exploit Pseudorandom Correlation Generators (PCG)[85], which require a one-time setup phase after which the data owners can jointly generate many secret shares without communication in later steps.

3.2. Trusted Execution Environments (TEEs)

Trusted Execution Environments (TEEs) represent a class of hardware devices characterized by a distinctive feature: an isolated area within the computer system. These devices showcase exceptional versatility, seamlessly integrating superior performance and expanded execution capacity when contrasted with alternative cryptographic techniques. They excel not only in securely storing data but also in executing a variety of code with enhanced efficiency compared to other PETs [86].

Although the term 'TEE' is relatively new, classes of TEEs such as Hardware Security Modules (HSMs) and Trusted Platform Modules (TPMs) have been available on the commercial market for some time[87]. However, in recent years, a new kind of TEE has emerged. This new category of TEE consists of general-purpose devices which offers a secure area isolated from the main operating system for processing sensitive data[88].

In general, this new kind of TEEs offers the following security guarantees:

- **Confidentiality**: The content within the trusted area is accessible exclusively to entities that have been authorized.
- **Integrity**: The content within the trusted area is exclusively modifiable by entities that have been granted authorization
- **Attestation**: Empowers external entities to verify the integrity of the code running within the secure enclave[89].

In the current landscape, leading processor manufacturers have integrated Trusted Execution Environments (TEEs) into their processors, with ARM pioneering TrustZone[90] technology in the early 2000s. TrustZone technology enables devices equipped with an ARM processor to execute isolated applications within a protected area known as the 'Secure World,' which operates on the foundation of the processor's firmware. However, the technology faced limitations, notably in the areas of confidentiality and integrity, which were safeguarded by firmware rather than hardware. This oversight permitted attackers with hardware access to compromise the information. Consequently, ARM has initiated the development of its new TEE, ARM CCA[91], which is poised to address these vulnerabilities but is not yet commercially available.

After ARM's initiative, other manufacturers have entered the arena, contributing to the ongoing evolution of this technology. Notably, Intel took a significant step forward with the introduction of SGX[92] in 2015, followed by introduced enhancements in SGXv2[93]. The SGX technology enables the creation of encrypted memory areas within the main computer memory, which are exclusively decryptable by the processor. These encrypted memory regions, commonly referred to as enclaves, ensure that no entity outside of the enclaves can access the data contained with them. Furthermore, in 2023, Intel introduced a new TEE technology known as Intel TDX[94]. TDX enables the deployment of a full virtual machine (VM) within a trusted area. This new trusted area, referred to by Intel as a 'trust domain', provides the same security guarantees as Intel SGX. However, it offers a significant advantage over SGX as it does not require changes at the application level.

AMD, on the other hand, has also invested in introducing TEEs in its processors by developing the SEV[95] (Secure Encrypted Virtualization) technology, along with its enhanced versions: SEV-ES (Encrypted State) and SEV-SNP (Secure Nested Paging). Like Intel's TDX, the SEV family enables the deployment of virtual machines within an isolated and trusted area.

Initially conceived as secure areas within the primary processor, Trusted Execution Environments have recently seen expansion beyond the processor itself. These secure areas now extend to other

trusted devices, including graphics cards. This extension enables TEEs to be utilized in scenarios where the computational power required was previously unfeasible, such as certain AI workloads. In this context, NVIDIA plays a pivotal role with its product Nvidia Confidential Computing[96]. This innovative solution allows sensitive workloads to be processed on graphics cards, effectively reducing the exposure of critical information. By bridging the gap between security and performance, Nvidia is advancing the landscape of secure computing.

In the context of FLUTE, a FL-based configuration, Trusted Execution Environments play a pivotal role in ensuring the security of the handled information. While FL mitigates data privacy risk, it encounters persistent challenges related to the confidentiality and integrity of computations, relying heavily on trust between involved parties[97]. Regarding this issue, TEEs can provide advantages such as code integrity in the untrusted parties as well as in the central aggregator. Some important examples of applying TEEs within the FL landscape includes SecFL[98], a confidential FL framework that performs the global and local training inside TEE enclaves. It also has a transparent remote attestation mechanism that allows the parties to verify all computations. Other important work is PPFL[99], a FL framework designed to limit data leakage and run on devices with limited resources. Like SecFL, PPFL uses TEEs on both sides, the central aggregator and the local training. Other significant research is Flatee[100], an efficient privacy-preserving FL framework based on TEEs, which considerably reduces training and communication time, However, this optimization comes at the cost of sacrificing protection against adversarial data poisoning.

4. Frameworks and tools

In this section, the different frameworks, tools and software that will be used in the FLUTE project will be defined. On the one hand, the tool that will be used on the FL environment will be discussed: Pysyft. On the other hand, we will define two specific tools that will be used within the framework of WP2 to process MRI images and extract the aforementioned QIBs for use as input in the csPCa prediction algorithms: the QP-Care® and the QP-Prostate®[101].

4.1. FL frameworks and tools: PySyft

This sub-section focuses on various software tools that enable FL model training. Here we will build upon the insights derived from research conducted in the TRUMPET project.

The research in the field of FL has been vast, complex and under continuous development. In the case of FLUTE project, we are bound to depend onto existing advancements and developments provided by existing tools and software to allow the project to make a solid trial on the pilot use cases, given the timeline we have. The existing software tools will allow us a foundational framework to begin with and browse our options while defining our own needs and adopt, update, and customize those tools based on those needs. We have learned a lot from the research on the tools and technologies available in the field of FL during our work in the TRUMPET project. The FL tools that are going to be used in FLUTE would basically be an extended version of this project. This extensibility was a base concern in the research and development of such software tools for TRUMPET.

