

Federated Learning for 2D Medical Image Registration: A Comprehensive Survey

Anubhab Parashar

Department of Computer Science and Engineering
Presidency University
Bengaluru, India
anubhab.20231COM0043@presidencyuniversity.in

Dr. Girija Jagannath

Department of Computer Science and Engineering
Bangalore Institute of Technology
Bengaluru, India
girija.bitcse@gmail.com

Abstract—Medical image registration—the task of aligning a moving image to a fixed reference by estimating a dense deformation field—is a foundational step in clinical workflows ranging from longitudinal disease monitoring to multi-modal fusion. Deep learning has substantially accelerated this task, yet training robust registration networks requires large, diverse datasets that individual institutions cannot amass without violating patient privacy regulations such as Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR). Federated Learning (FL) has emerged as a compelling paradigm that allows multiple hospitals to collaboratively train a shared model while keeping all raw images local. Despite rapid progress, a dedicated survey of FL applied specifically to 2D medical image registration is lacking. This paper fills that gap. We systematically review: (i) the problem formulation and deep-learning advances in 2D image registration, (ii) the fundamentals of FL and the challenges introduced by non-Independent and Identically Distributed (IID) medical data, (iii) privacy and security mechanisms including differential privacy, secure aggregation, and defences against gradient inversion attacks, and (iv) the body of methods that combine FL with deformable registration networks. We further survey benchmark datasets, evaluation protocols, and open problems. Our analysis finds that while FL offers a principled solution to the data-sharing barrier, domain shift, communication efficiency, hyperparameter personalisation, and the privacy–utility trade-off remain largely unsolved. We conclude with a research roadmap.

Index Terms—Federated learning, medical image registration, deformable registration, privacy-preserving learning, differential privacy, domain adaptation, deep learning.

I. INTRODUCTION

Medical image registration is the process of establishing spatial correspondences between two or more images, typically a *fixed* (reference) image I_F and a *moving* image I_M . The goal is to estimate a transformation ϕ such that the warped image $I_M \circ \phi$ aligns as closely as possible with I_F . Registration is central to a wide range of clinical tasks: longitudinal tracking of tumour progression, atlas-based segmentation, motion correction in cardiac imaging, and radiation therapy planning [1].

For many years, registration was solved through iterative numerical optimisation—procedures that could take minutes to hours per image pair. The VoxelMorph framework [2] and the

Spatial Transformer Network (STN) [3] shifted this to amortised inference: a network trained once can register unseen pairs in milliseconds. That shift, however, created a strong dependency on *large, representative training sets*—something individual healthcare institutions struggle to provide.

Patient data is tightly regulated. In the United States, HIPAA prohibits sharing identifiable health information without patient consent. In Europe, GDPR imposes comparable restrictions on cross-border data transfer. Beyond regulation, clinical imaging data is fragmented: each institution holds a dataset that may be too small, scanner-specific, or protocol-specific to train a generalisable model on its own. Better models need more data, but aggregating that data is legally and logistically blocked.

FL [4] offers a direct solution. A central server coordinates training across K institutions. Each institution trains a local model on its private data and uploads only weight updates—never raw images—to the server. The server aggregates these updates and broadcasts an improved global model. After many such rounds, the global model has seen the diversity of all clients' data without any image leaving its source.

Applying FL to medical image registration raises challenges that do not appear in centralised registration or in standard FL for classification. Registration losses are computed directly on image intensities, making them sensitive to scanner heterogeneity across sites. Deformation smoothness hyperparameters may need per-site tuning, which conflicts with a single shared model. Gradient updates from registration networks carry enough latent image information to be reconstructed by inversion attacks. Yet despite these known difficulties, no survey has focused specifically on FL for 2D medical image registration.

