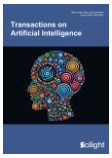
Scilight

Transactions on Artificial Intelligence

https://www.sciltp.com/journals/tai



Review

Federated Learning for Medical Image Analysis: Privacy-Preserving Paradigms and Clinical Challenges

Juntao Hu [1,2], Zhengjie Yang [1,3,*], Peng Wang [1], Guanyi Zhao [1], Hong Huang [1], Zhimin Zong [4] and Dapeng Oliver Wu [1]

[1] Department of Computer Science, City University of Hong Kong, Hong Kong
[2] School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China
[3] Hong Kong Generative AI Research and Development Center, The Hong Kong University of Science and Technology, Hong Kong
[4] Department of Electrical & Computer Engineering, University of Florida, Gainesville, FL 32611, USA
* Correspondence: zhengjie.yang@sydney.edu.au

How To Cite: Hu, J.; Yang, Z.; Wang, P.; et al. Federated Learning for Medical Image Analysis: Privacy-Preserving Paradigms and Clinical Challenges. *Transactions on Artificial Intelligence* **2025**, *1*(1), 153–169. https://doi.org/10.53941/tai.2025.100010.

Received: 14 June 2025
Revised: 1 August 2025
Accepted: 13 August 2025
Published: 18 August 2025

Abstract: Federated Learning (FL) has emerged as a transformative paradigm in medical image analysis, addressing the critical challenges of data scarcity and patient privacy. By enabling collaborative model training across decentralized datasets without requiring data sharing, FL aligns with stringent privacy regulations like HIPAA and GDPR. However, existing surveys on FL for medical image analysis often focus narrowly on aspects like privacy and security or fail to categorize methods within a clear taxonomy. Our survey bridges these gaps by systematically organizing FL methodologies for medical image analysis around three core pillars: training, architecture, and unlearning. We emphasize the unique demands of the medical domain, such as handling heterogeneous imaging modalities and annotations. Unlike prior works, our survey strikes a balance between technical rigor and clinical practicality, covering approaches not only for privacy and security but also for accuracy and efficiency. By synthesizing insights from various studies, we provide a comprehensive roadmap to guide researchers and practitioners in leveraging FL's potential to advance AI-driven healthcare.

Keywords: federated learning; medical image analysis; trustworthy machine learning; machine unlearning

1. Introduction

The rapid advancement of Artificial Intelligence (AI) in medical image analysis has revolutionized the field, enabling breakthroughs in disease diagnosis. Deep learning models, in particular, achieve remarkable accuracy in tasks such as tumor detection, organ segmentation, and pathology classification. However, these models traditionally rely on centralized training with large-scale datasets, which in healthcare are often siloed across institutions due to privacy regulations, ethical concerns, and logistical challenges. The sensitive nature of medical data, governed by strict policies like the Health Insurance Portability and Accountability Act (HIPAA) [1] and the General Data Protection Regulation (GDPR) [2], renders sharing raw patient information impractical, exposing centralized approaches to risks of privacy leaks, legal noncompliance, and data breaches. Furthermore, medical data is often fragmented across hospitals, clinics, and research centers, especially for rare diseases or underrepresented populations, resulting in small-scale local datasets and site-specific biases. These limitations ultimately compromise model performance.

Federated Learning (FL) has emerged as a transformative paradigm in medical image analysis to overcome these limitations. By enabling collaborative model training across distributed datasets without exchanging sensitive raw data, FL preserves patient privacy while leveraging diverse data sources—a critical advantage in

medical imaging, where data scarcity and heterogeneity hinder generalization. Despite its promise, implementing FL for medical image analysis introduces challenges, including performance degradation, communication overheads, requirements for robust trust and security, and demands for the right to be forgotten.

There have been several survey papers on FL for medical image analysis [3–8]. However, some surveys [8] overemphasize privacy and security concerns at the expense of critical methods for improving model performance and efficiency. Some [3,7] adopt a disease-centric taxonomy when presenting learning methods, leading to cumbersome repetitions of core FL algorithms rather than synthesizing how and why specific techniques address fundamental FL challenges like data heterogeneity or communication bottlenecks. In some surveys [5,6], the absence of a clear, unified methodological taxonomy results in overlapping content, increased complexity, and reduced clarity. Only a few of the surveys [4] comprehensively address the intertwined concerns of performance, efficiency, trust, and security within a coherent taxonomic framework. Furthermore, a near-universal omission is the discussion of unlearning techniques which are essential for compliance with data deletion in medical contexts.

To bridge these gaps, our survey systematically organizes FL methodologies for medical image analysis around three core pillars: training, architecture, and unlearning. We place paramount emphasis on the medical domain's unique demands, such as handling heterogeneous imaging modalities and annotations. Unlike prior works, our survey strikes a deliberate balance between technical rigor and clinical practicality, including approaches not only for privacy and security but also for accuracy and efficiency. By synthesizing insights from various studies, we provide a comprehensive roadmap empowering researchers and practitioners to harness FL's potential in advancing trustworthy, AI-driven healthcare.

2. Background

Recent advances in AI have demonstrated great potential in medical imaging, yet data privacy regulations and institutional silos hinder centralized model training. FL has emerged as a privacy-preserving alternative, enabling collaborative model development without sharing raw patient data. In this section, we first discuss the motivation behind FL for medical image analysis, highlighting challenges in privacy protection and local data scarcity that FL can address. Next, we detail the implementation of FL, including its problem formulation for medical image analysis and the client-server workflow, focusing on the widely adopted FedAvg [9] algorithm. Together, these subsections outline how FL can overcome critical barriers in medical AI while ensuring compliance and scalability.

2.1. Motivation

The implementation of AI in medical imaging relies on large-scale and diverse datasets to train robust and generalizable models. To acquire such datasets, traditional centralized machine learning paradigms pool data from multiple institutions into a single repository. However, medical images such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) scans, and X-rays are highly sensitive, often containing personally identifiable information. Improper use or sharing of these data can trigger ethical controversies and trust crises, threatening patient privacy and undermining public trust in medical institutions. To address these problems, governments have introduced stringent policies like HIPAA [1] and GDPR [2], imposing strict limits on the use and sharing of personal data. Thus, the data pooling inherent in centralized learning paradigms poses privacy breach risks, compliance challenges, and trust erosion. To tackle these issues, FL, a distributed machine learning paradigm, offers an innovative solution where only encrypted parameter updates are shared to iteratively improve the global model, resolving the dilemma between data sharing and privacy protection and laying the foundation for compliant development of medical AI.

Beyond centralized learning, training machine learning models using data from only a single site performs poorly due to limited sample sizes and data bias. Medical data is often fragmented across hospitals, clinics, and research centers, especially for rare diseases or underrepresented populations. Local datasets frequently contain only a small number of images, compounded by the high costs of accurate labeling. This fragmentation creates data silos, severely hindering model effectiveness. Moreover, datasets from a specific site often fail to reflect real-world data distributions. For example, one site may possess far more images for a particular disease than others, skewing model predictions and reducing clinical accuracy. Local datasets may also overrepresent regional patient characteristics or site-specific imaging protocols (e.g., scanner types and resolution), limiting model generalizability to broader populations or clinical environments. FL addresses these challenges by enabling multi-site collaboration without data centralization. By aggregating model updates from diverse institutions, FL leverages collective data volume and variability, significantly enhancing statistical power and generalizability to unseen datasets.

2.2. Implementation

2.2.1. Problem Formulation

In a FL scenario for medical image analysis, consider $K \in \mathrm{N}$ independent clients (e.g., hospitals, imaging centers, and research labs), each equipped with its own medical image dataset $\{D_1, D_2, ..., D_K\}$ (e.g., MRI slices, CT volumes, and X-rays). These clients cooperate to train a machine learning model (e.g., a U-Net for segmentation and a ResNet for classification) parameterized by ω. Crucially, no client can access other clients' datasets. The objective is to find an optimal model that minimizes the global loss function:

$$\begin{aligned} \min_{\omega} f(\omega) \quad &:= \sum_{k=1}^{K} p_k \, f_k(\omega, D_k), \\ \text{s.t.} \quad &\sum_{k=1}^{K} p_k = 1, \end{aligned} \tag{1}$$

where $f_k(\cdot)$ denotes the local loss function for client k (e.g., Dice loss for segmentation or cross-entropy for classification), evaluated over its private dataset D_k with $k \in \{1, 2, ..., K\}$. The term p_k represents the aggregation weight assigned to client k by the central server, typically reflecting the client's relative importance in the federated optimization process.