The initial research decision for TRUMPET FL was to use PySyft[102] as an FL library. The practicality of this decision is well elaborated in the TRUMPET D2.1 deliverable, section 2.4. A comparison between around 20 FL tools that are either production grade or research oriented was conducted. Based on our research on the usability of those tools for our need, vwe favoured PySyft the most. There were many reasons for this, but most importantly the design of PySyft makes one appreciate the possible scalability of this opensource platform by OpenMined [103]. However, PySyft has quite a steep learning curve, and may not be best-suited for industrial use. Neverthless given our context of use, it provides modularity and flexibility to work with other DL tools, an extensive library of privacy tools as well as a great community support.

But the decision of primarily using PySyft as the FL tool has been ever evolving, with research and development continuously revealing unforeseen or unconsidered perspectives.

The first glitch was the issue that arose around the use of PySyft as a lib or as a service. The decision was to use PySyft as a separately deployable service.

At this point it was clear that PySyft alone cannot deliver the infrastructural demand of the TRUMPET or FLUTE project. This led to the rise of the concept of FL wrapper project, that would partially use PySyft, but can also introduce other functionalities that do not exist in the current

version of PySyft. Basically, the wrapper project may contain custom implementations and several other services or libraries. The key abstraction considerations for the development of FL wrapper project are outlined below:

1. Continuously supporting integration of various FL tools.
2. Considering exposed APIs by the integrated dependency tool as a black box.
3. The wrapper defines the functionalities or features to support by its own custom APIs, abstracting any or all the wrapped tools.
4. An integrated or dependency tool may support only a subset of the decided key features.
5. One or more tools may be incorporated (wrapped)
6. One or more features provided by a given tool can be extended with custom or changed functionalities.
7. One or more features can have complete self-dependent implementation within the wrapper tool.
8. Where necessary custom algorithms can be introduced from outside complying appropriate interface protocols.

As we progressed, the need for this wrapper tool is better realized, and the realization is outlined below –

1. We want to have our own FL training flow.
2. We want to have custom PET methods.
3. We want to have custom budgeting functions, study agreement and in the future many more as we suppose.
4. We want to have custom AI algorithms.
5. We want to hide use of dependency tools (e.g. PySyft)
6. We want to be able to replace, extend, and add more dependency tools or libs (e.g. fedML)

The demand for strict privacy made us replace the communication mechanism developed by PySyft with our own custom VPN, using tailscale.

The latest change is the outcome of Gradiant's research on implementing PET's using PySyft. Their research and development gave rise to a set of new HTTP services that will work parallel to the PySyft services. These new services were introduced because PySyft did not allow implementation of custom PETs (Lagrange Coded Computing technique).

As the research continues, these new findings will evolve the landscape of the used Frameworks and tools and software for FL in the development of FLUTE. But the need for a scalable wrapper lib is not going to change. The platform requires to define a stable interface for Researchers and Data Owners to integrate. On the other hand, the wrapper lib should be scalable to accommodate the introduction of any new FL tools, libs and frameworks. This would allow us to continuously build for the FLUTE project without breaking the software or getting stuck into the implementation of some new PET or AI algorithm.

4.2. MRI processing and QIBs extraction tools: QP-Care® and QP-Prostate®

The AI algorithms developed in FLUTE aim to predict csPCa, and for this purpose, they will take two types of data as input: clinical variables and QIBs. The clinical variables will be provided directly by the three hospitals in the consortium (VHIR, CHU and IRST). Regarding the QIBs, these will be extracted from MRI images from the same hospitals using the tools provided by the partner QBIM.

All MRI image analysis provided by QBIM takes place on the Microsoft Azure cloud. QBIM's cloud is commonly known as QP-Care®, which provides a portal to a results viewer and behind which analysis tools are located. Once the studies are available in QP-Care®, they will be analysed and new series will be generated. Users can access a viewer through the QP-Care® portal to view the results.

QBIM currently offers an analysis module for prostate gland characterization integrated into the platform, called QP-Prostate®. This tool is a Food and Drug Administration (FDA) 510(k)-cleared software that has been clinically validated in installations in the US, Spain, Poland, Slovakia and South America as a safe and accurate decision support tool for PCa detection. It allows to process and analyse both bpMRI and mpMRI, segmenting the prostate gland images into zones and extracting those QIBs. This analysis will be performed whenever the T2-weighted (T2w), DWI sequences are available, and optionally, DCE sequences, provided that they meet the PI-RADS v2.1 criteria (or following the adapted acquisition protocol described in Annex 1, p. 43).

For prostate characterization, QP-Prostate® will perform automatic processing according to the following steps:

1. **Spatial Smoothing:** The images are smoothed to preserve the image edges and remove noise.
2. **Motion Correction:** The DWI images are correlated with the DCE images to correct artifacts.
3. **Spatial Alignment:** Using the T2-weighted series, the DWI and DCE images are aligned.
4. **Automatic segmentation of the prostate into three regions:** central/transitional zone, peripheral zone, and seminal vesicles.
5. **Generation of parametric maps (quantification):** by calculating the ADC from the DWI sequence or the vascular permeability rate (*Ktrans*) and other parameters extracted from the DCE sequence, such as *kep* or *ve*.

