Advancing oncology with federated learning: transcending boundaries in breast, lung, and prostate cancer. A systematic review

Anshu Ankolekar[1*], Sebastian Boie[2], Maryam Abdollahyan[3], Emanuela Gadaleta[3], Seyed Alireza Hasheminasab[4], Guang Yang[5], Charles Beauville[6], Nikolaos Dikaios[7], George Anthony Kastis[7], Michael Bussmann[8], Sara Khalid[4], Hagen Kruger[2], Philippe Lambin[1], Giorgos Papanastasiou[9*]

[1] Department of Precision Medicine, GROW Research Institute for Oncology and Reproduction, Maastricht University, Maastricht, the Netherlands

[2] Pfizer Pharma GmbH, Berlin, Germany

[3] Centre for Cancer Biomarkers and Biotherapeutics, Barts Cancer Institute, Queen Mary University of London, United Kingdom

[4] Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom

[5] Bioengineering Department and Imperial-X, Imperial College London, London, UK

[6] Flower Labs, Hamburg, Germany

[7] Mathematics Research Center, Academy of Athens, Athens, Greece

[8] Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany

[9] Pfizer Inc, New York, New York, USA

* Corresponding authors

Abstract

Federated Learning (FL) has emerged as a promising solution to address the limitations of centralised machine learning (ML) in oncology, particularly in overcoming privacy concerns and harnessing the power of diverse, multi-center data. This systematic review synthesises current knowledge on the state-of-the-art FL in oncology, focusing on breast, lung, and prostate cancer. Distinct from previous surveys, our comprehensive review critically evaluates the real-world implementation and impact of FL on cancer care, demonstrating its effectiveness in enhancing ML generalisability, performance and data privacy in clinical settings and data. We evaluated state-of-the-art advances in FL, demonstrating its growing adoption amid tightening data privacy regulations. FL outperformed centralised ML in 15 out of the 25 studies reviewed, spanning diverse ML models and clinical applications, and facilitating integration of multi-modal information for precision medicine. Despite the current challenges identified in reproducibility, standardisation and methodology across studies, the demonstrable benefits of FL in harnessing real-world data and addressing clinical needs highlight its significant potential for advancing cancer research. We propose that future research should focus on addressing these limitations and investigating further advanced FL methods, to fully harness data diversity and realise the transformative power of cutting-edge FL in cancer care.

Introduction

Oncology is undergoing rapid transformation due to the integration of machine learning (ML), which can enrich clinical evidence from large-scale datasets, surpassing traditional analytics [1-4]. However, as of today, ML models have predominantly been centralised within data silos [1, 2]. While centralised ML models have substantially advanced cancer research [3], the exponential growth and diversification of clinical data such as imaging, health records and molecular profiles now pose considerable challenges [4]. This surge in data, coupled with a trend toward international collaboration and standardised datasets, highlights the limitations of single-centre studies confined by local data acquisition practices and demographics. Multi-centre studies, drawing from diverse regions, offer a more comprehensive ML modelling approach. However, centralised models struggle to effectively exploit this increasingly complex data landscape, potentially compromising ML generalisability, performance, global applicability and trustworthiness [5]. While aggregating data from various sources in centralised data lakes potentially offers an alternative, it is susceptible to privacy breaches, complex data-sharing agreements and legal restrictions on data transfers [1].

Federated learning (FL) has emerged as a potential solution to these limitations. With FL, ML algorithms can be trained simultaneously on local datasets without data leaving their environment [6]. This decentralised approach allows hospitals and institutes to retain control over their data, addressing privacy concerns and regulatory restrictions, while benefiting from collective insights [7]. FL is particularly promising in oncology, where the data pertains to sensitive patient information and where timely collaborative analysis can have a significant impact on patient outcomes [8]. However, the adoption of FL is not without its challenges. Balancing effective model training with patient privacy techniques that can add computational overheads and may affect data contents, ensuring data quality and consistency across multiple centres, and maintaining robust model performance and trustworthiness are pressing concerns [9].

Given the rapid evolution and potential of FL in oncology, we conducted a systematic review to synthesise current knowledge, identify best practices, and highlight gaps in the existing literature. This review will provide researchers and clinicians with a

comprehensive understanding of how state-of-the-art FL can be effectively implemented to overcome the limitations of centralised ML. We analyse key FL/ML architectural and implementation designs, critically evaluate the effectiveness and scope of FL in breast, lung, and prostate cancer, and discuss best practices and considerations for future work. Further, we assess FL rigour using 2 objective criteria: a) inclusion of a comparative framework to evaluate the proposed FL/ML model against centralised ML baselines developed on the same datasets (either through direct evaluation in the study or by referencing reported literature values), and/or b) whether the proposed FL/ML model surpasses or demonstrates comparable performance to these baselines.

We evaluated state-of-the-art advances in FL, demonstrating its growing adoption amid tightening data privacy regulations. We perform this review analysis in the framework of the OPTIMA IMI2 project (OPTIMA | IMI Innovative Medicines Initiative), which focuses on combining ML and FL to enhance personalised diagnosis and treatment in breast, lung, and prostate cancer, and forms a systematic initiative to address global clinical challenges and unmet needs in cancer research. We aim to provide a comprehensive review of the current state of cutting-edge FL in these 3 oncology areas, to inform future research and clinical practices in the framework of the OPTIMA consortium and beyond.

Methods

Literature review strategy

A comprehensive literature search was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Figure 1) [10]. We systematically reviewed the literature published on FL techniques in breast, lung, and prostate cancer ML analysis, from January 1, 2020 to September 1, 2023. This period captures recent developments in FL, reflecting its transition to an established practice amid tightening data privacy regulations [6-9]. Despite initiating our literature review in early 2020, we found no publications matching our criteria until 2021 (Figures 2-3). Database searches included PubMed, Scopus, and Web of Science, utilising a combination of keywords and medical subject headings (MeSH) terms related to oncology, FL and ML. An expert librarian from the University of Oxford specified the keyword space.

Initial searches using broad keywords in abstracts, titles, and manuscript keywords related to oncology and ML yielded 5,766 papers. These keywords were: Oncology OR Cancer OR Carcinoma OR Malignant OR Neoplasm OR Tumor OR Tumour; Machine Learning OR Deep Learning OR Convolutional Neural Network OR CNN OR Generative Adversarial Network OR GAN OR Variational Autoencoder OR VAE OR Diffusion OR Transformer.

Subsequently, to focus the search, we mined only publications relevant to the 3 oncology areas of interest, FL and all possible patient-level data types involved in oncology. Specifically, we added the following keywords in the abstract, title, and manuscript keywords: Breast OR Mammary OR Prostate OR Lung Oncology OR Cancer OR Carcinoma OR Malignant OR Neoplasm OR Tumor OR Tumour; Federated Learning; Real World Evidence OR Real World Data OR Medical Imaging OR Magnetic Resonance Imaging OR Computed Tomography OR Positron Emission Tomography OR Ultrasound OR Echo OR Digital Breast Tomography OR Digital Pathology OR Genetic OR Genomics OR Transcriptomics OR Electronic Health Records OR Clinical Data OR Hospital Data OR Primary Care Data OR Secondary Care Data OR Computational Biology. By adding

these terms, we removed all irrelevant publications to the 3 oncology areas, FL and patient-level data types, leaving 81 papers for screening (Figure 1).

Non-journal publications and duplicates were removed by 6 authors (AA, MA, EG, SB, SAH, GP), resulting in 40 papers. These authors screened further for relevance focusing on FL methods in the specified cancer types using ML for analysis. We excluded non-English publications, studies outside the specified data types and oncology areas, and studies not focusing on FL applications, leaving 31 papers for full-text review. Following full-text review by the same 6 authors (in groups of 2 researchers per publication) and inclusion/exclusion criteria, we removed 6 more papers. Conflicts between authors were resolved through consensus with the rest of the review team. In total, 25 journal papers were included in our review analysis.

Review aspects

During full paper review, we considered the following aspects: (1) year of publication; (2) central ML technical task addressed; (3) clinical application addressed; (4) central ML model architecture used; (5) FL method; (6) aggregation strategies for FL; (7) device types; (8) datasets used for analysis (imaging, electronic health records, and other); (9) privacy method; (10) FL evaluation; (11) scope of FL; (12) oncology area. We also evaluated data diversity by (13) patient size and (14) data size. Furthermore, to assess whether it is possible to reproduce these FL techniques, we considered (15) data type per publication (private, public or both).

For clinical applications (3), we categorised: disease type differentiation, tumour identification, treatment response prediction, severity assessment, side effect prediction, survival analysis, and assessment of tumour recurrence. To improve clarity, we categorise all published work on cancer diagnosis against benign tumours and diseases other than cancer as “disease type differentiation”, on tumour classification, segmentation or detection from (bio-)medical imaging data as “tumour identification” and on staging as “severity assessment”. For ML models (4), we categorised: Classic ML: logistic regression, support vector machine, extremely randomised tree, random survival forest, random forest; Convolutional Neural Network (other than UNets and pre-trained models) (CNN ALL): CNN (Encoder-Decoder, E-D), 2D CNN (E-D), 3D CNN (E-D), CNN (E-D)

with multiple instance learning; Large Pre-trained (ALL): 2D ResNet, 3D ResNet, ResNet-201, ResNet-18, Mobilenet-v3, Xception-v3; GAN (ALL): GAN, WGAN, CycleGAN; UNet (ALL): 2D UNet, nnUNet, 3D UNet; Other: capsule neural network, recurrent neural network and region-based CNN.