The contributions of this survey are as follows:

- A structured review of 2D medical image registration from classical to deep-learning methods, with emphasis on loss functions and regularisation strategies relevant to the federated setting.
- A treatment of FL fundamentals—aggregation algorithms, non-IID challenges, and personalisation strategies—in the medical imaging context.
- An analysis of privacy and security mechanisms for federated registration, covering Differential Privacy (DP),

secure aggregation, and gradient inversion defences.

- A survey of existing works at the intersection of FL and 2D image registration, organised by architecture, loss design, and domain adaptation strategy.
- A comparative overview of benchmark datasets and evaluation metrics, with a discussion of open problems and a research roadmap.

Section II reviews 2D medical image registration. Section III covers FL fundamentals. Section IV addresses privacy and security. Section V surveys FL-based registration methods. Section VI presents datasets and evaluation. Section VII discusses open challenges, and Section VIII concludes.

II. BACKGROUND: 2D MEDICAL IMAGE REGISTRATION

A. Problem Formulation

Given a fixed image $I_F : \Omega \rightarrow \mathbb{R}$ and a moving image $I_M : \Omega \rightarrow \mathbb{R}$ on a spatial domain $\Omega \subset \mathbb{R}^2$, deformable registration seeks a displacement field $\mathbf{u} : \Omega \rightarrow \mathbb{R}^2$ such that $\phi = \text{id} + \mathbf{u}$ brings I_M into alignment with I_F :

$$\hat{\phi} = \arg \min_{\phi} \mathcal{L}_{\text{sim}}(I_F, I_M \circ \phi) + \lambda \mathcal{L}_{\text{reg}}(\phi) \quad (1)$$

where $\mathcal{L}_{\text{sim}}$ measures image dissimilarity, $\mathcal{L}_{\text{reg}}$ penalises implausible deformations, and $\lambda > 0$ controls the smoothness–accuracy trade-off. The estimated $\hat{\phi}$ can then propagate labels or segment structures from moving to fixed image space.

B. Traditional Registration Methods

Classical methods solve (1) iteratively for each image pair. Optical-flow methods [6] assume brightness constancy and spatial smoothness. The demons algorithm [7] treats the fixed image as a diffusing body that attracts the moving image via virtual forces. Free-form deformation models parameterise $\mathbf{u}$ using B-spline control points [8]. Atlas-based and statistical shape methods add population-level priors.

These methods are mathematically principled but slow at inference ($O(\text{minutes per pair})$), prone to local minima, and poorly generalisable across imaging protocols.

C. Deep Learning Approaches

1) *Spatial Transformer Networks*: Jaderberg et al. [3] introduced the STN: a differentiable module composed of a localisation network, a grid generator, and a bilinear sampler. Because the entire pipeline is differentiable, the transformation can be learned end-to-end as part of a larger network.

2) *VoxelMorph*: Balakrishnan et al. [2] proposed VoxelMorph, in which a U-Net backbone [9] takes the concatenated fixed and moving images and directly regresses a dense displacement field. Training is unsupervised: the loss is defined entirely on image similarity and regularisation, with no ground-truth deformation required. At test time, inference takes milliseconds.

3) *U-Net Architectures for Registration*: The encoder–decoder with skip connections is now the standard backbone for registration networks. The encoder downsamples through convolutional blocks to capture multi-scale features; the decoder upsamples and restores spatial resolution via skip connections; and the final layer produces a two-channel field $\mathbf{u}(x, y) = [u_x, u_y]$.

D. Loss Functions

1) Similarity Metrics:

- **Mean Squared Error (MSE)**: $\mathcal{L}_{\text{MSE}} = \frac{1}{|\Omega|} \sum_{p \in \Omega} (I_F(p) - I_M(\phi(p)))^2$. Suited to mono-modal pairs with similar intensity ranges.
- **Normalized Cross-Correlation (NCC)**: Normalised linear correlation computed in local patches. More robust to global intensity offsets caused by different scanners.
- **Mutual Information (MI)**: Measures statistical dependence between intensity distributions. Standard choice for multi-modal registration (e.g., CT–MRI).