Since clients with larger datasets (e.g., tertiary hospitals) contribute more to the global model, p_k is usually set as the proportion of the dataset size of client k, following the popular federated learning algorithm FedAvg [9], i.e.,

$$p_k = \frac{|D_k|}{\sum_{i=1}^{k} |D_i|}. \tag{2}$$

The ultimate output of FL is the learned model parameters, which are broadcast to each client for deployment. The shared model achieves relatively acceptable performance on all client datasets, demonstrating superior generalizability compared to locally trained models.

2.2.2. Client-Server Architecture

A FL algorithm often orchestrates training via iterative synchronization between a central server (e.g., cloud-based coordinator) and clinical clients. Below, we expand each step with medical imaging-specific considerations based on the popular Federated Averaging (FedAvg) algorithm [9] proposed by McMahan et al. It serves as the basis for most widely used algorithms for FL. As shown in Figure 1, the server initiates and coordinates the entire training process (without accessing clients' private data) until a predefined termination criterion is satisfied. A typical FL workflow for medical imaging can be summarized as follows:

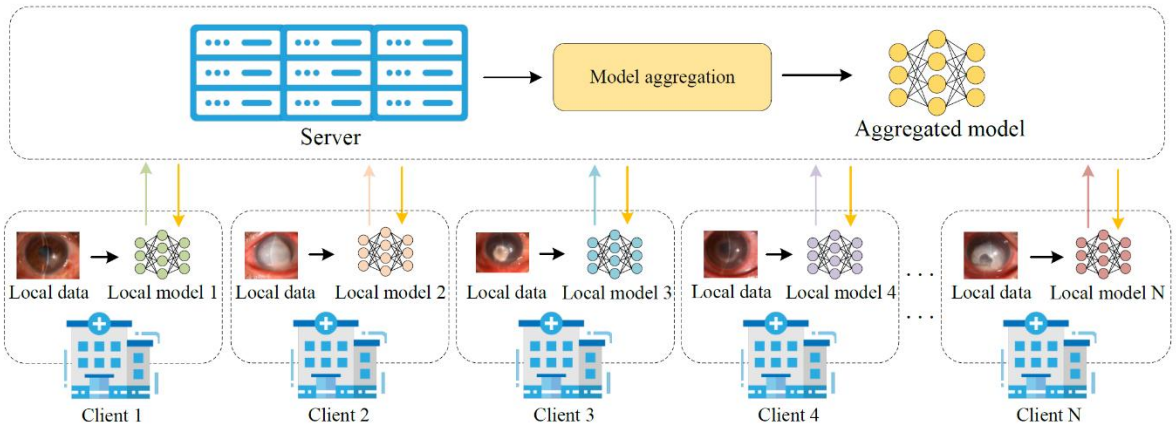


Figure 1. Overview of federated learning. The server aggregates clients' models (up arrow) to. refine the global model, and distributes the global model to clients (down arrow).

Initialization. First, a central server initializes a global model with parameter ω_0 suitable for medical tasks and distributes it to participating clients. Random initialization generates initial model parameters from statistical

distributions (like Gaussian distribution) for convolutional layers in imaging tasks. Other methods like Xavier Initialization [10] and Kaiming Initialization [11] also yield substantially faster convergence and better performance. Besides, pre-training is commonly applied in computer vision, where foundation models pre-trained on vast datasets carry extensive knowledge like in [12], which can be beneficial for FL model initialization.

Local training. Among T total training round (indexed by $t \in [0, T-1]$), in the t-th round, after receiving the global model parameter ω_t from the server (i.e., $\omega_{t,k}^0 \leftarrow \omega_t$) client k updates the local model parameter and trains the local model on its private dataset, performing forward and backward passes to compute model updates. Crucially, raw data remain localized; only model parameters or gradients are shared. During the forward pass, the loss is calculated via the loss function using the model output (and labels) from training data. During the backward pass, gradients derived from the loss update model parameters through backpropagation. For instance, in FedAvg, client k updates the local model parameter for E epochs (indexed by $e \in [0, E-1]$), and the update in the e-th epoch of the t-th round can be formulated as:

$$\omega_{t,k}^{e+1} \leftarrow \omega_{t,k}^{e} - \eta \nabla f_k\left(\omega_{t,k}^{e}, D_k\right), \tag{3}$$

where η is the learning rate. In medical image analysis, local training can help the model of each client to learn about local data (e.g., classify accurately for medical images in the local dataset). However, since medical images in different clients' datasets may vary significantly, the local model may not perform well for all data.

Aggregation. After E epochs of local training, the server collects updates from clients and aggregates them to refine the global model with parameter ω_{t+1}. For instance, in FedAvg, the server takes a weighted average of the local models:

$$\omega_{t+1} \leftarrow \sum_{k=1}^{K} p_k \, \omega_{t,k}^{E}. \tag{4}$$

Advanced techniques may incorporate differential privacy or non-uniform weighting to enhance security or fairness. After aggregation, the global model with parameter ω_{t+1} can retain some capabilities of each local model. Compared with a client's local model, the global model can behave better on other clients' data.

Distribution. The updated global model with parameter ω_{t+1} is redistributed to all participating clients for subsequent local training. This iterative process repeats over multiple rounds. With each round, the global model becomes increasingly refined and accurate through diverse data exposure. Training continues until meeting a termination criterion, such as achieving target accuracy or reaching a predefined iteration count.

Deployment. Once training concludes, the final global model is deployed. In medical image analysis, this final global model is expected to achieve an acceptable performance on all clients' data (e.g., classify accurately for medical images in all clients' local datasets). Key considerations for medical imaging include ensuring compatibility with hospital IT systems, meeting real-time requirements, and implementing continuous monitoring.

3. Methods

In Section 2, we explore a generic FL framework for medical image analysis, enabling distributed learning with privacy preservation. This framework accommodates most existing machine learning methods. For example, clients can employ U-Net models for medical image segmentation, training them collaboratively within the FL framework.

Federated learning for medical image analysis must simultaneously address model performance, communication efficiency, trust and security, and the right to be forgotten (challenging requirements). As shown in Figure 2, we categorize existing FL approaches into three groups to meet these challenges: (1) training, (2) architecture, and (3) unlearning.

In Section 3.1, we discuss training methods to enhance model performance. In Section 3.2, we examine architectural approaches to improve communication efficiency, trust, and security. Additionally, to address the right to be forgotten, Section 3.3 introduces unlearning methods.

3.1. Training

In a FL framework for medical image analysis, variations in disease prevalence, imaging protocols, or device types across institutions lead to skewed data distributions. Moreover, local datasets are frequently small-sized and deficient in label information. These issues may reduce the model performance. In this section, we focus on methods used to tackle these issues and ensure model performance. According to the type of labels the model learns from, we divide methods into supervised learning, weakly supervised learning, and unsupervised learning methods.

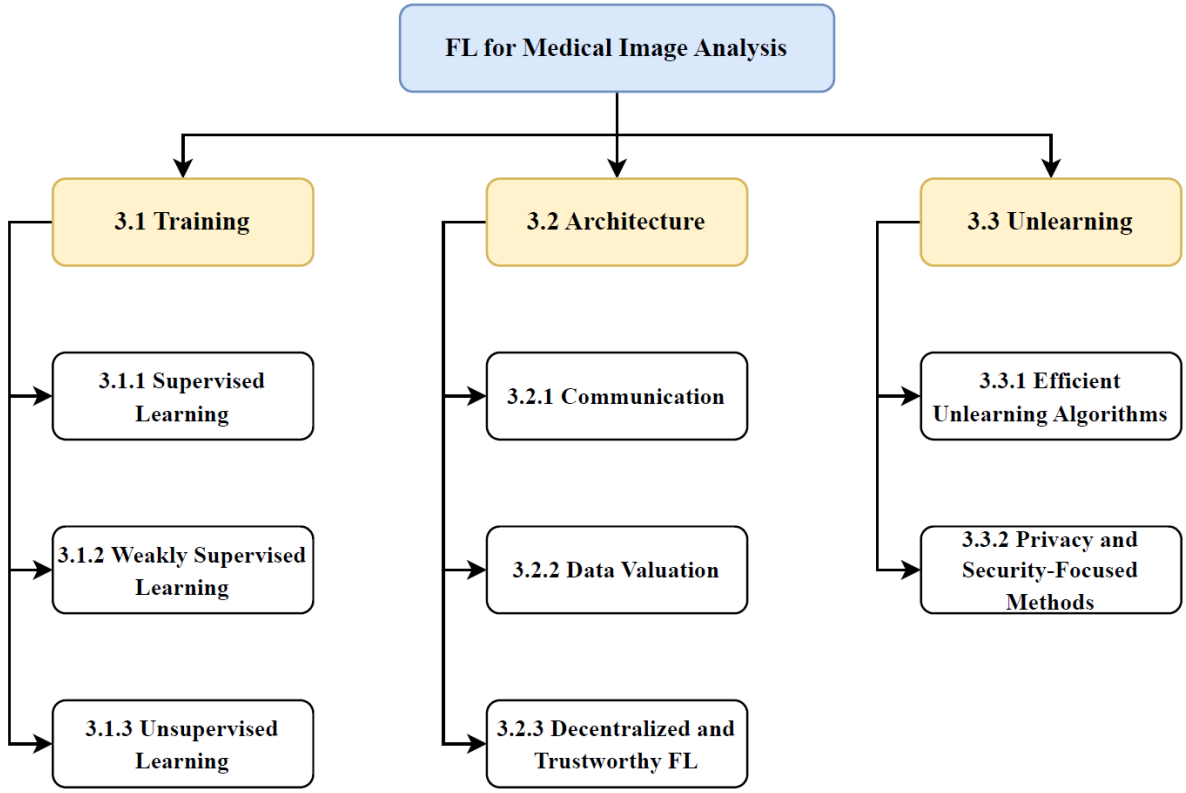


Figure 2. Our taxonomy. Existing FL approaches are categorized into training (Section 3.1), architecture (Section 3.2), and unlearning (Section 3.3).