In (5), we considered the 2 major FL methods: model-centric and data-centric [1]. In (6), aggregation strategy refers to the method by which updates or model parameters from participating devices are combined to generate a global FL model [1, 6-9]. This process involves aggregating local model updates or parameters, to balance the contributions of individual devices while preserving the privacy of the local data. Aggregation strategies play a pivotal role in FL by ensuring that the global model accurately represents the collective knowledge of all participating devices [1, 6]. For (7), the term “device types” refers to different categorisations of “central” devices based on their data distribution and functionality within the FL framework [1, 6]. In our reviewed work, we identified 2 device types: cross-device and cross-silo. In (11), the “scope of FL” encompasses all major end goals of scientific and clinical impact identified in our review: ML model generalisability (to unseen data instances from various local devices), ML model prediction improvement (by potentially learning from a broad range of data patterns and insights across devices), data privacy, disease understanding improvement (through the analysis of multiple device data), domain adaptation and training time reduction. Note that apart from data privacy, all other scopes benefit from increased data heterogeneity and patterns provided by FL/ML.

We assessed FL rigour based on 2 objective criteria across all FL scopes: a) whether a comparison framework was involved to evaluate the proposed FL technique against central ML baselines developed on the same datasets (either directly evaluated in the study or by reporting literature values); and b) if they outperformed or showed comparable results to these baselines.

Results

Research trends

Our review of FL publications in breast, lung, and prostate cancer research from 2020 to 2023 reveals a growing interest, with none, 4, 8 and 13 papers identified respectively.

We observed a diverse exploration of ML techniques including Large Pre-trained (10 papers), UNet (7), CNN (6), Classic ML (6), GAN (2), and Other (3). The majority addressed classification (14 papers), followed by segmentation (5), detection (5), and regression (1) (Figure 2b). Table 1 details technical tasks, ML algorithms, and evaluation metrics across studies.

In terms of clinical application, most studies focused on tumour identification (8 papers) and disease type differentiation (7), followed by severity assessment (4), treatment response prediction (2), survival analysis (2), and Other (2) (Figure 3a). Five studies did not mention their clinical application. On FL scope, the majority of FL techniques focused on improving ML model generalisability (14 papers), followed by ML prediction improvement (13), data privacy (8), disease understanding improvement (2), and Other (2) (Figure 3b). Table 2 details FL scopes and oncology areas.

Federated machine learning

We analysed combinations of datasets, central ML models, technical tasks, and FL methods (Figure 4). The Large Pre-trained models were used for (bio-)medical imaging data: magnetic resonance imaging (MRI), computed tomography (CT), whole-slide imaging (WSI), and X-ray. The UNet models were developed for MRI, CT, hybrid positron emission tomography-CT (PET-CT), and X-ray. The CNN models analysed diverse datasets including WSI, digital mammography, MRI, CT, and electronic health records (EHR). The Classic ML models were primarily used for EHR and imaging features extracted from CT [11, 13, 22] and MRI [22]. GAN models were only developed for MRI and CT [19, 22] and Other models for EHR, CT, and digital mammography [18, 31].

All ML model types addressed classification, with Large Pre-trained, Classic ML, and CNN (ALL), being the 3 most frequently used models (Figure 4, Table 1). Classification tasks were mainly addressed for (bio-)medical imaging data followed by EHR (Figure 4)

[11-14, 16, 18-21, 24, 26, 27, 29, 30, 32-35]. Segmentation and detection tasks were mainly approached through UNet variants [15, 22, 25, 28, 34] and Large Pre-trained models [17, 20], respectively. Some studies addressed multiple tasks (Table 1).

Federated cancer research

Subsequently, we analysed combinations of clinical applications, datasets, FL scope, and organ areas (Figure 5).

Tumour identification was addressed using (bio-)medical imaging data: CT [17, 18, 22], MRI [22, 34], X-ray [25, 35], PET-CT [15] and WSI [16]. Disease type differentiation was addressed using X-ray [35], WSI [16, 30], digital mammography [29, 31] and EHR data [14]. Severity assessment (staging) was investigated through (bio-)medical imaging data [18-20, 34]. Two pairs of studies explored the combination of tumour identification with other clinical applications: 1 pair focused on disease type differentiation (using WSI and X-ray respectively) [16, 35], while the other examined severity assessment (CT and MRI respectively) [18, 34]. Treatment response prediction (CT, MRI, WSI) [12] and survival analysis (clinical data, genomics) [27] were also examined as standalone tasks, and in combination (EHR, CT) [13]. Side effect prediction (EHR, CT) [11] and tumour recurrence (EHR) [26] were examined in 1 paper each.

In terms of FL scopes, most articles focused on model generalisability and ML prediction improvement (Figures 3b and 5, Table 2). For both ML model generalisability and prediction improvement, tumour identification was the predominant task (Figure 5). Breast and lung cancer research were the most frequent (8 studies each, Table 2).

Data diversity

The size of patient cohorts and data volumes varied considerably (Table 3). Most studies had cohorts of 100-1,500 patients, with data sizes (images or samples) in the range 1-5,000 [12, 19] and 5,000-1,000,000 [17, 27, 28]. Six papers analysed data from ≥1,500 patients, with 3 of these having explored data from >10,000 [13, 24] and >100,000 [32] patients, respectively. Two papers had small patient sizes of ≤100, corresponding to data sizes of 5,000-10,000 [16, 30]. Seven papers did not mention their patient size.

Datasets across studies were mostly public (18 papers) than private (9 papers), with 5 studies using a mixture of the 2 (Table 3; [15, 18, 22, 29, 32]).

FL implementation details

Most papers did not specify the FL method used (Figure 4). Among those that did, horizontal FL was the most frequent (11 papers), with only 2 papers referring to vertical FL [25, 27]. The remaining papers did not specify the model-centric technique employed. Horizontal FL was used mainly for classification tasks [14, 18-20, 24, 26, 33, 35], followed by detection [17, 20, 22], segmentation [22, 23] and regression [26].

Most papers did not report their aggregation strategy, with only 5 studies explicitly mentioning the techniques used: federated averaging [12, 16, 17, 29] and consensus model ensemble [11]. Device subtype information was limited, with only 4 papers referencing cross-silo FL [31, 32, 34, 35] and 1 referencing cross-device FL [30]. Only 5 papers explicitly reported their privacy method: differential privacy [26, 29], secure aggregation [14], secure multi-party computation [33] and homomorphic encryption [28].

Evaluating FL scope rigour

We evaluated articles for FL rigour based on 2 objective evaluation criteria: a) whether a comparison framework was used to evaluate FL against with central ML baselines on the same datasets, and/or b) whether FL outperformed or showed comparable results to these baselines. The following paragraphs are organised as follows: we start by describing the most frequent FL scopes identified in our review. When more than 1 FL scope is involved across papers, all scopes are explicitly detailed at the time each paper is first introduced under the relevant subsection.

ML model generalisability

FL can enhance model generalisability by using diverse data sources. For example, Agbley et al. used federated averaging with a pre-trained ResNet on histopathology images to classify breast tumours [16]. To overcome the challenge of dataset variability, the authors integrated various image magnification factors using self-attention. Their FL approach achieved 95.95% accuracy, surpassing various baseline models and demonstrating improved ML generalisability and prediction while protecting data privacy.

In the context of lung cancer, a paper by Zhu et al. presents a novel knowledge-sharing model for pulmonary nodule pre-segmentation and detection from CT data [22]. It uses a 3-stage framework with a) a UNet based mask generator, b) a discriminator with knowledge from electronic medical records and c) a random forest-based lung nodule detector. The system iteratively shares knowledge between a central server and client devices to improve the quality of generated masks and to normalise data distribution, addressing the challenge of non-independent and identically distributed (non-IID) data. Their FL technique outperformed a number of central ML baselines reaching a mean competition performance metric of 89% and a mean Dice score of 76% on non-IID data across each client, therefore improving model generalisability. In [24], the authors explored EHR for lung cancer classification, by developing single and cyclic weight FL using 2 underlying ML models: artificial neural network and logistic regression. By comparing their FL models against the same central ML models across 2 institutions, they showed that FL improved only the artificial neural network-derived results (accuracy ranged from 68-74%). Another collaborative learning method used a federated ResNet model for image classification of lung cancer using large data from 5 institutions with differing labels [32]. In total, they analysed >695,000 thoracic radiographs. The authors proposed a “flexible” FL architecture in which they divide the ResNet model into a classification head and a feature extraction backbone. The feature extraction backbone was shared across all sites, with weights jointly trained under a single FL scheme. Using their “flexible” FL method, the authors showed that model generalisability and classification accuracy were improved against locally trained ResNet models and conventional FL that used only uniformly annotated images.