2) *Regularisation Losses*: Regularisation keeps the predicted deformation field physically plausible. Three principal forms are used.

Diffusion regulariser (L2). Penalises large spatial gradients in $\mathbf{u}$:

$$\mathcal{L}_{\text{diff}} = \sum_{p \in \Omega} |\nabla \mathbf{u}(p)|^2 \quad (2)$$

Neighbouring pixels are pushed towards similar displacements, producing smooth fields. This is the default in VoxelMorph.

Total variation regulariser (L1). Uses the L1 norm of the gradient:

$$\mathcal{L}_{\text{TV}} = \sum_{p \in \Omega} |\nabla \mathbf{u}(p)| \quad (3)$$

TV regularisation tolerates piecewise-smooth fields with sharp transitions, which is useful when discontinuities at organ boundaries are expected.

Jacobian-based regulariser. Directly prevents folding by penalising locations where the Jacobian determinant $\det(J_{\phi})$ is non-positive:

$$\mathcal{L}_{\text{jac}} = \sum_{p \in \Omega} \max(0, -\det(J_{\phi}(p))) \quad (4)$$

A negative Jacobian determinant is a local fold—the mapping is no longer injective at that point. This regulariser is therefore the most anatomically grounded of the three.

III. FEDERATED LEARNING: FUNDAMENTALS

A. Federated Learning Setup

In the cross-silo FL setting relevant to healthcare, K institutions each hold a private local dataset $\mathcal{D}_k$. The shared objective is:

$$\min_{\mathbf{w}} F(\mathbf{w}) = \sum_{k=1}^K \frac{n_k}{n} F_k(\mathbf{w}) \quad (5)$$

where $n_k = |\mathcal{D}_k|$, $n = \sum_{k=1}^K n_k$, and $F_k(\mathbf{w}) = \mathbb{E}_{(x,y) \sim \mathcal{D}_k} [\ell(\mathbf{w}; x, y)]$ is the local empirical loss. Raw data

never leaves the client; only model parameters travel over the network.

B. Aggregation Algorithms

A growing number of aggregation strategies have been proposed for FL, each making different assumptions about data heterogeneity, communication cost, and privacy. We survey the most relevant methods and, in Section III-C, argue for the one best suited to 2D medical image registration.

1) *FedAvg*: McMahan et al. [4] proposed Federated Averaging (FedAvg), the foundational FL algorithm. In each communication round t :

- 1) The server broadcasts current global weights $\mathbf{w}^t$ to all clients.
- 2) Each client k runs E local SGD steps from $\mathbf{w}^t$, producing $\mathbf{w}_k^{t+1}$.
- 3) The server aggregates: $\mathbf{w}^{t+1} = \sum_k \frac{n_k}{n} \mathbf{w}_k^{t+1}$.

FedAvg is simple and communication-efficient but suffers from client drift when local data distributions differ substantially, leading to unstable convergence under strong non-IID conditions.

2) *FedProx*: Li et al. [5] addressed client drift by adding a proximal regularisation term to each client's local objective:

$$F_k^{\text{prox}}(\mathbf{w}) = F_k(\mathbf{w}) + \frac{\mu}{2} \|\mathbf{w} - \mathbf{w}^t\|^2 \quad (6)$$

The proximal term constrains local updates to stay near the global model, stabilising convergence in heterogeneous settings. The hyperparameter $\mu \geq 0$ interpolates between pure local training ($\mu = 0$, equivalent to FedAvg) and no local movement ($\mu \rightarrow \infty$). FedProx is particularly relevant in medical imaging, where inter-site heterogeneity is pronounced.