3.1.1. Supervised Learning

Supervised learning aims to generalize from labeled examples to predict labels for unseen data by training models using labeled datasets, where each input example is paired with a corresponding target output (label). The model learns a mapping from inputs to outputs by minimizing the value of a loss function. While supervised learning relies on labeled data, it varies significantly based on the nature of the input data. This leads us to categorize it into single-modal and multi-modal learning.

Single-modal learning. In the realm of medical image analysis, centralized supervised learning methods have been widely explored. In most early works, only one modality of data is used to train models, such as X-rays, Magnetic Resonance Imaging (MRI), or Computed Tomography (CT) scans. Since AlexNet [13] achieves a historic breakthrough in the ImageNet competition by introducing the ReLU activation function, Dropout regularization, and multi-GPU parallel training, Convolutional Neural Networks (CNNs) have continuously deepened their architectures through various designs to improve classification accuracy. For example, ResNet [14] proposes residual connections, addressing the vanishing gradient problem through identity mapping, making it possible for network depth to exceed a thousand layers. DenseNet [15] further promotes feature reuse through cross-layer dense connections. CNNs are also used in medical image analysis [16–18]. Cheng et al. [16] propose a medical image segmentation network with a dual U-structure, using the pre-trained DenseNet as the first feature encoder. In [18], a hybrid deep learning model demonstrates improved accuracy in bladder cancer staging. In recent years, unlike traditional CNN designs, the Transformer architecture [19] divides images into sequences of 16 × 16 pixel patches and uses self-attention mechanisms to capture global dependencies. In [20], the authors propose a medical image segmentation framework, employing a hybrid CNN-Transformer architecture.

According to [21], core challenges of FL include statistical heterogeneity and system heterogeneity. Statistical heterogeneity (e.g., skewed label distributions, feature space variations, the non-independent and identically distributed (non-i.i.d.) nature of participant data) may make theoretical analysis and experimental evaluation more complex. And system heterogeneity (e.g., device storage, computing power, network latency differences) may lead to higher demands for robustness. Considering statistical heterogeneity, research by Hsu et al. [22] demonstrates that data distribution disparities can reduce global model accuracy by over 20%. In medical imaging scenarios, variations in pathological slide annotation standards and imaging equipment across hospitals may lead to gradient direction conflicts during model aggregation. To address this issue, personalized FL methods synthesize personalized models for clients. Some methods improve model alignment. For instance, FedProx [23] introduces

an *L2* regularization term for model divergence in local loss functions, constraining client updates from deviating significantly from the global model. FedCL [24] integrates the core idea of contrastive learning to align feature representations of the local model to the global model, improving the generalization ability of the model. Besides, Shi et al. [25] employ Bayesian modeling and meta-learning to generate personalized models for clients, achieving SOTA performance across 5 datasets. Fu et al. [26] develop a contrast-based FL algorithm (pFedMo) to personalize both model parameters and momentum [27,28], enabling a contribution-aware and accelerated training process. To deal with system heterogeneity, compressed models can be deployed on devices with limited resources. For instance, AQFL [29] leverages adaptive model quantization to represent model parameters with lower bits for some clients, effectively reducing communication and computational costs. Finally, in [30], the authors conduct a comprehensive evaluation of several state-of-the-art federated learning algorithms in the context of medical imaging.

Multi-modal learning. Since single-modality data often lacks sufficient context for accurate representation of complex scenes, multi-modal learning emerges, effectively fusing data from diverse modalities to enhance model performance and enable a more comprehensive understanding of complex real-world scenarios. Theoretical analysis [31] first explains the superiority of multi-modal learning from the generalization perspective, demonstrating that multimodal learning improves generalization by learning more accurate latent space representations. Empirical studies further validate these advantages in cancer prognosis [32–35]. In [32], integrating histopathology, genomics, and clinical data significantly improves staging accuracy. In [35], combining CT scans, clinical data, and radiomics features enhances recurrence prediction. However, these approaches predominantly rely on centralized architectures requiring a central server with access to comprehensive patient data repositories. This assumption hinders real-world applicability where medical data is inherently distributed across institutions.

To overcome this constraint, Multi-modal Federated Learning (MFL) has emerged as a distributed paradigm extending FL principles to multi-modal contexts. Many previous MFL frameworks [36–38] primarily focus on homogeneous settings with uniform client modalities. For instance, ref. [36] fuses dermatology images and clinical data for melanoma detection, and ref. [38] introduces a co-attention mechanism to address modality discrepancy in FL. In practice, however, modality distribution varies drastically across sites due to resource constraints (e.g., cost of MRI vs. CT scanners) and technical limitations (e.g., unavailable imaging protocols). For instance, advanced hospitals might offer MRI, CT, and Positron Emission Tomography (PET) scans, while a rural clinic may only have X-ray and ultrasound capabilities. This necessitates heterogeneous MFL frameworks that support modality-incongruent clients. In [39], an FL system effectively integrating X-ray and ultrasound data is developed, allowing clinics with distinct imaging capabilities to jointly train models for disease diagnosis. For robust Alzheimer's disease diagnosis, Chen et al. propose a nonconvex minimax penalty framework to select informative multi-modal features [40].

3.1.2. Weakly Supervised Learning

Weakly supervised learning encompasses methods that train models using incomplete, inexact, or inaccurate labels. This approach addresses critical challenges endemic to medical imaging, where high-quality annotated data is scarce, costly, and time-consuming to produce. In FL settings, these issues are compounded by data silos across institutions, making label scarcity a critical bottleneck. This section explores strategies to mitigate the impact of weak supervision in FL.

Incomplete labels. This category maximizes utilization of limited labeled images and abundant unlabeled data. Key methodologies include consistency normalization [41–43] and pseudo-labeling [44–46]. Consistency regularization enforces prediction invariance for unlabeled data under varied perturbations. Pseudo-labeling assigns provisional labels to unlabeled data, integrates them with manual labels, and refines these pseudo-labels through iterative training. For instance, Sohn et al. propose FixMatch [47], combining weak augmentation for pseudo-label generation and strong augmentation for training, achieving state-of-the-art results in classification tasks with incomplete labels. Furthermore, many similar techniques show promise in medical image segmentation [48–50].

However, direct application of these methods to FL remains challenging due to distributed data heterogeneity. Many approaches have tried to use a FL framework to solve problems with incomplete labels in specific scenarios. For example, in [51–53], some clients only have labeled data and others only have unlabeled data; in [54], each client has both labeled and unlabeled data; and in [55], labels exist only at the server, with clients holding only unlabeled data. There are also recent advances tackling data heterogeneity in FL with incomplete labels: Wu et al. [56] leverage client-level prototype similarity, prototypical contrastive learning, and consistency-aware aggregation to dynamically integrate knowledge from labeled and unlabeled datasets; Ma et al. introduce self-assessment confidence generation and reliable pseudolabel generation for personalized knowledge transfer across sites [57].

Inexact and inaccurate labels. This approach addresses pixel-level annotation scarcity by leveraging weaker supervision forms including image-level labels [58,59], bounding boxes [60–66], points [67], and scribbles [68–71]. These methods provide viable alternatives to fully supervised paradigms, reducing annotation burden while maintaining clinical utility in annotation-constrained medical contexts. Methods using image-level labels often train classification models first, then generate pseudo-masks via class activation maps to localize regions of interest for segmentation. Methods using bounding boxes often regularize the model by applying tightness constraints to avoid overfitting to noisy labels [60–64] or enhance initial pseudo-labels for segmentation using conditional random fields to align with box constraints [65,66]. Moreover, researchers have successfully utilized points and scribbles for medical image segmentation, significantly reducing labeling effort.