In prostate cancer, Yan et al. developed a variation-aware FL (VAFL) method which aimed to assess tumour severity, through MRI classification [19]. The authors introduced VAFL to mitigate cross-client image variation. To perform VAFL, the client with the least complex data was selected to define the target image common space. A privacy-preserving GAN model was trained on these data to synthesise target MR images. A subset of these synthesised MR images was then shared with all clients and each client utilised a modified CycleGAN to map their own images onto this standardised target image space. Using synthetic MRI data, the authors improved both model generalisability

and ML classification accuracy in identifying clinically significant prostate cancer, outperforming local classifiers by reaching an accuracy of 98.75%. Data privacy also remained secured. In a multi-cancer work, the authors used a pre-trained ResNet-18 model to perform image detection and a CNN model to classify clinically significant against non-significant prostate cancer (from MRI data) and malignant against benign skin lesions [20]. They report that their FL model secured data privacy, and enhanced both model generalisability and ML classification accuracy, by outperforming locally trained classifiers and other FL model techniques. The diagnostic accuracy ranged from 95.6% to 82.9% on private data and from 88.7% to 73.4% on public data when their FL method was evaluated on 2 up to 32 clients, respectively. Another work focused on developing a federated 3D Anisotropic Hybrid Network for CT image segmentation in prostate cancer [23]. When tested on unseen data, this FL model consistently outperformed 3 local (private) models trained at individual institutions, reaching a Dice score of 89,5% on private data and 88.9% on public data, demonstrating superior generalisability and performance. Gao et al. proposed a novel swarm learning method for MRI and CT image segmentation in multi-disease data (cardiac, brain and prostate tumours), which could train UNet models using partially labelled images from multiple centres [28]. To tackle inhomogeneous label distributions across centres, they introduced a label skew-awared loss by consolidating global from local knowledge of partial labels. The authors demonstrated Dice scores in the range of 81.1-92.5% on non-IID data (across all disease areas), outperforming other FL methods and showing comparable results to centralised fully-supervised UNet models. In [34], the authors employed a FL approach to train a 3D UNet model with a region-of-interest classification head on diverse annotated prostate MRI data. This collaborative training approach led to substantial improvements in performance compared to local training, boosting lesion classification accuracy by 9.5-14.8% and nearly doubling lesion segmentation accuracy.

Improving ML prediction

FL harnesses the collective strength of diverse data sources, mitigating data biases and enabling the model to learn intricate patterns, potentially deriving more accurate predictions. In the following paragraphs, we detail articles which were designed to improve ML predictions through FL. Since there were studies that evaluated both FL/ML

model predictions and generalisability, we have already detailed these overlaps in the previous subsection [16, 19, 20, 32]. Here below, we highlight the remaining articles that satisfied our 2 criteria.

Breast cancer was the most frequent area in which FL was used for ML performance improvements [12, 14, 16, 27, 31, 33]. Study [12] applied federated averaging and multiple instance learning to WSI and clinical data, to predict the histological response to chemotherapy in early-stage triple-negative breast cancer. The authors aimed to improve ML predictions as well as disease understanding and achieved a mean AUC of 66% (range: 57-78%), outperforming central ML baselines. Study [14] focused on federating extremely randomised trees, to analyse distributed structured health data for disease classification, reporting an accuracy of 95.3% and an F1 score of 95.4%. It showed comparable results to central ML baselines. Another work expanded the application of privacy-preserving FL to survival analysis, assessing its potential in breast cancer genomics, with reported accuracies ranging from 81-92.5% across all 4 datasets explored [27]. The proposed FL outperformed all central ML baselines. In [31], the authors used a DenseNet feature extractor which fed an RNN-based classifier for breast cancer classification, from digital mammography data. The authors trained the model using a hybrid optimisation algorithm, achieving an accuracy of 94.22%, outperforming 3 central ML and 1 FL baselines. Another breast cancer work focused on enhancing model prediction and protecting data privacy by using a simple neural network classifier on clinical datasets, reaching an accuracy of 99% and outperforming a number of central ML baselines [33].

In lung cancer, we found 2 papers that applied FL to enhance ML predictions [17, 18]. The first study showcased an FL model employing a 3D ResNet18 for lung nodule detection, achieving an accuracy of 83.41% [17]. Although the proposed FL model was compared against previous FL methods, there was no comparison against central ML baselines. Study [18] combined FL with blockchain to create a collaborative, privacy-preserving model for lung cancer classification, demonstrating a high detection accuracy of 99.69% which outperformed central ML methods. The authors also reported that their FL method led to reduced training time, compared to central ML models.

In a multi-cancer study (including breast cancer) [26], the authors performed differential privacy to predict tumour recurrence from clinical data, by using a CNN model. The authors focused to improve ML predictions and disease understanding, demonstrating high FL accuracy (>90%) and outperforming central ML baselines.

Data privacy

Although data privacy is a core aspect of FL, only 8 papers explicitly mentioned it within their scope. Most of these papers were covered in the subsections “Model generalisability” [16, 19, 20] and “Improving ML prediction” [18, 27, 33], as their data privacy considerations were intertwined with those topics. In 2 publications, data privacy was addressed as a standalone focus and these are further discussed below [11, 30].

Peta and Koppu developed an automated breast cancer classification system using FL combined with deep learning to enhance diagnostic accuracy from histopathological images [30]. The authors compared the performance of their proposed Convolutional Capsule Twin Attention Tuna Optimal Network against existing deep learning models (i.e., CNN, BiLSTM, DNN, CapsuleNet), using the public BreakHis dataset. The authors reported superior performance with an accuracy of 95.68%, outperforming all previous centralised ML models.

In their lung cancer classification work, Field et al. developed a federated consensus model ensemble using logistic regression on private clinical data [11]. Their technique demonstrated comparable accuracy to their central baselines, reaching a mean AUC of 70%.

Improving disease understanding

Three of the papers reviewed used FL to improve disease understanding: 2 papers were about breast cancer [12, 29] and 1 was on multiple cancers [26]. Two of these papers have already been detailed in the “Improving ML prediction” subsection [12, 26]. In the remaining paper, the authors used federated averaging and a memory-aware CNN to refine breast cancer classification from digital mammography data [29]. This approach was shown to prioritise challenging instances that often experience prediction inconsistencies, thereby contributing to a more refined understanding of disease characteristics and diagnostic accuracy.

Other FL scopes

Two further FL scopes, reduced training time [18] and domain adaptation [21], were each the subject of 1 paper in our review. The study by Heidary et al has been previously described in the “Model generalisability” and “Improving ML prediction” subsections [11]. Although the authors in [21] describe that their FL methods outperformed conventional deep learning models, there was no such comparison identified in their study [21]. Therefore, their work did not meet our criteria for rigorous FL research and are not further described herein.

Discussion

In this systematic review, we examined FL methods in breast, lung and prostate cancer from 2020-2023. Distinct from previous surveys that focused on the theoretical and technological aspects of FL [1, 6-9, 36], our analysis investigates its practical implementation and impact in real-world breast, lung and prostate cancer settings. We examined the scope and evaluated the methodological rigour and effectiveness of FL in improving major ML domains such as model generalisability, predictive accuracy and data privacy, by comparing FL to centralised models and assessing its scientific and clinical impact. Most papers (18 out of 25 reviewed) met our objective criteria for FL rigour, including extensive comparisons against centralised ML baselines. Notably, FL methods outperformed centralised ML in 15 papers and showed comparable results in 3 papers. A diverse range of FL/ML techniques and their key clinical applications were also comprehensively explored and revealed. This review contributes meaningful insights into the real-world application of FL in cancer research, supporting the transition from promising proof-of-concepts to widespread implementation in clinical settings.

Among studies meeting our rigorous FL criteria, a diverse range of ML models were employed, including Large Pre-trained models, Classic ML, UNets and CNNs. Classification was the most common task, followed by segmentation and detection. While tumour identification and disease type differentiation were the dominant clinical applications, all clinical applications were represented in this group of studies. Improving ML generalisability and predictive accuracy were the primary focus of most studies meeting our FL criteria. Nevertheless, all FL scopes were represented, except for domain adaptation. Breast and lung cancer were the primary areas of focus, but studies on all organ areas were observed. Of these studies, 2 involved large patient populations (>1,500 patients) [20, 29] and 2 used very large datasets with over 20,000 and 900,000 patients, respectively (Supplementary Table 1) [24, 32]. The majority (6 papers) used moderate to large patient sizes (300-1,500 patients) [11, 12, 15, 19, 23, 26-28]. These FL techniques can potentially have merit to be further validated as generalised solutions for real-world cancer research.

By training models on large, diverse datasets from multiple sources, FL can uncover hidden patterns that may not be evident in smaller, isolated datasets. This can lead to the development of more precise diagnostic and treatment strategies tailored to specific patient groups (precision medicine) [37, 38]. FL can further support precision medicine, by potentially integrating multi-modal information from genetic, clinical, and (bio-)medical imaging data across institutions [39-46]. In fact, among studies that satisfied our FL rigour criteria, there were 7 papers which performed multi-modal ML analysis, combining information from various sources such as EHR, (bio-)medical imaging and/or genomics [11, 12, 21, 22, 26, 27, 28]. Of these studies, 1 focused on predicting response to treatment using WSI and clinical data [12], while the others aimed to develop diagnostic or identification methods. Thus, multi-modal FL offers a unique advantage for advancing precision medicine by integrating diverse patient-level data types across institutions, potentially enhancing patient outcomes.