3) *SCAFFOLD*: Karimireddy et al. [14] proposed SCAFFOLD (Stochastic Controlled Averaging), which uses control variates to correct for client drift rather than constraining the update magnitude. Each client k maintains a local control variate $\mathbf{c}_k$ and the server maintains a global variate $\mathbf{c}$. The corrected local gradient update is:

$$\mathbf{g}_k^{\text{corr}} = \nabla F_k(\mathbf{w}) - \mathbf{c}_k + \mathbf{c} \quad (7)$$

SCAFFOLD provably converges at the same rate as centralised SGD under non-IID conditions, but requires transmitting control variates alongside model weights, doubling communication cost per round.

4) *FedNova*: Wang et al. [19] observed that FedAvg implicitly assumes all clients perform the same number of local steps. When clients have unequal data sizes or compute budgets—common in real healthcare deployments—this leads to objective inconsistency. FedNova normalises client updates by the number of local steps before aggregation:

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \sum_{k=1}^K \frac{n_k \Delta_k^t}{n \tau_k} \quad (8)$$

where $\Delta_k^t = \mathbf{w}^t - \mathbf{w}_k^{t+1}$ is the local update and τ_k is the number of local steps taken by client k . This correction eliminates the bias introduced by heterogeneous local computation.

5) *FedBN*: Li et al. [20] proposed FedBN specifically to handle feature shift across clients. The key insight is that batch normalisation (BN) statistics capture site-specific intensity distributions. FedBN keeps BN layers local—never sharing or aggregating them—while synchronising all other parameters globally. Formally, the shared parameter set is $\mathbf{w} \setminus \mathbf{w}^{\text{BN}}$, and each client retains its own $\mathbf{w}_k^{\text{BN}}$. This is directly applicable to registration U-Nets trained on images from scanners with different intensity profiles.

6) *FedMA*: Yurochkin et al. [21] and Wang et al. [22] proposed Federated Matched Averaging (FedMA), which aggregates networks by first matching neurons across clients using the Beta-Bernoulli process before averaging. Unlike FedAvg, which averages weights at the same index regardless of whether they represent the same feature, FedMA aligns functionally equivalent neurons before aggregation. This approach is well-suited to networks with permutation symmetry—such as convolutional encoders—but adds significant computational overhead at the server.

7) *Moon*: Li et al. [23] proposed MOON (Model-Contrastive Federated Learning), which uses contrastive learning at the representation level to correct local training. Each client's local loss is augmented with a contrastive term:

$$\mathcal{L}_{\text{moon}} = \mathcal{L}_{\text{sup}} - \frac{\tau}{2} \log \frac{e^{\text{sim}(\mathbf{z}, \mathbf{z}^g)/\tau}}{e^{\text{sim}(\mathbf{z}, \mathbf{z}^g)/\tau} + e^{\text{sim}(\mathbf{z}, \mathbf{z}^{\text{prev}})/\tau}} \quad (9)$$

where $\mathbf{z}$ is the local representation, $\mathbf{z}^g$ the global model representation, and $\mathbf{z}^{\text{prev}}$ the previous round's local representation. MOON pulls local representations closer to the global model and pushes them away from the previous local model, reducing representation drift.

8) *Ditto*: Li et al. [24] proposed Ditto, a personalised FL framework that simultaneously trains a global model and a per-client personalised model. The global model is trained via FedAvg; each client then solves a regularised local problem:

$$\min_{\mathbf{v}_k} F_k(\mathbf{v}_k) + \frac{\lambda_d}{2} \|\mathbf{v}_k - \mathbf{w}\|^2 \quad (10)$$

where $\mathbf{v}_k$ is the personalised model and $\mathbf{w}$ is the global model. Ditto is particularly relevant when sites have strongly different patient populations and a single global model cannot serve all sites equally well.

9) *Comparison Summary*: Table I summarises the aggregation methods reviewed above across the dimensions most relevant to federated 2D medical image registration.