QP-Prostate® generates prostate segmentation into its different regions, including the central and transitional zone, the peripheral zone, and the seminal vesicles. Additionally, it automatically identifies regions suspected to be pathological in the prostate gland based on bpMRI. The results of these segmentations are saved in DICOM-Seg format and can be viewed by users on the platform.

Thanks to this software, in FLUTE we will be able to extract QIBs maps from MRI scans that will be used, together with clinical variables, as input to the fusion models for csPCa prediction.

5. Conclusions

This document reported the research results corresponding to Task 2.1 of the FLUTE project. An exhaustive analysis has been conducted on current research and available technologies related to the use case of this project: training a clinically significant prostate cancer (csPCa) prediction AI algorithm in a secure, scalable Federated Learning (FL) environment while also maintaining AI performance.

AI adoption across several sectors of society has been rapidly increasing in recent years, with its integration into the clinic representing a particular notable example. The intersection of AI and medical domain signifies a pivotal moment in the evolution of healthcare. In this context, the FLUTE project aims to contribute to this ongoing innovation by developing an AI model for csPCa prediction. This document presents a comprehensive review of current research on the use of AI technologies to detect this ailment. This has helped to better understand the use case and the types of data typically used in such tasks, including clinical data like prostate-specific antigen (PSA) or Quantitative Image Biomarkers (QIBs). It has been concluded that most studies in this area focus on predicting csPCa using MRI images and clinical data separately, however a fusion of both information sources provide higher results in prediction task. In the FLUTE project, we will use a combination of both types of data to improve results, employing advanced Deep Learning (DL) techniques for this task. We will build upon the previous study conducted by our partner VHIR, where they predict csPCa using ML techniques on clinical data. Our aim is to enhance the algorithms by training them in a federated manner, incorporating data from our other two clinical partners: CHU and IRST. Additionally, we will train a fusion DL model using QIBs extracted by QBIM using their QP-Care® and QP-Prostate® tools. All this work will be developed in Task 2.2.

Throughout this document, it has been also shown that AI, despite being an extremely useful tool, has some drawbacks. Among them is the fact that to train algorithms with good performance and able to generalise correctly, it is necessary to use large and high-quality datasets, and often this is a challenge in the real world. Clinical data is often located in different sites, and due to the sensitivity of this data, sharing it raises clear concerns related to privacy and security, which are regulated under strict policies such as GDPR at the European level, or internal Ethical Committees of each hospital. One solution to these issues is the use of FL, a technology that allows global training of algorithms at each data location, making sharing data no longer necessary. This document includes numerous examples of the use of this technology in the clinical field and how it is possible to apply different aggregation techniques to generate these global models through sparse training. Additionally, as one of the FLUTE objectives is to maintain AI models performance, possible techniques to improve this aspect, such as model initialisation or data valuation, are also being studied. These identified techniques will be analysed in order to choose the best option to be used in the FLUTE setting and will be developed within the framework of Task 2.2.

In this way, FL represents a paradigm shift in Machine Learning (ML), offering a framework where data privacy is inherently respected as the data remains within the confines of its owner's infrastructure. Despite this, research indicates that privacy risks persist, necessitating additional safeguards. This project proposes the integration of Privacy-Enhancing Technologies (PETs), specifically Secure Multi-Party Computation (SMPC) and Trusted Execution Environments (TEEs), to enhance privacy protections. By combining these PETs, one software-based, the other hardware-based, we aim to improve security guarantees. However, it is crucial to acknowledge that the adoption of these technologies may impact on system performance and limit the selection of deployable AI models. Therefore, a balanced approach must be adopted to ensure that the benefits of enhanced privacy do not compromise the system's efficiency and the AI's capabilities. This document presents an analysis of the state-of-the-art of PET technologies, both software and hardware, to be used in FLUTE, specifically in tasks Tasks 2.3 and 2.4. Furthermore, the combined performance of these techniques, considering privacy, computation, and communication cost trade-offs, will be studied in task T2.5.

REFERENCES

[1] OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat

[2] OpenAI. (2023). DALL-E 3. https://openai.com/dall-e-3

[3] TRUMPET project: https://cordis.europa.eu/project/id/101070038

[4] Chauhan, Rahul and Ghanshala, Kamal Kumar and Joshi, R.C. Convolutional Neural Network (CNN) for Image Detection and Recognition, 2018. Avaliable at https://ieeexplore.ieee.org/document/8703316

[5] Yang, Q., Liu, Y., Chen, T. & Tong, Y. Federated machine learning: concept and applications. ACM Trans. Intell. Syst. Technol. (TIST) 10, 12 (2019).

[6] Li, T., Sahu, A. K., Talwalkar, A. & Smith, V. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine 37, 50–60 (IEEE, 2020).

[7] Secinaro, S., Calandra, D., Secinaro, A. et al. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak 21, 125 (2021). https://doi.org/10.1186/s12911-021-01488-9

[8] Yang, X., Wang, Y., Byrne, R., Schneider, G., & Yang, S. (2019). Concepts of artificial intelligence for computer-assisted drug discovery. Chemical reviews, 119(18), 10520-10594.

[9] Burton, R. J., Albur, M., Eberl, M., & Cuff, S. M. (2019). Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. BMC medical informatics and decision making, 19, 1-11.