Our analysis reveals a strong emphasis on generalisable models within FL frameworks across multiple oncology domains. Most of these publications met our FL rigour criteria [16, 19, 20, 22-24, 28, 32, 34]. The inclusion of heterogeneous data across institutions can potentially enhance the ability of ML models to learn more representative patterns, reduce overfitting and perform robustly across various clinical settings [37]. We also observed a large variability in data types across studies in this area, spanning from histopathological images to MRI, CT and EHR data. This shows that FL has considerable scope for developing generalisable ML models with broad applicability. Of note, 2 of these publications report results from non-IID data, which is central to ML model generalisability [2, 37]. Next to generalisability, learning from broad and heterogeneous data patterns across institutions can potentially improve ML performance. We observed a substantial focus on improving ML performance across oncology domains and data types. Most of these papers satisfied our FL rigour criteria [12, 14-16, 18-20, 26, 27, 31, 32]. The surge in publications from 2021 to 2023 underscores a growing awareness of these FL opportunities within the cancer research community.

Our findings highlighted potential advancements in evaluating tumour severity [18-20, 34] and enhancing disease understanding and characterisation [12, 26, 29]. These advancements can make clinical workflows more efficient by potentially assisting in

patient stratification, risk assessment as well as enhancing early diagnosis and treatment strategies. FL offers a notable advantage in accelerating ML model training through parallel processing, data privacy (eliminating the need for data sharing) and fast convergence due to model averaging [1, 6, 9]. A lung cancer study reported shortened training times without compromising data integrity [18], but other studies did not evaluate this aspect. These benefits may be offset by communication overheads due to sharing model updates (especially with large models or frequent updates) and the complexities of coordinating heterogeneous devices and privacy preservation techniques [36]. Further work is required to potentially demonstrate the benefits of FL in expediting ML model training and data-driven decision making.

Moreover, there are limitations and challenges that must be considered. The major limitation identified was the lack of FL implementation details across most publications. Specifically, only 13, 5 and 5 out of the 25 papers explicitly described the FL method, aggregation strategy and device subtype used, respectively (see Results). Of note, only 5 papers detailed their privacy method. Although most FL papers involved at least 1 public dataset, 4 papers involved only private data (Table 3). Moreover, among the 18 papers that met our FL rigour criteria, only 8 provided accessible code [11, 12, 14, 15, 28, 29, 32, 34]. These limitations hamper reproducibility and wider FL model adoption. We strongly advocate for increased transparency and open access practices to foster wider exploration and validation of these approaches in future large-scale cancer research. Comprehensive documentation of FL methods and code is important to ensure that findings can be reliably reproduced and externally validated.

Another major limitation was the use of datasets with low or moderate patient diversity in some studies (Table 3, Supplementary Table 1), which may be associated with limited generalisability, reduced robustness and susceptibility to biases. In addition, 7 papers did not meet our FL rigour criteria, which means that FL was not cross-validated against central ML baselines. Hence, the presence of systematic biases and the effectiveness of these FL methods in achieving their intended scope and clinical application remain unclear.

The underlying ML models used varied considerably even across the same technical tasks (Figure 4, Table 1). For instance, all ML model families were used to address either classification or detection tasks. This heterogeneity in both problems (data, oncology domains) and scientific solutions (ML models, FL methods) makes establishing best practices and method scalability challenging. In addition, many studies did not use a consistent set of metrics (Table 1). For example, while some studies report improvements over central ML models using accuracy and AUC metrics, others relied on precision, recall and F1-scores without clear justification. Another potential hurdle is the lack of universally accepted benchmark datasets for evaluating FL. Benchmark data with appropriate ground truths and evaluation metrics [47], are essential for unbiased comparison of different FL approaches. The lack of such datasets complicates transparent FL evaluations. However, most FL papers in our review involved at least 1 public dataset in their evaluation, which can help to establish frameworks for thorough benchmark studies.

Data privacy is crucial in FL, necessitating techniques like differential privacy and secure computation [1, 6]. All other FL scopes benefit primarily from increased data heterogeneity [48]. In our review, most of the papers reporting implementation details, focused on federated averaging [12, 16, 17, 29] which mitigates overfitting and enhances model generalisation by aggregating models trained on diverse local datasets [49]. As future work, methods like FedProx may be promising for biomedical data, since it regularises the local objective function and expedites training across diverse data distributions [50]. FedNova further refines this optimisation by normalising contributions based on local training steps [51], whilst FedDyn adapts the learning process to the unique characteristics of each local dataset [52]. Unlike other FL methods that prioritise either global or local model performance, Fed-ROD uniquely addresses fairness by optimising for both simultaneously, ensuring equitable outcomes for all clients regardless of the data distribution [53]. Furthermore, the FedOpt framework enables the integration of various FL optimisation algorithms [54]. Federated transfer learning, where pre-trained models are fine-tuned with federated data, can also potentially enhance ML predictions and disease understanding by improving adaptation to new oncology data [52]. Future work could investigate further advanced FL methods such as the aforementioned to

harness data heterogeneity for improved model generalisation, performance and fairness.

In conclusion, this comprehensive review serves as a foundational resource for the broader oncology community, demonstrating the burgeoning potential of cutting-edge FL to revolutionise breast, lung and prostate cancer research by leveraging diverse, real-world datasets while maintaining patient privacy. It also aligns perfectly with the overarching goals of the OPTIMA IMI2 project and provides crucial insights that will guide future research. Despite current challenges in reproducibility, standardisation and methodology, the clear advantages of FL in enhancing model generalisability, performance and addressing clinical needs across various cancer types, highlight its immense promise for harnessing diverse real-world clinical data and transforming cancer care.

Tables and Figures

Table 1) ML algorithm and evaluation metrics per technical task are presented. Publications occurring in multiple cells represent overlapping tasks and/or evaluation metrics. In papers where multiple (>1) tasks were involved, we only present the ML model for which evaluation metrics are reported. Note that we refer to technical tasks, ML algorithms and evaluation metrics, as reported in each publication. The term "Accuracy" presents a standalone metric (separate to ROC analysis), as reported by each paper authors. The term "Other" represents metrics that occurred once per ML model/technical task. ML: machine learning; ROC: receiver operating characteristic curve; MCC: Matthew's correlation coefficient; AUC: area under the curve; IoU: intersection-over-union.

Technical task	ML algorithm	Evaluation	Studies
Classification	Classic ML	ROC	[11, 13, 24]
		Accuracy	[14, 26]
		Other (MCC, Precision, Recall, F1)	[14, 27]
	CNN	ROC	[12, 29, 33]
		Other (Precision-Recall AUC, F1, Accuracy)	[29, 33]
	Large Pre-trained	ROC	[16, 20, 32]
		F1 score	[21, 35]
		Accuracy	[20, 21, 30, 32, 35]
		Other (Confusion matrices, Precision, Recall, Kappa coefficient)	[21, 30, 32 35]
	GAN	ROC	[19, 21]
Detection	Large Pre-trained	ROC	[17, 20, 31]
	Other	Other (Precision, F1, MCC)	[31]
Segmentation	UNet	Dice coefficient	[15, 22, 28, 34]
		Other (IoU)	[15, 34]
Regression	Other (CNN)	Other (Dice coefficient, Accuracy)	[23, 25]

Table 2) FL scope and oncology area examined per study. Multi-cancer refers to publications that examine more than 1 oncology area. The term "multi-disease" corresponds to studies in which data from at least 1 of the oncology areas of interest and from other diseases were analysed using FL. FL: federated learning.

FL Scope	Oncology Area	Studies
Data privacy	Breast	[16, 27, 30, 33]
	Lung	[11, 18]
	Prostate	[19]
	Multi-cancer	[20]
Domain adaptation	Multi-cancer	[21]
ML prediction improvement	Breast	[12, 14, 16, 27, 31, 33]
	Lung	[17, 18, 32]
	Prostate	[19]
	Multi-cancer	[15, 20, 26]
	Multi-disease	[35]
Disease understanding improvement	Breast	[12, 29]
ML model generalisability	Breast	[16]
	Lung	[13, 17, 22, 24, 25, 32]
	Prostate	[19, 23, 34]
	Multi-cancer	[20, 21]
	Multi-disease	[28, 35]
Training time reduction	Lung	[18]

Table 3) Data diversity in terms of patient and data size. Patient size corresponds to the number of individual patients included in the analysis and data size to the number of images or samples analysed. Studies using private and/or public data are also presented at the bottom of the table.