TABLE I
COMPARISON OF FL AGGREGATION METHODS FOR MEDICAL IMAGE REGISTRATION

Method	Non-IID	Comm. cost	Site heterogeneity	Personalisation
FedAvg [4]	Low	Low	Poor	None
FedProx [5]	Medium	Low	Moderate	None
SCAFFOLD [14]	High	High	Good	None
FedNova [19]	Medium	Low	Moderate	None
FedBN [20]	High	Low	Excellent	Partial
FedMA [22]	Medium	Medium	Good	None
MOON [23]	Medium	Low	Moderate	None
Ditto [24]	High	Low	Good	Full

C. Recommended Aggregation Method for 2D Image Registration

Of the methods surveyed, **FedBN** [20] combined with **FedProx**-style proximal regularisation is the most suitable strategy for federated 2D medical image registration. The reasoning is as follows.

The dominant challenge in multi-site registration is scanner-induced intensity shift: the same anatomy appears with systematically different pixel statistics across institutions. This is precisely the problem FedBN was designed for. By keeping batch normalisation layers local, each site's encoder adapts to its own intensity distribution, while the shared convolutional weights—which capture structural, anatomy-level features—converge towards a common global prior. This separation of site-specific statistics from shared anatomical knowledge is architecturally natural for registration U-Nets.

Adding a proximal term (FedProx) prevents the shared convolutional weights from drifting too far during local training, which is important when the number of local epochs E is large or when some sites have significantly more data than others. The combined objective at each client becomes:

$$F_k^*(\mathbf{w}) = \mathcal{L}_{\text{sim}}(I_F, I_M \circ \phi \mathbf{w}) + \lambda \mathcal{L}_{\text{reg}}(\phi \mathbf{w}) + \frac{\mu}{2} \|\mathbf{w} \setminus \mathbf{w}^{\text{BN}} - \mathbf{w}^t \setminus \mathbf{w}^{t, \text{BN}}\|^2 \quad \Pr[\mathbf{M}(D) \in S] \leq e^\epsilon \cdot \Pr[\mathbf{M}(D') \in S] + \delta \quad (12)$$

where the proximal term operates only on the shared (non-BN) parameters. This formulation jointly addresses scanner heterogeneity (via local BN), client drift (via proximal regularisation), and deformation plausibility (via $\mathcal{L}_{\text{reg}}$)—the three principal challenges in federated 2D registration.

In scenarios where sites also have substantially different patient populations (e.g., paediatric vs. adult imaging), adding a Ditto-style personalisation layer on top of FedBN+FedProx provides further per-site adaptation without sacrificing the shared global backbone.

D. Non-IID Challenges in Medical Data

The IID assumption rarely holds in healthcare. Medical datasets exhibit several forms of heterogeneity:

- **Feature distribution shift:** Different scanners, field strengths, and acquisition protocols produce different intensity distributions for the same anatomy.
- **Label distribution shift:** Patient demographics and disease prevalence vary across hospitals and geographic regions.
- **Quantity imbalance:** Large academic centres may hold orders of magnitude more data than smaller community hospitals.

In registration specifically, non-IID data means that the degree of anatomical deformation varies across sites—a problem for any single global regularisation weight λ .

E. Personalised Federated Learning

Personalised FL lets each client maintain a model suited to its local distribution while still benefiting from the global training signal. Common approaches include:

- **Local fine-tuning:** After each aggregation round, the global model is fine-tuned on local data for a small number of gradient steps.
- **pFedMe** [13]: Separates the personalised and global model objectives using a Moreau envelope.
- **SCAFFOLD** [14]: Uses control variates to correct for client drift without a proximal penalty.
- **Per-layer personalisation:** Encoder layers—which learn general, often scanner-specific features—are shared globally, while decoder layers are fine-tuned locally.

IV. PRIVACY AND SECURITY IN FEDERATED MEDICAL IMAGING

FL keeps raw data local, but model updates can still leak patient information. This section covers the main threat models and their defences.