For medical image analysis, the imperative to reduce annotation costs necessitates FL frameworks accommodating heterogeneous weak supervision across institutions. Recently, focusing on bounding boxes, Zhu et al. [72] used Monte Carlo sampling to generate peer models for each client to correct noisy labels, tackling data heterogeneity in FL. To make full use of data labeled in any form, Wicaksana et al. [73] employ iterative pseudo-label generation and refinement for various weakly labeled data, followed by an adaptive aggregation strategy. Additionally, Lin et al. [74] utilize three contextual prompts and dual-decoder to enable personalization in the FL framework, addressing federated medical image segmentation with heterogeneous weak supervision.

3.1.3. Unsupervised Learning

Unsupervised learning in medical image analysis aims to extract meaningful patterns, features, or structures from medical images without explicit labels. Given the high cost and time-consuming nature of obtaining labeled medical data, unsupervised methods offer a valuable alternative. These methods rely on the inherent statistical properties and relationships within the data itself to uncover useful information. In this part, we focus on unsupervised anomaly detection and unsupervised image registration.

Unsupervised anomaly detection. Anomaly detection in medical imaging aims to identify abnormal samples deviating from expected normal patterns [75], crucial for spotting rare diseases, abnormal anatomical structures, or unexpected changes in patient conditions. Unsupervised anomaly detection works under the assumption that abundant normal samples with similar patterns are available, while abnormal samples with unknown patterns are scarce or impractical to collect comprehensively. Consequently, models are typically trained solely on normal samples. Cai et al. [76] categorize anomaly detection methods into reconstruction-, self-supervised learning-, and feature reference-based methods. Reconstruction-based methods [77,78] typically utilize generative models such as Generative Adversarial Networks (GANs) [79], Auto-Encoders (AEs) trained to reconstruct normal images. Anomalies are then detected during inference based on high reconstruction error. Self-supervised learning-based methods [80,81] directly leverage self-supervised learning to train models or learn related representations. Feature reference-based methods [82,83] detect anomalies by measuring deviations between a sample's features and reference features (e.g., feature maps or prototypes) derived from normal training data.

In unsupervised anomaly detection within FL frameworks, no labeled data is available to guide the learning process. Additionally, data collected from clients are often noisy, and distribution drift is common due to variations in device characteristics, acquisition tools, and target objects. Inspired by disentangled representation studies and brain MRI anomaly detection, Bercea et al. [84] collaboratively train an unsupervised deep convolutional autoencoder to identify pathologies like multiple sclerosis and vascular lesions. Moreover, Gupta et al. [85] propose a hierarchical FL framework using Long Short-Term Memory (LSTM) networks for anomaly detection, demonstrated in a remote patient monitoring scenario with digital twins and edge cloudlets. Recently, He et al. [86] introduce feature decoupling FL with a contrastive learning mechanism and a self-attention block, highlighting its effectiveness and superiority in multivariate time series anomaly detection.

Unsupervised domain adaptation. The domain shift problem [87] can adversely degrade model performance when using images with different appearances as clinical guidance. Unsupervised Domain Adaptation (UDA) addresses this by adapting models trained on a labeled source domain to an unlabeled target domain, primarily by minimizing distribution discrepancies between domains. UDA methods enforce information alignment from feature-level or pixel-level perspectives. Feature-level methods [88,89] focus on aligning domain distributions by adjusting discriminative feature spaces. Pixel-level methods [90,91] typically generate images combining source-domain content with target-domain style via adversarial learning. To fully utilize Transformer-derived attention in UDA, Ji et al. [92] propose meta-attention for alignment. Recently, Deng et al. [93] present the first electron microscopy image denoising method from a UDA perspective, exploiting electron microscopy image characteristics to bridge synthetic source and real target domains.

Inspired by FL and UDA, Peng et al. [94] introduced the concept of Unsupervised Federated Domain Adaptation (UFDA). The non-i.i.d. nature of participant data can lead to domain shift between clients. In [94], the authors propose Federated Adversarial Domain Adaptation (FADA), transferring knowledge from distributed source domains to an unlabeled target domain using a dynamic attention mechanism. Recently, in [95], a knowledge distillation-based multisource domain adaptation method for FL is proposed, using contrastive learning to control single-source domain drift and align local model representations with the global model.

3.1.4. Discussion

In training methods of FL for medical image analysis, current challenges mainly include statistical and system heterogeneity, label scarcity, and unsupervised learning complexity. For statistical heterogeneity, non-i.i.d. data distributions across clients (e.g., varying imaging protocols and disease prevalence) lead to gradient conflicts during model aggregation, degrading global model performance. Existing methods like FedProx ($L2$ regularization) or FedCL (contrastive learning) improve model alignment but may compromise local model adaptivity or introduce computational overhead. Also, system heterogeneity may challenge robustness. Model compression methods like quantization can help to tackle this issue. For label scarcity, medical imaging often lacks high-quality, pixel-level annotations. Pseudo-labeling and consistency regularization for incomplete labels in FL require careful thresholding and iterative refinement, which are sensitive to client-specific data biases. Weakly supervised methods rely on inexact and inaccurate labels (e.g., image-level annotations and bounding boxes), which introduce bias and reduce segmentation accuracy. For unsupervised learning complexity, unsupervised anomaly detection in FL struggles with noisy client data and distribution drift (e.g., device-specific imaging variations), requiring robust feature representations that capture cross-institutional consistency. UDA methods like FADA face challenges in scaling to multi-modal, multi-center settings with diverse domain shifts.

To ensure model performance in FL for medical image analysis, we can explore metalearning frameworks to rapidly adapt global models to heterogeneous client data, reducing performance gaps caused by non-i.i.d. distributions. To tackle label scarcity, we can explore generative models (e.g., GANs) to synthesize desensitized, cross-modal data on the server, augmenting scarce local datasets without privacy risks. Additionally, we can leverage foundation models like in [12] pre-trained on large-scale medical datasets to initialize FL models, improving convergence speed and generalization.

3.2. Architecture

In a FL framework for medical image analysis, optimizing the transmission of large model updates is essential to reduce bandwidth costs. Equally important is ensuring robust security through protected communication channels and defenses against malicious actors, such as those attempting model poisoning. This section focuses on the methods employed to address these challenges and ensure efficiency, trust, and security within the FL system.

3.2.1. Communication

Client-server communication is a vital research topic in FL for medical image analysis. There are two research tracks: (1) secure communication focusing on content privacy (raw training data, model parameters, etc.) and (2) efficient communication between clients and the server to decrease communication overheads. We mainly gather recent emerging works (2024–2025) to fill the gap that previous surveys have not covered.

Secure communication. The goal of secure communication is to let clients collaboratively train a global model with distributed data while preventing sensitive data leakage. Techniques including Differential Privacy (DP) [96] and Homomorphic Encryption (HE) [97] can be leveraged to secure content during communication. In [98], the authors propose a general FL framework to encrypt model parameters on the client side before communication and transmit the encrypted data via secure communication protocols. On the server side, the encrypted data is decrypted by a private key. The framework is trained and tested on the MedMNIST dataset [99]. In [100], the authors develop a new data partition technique tailored to mitigate the label skew problem in FL and utilize CKKS homomorphic encryption to improve communication security. The system is implemented to help analyze Parkinson's disease using the PD–BioStampRC21 dataset [101]. In [102], the authors propose an adaptive differential privacy-based FL model by adding random noise to client/local models before uploading to the server. The system is implemented to predict COVID-19 disease on COVID-19-Pneumonia-Normal chest X-ray images [103]. Besides, in [104], the authors propose a privacy-preserving FL framework based on trusted execution environments to improve privacy against various attacks. Other works [105–107] related to homomorphic encryption and random processes using Markov chains to secure model parameters have also emerged to help diagnose lung abnormalities and brain tumors.

Efficient communication. The goal of efficient communication is to accelerate the training process. To achieve this, we can either improve the training performance between two consecutive communication rounds or decrease the communication overheads per communication round. In this way, to achieve the target model performance goal (e.g., model testing accuracy), the total resource consumption is decreased. In [108], the authors propose an adaptive model aggregation mechanism to employ either FedAvg for computationally efficient aggregation or FedSGD for finer gradient updates to guarantee optimized aggregation performance in different non-i.i.d. data distributions. In [109], by utilizing Knowledge Distillation (KD) [110], the proposed framework FKD-Med distills the knowledge from the rich teacher model and trains a light-weight student model in a distributed manner, significantly decreasing both computation and communication costs while maintaining model performance. In [111–113], these methods apply sparsification techniques to compress model weights, improving computational, memory, and communication efficiency in FL. In [114], the authors adapt their framework from a traditional two-tier FL architecture to a three-tier hierarchical architecture [115], replacing expensive client-cloud communication with low-cost client-edge communication to significantly decrease communication overheads. In [116], the proposed pFLFE framework develops a newly-designed feature enhancement network trained exclusively on positive samples to tackle client drift in medical image segmentation. It also introduces an alternative fast-converging framework to achieve comparable model performance with fewer communication rounds.

