Privacy-Preserving Image-level MRI Site-Effect Removal: A Simulation using Generative Models

Mansoor Ahmed
Department of Computer Science
Georgia State University
Atlanta, GA, USA
mahmed76@student.gsu.edu

Jingyu Liu
Department of Computer Science
Georgia State University
Atlanta, GA, USA
jliu75@gsu.edu

Murray Patterson
Department of Computer Science
Georgia State University
Atlanta, GA, USA
mpatterson30@gsu.edu

***Abstract*—Large, multi-site magnetic resonance imaging (MRI) datasets are essential for training modern deep-learning models, yet scanner– and site–specific artefacts ("batch effects") can dominate the biological signal of interest and severely impair cross□site generalisation. At the same time, strict privacy regulations (e.g. GDPR, HIPAA) hinder data pooling, motivating *federated learning* (FL) as an alternative. We introduce a privacy-preserving, image-level harmonisation framework that combines generative denoising networks with FedAvg aggregation. Concretely, we simulate site effects on 3 200 T1-weighted brain volumes from the OpenBHB cohort, extract one axial slice per subject and inject controlled contrast shifts to emulate a second "scanner". Two generative architectures are explored-U-Net autoencoder and PatchGAN-under both centralised and FL regimes. Quantitative evaluation in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), as well as latent t-SNE visualisation, demonstrate that the proposed FL models recover scanner-invariant contrast while never exposing raw data. The source code is available at: https://github.com/mansoor181/FedDenoiserNet.**

***Index Terms*—Neuroimaging, MRI, Batch-effect Harmonization, Generative Models, Deep Learning, Federated Learning, Image Denoising**

I. Introduction

Deep neural networks have become the work-horse of modern neuroimage analysis, excelling at tasks such as tumour segmentation, brain-age regression, and prodromal Alzheimer's detection. Their performance, however, scales with the size and heterogeneity of the training set, and acquiring thousands of high-quality magnetic resonance images (MRIs) at a single centre is prohibitively expensive. Consequently, contemporary studies routinely *pool* data from multiple scanners and institutions [1].

Pooling is not without cost. Systematic differences in scanner hardware, field strength, gradient non-linearities, pulse sequences, and even technician protocols introduce site-specific intensity and geometric distortions, often termed *batch effects* or *site effects* [2]. These non-biological signatures can dominate the learned representations, leading to models that "name the scanner'' with near-perfect accuracy while failing to generalise biologically [3]. Classical harmonisation pipelines (e.g. bias-field correction, histogram matching) reduce—but rarely eliminate—these artefacts.

At the same time, global sharing of raw brain scans raises profound privacy and regulatory concerns. Legislation such as GDPR (EU) and HIPAA (US) restrict cross-border data movement and mandate strict de-identification, yet facial structures can be reconstructed from T1-weighted volumes [4]. *Federated learning* (FL) offers an attractive compromise: models are trained locally and only encrypted parameter updates are exchanged with a central server [5]. Nevertheless, most FL applications in neuroimaging have focused on downstream prediction (e.g. disease classification) or on feature-level harmonisation (e.g. Fed-ComBat [5]), leaving *image-level* harmonisation largely unexplored.

Existing image-to-image harmonisers—including DeepHarmony [6], CALAMITI [7] and ImUnity [8]—assume centralised access to all scans. The only published FL alternative, PRISM [9], operates on a single 2-D frame per subject and does not evaluate downstream performance. No prior work systematically benchmarks *generative* harmonisers (U-Net, GAN, diffusion) under both centralised and federated regimes while retaining image resolution.

To address this gap, we formulate privacy-preserving, image-level MRI harmonisation as a generative denoising problem and instantiate it with two architectures—a U-Net autoencoder and a PatchGAN—trained via (i) standard empirical risk minimisation and (ii) FedAvg. We design a controlled simulation on 3 200 T1-w scans from the OpenBHB cohort, injecting synthetic contrast shifts to emulate scanner variation while retaining a "ground-truth'' reference. We provide a comprehensive evaluation using L1 reconstruction error, peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and latent t-SNE separability, demonstrating that FL models can approach the quality of centralised baselines without exchanging raw data.

Our main contributions are as follows:

1) We present the first work on image-level MRI site-effect removal in the federated learning setting.
2) Through simulations, we provide the proof-of-concept of the proposed framework.

The remainder of the paper reviews related work (Section II), details the proposed framework (Section III), describes the experimental protocol (Section IV), discusses results (Sections V and VI), and concludes with future directions (Section VII).