Patient size	Data size	Publications
1-100	5,000-10,000	[16, 30]
	NM	[14]
100-1,500	1-1,000	[12]
	1,000-5,000	[19]
	5,000-10,000	[28]
	10,000-100,000	[27]
	100,000-1,000,000	[17]
	NM	[11, 15, 23, 26]
1,500-5,000	NM	[29, 34]
5,000-10,000	10,000-100,000	[20]
>10,000	NM	[13, 24]
>100,000	100,000-1,000,000	[32]
NM	1-1,000	[33]
	1,000-5,000	[31]
	5,000-10,000	[21]
	10,000-100,000	[18, 22, 25, 35]

Data type	Private	Public
	[11-13, 15, 18, 22, 29, 32, 34]	[14-22, 25, 27-33, 35]

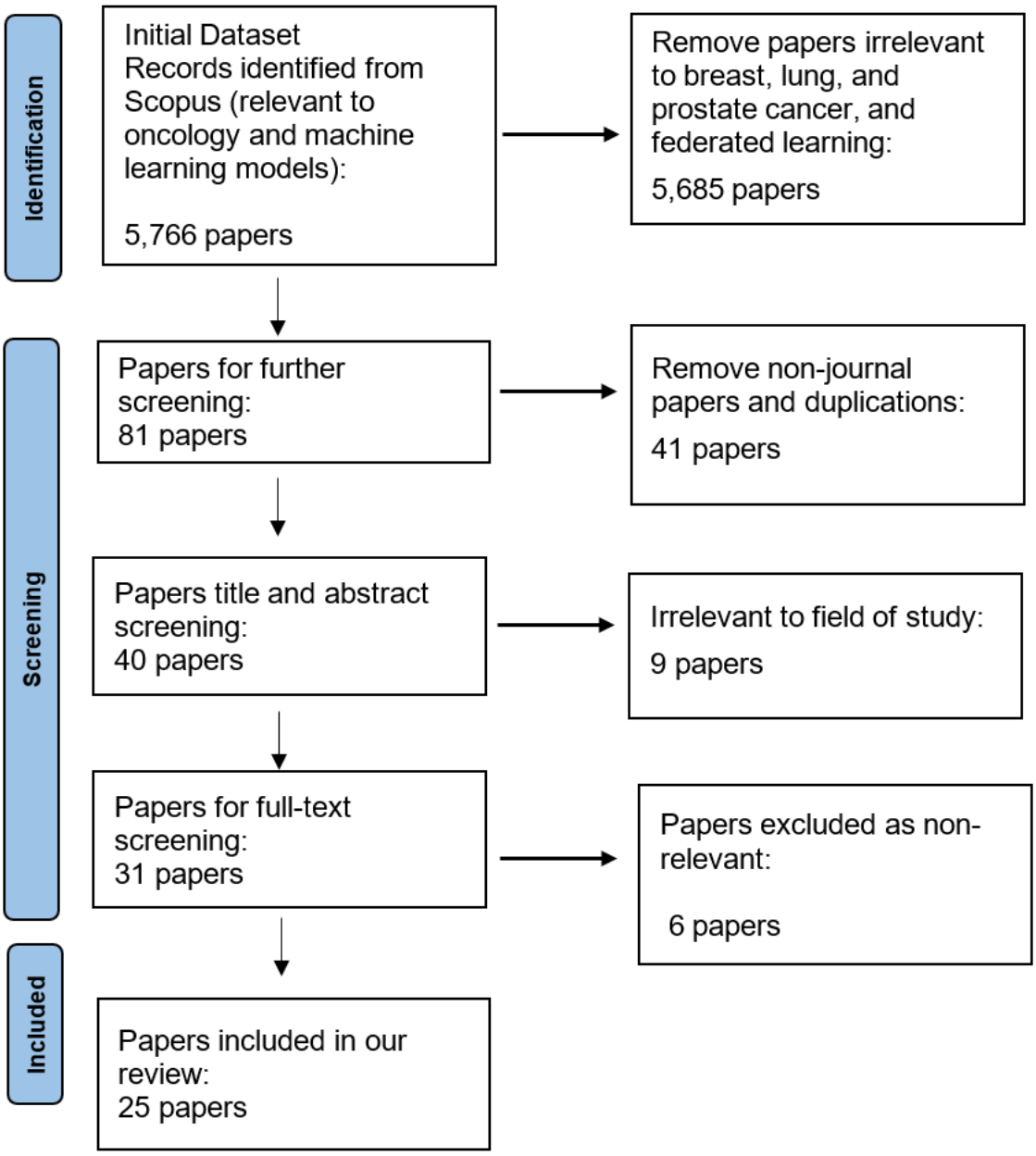


Figure 1) PRISMA flow of the systematic review process. The flow presents inclusion and exclusion of papers at each review stage.

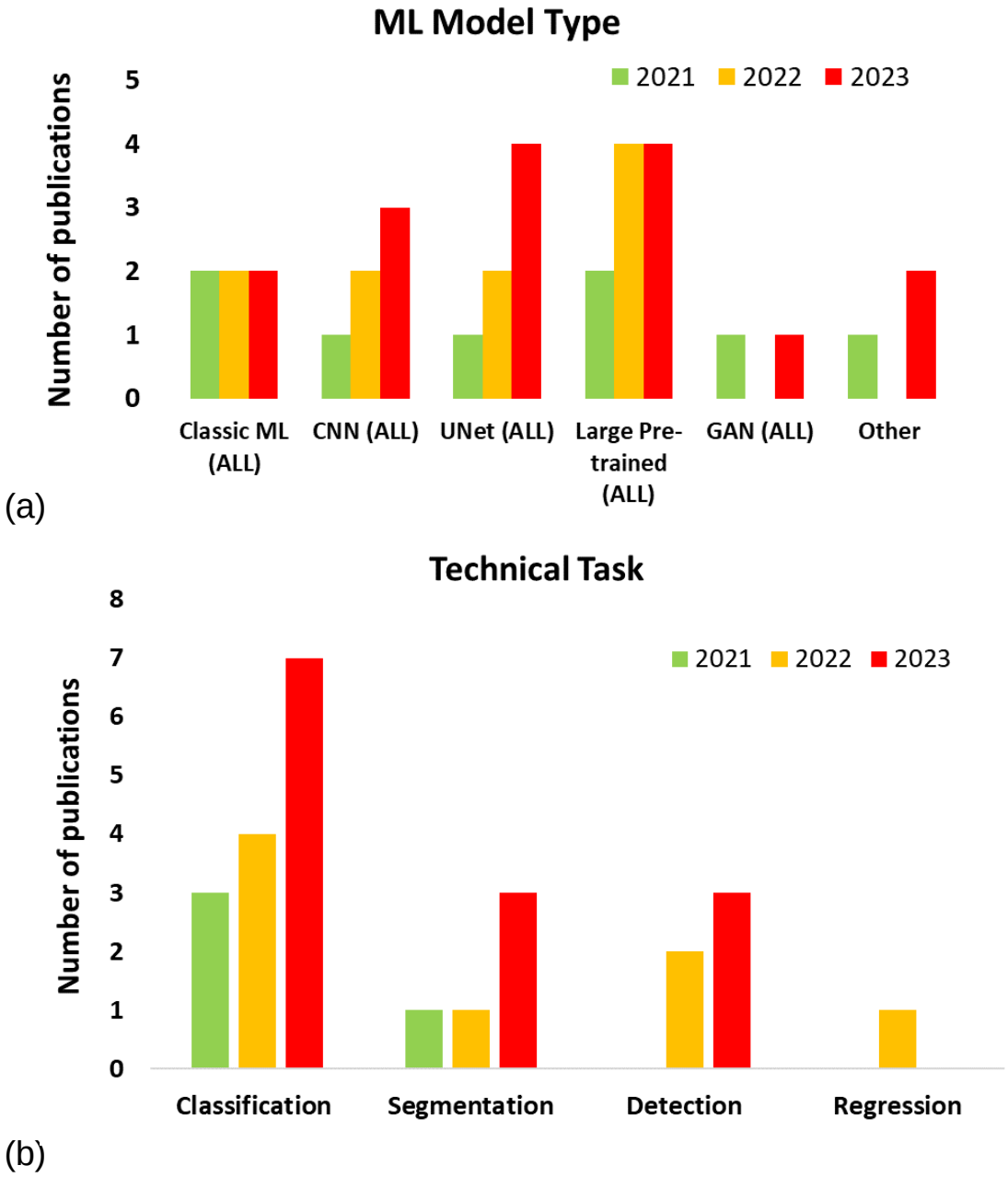
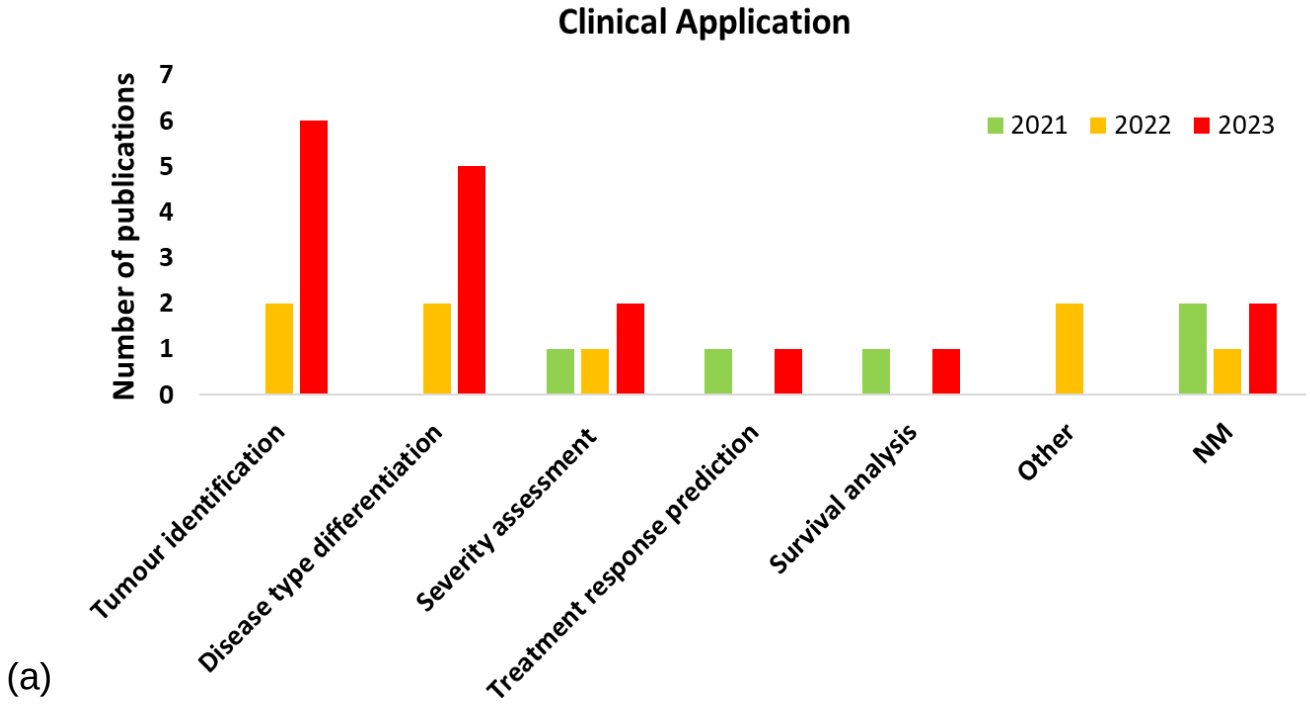
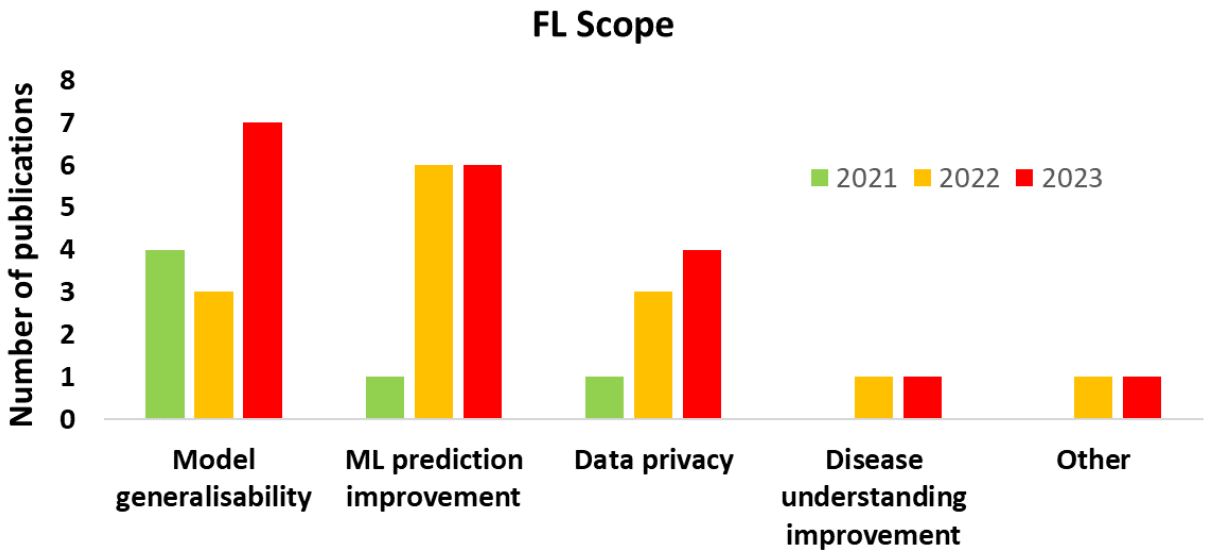