[10] World Health Organization: Cancer (https://www.who.int/news-room/fact-sheets/detail/cancer , February 2022)

[11] Van Booven, D. J., Kuchakulla, M., Pai, R., Frech, F. S., Ramasahayam, R., Reddy, P., … Arora, H. (2021). A Systematic Review of Artificial Intelligence in Prostate Cancer. Research and Reports in Urology, 13, 31–39. https://doi.org/10.2147/RRU.S268596

[12] Gentile F, Ferro M, Della Ventura B, La Civita E, Liotti A, Cennamo M, Bruzzese D, Velotta R, Terracciano D. Optimized Identification of High-Grade Prostate Cancer by Combining Different PSA Molecular Forms and PSA Density in a Deep Learning Model. Diagnostics (Basel). 2021 Feb 18;11(2):335. doi: 10.3390/diagnostics11020335. PMID: 33670632; PMCID: PMC7922417.

[13] Morote J, Borque-Fernando A, Triquell M, Celma A, Regis L, Escobar M, Mast R, de Torres IM, Semidey ME, Abascal JM, Sola C, Servian P, Salvador D, Santamaría A, Planas J, Esteban LM, Trilla E. The Barcelona Predictive Model of Clinically Significant Prostate Cancer. Cancers (Basel). 2022 Mar 21;14(6):1589. doi: 10.3390/cancers14061589. PMID: 35326740; PMCID: PMC8946272.

[14] Stojadinovic, M., Stojadinovic, M., & Pantic, D. (2019). Decision tree analysis for prostate cancer prediction.

[15] Djavan, B., Remzi, M., Zlotta, A., Seitz, C., Snow, P., & Marberger, M. (2002). Novel artificial neural network for early detection of prostate cancer. Journal of Clinical Oncology, 20(4), 921-929.

[16] Finne, P., Finne, R., Auvinen, A., Juusela, H., Aro, J., Määttänen, L., ... & Stenman, U. H. (2000). Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. Urology, 56(3), 418-422.

[17] Gentile F, Ferro M, Della Ventura B, La Civita E, Liotti A, Cennamo M, Bruzzese D, Velotta R, Terracciano D. Optimized Identification of High-Grade Prostate Cancer by Combining Different PSA Molecular Forms and PSA Density in a Deep Learning Model. Diagnostics (Basel). 2021 Feb 18;11(2):335. doi: 10.3390/diagnostics11020335. PMID: 33670632; PMCID: PMC7922417.

[18] Bass, E.J., Pantovic, A., Connor, M. et al. A systematic review and meta-analysis of the diagnostic accuracy of biparametric prostate MRI for prostate cancer in men at risk. Prostate Cancer Prostatic Dis 24, 596–611 (2021). https://doi.org/10.1038/s41391-020-00298-w

[19] Lu H, Parra NA, Qi J, Gage K, Li Q, Fan S, Feuerlein S, Pow-Sang J, Gillies R, Choi JW, Balagurunathan Y. Repeatability of Quantitative Imaging Features in Prostate Magnetic Resonance Imaging. Front Oncol. 2020 May 7;10:551. doi: 10.3389/fonc.2020.00551. PMID: 32457827; PMCID: PMC7221156.

[20] Radiology ACo. Prostate Imaging Reporting and Data System (PIRADS) version 2. (2015). Available online at: https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2.pdf (accessed January 10, 2020).

[21] Rosenkrantz, A. B., Ginocchio, L. A., Cornfeld, D., Froemming, A. T., Gupta, R. T., Turkbey, B., ... & Margolis, D. J. (2016). Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. Radiology, 280(3), 793-804.

[22] Wang, J., Wu, C. J., Bao, M. L., Zhang, J., Wang, X. N., & Zhang, Y. D. (2017). Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. European radiology, 27, 4082-4090.

[23] Cindil E, Oner Y, Sendur HN, Ozdemir H, Gazel E, Tunc L, Cerit MN. The Utility of Diffusion-Weighted Imaging and Perfusion Magnetic Resonance Imaging Parameters for Detecting Clinically Significant Prostate Cancer. Can Assoc Radiol J. 2019 Nov;70(4):441-451. doi: 10.1016/j.carj.2019.07.005. Epub 2019 Sep 24. PMID: 31561925.

[24] Algohary, A., Viswanath, S., Shiradkar, R., Ghose, S., Pahwa, S., Moses, D., ... & Madabhushi, A. (2018). Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. Journal of Magnetic Resonance Imaging, 48(3), 818-828.

[25] Sánchez Iglesias Á, Morillo Macías V, Picó Peris A, Fuster-Matanzo A, Nogué Infante A, Muelas Soria R, Bellvís Bataller F, Domingo Pomar M, Casillas Meléndez C, Yébana Huertas R, Ferrer Albiach C. Prostate Region-Wise Imaging Biomarker Profiles for Risk Stratification and Biochemical Recurrence Prediction. Cancers (Basel). 2023 Aug 18;15(16):4163. doi: 10.3390/cancers15164163. PMID: 37627191; PMCID: PMC10453281.

[26] Winkel DJ, Breit HC, Shi B, Boll DT, Seifert HH, Wetterauer C. Predicting clinically significant prostate cancer from quantitative image features including compressed sensing radial MRI of prostate perfusion using machine learning: comparison with PI-RADS v2 assessment scores. Quant Imaging Med Surg. 2020 Apr;10(4):808-823. doi: 10.21037/qims.2020.03.08. PMID: 32355645; PMCID: PMC7188610.