3.2.2. Data Valuation

Accurate data valuation is vital for establishing fair incentive mechanisms, with significant potential to motivate healthcare institutions to contribute high-quality data to FL models for applications like precision medicine and population health analytics. The Shapley value [117] offers a theoretical basis for assessing contributions based on model performance gains. Wei et al. [118] propose a gradient-based method to efficiently estimate contribution indices, potentially reducing computational overhead for large medical datasets like imaging archives. Liu et al. [119] introduce a guided truncation gradient Shapley approach with advanced sampling, improving efficiency in dynamic FL settings, such as real-time telemedicine platforms. Fan et al. [120] develop a matrix-based federated Shapley scheme, capturing contributions across data subsets, which could support equitable valuation in healthcare collaborations. Fair incentives extend beyond compensation, fostering data quality and diversity—critical in healthcare where datasets may be limited or skewed. By encouraging participation from diverse institutions, such as rural hospitals or specialized clinics, FL models could improve robustness and generalizability, particularly for rare diseases or underrepresented demographics. However, challenges include high computational complexity and risks of manipulation, such as data poisoning, which could undermine fairness in heterogeneous medical FL environments.

3.2.3. Decentralized and Trustworthy FL

Blockchain-enhanced FL frameworks provide robust security and trust, with promising applications for safeguarding sensitive medical data, such as EHRs and outputs from Internet of Medical Things (IoMT) devices. Kim et al. [121] introduce a blockchain-based FL system that encrypts local model updates, potentially preventing tampering in collaborative medical research across institutions. Xian et al. [122] propose a smart contract-based FL framework, enabling automated verification of updates to ensure transparency in multi-institutional health studies, such as clinical trials. Otoum et al. [123] develop a trust-scoring mechanism with Byzantine fault tolerance, adaptable for securing IoMT-based FL in resource-constrained environments like remote patient monitoring. Myrzashova et al. [124] explore blockchain applications in FL, emphasizing their potential for secure data management in healthcare scenarios like telemedicine networks. Zhu et al. [125] classify blockchain-FL models, highlighting privacy-preserving mechanisms suitable for cross-border health data sharing. Beyond security, blockchain's decentralized structure aligns with FL's distributed nature, offering a framework for managing patient consent across healthcare providers. For instance, a blockchain-based system could track and verify patient authorization in real-time, enhancing trust and accountability in data-sharing networks. Nevertheless, scalability remains a concern due to the high computational and storage demands of consensus mechanisms, hindering widespread adoption in large-scale medical FL systems.

3.2.4. Discussion

In methods regarding the architecture of FL for medical image analysis, current challenges mainly include communication overhead, security risks, and difficulties in fair data valuation. For communication overhead, transmitting large model updates (e.g., in FedAvg) between clients and servers incurs high bandwidth costs. Besides, low-power edge devices (e.g., mobile clinics) may struggle with complex aggregation algorithms, thus

requiring lightweight architectures. For security risks, model poisoning attacks and privacy leaks during aggregation require robust defenses (e.g., differential privacy and homomorphic encryption), which often increase computational complexity. Blockchain-based FL frameworks improve transparency but face scalability issues due to high storage and consensus overheads. For fair data valuation, quantifying each client's data contribution (e.g., using Shapley values) is computationally intensive for large-scale FL systems, hindering the establishment of fair incentive mechanisms.

In the future, to tackle communication overhead, we can develop compressed model update techniques (e.g., knowledge distillation and sparse gradient transmission) to reduce data transfer volume. To address security risks, we can combine HE for critical updates with DP for non-sensitive data to balance privacy and efficiency in multi-tier FL architectures (e.g., client-edge-cloud) [115,126]. To achieve fair data valuation, we can integrate real-time performance metrics (e.g., model improvement per client) into Shapley value calculations, enabling adaptive incentive mechanisms for medical institutions.

3.3. Unlearning

In recent years, the escalating demand for the "right to be forgotten"—as enshrined in regulations such as GDPR and HIPAA—has positioned unlearning in FL as a vital research area for privacy protection, particularly in medical scenarios. For instance, in healthcare FL systems, patients may request the deletion of their sensitive data (e.g., medical imaging or electronic health records) due to withdrawal of consent, necessitating efficient removal without disrupting the global model's performance on other clients' data, shown in Figure 3. Unlearning seeks to eliminate the contributions of specific clients or their data from the global model without necessitating full retraining, thereby mitigating privacy risks and improving model adaptability in clinical settings. Current studies explore a spectrum of challenges, including the development of efficient unlearning algorithms, the assessment of privacy and security threats, the allocation of unlearning responsibilities between servers and clients, and the accommodation of diverse unlearning objectives. This multifaceted research landscape underscores the complexity of balancing computational efficiency, privacy preservation, and model performance in FL systems.

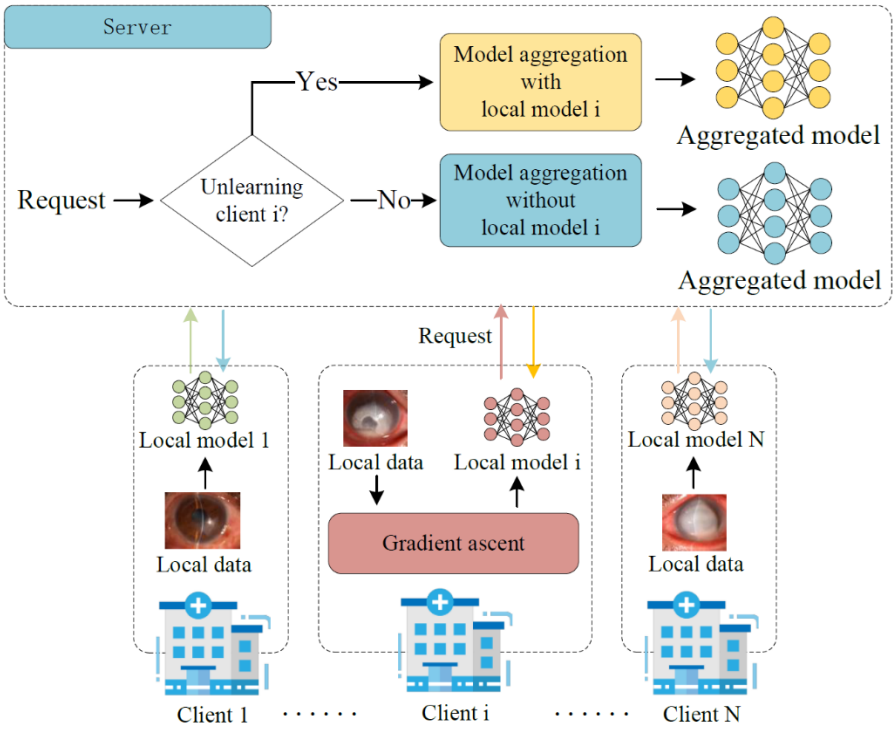


Figure 3. Overview of federated gradient ascent-based approximation unlearning. The server decides whether accepting the unlearning request from clients. If yes, aggregate unlearned clients' models. If no, ignore them.

3.3.1. Efficient Unlearning Algorithms

A significant thrust of unlearning research in FL focuses on crafting efficient algorithms that minimize computational overhead. For instance, FedEraser [127] employs an approximate unlearning technique that utilizes historical parameter replay to reconstruct and nullify the local updates of targeted clients, circumventing the need

for complete model retraining. However, this method imposes a substantial burden on server storage due to its reliance on extensive historical data. Similarly, Liu et al. [128] propose an approach leveraging a first-order Taylor expansion to approximate alterations in the loss function, enabling rapid retraining through an optimization process that engages all clients. While this method markedly reduces computational demands, it resembles "fast substitute training" more than traditional unlearning, highlighting a trade-off between efficiency and precision in data removal.

3.3.2. Privacy and Security-Focused Methods

Privacy and security considerations are paramount in FL unlearning, as the process must not only remove data influences but also guard against emerging vulnerabilities. Chen et al. [129] underscore this by developing an enhanced membership inference attack framework that incorporates historical model trajectories and intermediate activation values, exposing progressive privacy leakage across multiple unlearning iterations. This reveals the necessity for robust unlearning strategies to counteract such risks. In response, Zhao et al. [130] introduce a knowledge erasure strategy termed Momentum Degradation (MoDe), which decouples unlearning from the training phase. This separation allows MoDe to be seamlessly integrated into any FL-trained model without altering the original training protocol, offering a flexible and privacy-preserving solution that maintains model integrity. Furthermore, unlearning processes may introduce vulnerabilities to data reconstruction attacks, where adversaries attempt to recover forgotten data from model differences before and after unlearning. For example, Bertran et al. [131] demonstrate that simple models are susceptible to such attacks, enabling the reconstruction of sensitive inputs. In medical FL contexts, this is particularly concerning for imaging data (e.g., reconstructing patient MRI scans), potentially exacerbating privacy leaks despite the distributed nature of FL. Although not the primary focus of FL attack/defense research, incorporating defenses against reconstruction attacks remains crucial for robust unlearning strategies.