Figure 2) Publication record over time for all machine learning (ML) types (a) and technical tasks (b) identified. In a), "Other" included recurrent neural network, capsule neural network and region-based CNN.



(a)



(b)

Figure 3) Publication record over time for all clinical applications addressed (a) and FL scopes (b) identified. In a), "Other" included side effect prediction (1) and tumour recurrence assessment (1). In b), "Other" involved domain adaptation (1) and training time reduction (1). NM: not mentioned; FL: federated learning.

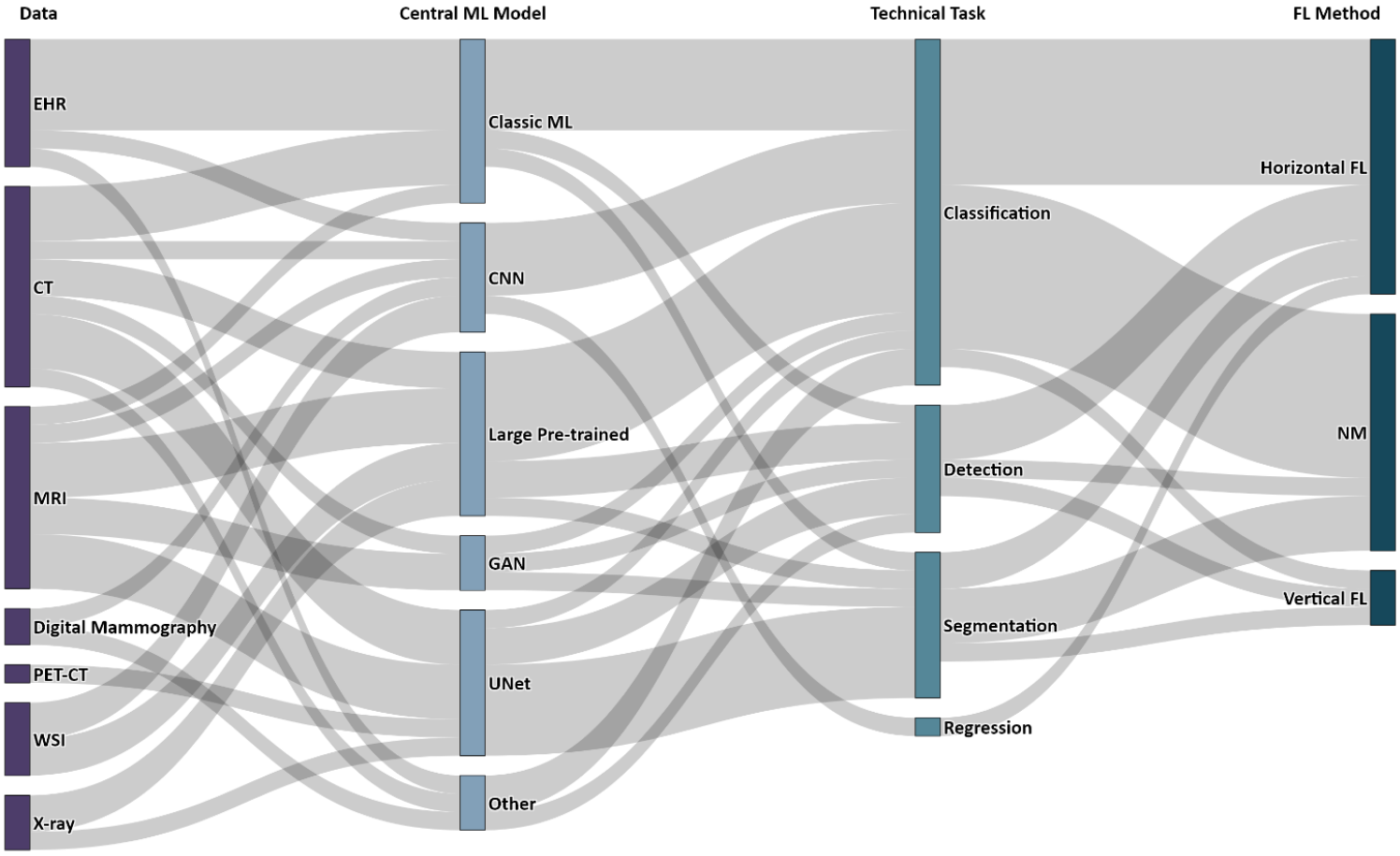


Figure 4) Sankey diagram depicting relationships (combinations) between the following technical aspects extracted across studies (represented as nodes): data, central ML model, technical task addressed and FL method. The width of each flow is proportional to the quantity being represented: thicker width corresponds to a higher combination prevalence across the reviewed papers, and vice versa. EHR: electronic health records; CT: computed tomography; MRI: magnetic resonance imaging; PET-CT: hybrid positron emission tomography-computed tomography; WSI: whole slide imaging; ML: machine learning; FL: federated learning; NM: not mentioned.