[27] Sun Z, Wang K, Kong Z, Xing Z, Chen Y, Luo N, Yu Y, Song B, Wu P, Wang X, Zhang X, Wang X. A multicenter study of artificial intelligence-aided software for detecting visible clinically significant prostate cancer on mpMRI. Insights Imaging. 2023 Apr 30;14(1):72. doi: 10.1186/s13244-023-01421-w. PMID: 37121983; PMCID: PMC10149551.

[28] Zhu, Y., Wei, R., Gao, G., Ding, L., Zhang, X., Wang, X., & Zhang, J. (2019). Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. Journal of Magnetic Resonance Imaging, 49(4), 1149-1156.

[29] Zhao L, Bao J, Qiao X, Jin P, Ji Y, Li Z, Zhang J, Su Y, Ji L, Shen J, Zhang Y, Niu L, Xie W, Hu C, Shen H, Wang X, Liu J, Tian J. Predicting clinically significant prostate cancer with a deep learning approach: a multicentre retrospective study. Eur J Nucl Med Mol Imaging. 2023 Feb;50(3):727-741. doi: 10.1007/s00259-022-06036-9. Epub 2022 Nov 21. PMID: 36409317; PMCID: PMC9852176.

[30] Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., & Paluri, M. (2017). A Closer Look at Spatiotemporal Convolutions for Action Recognition. ArXiv. /abs/1711.11248

[31] Hosseinzadeh M, Saha A, Brand P, Slootweg I, de Rooij M, Huisman H. Deep learning-assisted prostate cancer detection on bi-parametric MRI: minimum training data size requirements and effect of prior knowledge. Eur Radiol. 2022 Apr;32(4):2224-2234. doi: 10.1007/s00330-021-08320-y. Epub 2021 Nov 16. PMID: 34786615; PMCID: PMC8921042.

[32] Xu, L., Zhang, G., Shi, B., Liu, Y., Zou, T., Yan, W., … Sun, H. (2019). Comparison of biparametric and multiparametric MRI in the diagnosis of prostate cancer. Cancer Imaging, 19(1). https://doi.org/10.1186/s40644-019-0274-9

[33] Sushentsev N, Moreira Da Silva N, Yeung M, Barrett T, Sala E, Roberts M, Rundo L. Comparative performance of fully-automated and semi-automated artificial intelligence methods for the detection of clinically significant prostate cancer on MRI: a systematic review. Insights Imaging. 2022 Mar 28;13(1):59. doi: 10.1186/s13244-022-01199-3. PMID: 35347462; PMCID: PMC8960511.

[34] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv. /abs/1505.04597

[35] SPIE-AAPM-NCI PROSTATEx Challenges (PROSTATEx): https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=23691656

[36] Cheng, X., Chen, Y., Xu, J. et al. Development and validation of a predictive model based on clinical and MpMRI findings to reduce additional systematic prostate biopsy. Insights Imaging 15, 3 (2024). https://doi.org/10.1186/s13244-023-01544-0

[37] Chen, Z., Zhang, J., Jin, D. et al. A novel clinically significant prostate cancer prediction system with multiparametric MRI and PSA: P.Z.A. score. BMC Cancer 23, 1138 (2023). https://doi.org/10.1186/s12885-023-11306-2

[38] Hiremath A, Shiradkar R, Fu P, Mahran A, Rastinehad A, Tewari A, Tirumani S, Purysko A, Ponsky L, Madabhushi . An integrated nomogram combining deep learning, Prostate Imaging–Reporting and Data System (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study. The Lancet Digital Health 2021 Jul; 3(7). doi: 10.1016/S2589-7500(21)00082-0

[39] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

[40] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[41] i, M., Chen, T., Zhao, W., Wei, C., Li, X., Duan, S., Ji, L., Lu, Z., & Shen, J. (2020). Radiomics prediction model for the improved diagnosis of clinically significant prostate cancer on biparametric MRI. Quantitative imaging in medicine and surgery, 10 2, 368-379.

[42] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2016). Communication-Efficient Learning of Deep Networks from Decentralized Data. ArXiv. /abs/1602.05629

[43] Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. npj Digit. Med. 3, 119 (2020). https://doi.org/10.1038/s41746-020-00323-1

[44] Dwork C. Differential privacy: a survey of results. In: Agrawal M, Du D, Duan Z, Li A, editors. Theory and applications of models of computation. Heidelberg, Germany: Springer; 2008. pp. 1–19.

[45] Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, et al. Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; 2016 Oct 24–28; Vienna, Austria. pp. 308–18.

[46] Kairouz, P. et al. Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977 (2019).

[47] Theodora S. Brisimi, Ruidi Chen, Theofanie Mela, Alex Olshevsky, Ioannis Ch. Paschalidis, Wei Shi, Federated learning of predictive models from federated Electronic Health Records, International Journal of Medical Informatics, Volume 112, 2018, Pages 59-67, ISSN 1386-5056, https://doi.org/10.1016/j.ijmedinf.2018.01.007.

[48] Dayan, I., Roth, H.R., Zhong, A. et al. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med 27, 1735–1743 (2021). https://doi.org/10.1038/s41591-021-01506-3

[49] Xu, Y. et al. A collaborative online AI engine for CT-based COVID-19 diagnosis. Preprint at https://www.medrxiv.org/content/10.1101/2020.05.10.20096073v2 (2020).

[50] Raisaro, J. L. et al. SCOR: A secure international informatics infrastructure to investigate COVID-19. J. Am. Med. Inform. Assoc. 27, 1721–1726 (2020).