3.3.3. Discussion

In unlearning methods for FL in medical image analysis, current challenges include difficulties in efficient data removal and privacy risks. For data removal efficiency, existing unlearning methods often impose a substantial burden on storage (e.g., FedEraser) or cause performance degradation to minimize computational overhead. For privacy risks, model updates during unlearning may expose sensitive information (e.g., through membership inference attacks). The existing MoDe framework decouples unlearning from the training phase to address this problem. In the future, we can develop client-specific unlearning policies (e.g., tiered deletion based on data sensitivity) using attention-based models to target only relevant parameters. Additionally, adversarial attacks can be employed to test the effectiveness of unlearning algorithms, ensuring no residual traces of deleted data remain in the global model.

4. Conclusions

Our survey systematically organizes FL methodologies for medical image analysis into three parts: training, architecture, and unlearning. We emphasize the medical domain's unique demands, such as handling heterogeneous imaging modalities and annotations. Unlike prior works, our review includes approaches not only for privacy and security but also for accuracy and efficiency. By synthesizing insights from various studies, we help researchers and practitioners navigate the opportunities and challenges of FL in advancing AI-driven healthcare.

In the future, to address data heterogeneity, label scarcity, and unsupervised learning complexity in training methods, we can explore meta-learning for personalization, use generative models to synthesize desensitized, cross-modal data on the server, and leverage foundation models pre-trained on large-scale medical datasets to initialize FL models—improving convergence speed and generalization. For communication overhead, security risks, and difficulties in fair data valuation within architectural methods, we can develop compressed model update techniques, combine HE for critical updates with DP for non-sensitive data in multi-tier FL architectures, and integrate real-time performance metrics into Shapley value calculations to enable adaptive incentive mechanisms for medical institutions. For challenges in efficient data removal and privacy risks in unlearning methods, we can design client-specific unlearning policies and use adversarial attacks to test algorithm effectiveness, ensuring no residual data traces remain.

Author Contributions

J.H.: Conceptualization, investigation, writing, and revision. Z.Y.: Conceptualization, investigation, writing, and revision. P.W.: Conceptualization, investigation, writing, and revision. G.Z.: Conceptualization, investigation, writing, and revision. H.H.: Conceptualization, investigation, writing, and revision. Z.Z.: Conceptualization,

investigation, writing, and revision. D.W.: Conceptualization, investigation, writing, and revision. All authors have read and agreed to the published version of the manuscript.

Funding

This paper is partially supported by Hong Kong Innovation and Technology Commission (ITC) grant #MHP/034/22.

Conflicts of Interest

The authors declare no conflict of interest. Given the role as Editor-in-Chief, Dapeng Oliver Wu had no involvement in the peer review of this paper and had no access to information regarding its peer-review process. Full responsibility for the editorial process of this paper was delegated to another editor of the journal.

References

1. Act, A. Health insurance portability and accountability act of 1996. *Public Law* **1996**, *104*, 191.
2. Regulation, P. General data protection regulation. *Intouch* **2018**, *25*, 1–5.
3. Nazir, S.; Kaleem, M. Federated learning for medical image analysis with deep neural networks. *Diagnostics* **2023**, *13*, 1532.
4. Guan, H.; Yap, P.-T.; Bozoki, A.; et al. Federated learning for medical image analysis: A survey. *Pattern Recognit.* **2024**, 110424.
5. Pfitzner, B.; Steckhan, N.; Arnrich, B. Federated learning in a medical context: A systematic literature review. *ACM Trans. Internet Technol.* **2021**, *21*, 50.
6. Ciupek, D.; Malawski, M.; Pieciak, T. Federated learning: A new frontier in the exploration of multi-institutional medical imaging data. *arXiv* **2025**, arXiv:2503.20107.
7. Sandhu, S.S.; Gorji, H.T.; Tavakolian, P.; et al. Medical imaging applications of federated learning. *Diagnostics* **2023**, *13*, 3140.
8. Kaissis, G.A.; Makowski, M.R.; Ru, D.; et al. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat. Mach. Intell.* **2020**, *2*, 305–311.
9. BMcMahan; Moore, E.; Ramage, D.; et al. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*; PMLR: New York, NY, USA, 2017; pp. 1273–1282.
10. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
11. He, K.; Zhang, X.; Ren, S.; et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
12. Kirillov, A.; Mintun, E.; Ravi, N.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 4015–4026.
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*.
14. He, K.; Zhang, X.; Ren, S.; et al. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Huang, G.; Liu, Z.; Van Der Maaten, L.; et al. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
16. Cheng, J.; Tian, S.; Yu, L.; et al. Ddu-net: A dual dense u-structure network for medical image segmentation. *Appl. Soft Comput.* **2022**, *126*, 109297.
17. Xu, Y.; Kong, M.; Xie, W.; et al. Deep sequential feature learning in clinical image classification of infectious keratitis. *Engineering* **2021**, *7*, 1002–1010.
18. Sarkar, S.; Min, K.; Ikram, W.; et al. Performing automatic identification and staging of urothelial carcinoma in bladder cancer patients using a hybrid deep-machine learning approach. *Cancers* **2023**, *15*, 1673.
19. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
20. Chen, J.; Lu, Y.; Yu, Q.; et al. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
21. Li, T.; Sahu, A.K.; Talwalkar, A.; et al. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60.
22. Hsu TM, H.; Qi, H.; Brown, M. Measuring the effects of non-identical data distribution for federated visual classification. *arXiv* **2019**, arXiv:1909.06335.

23. Li, T.; Sahu, A.K.; Zaheer, M.; et al. Federated optimization in heterogeneous networks. *Proc. Mach. Learn. Syst.* **2020**, *2*, 429–450.
24. Liu, Z.; Wu, F.; Wang, Y.; et al. Fedcl: Federated contrastive learning for multi-center medical image classification. *Pattern Recognit.* **2023**, *143*, 109739.
25. Shi, M.; Zhou, Y.; Wang, K.; et al. Prior: Personalized prior for reactivating the information overlooked in federated learning. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 28378–28392.
26. Fu, S.; Yang, Z.; Hu, C.; et al. Personalized federated learning with contrastive momentum. *IEEE Trans. Big Data* **2024**. https://doi.org/10.1109/TBDATA.2024.3403387.
27. Yang, Z.; Bao, W.; Yuan, D.; et al. Federated learning with nesterov accelerated gradient. *IEEE Trans. Parallel Distrib. Syst.* **2022**, *33*, 4863–4873.
28. Yang, Z.; Fu, S.; Bao, W.; et al. FastSlowMo: Federated learning with combined worker and aggregator momenta. *IEEE Trans. Artif. Intell.* **2022**, *4*, 1041–1050.
29. Abdelmoniem, A.M.; Canini, M. Towards mitigating device heterogeneity in federated learning via adaptive model quantization. In Proceedings of the 1st Workshop on Machine Learning and Systems, Online, UK, 26 April 2021; pp. 96–103.
30. Zhou, Z.; Luo, G.; Chen, M.; et al. Federated learning for medical image classification: A comprehensive benchmark. *arXiv* **2025**, arXiv:2504.05238.
31. Huang, Y.; Du, C.; Xue, Z.; et al. What makes multi-modal learning better than single (provably). *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 10944–10956.
32. Shao, W.; Wang, T.; Sun, L.; et al. Multi-task multi-modal learning for joint diagnosis and prognosis of human cancers. *Med. Image Anal.* **2020**, *65*, 101795.
33. Boehm, K.M.; Aherne, E.A.; Ellenson, L.; et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat. Cancer* **2022**, *3*, 723–733.
34. Olatunji, I.; Cui, F. Multimodal ai for prediction of distant metastasis in carcinoma patients. *Front. Bioinform.* **2023**, *3*, 1131021.
35. Kim, G.; Moon, S.; Choi, J.-H. Deep learning with multimodal integration for predicting recurrence in patients with non-small cell lung cancer. *Sensors* **2022**, *22*, 6594.
36. Agbley, B.L.Y.; Li, J.; Haq, A.U.; et al. Multimodal melanoma detection with federated learning. In Proceedings of the 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 17–19 December 2021; pp. 238–244.
37. Cassará; P; Gotta, A.; Valerio, L. Federated feature selection for cyber-physical systems of systems. *IEEE Trans. Veh. Technol.* **2022**, *71*, 9937–9950.
38. Xiong, B.; Yang, X.; Qi, F.; et al. A unified framework for multi-modal federated learning. *Neurocomputing* **2022**, *480*, 110–118.
39. Qayyum, A.; Ahmad, K.; Ahsan, M.A.; et al. Collaborative federated learning for healthcare: Multi-modal COVID-19 diagnosis at the edge. *IEEE Open J. Comput. Soc.* **2022**, *3*, 172–184.
40. Chen, Z.; Liu, Y.; Zhang, Y.; et al. Enhanced multimodal low-rank embedding based feature selection model for multimodal Alzheimer's disease diagnosis. *IEEE Trans. Med. Imaging* **2024**, *44*, 815–827.
41. Laine, S.; Aila, T. Temporal ensembling for semi-supervised learning. *arXiv* **2016**, arXiv:1610.02242.
42. Xie, Q.; Dai, Z.; Hovy, E.; et al. Unsupervised data augmentation for consistency training. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6256–6268.
43. Li, J.; Xiong, C.; Hoi, S.C. Comatch: Semi-supervised learning with contrastive graph regularization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 9475–9484.
44. Lee, D.H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning*; ICML: Atlanta, GA, USA, 2013; Volume 3, p. 896.
45. Xie, Q.; Luong, M.-T.; Hovy, E.; et al. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10687–10698.
46. Pham, H.; Dai, Z.; Xie, Q.; et al. Meta pseudo labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 11557–11568.
47. Sohn, K.; Berthelot, D.; Carlini, N.; et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 596–608.
48. Chen, J.; Zhang, H.; Mohiaddin, R.; et al. Adaptive hierarchical dual consistency for semi-supervised left atrium segmentation on cross-domain data. *IEEE Trans. Med. Imaging* **2021**, *41*, 420–433.
49. Wang, G.; Zhai, S.; Lasio, G.; et al. Semi-supervised segmentation of radiation-induced pulmonary fibrosis from lung CT scans with multi-scale guided dense attention. *IEEE Trans. Med. Imaging* **2021**, *41*, 531–542.
50. Shi, Y.; Zhang, J.; Ling, T.; et al. Inconsistency-aware uncertainty estimation for semi-supervised medical image