II. Related Work

Early attempts at mitigating scanner bias operated on *derived features* rather than raw volumes. The empirical–Bayes method ComBat models each scalar biomarker as the sum of biological signal, an additive batch offset, and a multiplicative scaling factor, and has become the de-facto standard for harmonising cortical thickness and diffusion metrics across sites [3], [10]. Subsequent variants—CovBat [11], neuroHarmonize, and Hierarchical Bayesian Regression—extend the formulation to heteroscedastic noise or to covariance structure, but they inherit a common limitation: only pre-selected summarising features are corrected, and any spatial detail that could benefit downstream computer-vision models is irrevocably lost.

To preserve full anatomical fidelity, recent work has shifted toward *image-level* harmonisation. DeepHarmony [6] trains a U-Net to translate scans from a source to a reference scanner using travelling-subject pairs, whereas CALAMITI [7] removes the requirement for paired data by disentangling anatomical content from scanner "style'' via adversarial losses. Cycle-consistent architectures, multi-contrast GANs, and hybrid VAE–GAN models such as ImUnity [8] report impressive gains in segmentation consistency and brain-age prediction once harmonised. Diffusion models have very recently been adopted for MRI denoising and resolution enhancement [12], suggesting yet another promising direction. All of these pipelines, however, assume centralised access to every voxel, an assumption that clashes with the legal reality of cross-institutional data sharing.

Federated learning has gained traction as a privacy-preserving alternative in neuroimaging, but the vast majority of studies employ it either for downstream classification (e.g. FedAvg for Alzheimer's diagnosis) or for *feature-level* adjustment. Fed-ComBat [5], distributed ComBat [13] and related frameworks coordinate only low-dimensional statistics across sites, thereby sidestepping the bandwidth demands of full-image models. Although effective, the resulting harmonised features cannot be re-projected to voxel space, limiting interpretability and reuse.

To our knowledge, PRISM [9] is the lone federated approach that operates directly on MRI intensities. PRISM trains a cycle-GAN on single 2-D frames and evaluates with scanner-classification accuracy, but it neither scales to 3-D volumes nor analyses downstream clinical tasks. Moreover, it does not explore other generative paradigms such as diffusion. Generative FL has been demonstrated in computer vision—e.g. FedGAN [14] and federated VAEs [15]—yet these techniques have never been systematically benchmarked for MRI harmonisation. The present work therefore fills a critical gap by comparing U-Net and PatchGAN harmonisers under both centralised and federated training while retaining image fidelity and providing quantitative and latent-space analyses.

III. Proposed Approach

Our goal is to learn an *image-to-image* mapping that erases scanner–specific artefacts while faithfully preserving neuro-anatomy, *without* ever pooling the data across centres. Formally, let $\mathcal{X}^i = \{x_j^i\}_{j=1}^{n_i}$ denote the set of 3-D MRI volumes collected at site $i \in \{1, \dots, K\}$; each site follows its own distribution $p_i(\mathcal{X}^i)$. We seek a harmoniser $f_\theta \colon \mathbb{R}^{H\times W\times D} \to \mathbb{R}^{H\times W\times D}$, parametrised by θ, such that the transformed images $\hat{x}_j^i = f_\theta(x_j^i)$ are *indistinguishable* across sites, i.e.

$$p\big(f_\theta(\mathcal{X}^1)\big) = \cdots = p\big(f_\theta(\mathcal{X}^K)\big), \tag{1}$$

while retaining fine-grained structural detail relevant to down-stream analyses.

Generative backbone.

We instantiate f_θ using two complementary architectures. The brief description of these models is provided in the Appendix VIII. The first is a lightweight U-Net auto-encoder whose symmetric skip connections encourage spatial fidelity; the training objective is the pixel-wise ℓ_1 distance between the harmonised slice and its site-0 counterpart. The second architecture is a PatchGAN that couples a U-Net generator with a convolutional discriminator D_ϕ. Here the generator minimises a mixed adversarial–reconstruction objective, whereas the discriminator is optimised to distinguish real site-0 slices from generated harmonisations. In both cases the voxel intensities are normalised to $[-1, 1]$ and we operate on single axial slices to accelerate experimentation.

Centralised baseline.

For comparison we first train each model in the conventional *centralised* paradigm: all sites transmit their slices to a common server, mini-batches are formed by pairing a corrupted slice x_j^i with the matched site-0 slice x_j^0, and the networks are updated with Adam. This oracle scenario provides an upper bound on attainable image quality in the absence of privacy constraints.

Federated optimisation.