[51] Vaid A, Jaladanki S, Xu J, Teng S, Kumar A, Lee S, Somani S, Paranjpe I, De Freitas J, Wanyan T, Johnson K, Bicak M, Klang E, Kwon Y, Costa A, Zhao S, Miotto R, Charney A, Böttinger E, Fayad Z, Nadkarni G, Wang F, Glicksberg B. Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach. JMIR Med Inform 2021;9(1):e24207- URL: https://medinform.jmir.org/2021/1/e24207. DOI: 10.2196/24207

[52] Jiménez-Sánchez, A., Tardy, M., Ballester, M. A. G., Mateus, D., & Piella, G. (2023). Memory-aware curriculum federated learning for breast cancer classification. Computer Methods and Programs in Biomedicine, 229, 107318.

[53] Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S. H., Reina, G. A., ... & Poisson, L. (2022). Federated learning enables big data for rare cancer boundary detection. Nature communications, 13(1), 7346.

[54] Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W., & Tizhoosh, H. R. (2022). Federated learning and differential privacy for medical image analysis. Scientific reports, 12(1), 1953.

[55] Lu, M. Y., Chen, R. J., Kong, D., Lipkova, J., Singh, R., Williamson, D. F., ... & Mahmood, F. (2022). Federated learning for computational pathology on gigapixel whole slide images. Medical image analysis, 76, 102298.

[56] Lee H, Chai Y, Joo H, Lee K, Hwang J, Kim S, Kim K, Nam I, Choi J, Yu H, Lee M, Masuoka H, Miyauchi A, Lee K, Kim S, Kong H. Federated Learning for Thyroid Ultrasound Image Analysis to Protect Personal Information: Validation Study in a Real Health Care Environment. JMIR Med Inform 2021;9(5):e25869. URL: https://medinform.jmir.org/2021/5/e25869. DOI: 10.2196/25869

[57] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv. /abs/1409.1556

[58] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. ArXiv. /abs/1512.03385

[59] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2016). Aggregated Residual Transformations for Deep Neural Networks. ArXiv. /abs/1611.05431

[60] Chowdhury, A., Kassem, H., Padoy, N., Umeton, R., Karargyris, A. (2022). A Review of Medical Federated Learning: Applications in Oncology and Cancer Research. In: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science, vol 12962. Springer, Cham. https://doi.org/10.1007/978-3-031-08999-2_1

[61] Rajagopal A, Redekop E, Kemisetti A, Kulkarni R, Raman S, Sarma K, Magudia K, Arnold CW, Larson PEZ. Federated Learning with Research Prototypes: Application to Multi-Center MRI-based Detection of Prostate Cancer with Diverse Histopathology. Acad Radiol. 2023 Apr;30(4):644-657. doi: 10.1016/j.acra.2023.02.012. Epub 2023 Mar 12. PMID: 36914501; PMCID: PMC10869141.

[62] Sarma, K. V., Harmon, S., Sanford, T., Roth, H. R., Xu, Z., Tetreault, J., ... & Arnold, C. W. (2021). Federated learning improves site performance in multicenter deep learning without data sharing. Journal of the American Medical Informatics Association, 28(6), 1259-1264.

[63] I. Shiri et al., "Collaborative Multi-Institutional Prostate Lesion Segmentation from MR images Using Deep Federated Learning Framework," 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Italy, 2022, pp. 1-3, doi: 10.1109/NSS/MIC44845.2022.10398941.

[64] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, Yuan Gao, A survey on federated learning, Knowledge-Based Systems, Volume 216, 2021, 106775, ISSN 0950-7051.

[65] Yu, Shuyang, et al. "Turning the curse of heterogeneity in federated learning into a blessing for out-of-distribution detection." The Eleventh International Conference on Learning Representations. 2022.

[66] L. Shapley, “A value for n-person games,” RAND CORP SANTA MONICA CA, Tech. Rep., 1952.

[67] Wang, Tianhao, et al. “A Principled Approach to Data Valuation for Federated Learning.” Federated Learning, 2020, pp. 153–67. Crossref, https://doi.org/10.1007/978-3-030-63076-8_11.

[68] Fan, Zhenan, et al. “Improving Fairness for Data Valuation in Horizontal Federated Learning.” 2022 IEEE 38th International Conference on Data Engineering (ICDE), May 2022. Crossref, https://doi.org/10.1109/icde53745.2022.00228.

[69] D. Makhija, N. Ho, J. Ghosh, “Federated self-supervised learning for heterogeneous clients.” arXiv preprint arXiv:2205.12493 (2022)

[70] S. A. Khowaja, K. Dev, S. M. Anwar, M. G. Linguraru, “SelfFed: self-supervised federated learning for data heterogeneity and label scarcity in IoMT”, arXiv preprint arXiv:2307.01514 (2023)

[71] Yuanquin He, et al. “A hybrid self-supervised learning framework for vertical federated learning”, arXiv preprint arXiv:2208.08934 (2022)

[72] Ahn, Jin-Hyun, et al. "Federated active learning (f-al): an efficient annotation strategy for federated learning." arXiv preprint arXiv:2202.00195 (2022).

[73] Jadhav, Adwaita Janardhan, and Ishmeet Kaur. “A Comprehensive Study on Model Initialization Techniques Ensuring Efficient Federated Learning.” 2023 International Conference on Intelligent Computing and Next Generation Networks（ICNGN), IEEE, 2023. Crossref, https://doi.org/10.1109/icngn59831.2023.10396802.

[74] Nguyen, J., Wang, J., Malik, K., Sanjabi, M., & Rabbat, M. (2023). Where to Begin? On the Impact of Pre-Training and Initialization in Federated Learning. arXiv preprint arXiv:2206.15387.