segmentation. *IEEE Trans. Med. Imaging* **2021**, *41*, 608–620.
51. Liu, Q.; Yang, H.; Dou, Q.; et al. Federated semi-supervised medical image classification via inter-client relation matching. In *Medical Image Computing and Computer Assisted Intervention. Proceedings of the MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021*; Springer: Cham, Switzerland, 2021; pp. 325–335.
52. Liang, X.; Lin, Y.; Fu, H.; et al. Rscfed: Random sampling consensus federated semi-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10154–10163.
53. Yang, D.; Xu, Z.; Li, W.; et al. Federated semi-supervised learning for covid region segmentation in chest CT using multi-national data from China, Italy, Japan. *Med. Image Anal.* **2021**, *70*, 101992.
54. Bdair, T.; Navab, N.; Albarqouni, S. Fedperl: Semi-supervised peer learning for skin lesion classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention, Proceedings of the 24th International Conference, Strasbourg, France, 27 September–1 October 2021*; Springer: Cham, Switzerland, 2021; pp. 336–346.
55. Jeong, W.; Yoon, J.; Yang, E.; et al. Federated semi-supervised learning with inter-client consistency & disjoint learning. *arXiv* **2020**, *arXiv:2006.12097*.
56. Wu, H.; Zhang, B.; Chen, C.; et al. Federated semi-supervised medical image segmentation via prototype-based pseudo-labeling and contrastive learning. *IEEE Trans. Med. Imaging* **2023**, *43*, 649–661.
57. Ma, Y.; Wang, J.; Yang, J.; et al. Model-heterogeneous semi-supervised federated learning for medical image segmentation. *IEEE Trans. Med. Imaging* **2024**, *43*, 1804–1815.
58. Wang, Y.; Zhang, J.; Kan, M.; et al. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12275–12284.
59. Patel, G.; Dolz, J. Weakly supervised segmentation with cross-modality equivariant constraints. *Med. Image Anal.* **2022**, *77*, 102374.
60. Lempitsky, V.; Kohli, P.; Rother, C.; et al. Image segmentation with a bounding box prior. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 277–284.
61. Kervadec, H.; Dolz, J.; Tang, M.; et al. Constrained-cnn losses for weakly supervised segmentation. *Med. Image Anal.* **2019**, *54*, 88–99.
62. Kervadec, H.; Dolz, J.; Wang, S.; et al. Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision. In *Medical Imaging with Deep Learning*; PMLR: New York, NY, USA, 2020; pp. 365–381.
63. Hsu, C.-C.; Hsu, K.-J.; Tsai, C.-C.; et al. Weakly supervised instance segmentation using the bounding box tightness prior. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
64. Wang, J.; Xia, B. Bounding box tightness prior for weakly supervised image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; pp. 526–536.
65. Rajchl, M.; Lee, M.C.; Oktay, O.; et al. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE Trans. Med. Imaging* **2016**, *36*, 674–683.
66. Song, C.; Huang, Y.; Ouyang, W.; et al. Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3136–3145.
67. Laradji, I.; Rodriguez, P.; Manas, O.; et al. A weakly supervised consistency-based learning method for COVID-19 segmentation in CT images. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtua, 5–9 January 2021; pp. 2453–2462.
68. Lin, D.; Dai, J.; Jia, J.; et al. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 3159–3167.
69. Liu, X.; Yuan, Q.; Gao, Y.; et al. Weakly supervised segmentation of COVID-19 infection with scribble annotation on CT images. *Pattern Recognit.* **2022**, *122*, 108341.
70. Luo, X.; Hu, M.; Liao, W.; et al. Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In Proceedings of the 25th International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; pp. 528–538.
71. Liu, X.; Wang, S.; Zhang, Y.; et al. Scribble-supervised meibomian glands segmentation in infrared images. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2022**, *18*, 88.
72. Zhu, M.; Chen, Z.; Yuan, Y. Feddm: Federated weakly supervised segmentation via annotation calibration and gradient de-conflicting. *IEEE Trans. Med. Imaging* **2023**, *42*, 1632–1643.
73. Wicaksana, J.; Yan, Z.; Zhang, D.; et al. Fedmix: Mixed supervised federated learning for medical image segmentation. *IEEE Trans. Med. Imaging* **2022**, *42*, 1955–1968.