A. Differential Privacy

Differential Privacy (DP) [10] provides a formal privacy guarantee. A randomised mechanism $\mathbf{M}$ is (ϵ, δ) -differentially private if, for any two datasets D and D' differing in a single record and any output set S :

In FL, this is enforced by clipping per-sample gradients to norm C and injecting calibrated Gaussian noise before each upload:

$$\tilde{\mathbf{g}}_k = \frac{1}{B} \sum_{i=1}^h \text{clip}(\mathbf{g}_i, C) + \mathbf{N}(0, \sigma^2 C^2 \mathbf{I}) \quad (13)$$

Smaller ϵ gives stronger privacy but degrades accuracy. This trade-off is particularly sharp in registration, where the similarity loss requires precise gradient information about image intensities.

B. Secure Aggregation

Secure aggregation [11] lets the server compute the sum of client updates without seeing any individual update in plaintext. Clients use secret-sharing to mask their updates; the masks cancel during aggregation. This defends against a semi-honest server.

Homomorphic encryption (HE) extends protection further by supporting arithmetic on encrypted weights. HE provides strong guarantees but is computationally expensive—a real constraint for the large U-Nets used in registration.

C. Gradient Inversion Attacks

Gradient inversion attacks reconstruct training images from shared gradient updates. Zhu et al. [12] showed that images can be recovered with high fidelity from gradients of shallow convolutional networks. For registration networks, the risk is higher: because the similarity loss depends directly on pixel intensities, gradients carry information about both the fixed and moving images simultaneously.

Practical defences include:

- **Gradient clipping and noise injection** (the DP mechanism described above).
- **Gradient compression**: Transmitting only the top- k gradient components reduces information leakage while also lowering communication cost.
- **Gradient obfuscation**: Adding structured perturbations that preserve aggregation correctness but degrade reconstruction.

V. FEDERATED LEARNING FOR 2D MEDICAL IMAGE REGISTRATION

A. Network Architectures in Federated Registration

The dominant architecture for federated registration is the U-Net with skip connections, instantiated in the VoxelMorph style. The network takes the concatenated fixed and moving images as a two-channel input and outputs a two-channel displacement field. In the FL framework, the full network—encoder, bottleneck, and decoder—is typically shared across all clients.

A key design question is what to share and what to personalise. Encoder layers capture low-level, scanner-specific statistics; personalising them (or using site-specific batch normalisation) while sharing the decoder can improve generalisation. Table II summarises representative architectures from the literature.

TABLE II
REPRESENTATIVE NETWORK ARCHITECTURES IN FEDERATED 2D MEDICAL IMAGE REGISTRATION

Architecture	FL Algorithm	Personalisation	Modality
VoxelMorph + U-Net	FedAvg	None	X-ray, retinal
U-Net + BN personalisation	FedProx	Site-specific BN	Chest X-ray
Encoder-decoder + STN	SCAFFOLD	Decoder fine-tuning	Fundus, OASIS
Lightweight CNN	FedAvg	None	Cardiac (ACDC)

B. Loss Functions and Regularisation under Federated Learning

The total registration loss in the federated setting follows (1). Several FL-specific considerations arise.

Per-site λ tuning. Sites imaging populations with greater anatomical variation need larger λ values than sites with more homogeneous anatomy. Current approaches either treat λ as a non-shared local hyperparameter or learn a site embedding that modulates λ as part of the personalised model component.

Unsupervised nature. Registration losses are computed on image pairs without any ground-truth deformation. Even a small hospital with no annotated data can therefore participate fully in federated training—an advantage that FL for supervised tasks does not have.

Consistency regularisation across clients. An additional inter-client consistency term can penalise discrepancy between the deformation predicted by the global model and the locally fine-tuned model. This keeps the global backbone from drifting too far from shared anatomical priors.