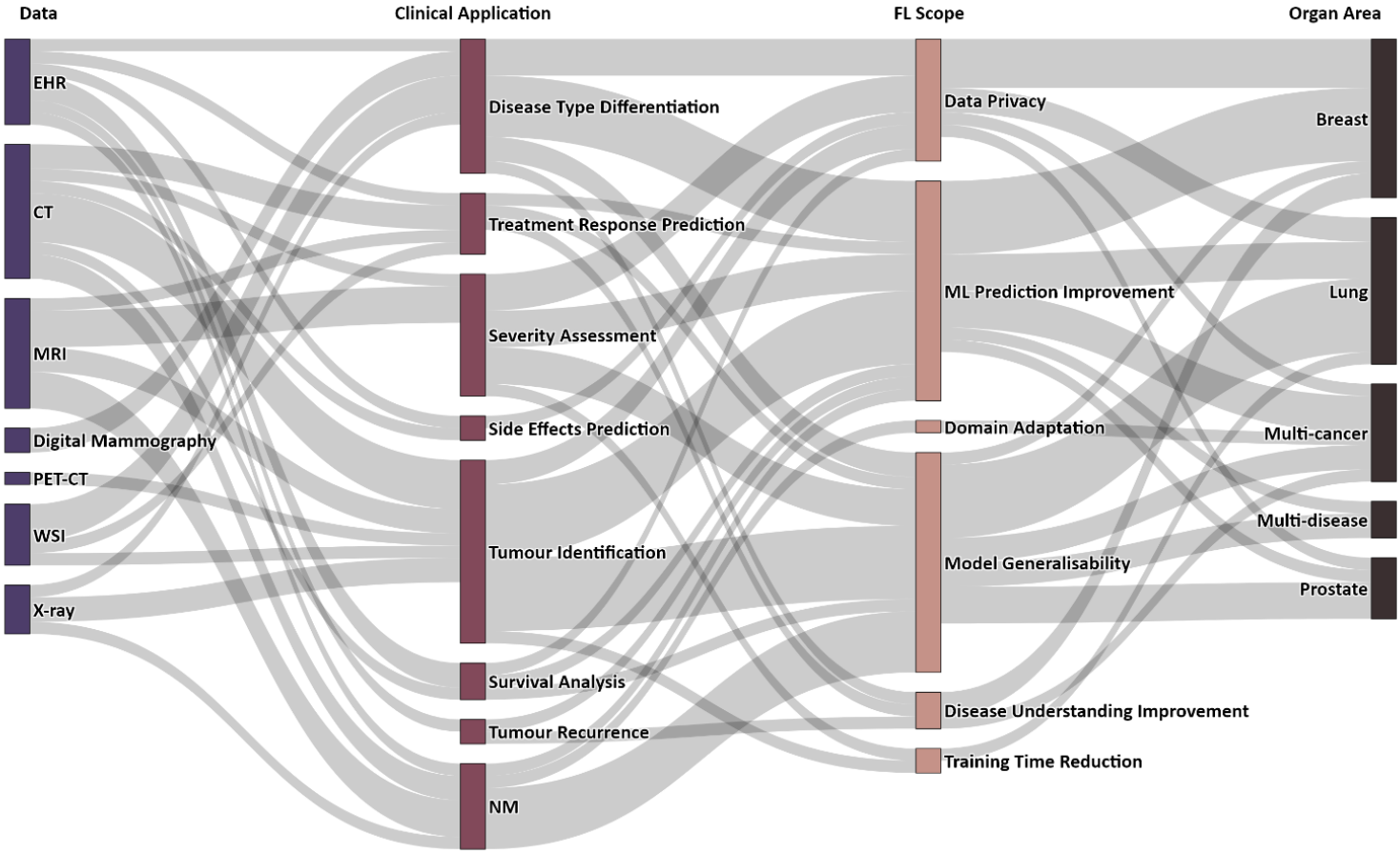


Figure 5) Sankey diagram illustrating combinations between the following application aspects extracted across studies (represented as nodes): data, clinical application, FL scope and organ area. EHR: electronic health records; CT: computed tomography; MRI: magnetic resonance imaging; PET-CT: hybrid positron emission tomography-computed tomography; WSI: whole slide imaging; FL: federated learning; NM: not mentioned.

Supplementary Table 1)

Detailed patient and data size information per study.

Data type	Patient size	Data size	Reference
Private	883 patients	NM	Field et al (2022) [11]
Private	676 patients	686 images	Terrail et al (2023) [12]
Private	12,047 patients	NM	Field et al (2021) [13]
Public	59 patients	NM	Aminifar et al (2022) [14]
Private, Public	534 patients	NM	Wang et al (2022) [15]
Public	82 patients	9,109 images	Agbley et al (2023) [16]
Public	1,010 patients	>300,000 images	Liu et al (2023) [17]
Private, Public	NM	>24,000 images	Heidari et al (2023) [18]
Public	323 patients	>2,000 images	Yan et al (2020) [19]
Public	7,802 patients	>10,000 images	Wicaksana et al (2022) [20]
Public	NM	>7,500 images	Subramanian et al (2022) [21]
Private, Public	NM	>10,000 images	Zhu et al (2022) [22]
Private	300 patients	NM	Sarma et al (2021) [23]
Private	23,000 patients	NM	Rajendran et al (2021) [24]
Public	311 patients	434 images	Horry et al (2023) [25]
NM	500 patients	NM	Ma et al (2022) [26]
Public	1,088 patients	15,848 samples	Archetti et al (2023) [27]
Public	1,367 patients	5,384 images	Gao et al (2022) [28]

Private, Public	2,722 patients	NM	Jimenez-Sanchez et al (2023) [29]
Public	82 patients	9,109 images	Peta and Koppu (2023) [30]
Public	NM	2,620 images	Kumbhare et al (2023) [31]
Private, Public	>100,000 patients	924,907 images	Tayebi Arasteh et al (2023) [32]
Public	NM	569 samples	Abou El Houda et al (2022) [33]
Private	NM	NM	Rajagopal et al (2023) [34]
Public	NM	17,017 images	Malik et al (2023) [35]

Funding

OPTIMA is funded through the IMI2 Joint Undertaking and is listed under grant agreement No. 101034347. IMI2 receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA). IMI supports collaborative research projects and builds networks of industrial and academic experts in order to boost pharmaceutical innovation in Europe.

The views communicated within are those of OPTIMA. Neither the IMI nor the European Union, EFPIA, or any Associated Partners are responsible for any use that may be made of the information contained herein.

Acknowledgements

We are grateful to Shona Kirtley, expert librarian from the University of Oxford, for her invaluable assistance in defining the keyword space.

Competing Interests statement

GP, SB and HK are full-time employees of Pfizer and hold stock/stock options. CB is a full-time employee of Flower. The other authors do not have any financial or non-financial competing interests to declare.

References

[1] N. Rieke et al., “The future of digital health with federated learning,” NPJ Digit. Med., vol. 3, no. 1, p. 1–7, 2020.

[2] A. S. Panayides et al., "AI in Medical Imaging Informatics: Current Challenges and Future Directions," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, p. 1837-1857, 2020.

[3] D. Painuli and S. Bhardwaj, “Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review,” *Comput. Biol. Med.*, vol. 146, p. 105580, 2022.

[4] P. Jiang, S. Sinha, K. Aldape, S. Hannenhalli, C. Sahinalp, and E. Ruppin, “Big data in basic and translational cancer research,” *Nat. Rev. Cancer*, vol. 22, no. 11, p. 625–639, 2022.

[5] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” *Brief. Bioinform.*, vol. 19, no. 6, p. 1236–1246, 2018.

[6] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, “A survey on federated learning,” *Knowl.-Based Syst.*, vol. 216, p. 106775, 2021.

[7] M. Ali, F. Naeem, M. Tariq, and G. Kaddoum, “Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, p. 778–789, 2022.

[8] A. Chowdhury, H. Kassem, N. Padoy, R. Umeton, and A. Karargyris, “A review of medical federated learning: Applications in oncology and cancer research,” presented at the International MICCAI Brainlesion Workshop, Springer, p. 3–24 2021.

[9] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, “Federated learning for healthcare informatics,” *J. Healthc. Inform. Res.*, vol. 5, p. 1–19, 2021.

[10] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *Bmj*, vol. 372, 2021.

[11] M. Field et al., “Infrastructure platform for privacy-preserving distributed machine learning development of computer-assisted theragnostics in cancer,” J. Biomed. Inform., vol. 134, p. 104181, 2022.

[12] J. Ogier du Terrail et al., “Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer,” Nat. Med., vol. 29, no. 1, p. 135–146, 2023.

[13] M. Field et al., “Implementation of the Australian Computer-Assisted Theragnostics (AusCAT) network for radiation oncology data extraction, reporting and distributed learning,” J. Med. Imaging Radiat. Oncol., vol. 65, no. 5, p. 627–636, 2021.