[75] Jadhav, Adwaita Janardhan, and Ishmeet Kaur. “A Comprehensive Study on Model Initialization Techniques Ensuring Efficient Federated Learning.” 2023 International Conference on Intelligent Computing and Next Generation Networks（ICNGN), IEEE, 2023. Crossref, https://doi.org/10.1109/icngn59831.2023.10396802.

[76] Maithra Raghu et al. “Transfusion: Understanding transfer learning for medical imaging”. In: Advances in neural information processing systems 32 (2019)

[77] Lisa Torrey and Jude Shavlik. “Transfer learning”. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, 2010, pp. 243–263.

[78] Constantin Philippenko, Aymeric Dieuleveut. Compressed and distributed least-squares regression: convergence rates with applications to Federated Learning.CoRRabs/2308.01358 (2023)

[79] Nutini, J., Schmidt, M., Laradji, I., Friedlander, M. & Koepke, H.. (2015). Coordinate Descent Converges Faster with the Gauss-Southwell Rule Than Random Selection. Proceedings of the 32nd International Conference on Machine Learning, in Proceedings of Machine Learning Research 37:1632-1641 Available from https://proceedings.mlr.press/v37/nutini15.html.

[80] Karimireddy, S.P., Rebjock, Q., Stich, S. & Jaggi, M.. (2019). Error Feedback Fixes SignSGD and other Gradient Compression Schemes. Proceedings of the 36th International Conference on Machine Learning, in Proceedings of Machine Learning Research. 97:3252-3261 Available from https://proceedings.mlr.press/v97/karimireddy19a.html.

[81] Bach, F., & Moulines, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate O(1/n). ArXiv. /abs/1306.2119

[82] A. Agarwal, P. L. Bartlett, P. Ravikumar and M. J. Wainwright, "Information-Theoretic Lower Bounds on the Oracle Complexity of Stochastic Convex Optimization," in IEEE Transactions on Information Theory, vol. 58, no. 5, pp. 3235-3249, May 2012, doi: 10.1109/TIT.2011.2182178.

[83] Safaryan, M., Islamov, R., Qian, X., & Richtárik, P. (2021). FedNL: Making Newton-Type Methods Applicable to Federated Learning. ArXiv. /abs/2106.02969

[84] Information Commissioner’s Office (September 2022). Chapter 5: Anonymisation and PETs. Retrieved April 8, 2024, from https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf

[85] Geoffroy Couteau, Elette Boyle, Niv Gilboa, Yuval Ishai, Lisa Kohl, Peter Scholl, Idc Herzliya. Efficient Pseudorandom Correlation Generators: Silent OT Extension and More. https://hal.science/hal-03373123/document

[86] A. Akram, A. Giannakou, V. Akella, J. Lowe-Power, and S. Peisert, “Performance Analysis of Scientific Computing Workloads on General Purpose TEEs.” Accessed: Mar. 19, 2024. [Online]. Available: https: //arch.cs.ucdavis.edu/assets/papers/ipdps21- hpc- tee- performance.pdf

[87] F. Kammel, M. Ylinen, and T. Feldman-Fitzthum, “Confidential Kubernetes: Use Confidential Virtual Machines and Enclaves to improve your cluster security,” Jul. 26, 2023. https://kubernetes.io/blog/2023/07/06/confidential-kubernetes/ (accessed Mar. 19, 2024).

[88] Sabt, Mohamed, Mohammed Achemlal and Abdelmadjid Bouabdallah. “Trusted Execution Environment: What It is, and What It is Not.” 2015 IEEE Trustcom/BigDataSE/ISPA 1 (2015): 57-64.

[89] Confidential Computing Consortium “A Technical Analysis of Confidential Computing”, 2021

[90] TRUSTZONE FOR CORTEX-A. [online] Available: https://www.arm.com/technologies/trustzone-for-cortex-a

[91] Weidner, Johannes. “ARM Confidential Compute Architecture A New Model of Trusted Execution Environment On The ARM Architecture.” (2023).

[92] Costan, Victor and Srinivas Devadas. “Intel SGX Explained.” IACR Cryptol. ePrint Arch. 2016 (2016): 86.

[93] El-Hindi, Muhammad, Tobias Ziegler, Matthias Heinrich, Adrian Lutsch, Zheguang Zhao and Carsten Binnig. “Benchmarking the Second Generation of Intel SGX Hardware.” Proceedings of the 18th International Workshop on Data Management on New Hardware (2022)

[94] "Trust domain extension", 2023, [online] Available: https://www.intel.com/content/www/us/en/developer/articles/technical/intel-trust-domain-extensions.html.

[95] Kaplan, David, Jeremy Powell, and Tom Woller. "AMD memory encryption." White paper (2016): 13.

[96] Rob Nertney, Confidential Compute on NVIDIA Hopper H100”, Nvidia, July 25, 2023. [Online]. Available: https://images.nvidia.com/aem-dam/en-zz/Solutions/data-center/HCC-Whitepaper-v1.0.pdf

[97] Liu, P., Xu, X., & Wang, W. (2022). Threats, attacks and defenses to federated learning: issues, taxonomy and perspectives. Cybersecurity, 5.

[98] Quoc, D.L., & Fetzer, C. (2021). SecFL: Confidential Federated Learning using TEEs. ArXiv, abs/2110.00981.