C. Domain Adaptation and Cross-Site Generalisation

Scanner heterogeneity is the main source of cross-site domain shift in 2D registration. Strategies used in the literature include:

- **Histogram matching**: Normalising intensity distributions to a shared reference before computing similarity. Simple and effective for mono-modal scenarios.
- **Instance normalisation**: Replacing batch normalisation with instance normalisation reduces sensitivity to global intensity statistics.
- **Adversarial domain adaptation**: A domain discriminator is trained to predict the site of origin from encoder features; the encoder is trained simultaneously to fool it, producing site-invariant representations.
- **Contrastive pre-training**: Self-supervised contrastive pre-training of the encoder across sites before federated fine-tuning for registration.

VI. BENCHMARKS AND EVALUATION

A. Benchmark Datasets

Table III lists the principal publicly available 2D medical image datasets used in registration studies. In the federated setting, a centralised dataset is usually partitioned across simulated clients by site, scanner, or demographic group to emulate the multi-institutional scenario.

TABLE III
BENCHMARK DATASETS FOR 2D MEDICAL IMAGE REGISTRATION IN FEDERATED LEARNING STUDIES

Dataset	Modality	Anatomy	Size	Annotation
OASIS [15]	MRI (slices)	Brain	416 subjects	Seg. masks
ACDC [16]	Cardiac MRI	Heart	150 patients	ED/ES labels
FIRE [17]	Retinal fundus	Retina	129 pairs	Landmarks
ChestX-ray14 [18]	Chest X-ray	Thorax	112,120	Disease labels
DRIVE [25]	Retinal fundus	Retina	40 images	Vessel seg.
Montgomery CXR	Chest X-ray	Lung fields	138 images	Lung masks

B. Evaluation Metrics

A complete evaluation covers both alignment quality and deformation plausibility.

1) Alignment Quality:

- **Dice Similarity Coefficient (DSC)**: Overlap between propagated and ground-truth segmentation masks after registration. $DSC = 2|A \cap B|/(|A| + |B|)$. Higher is better.
- **Target Registration Error (TRE)**: Mean Euclidean distance between corresponding anatomical landmarks in the registered and fixed image. Lower is better; reported in mm or pixels.
- **Similarity score**: NCC or MI on the registered pair, providing a direct image-level measure.

2) Deformation Plausibility:

- **Non-positive Jacobian fraction** ($\{p : \det J_\phi(p) \leq 0\}$): Fraction of spatial locations with folding. Lower is better; ideally zero.

- **Standard deviation of $\det(J_\phi)$:** Variation in local volume change. Values near zero indicate near-diffeomorphic deformations.

3) *Federated-Specific Metrics:*

- **Communication rounds to convergence:** Rounds needed for the global model to reach a target DSC.
- **Cross-site generalisation gap:** Difference in DSC between in-distribution (same site) and held-out site evaluation.
- **Privacy–utility curve:** DSC as a function of ϵ under DP.

C. Evaluation Protocol in the Federated Setting

A rigorous evaluation should report:

- 1) Global model performance at a held-out site not seen during training.
- 2) Performance after local fine-tuning of the global model, to quantify personalisation gain.
- 3) An ablation over the number of clients K , local epochs E , and regularisation weight λ .
- 4) A privacy ablation reporting DSC versus ϵ .

VII. DISCUSSION

A. Open Challenges

Several problems in this area remain genuinely unsolved.

Non-IID deformation priors. The weight λ encodes an implicit prior over how much anatomy should deform. That prior differs across sites. Treating λ as a fixed global hyperparameter is wrong in principle, but per-site tuning without guidance can be arbitrary. Learning site-adaptive regularisation within FL—in a principled way—is an open problem.

Communication efficiency. Large U-Net registration models contain tens of millions of parameters. Gradient compression, low-rank update approximations, and partial sharing (e.g., encoder weights only) are promising, but none has been studied carefully in the registration context.

Privacy–utility trade-off in registration. Standard DP-SGD hurts registration more than classification, because the similarity loss needs precise gradient information about image intensities. Noise mechanisms calibrated specifically to unsupervised registration losses are needed.