[14] A. Aminifar, M. Shokri, F. Rabbi, V. K. I. Pun, and Y. Lamo, “Extremely randomized trees with privacy preservation for distributed structured health data,” IEEE Access, vol. 10, p. 6010–6027, 2022.

[15] M. Wang, H. Jiang, T. Shi, and Y. Yao, “SCL-Net: Structured collaborative learning for PET/CT based tumor segmentation,” IEEE J. Biomed. Health Inform., vol. 27, no. 2, p. 1048–1059, 2022.

[16] B. L. Y. Agbley et al., “Federated fusion of magnified histopathological images for breast tumor classification in the internet of medical things,” IEEE J. Biomed. Health Inform., vol. 28, no. 6, p. 3389-3400, 2024.

[17] L. Liu, K. Fan, and M. Yang, “Federated learning: a deep learning model based on

resnet18 dual path for lung nodule detection," Multimed. Tools Appl., vol. 82, no. 11, p. 17437–17450, 2023.

[18] A. Heidari, D. Javaheri, S. Toumaj, N. J. Navimipour, M. Rezaei, and M. Unal, "A new lung cancer detection method based on the chest CT images using Federated Learning and blockchain systems," Artif. Intell. Med., vol. 141, p. 102572, 2023.

[19] Z. Yan, J. Wicaksana, Z. Wang, X. Yang, and K.-T. Cheng, "Variation-aware federated learning with multi-source decentralized medical image data," IEEE J. Biomed. Health Inform., vol. 25, no. 7, p. 2615–2628, 2020.

[20] J. Wicaksana, Z. Yan, X. Yang, Y. Liu, L. Fan, and K.-T. Cheng, "Customized federated learning for multi-source decentralized medical image classification," IEEE J. Biomed. Health Inform., vol. 26, no. 11, p. 5596–5607, 2022.

[21] M. Subramanian, V. Rajasekar, S. VE, K. Shanmugavadivel, and P. Nandhini, "Effectiveness of decentralized federated learning algorithms in healthcare: a case study on cancer classification," Electronics, vol. 11, no. 24, p. 4117, 2022.

[22] H. Zhu, G. Han, J. Hou, X. Liu, and Y. Ma, "Knowledge Sharing for Pulmonary Nodule Detection in Medical Cyber-Physical Systems," IEEE J. Biomed. Health Inform., vol. 27, no. 2, p. 625–635, 2022.

[23] K. V. Sarma, et al., "Federated learning improves site performance in multicenter deep learning without data sharing," J. Am. Med. Inform. Assoc., vol. 28, no. 6, p. 1259–1264, 2021.

[24] S. Rajendran, et al., "Cloud-based federated learning implementation across medical centers," JCO Clin. Cancer Inform., vol. 5, p. 1–11, 2021.

[25] M. J. Horry, et al., "Development of debiasing technique for lung nodule chest X-ray datasets to generalize deep learning models," Sensors, vol. 23, no. 14, p. 6585, 2023.

[26] Z. Ma, et al., "An assisted diagnosis model for cancer patients based on federated learning," Front. Oncol., vol. 12, p. 860532, 2022.

[27] A. Archetti, F. Ieva, and M. Matteucci, "Scaling survival analysis in healthcare with federated survival forests: A comparative study on heart failure and breast cancer genomics," Future Gener. Comput. Syst., vol. 149, p. 343–358, 2023.

[28] Z. Gao, F. Wu, W. Gao, and X. Zhuang, "A new framework of swarm learning consolidating knowledge from multi-center non-iid data for medical image segmentation," IEEE Trans. Med. Imaging, vol. 42, no. 7, p. 2118-2129, 2023.

[29] A. Jiménez-Sánchez, M. Tardy, M. A. G. Ballester, D. Mateus, and G. Piella, "Memory-aware curriculum federated learning for breast cancer classification," Comput. Methods Programs Biomed., vol. 229, p. 107318, 2023.

[30] J. Peta and S. Koppu, "Breast Cancer Classification In Histopathological Images Using Federated Learning Framework," IEEE Access, vol. 11, p. 61866-81880, 2023.

[31] S. Kumbhare, A. B. Kathole, and S. Shinde, "Federated learning aided breast cancer detection with intelligent Heuristic-based deep learning framework," Biomed. Signal Process. Control, vol. 86, p. 105080, 2023.

[32] S. Tayebi Arasteh, et al., "Collaborative training of medical artificial intelligence models with non-uniform labels," Sci. Rep., vol. 13, no. 1, p. 6046, 2023.

[33] Z. Abou El Houda, A. S. Hafid, L. Khoukhi, and B. Brik, "When collaborative federated learning meets blockchain to preserve privacy in healthcare," IEEE Trans. Netw. Sci. Eng., vol. 10, no. 5, p. 2455-2465, 2023.

[34] A. Rajagopal, et al., “Federated learning with research prototypes: Application to multi-center MRI-based detection of prostate cancer with diverse histopathology,” Acad. Radiol., vol. 30, no. 4, p. 644–657, 2023.
[35] H. Malik, A. Naeem, R. A. Naqvi, and W.-K. Loh, “Dmfl_net: A federated learning-based framework for the classification of covid-19 from multiple chest diseases using x-rays,” Sensors, vol. 23, no. 2, p. 743, 2023.
[36] R. Liu, P. Xing, Z. Deng, A. Li, C. Guan, H. Yu, "Federated Graph Neural Networks: Overview, Techniques, and Challenges," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2024.3360429, 2024.
[37] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, G. Yang, "Is Attention all You Need in Medical Image Analysis? A Review," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, p. 1398-1411, 2024.
[38] G. Huang, Y. Li, S. Jameel, Y. Long, G. Papanastasiou, “From explainable to interpretable deep learning for natural language processing in healthcare: How far from reality?”, Computational and Structural Biotechnology Journal, vol. 24, p. 362-373, 2024.
[39] N.B. Spath, T. Singh, G. Papanastasiou, et al. Assessment of stunned and viable myocardium using manganese-enhanced MRI, Open Heart, vol. 8, e001646, 2021.
[40] G. Papanastasiou, et al., "Multimodality Quantitative Assessments of Myocardial Perfusion Using Dynamic Contrast Enhanced Magnetic Resonance and 15O-Labeled Water Positron Emission Tomography Imaging," in IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 2, no. 3, p. 259-271, 2018.
[41] H. Jiang, et al. Semi-supervised Pathology Segmentation with Disentangled Representations. In: Albarqouni, S., et al. Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning. DART DCL, 2020.
[42] C. Wang, G. Yang, G. Papanastasiou, "FIRE: Unsupervised bi-directional inter- and intra-modality registration using deep networks," 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, p. 510-514, 2021.
[43] X. Xing, G. Papanastasiou, S. Walsh, G. Yang, "Less Is More: Unsupervised Mask-Guided Annotated CT Image Synthesis With Minimum Manual Segmentations," in IEEE Transactions on Medical Imaging, vol. 42, no. 9, p. 2566-2576, 2023.
[44] G. Papanastasiou, et al. Pharmacokinetic modelling for the simultaneous assessment of perfusion and 18F-flutemetamol uptake in cerebral amyloid angiopathy using a reduced PET-MR acquisition time: Proof of concept, NeuroImage, vol. 225, 117482, 2021.
[45] G. Papanastasiou, et al. “Focus on machine learning models in medical imaging”, Phys. Med. Biol., vol. 68, 010301, 2023.
[46] G. Papanastasiou, et al. Multidimensional Assessments of Abdominal Aortic Aneurysms by Magnetic Resonance Against Ultrasound Diameter Measurements. In: Valdés Hernández, M., González-Castro, V. (eds) Medical Image Understanding and Analysis. MIUA, 2017.
[47] T. Melistas, et al., “Benchmarking Counterfactual Image Generation”. arXiv preprint. arXiv:2403.20287v2, 2024.
[48] P. Kairouz, et al., "Advances and open problems in federated learning." Foundations and trends in machine learning, vol. 14, p. 1-210, 2021.

[49] H.B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, "Communication-efficient learning of deep networks from decentralized data", Artificial Intelligence and Statistics, vol. 54, p. 1273-1282, 2017.

[50] T. Li, A. K. Sahu, A. Talwalkar, V. Smith, "Federated learning: Challenges, methods, and future directions", IEEE Signal Processing Magazine, vol. 37, no. 3, p. 50-60, 2020.

[51] J. Wang, Z. Charles, Z. Xu, et al. Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization. NeurIPS, 2020.

[52] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, V. Saligrama, "Federated learning based on dynamic regularization." In International Conference on Learning Representations, 2021.

[53] H-Y. Chen, W-L Chao, "On Bridging Generic and Personalized Federated Learning for Image Classification." International Conference on Learning Representations, 2022.

[54] S. Reddi, et al., "Adaptive federated optimization", arXiv preprint arXiv:2003.00295, 2020.

[55] Q. Yang, Y. Liu, T. Chen, Y. Tong, "Federated machine learning: Concept and applications", ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, no. 2, p. 1-19, 2019.