74. Lin, L.; Liu, Y.; Wu, J.; et al. Fedlppa: Learning personalized prompt and aggregation for federated weakly-supervised medical image segmentation. *IEEE Trans. Med. Imaging* **2024**, *44*, 1127–1139.
75. Pang, G.; Shen, C.; Cao, L.; et al. Deep learning for anomaly detection: A review. *ACM Comput. Surv.* **2021**, *54*, 38.
76. Cai, Y.; Zhang, W.; Chen, H.; et al. Medianomaly: A comparative study of anomaly detection in medical images. *Med. Image Anal.* **2025**, *102*, 103500.
77. Schlegl, T.; Seebo, P.; Waldstein, S.M.; et al. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.* **2019**, *54*, 30–44.
78. Baur, C.; Denner, S.; Wiestler, B.; et al. Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. *Med. Image Anal.* **2021**, *69*, 101952.
79. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; et al. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*.
80. Schlu, H.M.; Tan, J.; Hou, B.; et al. Natural synthetic anomalies for self-supervised anomaly detection and localization. In Proceedings of the 17th European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 474–489.
81. Tian, Y.; Liu, F.; Pang, G.; et al. Self-supervised pseudo multi-class pre-training for unsupervised anomaly detection and segmentation in medical images. *Med. Image Anal.* **2023**, *90*, 102930.
82. Zhang, X.; Li, S.; Li, X.; et al. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 3914–3923.
83. Roth, K.; Pemula, L.; Zepeda, J.; et al. Towards total recall in industrial anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 14318–14328.
84. Bercea, C.I.; Wiestler, B.; Rueckert, D.; et al. Federated disentangled representation learning for unsupervised brain anomaly detection. *Nat. Mach. Intell.* **2022**, *4*, 685–695.
85. Gupta, D.; Kayode, O.; Bhatt, S.; et al. Hierarchical federated learning based anomaly detection using digital twins for smart healthcare. In Proceedings of the 2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC), Atlanta, GA, USA, 13–15 December 2021; pp. 16–25.
86. He, Y.; Ding, X.; Tang, Y.; et al. Unsupervised multivariate time series anomaly detection by feature decoupling in federated learning scenarios. *IEEE Trans. Artif. Intell.* **2025**, *6*, 2013–2026.
87. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 1180–1189.
88. Tzeng, E.; Hoffman, J.; Saenko, K.; et al. Adversarial discriminative domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7167–7176.
89. Volpi, R.; Morerio, P.; Savarese, S.; et al. Adversarial feature augmentation for unsupervised domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5495–5504.
90. Taigman, Y.; Polyak, A.; Wolf, L. Unsupervised cross-domain image generation. *arXiv* **2016**, arXiv:1611.02200.
91. Liu, M.-Y.; Tuzel, O. Coupled generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2016**, 29.
92. Ji, W.; Chung, A.C. Unsupervised domain adaptation for medical image segmentation using transformer with meta attention. *IEEE Trans. Med. Imaging* **2023**, *43*, 820–831.
93. Deng, S.; Chen, Y.; Huang, W.; et al. Unsupervised domain adaptation for EM image denoising with invertible networks. *IEEE Trans. Med. Imaging* **2024**, *44*, 92–105.
94. Peng, X.; Huang, Z.; Zhu, Y.; et al. Federated adversarial domain adaptation. *arXiv* **2019**, arXiv:1911.02054.
95. Xiao, Y.; Guo, Y.; Zhu, H.; et al. An unsupervised federated domain adaptation method based on knowledge distillation. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *36*, 10993–11007.
96. El Ouadrhiri, A.; Abdelhadi, A. Differential privacy for deep and federated learning: A survey. *IEEE Access* **2022**, *10*, 22359–22380.
97. Acar, A.; Aksu, H.; Uluagac, A.S.; et al. A survey on homomorphic encryption schemes: Theory and implementation. *ACM Comput. Surv.* **2018**, *51*, 79.
98. Muthalakshmi, M.; Jeyapal, K.; Vinoth, M.; et al. Federated learning for secure and privacy-preserving medical image analysis in decentralized healthcare systems. In Proceedings of the 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 7–9 August 2024; pp. 1442–1447.
99. Yang, J.; Shi, R.; Wei, D.; et al. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Sci. Data* **2023**, *10*, 41.
100. Tanim, S.A.; Mridha, M.; Safran, M.; et al. Secure federated learning for parkinson's disease: Non-iid data partitioning and homomorphic encryption strategies. *IEEE Access* **2024**, *12*, 127309–127327.
101. Adams, J.L.; Dinesh, K.; Snyder, C.W.; et al. Pd-biostamprc21: Parkinson's disease accelerometry dataset from five wearable sensor study. *IEEE Dataport* **2020**. https://doi.org/10.21227/g2g8-1503.

102. RAhmed; Maddikunta, P.K.R.; Gadekallu, T.R.; et al. Efficient differential privacy enabled federated learning model for detecting covid-19 disease using chest x-ray images. *Front. Med.* **2024**, *11*, 1409314.
103. Shastri, S.; Kansal, I.; Kumar, S.; et al. Cheximagenet: A novel architecture for accurate classification of covid-19 with chest x-ray digital images using deep convolutional neural networks. *Health Technol.* **2022**, *12*, 193–204.
104. Mo, F.; Haddadi, H.; Katevas, K.; et al. Ppfl: Privacypreserving federated learning with trusted execution environments. In Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, Virtual, 24 June–2 July 2021; pp. 94–108.
105. Makkar, A.; Santosh, K. Securefed: Federated learning empowered medical imaging technique to analyze lung abnormalities in chest X-rays. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 2659–2670.
106. Lessage, X.; Collier, L.; Van Ouytsel, C.-H.B.; et al. Secure federated learning applied to medical imaging with fully homomorphic encryption. In Proceedings of the 2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC), Houston, TX, USA, 7–9 February 2024; pp. 1–12.
107. Rieyan, S.A.; News, M.R.K.; Rahman, A.M.; et al. An advanced data fabric architecture leveraging homomorphic encryption and federated learning. *Inf. Fusion* **2024**, *102*, 102004.
108. Haripriya, R.; Khare, N.; Pandey, M. Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings. *Sci. Rep.* **2025**, *15*, 12482.
109. Sun, G.; Shu, H.; Shao, F.; et al. Fkd-med: Privacy-aware, communication-optimized medical image segmentation via federated learning and model lightweighting through knowledge distillation. *IEEE Access* **2024**, *12*, 33687–33704.
110. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
111. Huang, H.; Zhang, L.; Sun, C.; et al. Distributed pruning towards tiny neural networks in federated learning. In Proceedings of the 2023 IEEE 43rd International Conference on Distributed Computing Systems (ICDCS), Hong Kong, China, 18–21 July 2023; pp. 190–201.
112. Huang, H.; Zhuang, W.; Chen, C.; et al. Fedmef: Towards memory-efficient federated dynamic pruning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 27548–27557.
113. Huang, H.; Yang, H.; Chen, Y.; et al. Fedrts: Federated robust pruning via combinatorial thompson sampling. *arXiv* **2025**, arXiv:2501.19122.
114. Xie, H.; Zhang, Y.; Zhou, Z.; et al. Privacy-preserving medical data collaborative modeling: A differential privacy enhanced federated learning framework. *J. Knowl. Learn. Sci. Technol.* **2024**, *3*, 340–350.
115. Yang, Z.; Fu, S.; Bao, W.; et al. Hierarchical federated learning with momentum acceleration in multi-tier networks. *IEEE Trans. Parallel Distrib. Syst.* **2023**, *34*, 2629–2641.
116. Xie, L.; Lin, M.; Liu, S.; et al. pflfe: Cross-silo personalized federated learning via feature enhancement on medical image segmentation. In Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention, Marrakesh, Morocco, 6–10 October 2024; pp. 599–610.
117. Shapley, L.S. A value for n-person games. In *Contributions to the Theory of Games, Volume II*; Princeton University Press: Princeton, NJ, USA, 2016; pp. 307–318.
118. Wei, S.; Tong, Y.; Zhou, Z.; et al. Efficient and fair data valuation for horizontal federated learning. In *Federated Learning: Privacy and Incentive*; Springer: Cham, Switzerland, 2020; pp. 139–152.
119. Liu, Z.; Chen, Y.; Yu, H.; et al. Gtg-shapley: Efficient and accurate participant contribution evaluation in federated learning. *ACM Trans. Intell. Syst. Technol.* **2022**, *13*, 60.
120. Fan, Z.; Fang, H.; Zhou, Z.; et al. Improving fairness for data valuation in horizontal federated learning. In Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022; pp. 2440–2453.
121. Kim, H.; Park, J.; Bennis, M.; et al. Blockchained on-device federated learning. *IEEE Commun. Lett.* **2020**, *24*, 1279–1283.
122. Lo, S.K.; Liu, Y.; Lu, Q.; et al. Toward trustworthy ai: Blockchain-based architecture design for accountability and fairness of federated learning systems. *IEEE Internet Things J.* **2023**, *10*, 3276–3284.
123. Otoum, S.; Ridhawi, I.A.; Mouftah, H. Securing critical IoT infrastructures with blockchain-supported federated learning. *IEEE Internet Things J.* **2022**, *9*, 2592–2601.
124. Myrzashova, R.; Alsamhi, S.H.; Shvetsov, A.V.; et al. Blockchain meets federated learning in healthcare: A systematic review with challenges and opportunities. *IEEE Internet Things J.* **2023**, *10*, 14418–14437.
125. Zhu, J.; Cao, J.; Saxena, D.; et al. Blockchain-empowered federated learning: Challenges, solutions, and future directions. *ACM Comput. Surv.* **2023**, *55*, 240.
126. Yang, Z.; Fu, S.; Bao, W.; et al. Hierarchical federated learning with adaptive momentum in multi-tier networks. In Proceedings of the 2023 IEEE 43rd International Conference on Distributed Computing Systems (ICDCS), Hong Kong, China, 18–21 July 2023; pp. 499–510.
127. Liu, G.; Ma, X.; Yang, Y.; et al. Federaser: Enabling efficient client-level data removal from federated learning models.

In Proceedings of the 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS), Tokyo, Japan, 25–28 June 2021; pp. 1–10.

128. Liu, Y.; Xu, L.; Yuan, X.; et al. The right to be forgotten in federated learning: An efficient realization with rapid retraining. In Proceedings of the IEEE INFOCOM 2022-IEEE Conference on Computer Communications, London, UK, 2–5 May 2022; pp. 1749–1758.
129. Chen, M.; Zhang, Z.; Wang, T.; et al. When machine unlearning jeopardizes privacy. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, Virtual, 15–19 November 2021; pp. 896–911.
130. Zhao, Y.; Wang, P.; Qi, H.; et al. Federated unlearning with momentum degradation. *IEEE Internet Things J.* **2023**, 11, 8860–8870.
131. Bertran, M.; Tang, S.; Kearns, M.; et al. Reconstruction attacks on machine unlearning: Simple models are vulnerable. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 104995–105016.