Lack of standardised federated benchmarks. Most published work simulates federation by partitioning a centralised dataset, which may not reflect the heterogeneity of real multi-site data. Benchmarks with genuine institutional provenance and standardised evaluation splits would make results more comparable and trustworthy.

Clinical translation. The path from research prototype to clinical deployment is not straightforward. Regulatory approval for FL-trained models, behaviour under distribution shift at deployment time, and the interpretability of predicted deformation fields for clinicians all need attention.

B. Research Roadmap

We propose five directions:

- 1) **Adaptive regularisation in FL:** Learn site-specific λ alongside shared backbone weights, rather than treating regularisation as a fixed global setting.
- 2) **Registration-aware DP:** Design noise mechanisms calibrated to the geometry of unsupervised registration losses, rather than applying generic gradient perturbation.
- 3) **Continual federated learning:** Allow new sites to join training incrementally without forgetting previously learned deformation priors.
- 4) **Standardised federated benchmarks:** Multi-institutional 2D registration benchmarks with genuine site-level splits, public leaderboards, and reproducible evaluation protocols.
- 5) **Foundation model encoders:** Investigate whether large-scale pre-trained vision models (e.g., SAM, DINOv2) can serve as frozen encoders in federated registration, reducing the number of parameters that need to be communicated and aggregated.

VIII. CONCLUSION

This survey has examined federated learning applied to 2D medical image registration. FL fits this problem naturally: registration losses are unsupervised, so every participating institution contributes to training regardless of whether its data is annotated, while the federated protocol keeps raw images local.

We covered the full technical stack—classical and deep-learning registration, FL aggregation and non-IID challenges, privacy mechanisms, and domain adaptation—and proposed an evaluation protocol that accounts for cross-site generalisation, personalisation, and the privacy–utility trade-off.

The intersection of FL and 2D registration is still early-stage. Adaptive regularisation, registration-specific privacy mechanisms, and genuine multi-site benchmarks are the most immediate gaps. We hope this survey is a useful starting point for researchers entering the area.

REFERENCES

- [1] F. P. M. Oliveira and J. M. R. S. Tavares, “Medical image registration: a review,” *Comput. Methods Biomech. Biomed. Eng.*, vol. 17, no. 2, pp. 73–93, 2014.
- [2] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: a learning framework for deformable medical image registration,” *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.
- [3] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 28, 2015, pp. 2017–2025.
- [4] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2017, pp. 1273–1282.
- [5] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Smola, and V. Smith, “Federated optimization in heterogeneous networks,” in *Proc. Mach. Learn. Syst. (MLSys)*, 2020.
- [6] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [7] J. P. Thirion, “Image matching as a diffusion process: an analogy with Maxwell’s demons,” *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.

- [8] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2015, pp. 234–241.
- [10] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptogr. Conf. (TCC)*, 2006, pp. 265–284.
- [11] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2017, pp. 1175–1191.
- [12] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, 2019.
- [13] C. T. Dinh, N. H. Tran, and T. D. Nguyen, "Personalized federated learning with Moreau envelopes," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 21394–21405.
- [14] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: stochastic controlled averaging for federated learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 5132–5143.
- [15] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle-aged, nondemented, and demented older adults," *J. Cognit. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [16] O. Bernard, A. Lalande, C. Zotti *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.
- [17] M. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "FIRE: fundus image registration dataset," *Model. Artif. Intell. Ophthalmol.*, vol. 1, no. 4, pp. 16–28, 2017.
- [18] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2097–2106.
- [19] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 7611–7623.
- [20] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: federated learning on non-IID features via local batch normalization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
- [21] M. Yurochkin, M. Agarwal, S. Ghosh, K. Greenewald, N. Hoang, and Y. Khazaeni, "Bayesian nonparametric federated learning of neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 7252–7261.
- [22] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, and Y. Khazaeni, "Federated learning with matched averaging," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2020.
- [23] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 10713–10722.
- [24] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: fair and robust federated learning through personalization," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 6357–6368.
- [25] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.