[99] Mo, F., Haddadi, H., Katevas, K., Marin, E., Perino, D., & Kourtellis, N. (2021). PPFL: privacy-preserving federated learning with trusted execution environments. Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services.

[100] Mondal, A.K., More, Y., Rooparaghunath, R.H., & Gupta, D. (2021). Poster: FLATEE: Federated Learning Across Trusted Execution Environments. 2021 IEEE European Symposium on Security and Privacy (EuroS&P), 707-709.

[101] QP-Prostate®: https://quibim.com/products/qp-prostate/

[102] PySyft: https://github.com/OpenMined/PySyft

[103] OpenMined: https://github.com/OpenMined

ANNEX 1

Quibim's required acquisition protocol for the use of QP-Prostate®

The QP-Prostate® analysis is based on T2-weighted magnetic resonance (MR), diffusion-weighted image (DWI) and dynamic contrast-enhanced (DCE) sequences. For a successful launch of QP-Prostate®, the study must include:

- The T2-weighted MR sequence and the DWI.
- The T2-weighted MR sequence, the DWI and DCE.

The inclusion criteria for the T2W, DWI and DCE are based on PI-RADS® v2.1 recommendations. The acceptance protocols for image acquisition are explained below.

1.1. T2-weighted required acquisition protocol for the use of QP-Prostate®

Parameter	PI-RADS® V2.1 COMPLIANT	Acceptable for analysis but not PI-RADS® V2.1 COMPLIANT	Stop Analysis
Field Strength	3T, 1.5T	3T, 1.5T	-
Acquisition sequence	Multiplanar (axial, coronal, and sagittal) T2-weighted images are usually obtained with 2D RARE (rapid acquisition with relaxation enhancement) pulse sequences, more commonly known as fast-spin echo (FSE) or turbo-spin-echo (TSE). To avoid blurring, excessive echo train lengths should be avoided.	Axial	Sagittal / Coronal
Receive Coil type	Without endorectal coil.	Without endorectal coil.	With endorectal coil.
Slice thickness	≤ 3 mm	≤ 4 mm	> 4 mm
Lipid suppression	OFF	OFF	ON
Gap thickness	≤ No gap	≤1 mm	>1mm
Field-of-view	Generally, 120-200 mm encompass the entire prostate gland and seminal vesicles.	≥120x120mm	<120x120mm
Pixel spacing	≤ 0.7 mm	≤ 1 mm	> 1 mm

DWI required acquisition protocol for the use of QP-Prostate®

Parameter	PI-RADS® V2.1 COMPLIANT	Acceptable for analysis but not PI-RADS® V2.1 COMPLIANT	Stop Analysis
Field Strength	3T, 1.5T	3T, 1.5T	-
Acquisition sequence	Diffusion-weighted Single-Shot Echo planar Imaging SS-EPI (free breathing).	SS-EPI	Other sequences
Receive Coil type	Without endorectal coil.	Without endorectal coil.	With endorectal coil.
Slice thickness	≤ 4 mm	≤ 5 mm	> 5 mm
Lipid suppression	ON	ON	OFF
Number of b-values	Two or more b-values. If only two b-values can be acquired, it is preferred that the lowest b-value should be set at 0–100 sec/mm^2 (preferably 50–100 sec/mm^2), and the highest should be 800 – 1000 sec/mm^2. Additional b-values between 100 and 1000 may provide more accurate ADC calculations. High b-value (≥ 1,400 sec/mm^2) is also mandatory. If it is not included in the sequence, it will be calculated from the low and intermediate b-value images (computed b-value).	2 (including one b-value < 50-100 s/mm^2 and one at highest b-value between 500-1500 s/mm^2)	< 2 b-values -no b-value between 0–100 sec/mm^2 -no b-value between 800 – 1500 sec/mm^2
Gap thickness	≤ No gap	≤ 1 mm	>1 mm
Field-of-view	160-220 mm	Covering the entire prostate	Non-covering the entire prostate
Pixel spacing	In-plane dimensions ≤ 2.5 mm phase and frequency	≤ 2.5 mm	>2.5 mm
TR	≥ 3000 msec	≥ 2000 msec	< 2000 msec
TE	≤ 90 msec	≤ 100 msec	>100 msec

DCE required acquisition protocol for the use of QP-Prostate®

Parameter	PI-RADS® V2.1 COMPLIANT	Acceptable for analysis but not PI-RADS® V2.1 COMPLIANT	Stop Analysis
Field Strength	3T, 1.5T	3T, 1.5T	-
Acquisition sequence	3D fast spoiled gradient recalled echo or equivalent.	3D fast spoiled gradient recalled echo or equivalent.	Other sequences
Receive Coil type	Without endorectal coil.	Without endorectal coil.	With endorectal coil.
Slice thickness	≤ 3 mm	≤ 4 mm	> 4 mm
Lipid suppression	ON	ON	OFF
Gap thickness	≤ No gap	≤ 1 mm	> 1 mm
Field-of-view	Encompass the entire prostate gland and seminal vesicles.	Encompass the entire prostate gland and seminal vesicles.	Not encompassing the entire prostate gland and seminal vesicles.
Pixel spacing	≤ 2 mm x ≤ 2 mm	≤ 2 mm x ≤ 2 mm	> 2 mm
Temporal resolution	≤ 15 sec	≤ 15 sec	> 15sec
Total observation rate	≥ 2min	≥ 2min	< 2min
TR	< 100 msec	< 100 msec	> 100 msec
TE	< 5 msec	< 5 msec	> 5 msec