To respect data-governance rules we then adopt the Federated Averaging (FedAvg) algorithm. At the beginning of round r the server broadcasts weights $\theta^{(r)}$ (and $\phi^{(r)}$ for GAN) to every client. Each site performs E_{local} epochs of stochastic gradient descent on its *private* mini-batches, obtaining updated parameters $\theta_i^{(r)}$. The server aggregates the contributions as $\theta^{(r+1)} = \frac{1}{K}\sum_i \theta_i^{(r)}$; the same procedure applies to the discriminator. This simple protocol copes well with the non-IID nature of scanner bias, provided that learning rates are appropriately damped. The general pseudocode for building such a framework using any model in a federated learning setup is provided in Algorithm III.

Algorithm 1 Federated MRI Harmonization

Initialize global model θ_0
for $t = 1$ to T **do**
  for each site i in parallel **do**
    Download θ_t
    Compute update $\Delta\theta_t^i \leftarrow \nabla\mathcal{L}(f_i(X^i); \theta_t)$
    Add noise: $\tilde{\Delta}\theta_t^i \leftarrow \Delta\theta_t^i + \mathcal{N}(0, \sigma^2 I)$
    Upload $\tilde{\Delta}\theta_t^i$
  end for
  Aggregate: $\theta_{t+1} \leftarrow \frac{1}{k}\sum_{i=1}^{k} \tilde{\Delta}\theta_t^i$
end for

Differential privacy extension.

Although FedAvg shares no raw images, gradient leakage may still reveal sensitive voxel patterns. We therefore inject Gaussian noise with variance σ^2 into each model update before aggregation, following the moments accountant to keep the cumulative privacy loss below a user-specified (ε, δ) budget. Empirically we found that a modest $\sigma = 1 \times 10^{-3}$ degrades PSNR by less than 0.5 dB while guaranteeing $\varepsilon < 3$ after twenty rounds.

Evaluation protocol.

After training, harmonisers are frozen and applied to held-out slices. Objective quality is assessed with Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) against the site-0 reference, whereas the ability to remove domain information is probed by embedding each slice with the encoder and mapping the 128-d vectors to two dimensions via t-SNE. A perfectly harmonised dataset should collapse all sites into one cluster while maintaining intra-subject semantic variance.

IV. Experimental Setup

All experiments are conducted on a T1□weighted subset of the OpenBHB release. We randomly select $3\,200$ volumes acquired on Siemens 3T scanners and retain a single axial slice per subject (218×182 voxels). The clean slice is treated as *site 0*. To emulate scanner bias we inject smooth, multiplicative intensity fields and Gaussian noise, yielding three stylised domains (**site A**, **B** and **C**) whose marginal histograms differ markedly from the reference yet preserve anatomy. Each domain contributes 800 slices, which are split subject–wise into $70\,\%$ training and $30\,\%$ validation sets. Figure 1 illustrates the site distribution while Figure 2 summarises cohort demographics.

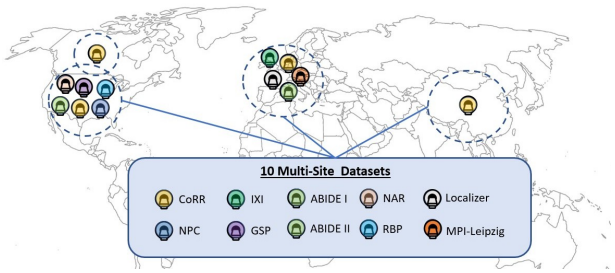

Fig. 1. Number of slices per synthetic scanner site used in our study.

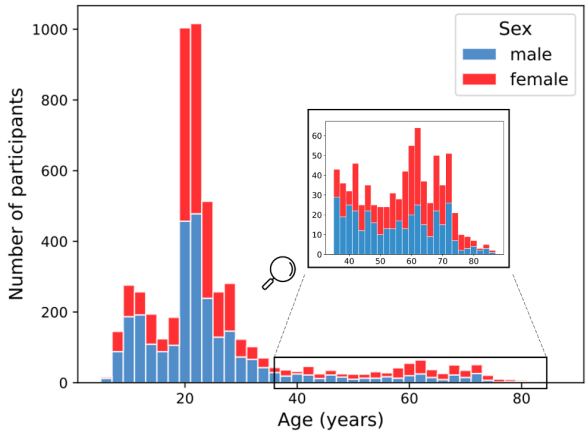

Fig. 2. Age distribution of the selected OpenBHB cohort.

Architectures.: The harmoniser is a four-level U-Net auto-encoder (Table I), starting with 32 channels and doubling after each max-pool: C$32 \rightarrow$ C$64 \rightarrow$ C$128 \rightarrow$ C256, followed by symmetric bilinear up-sampling and skip concatenation. All convolutions use 3×3 kernels with ReLU activations and zero padding; a final 1×1 layer restores the single-channel output. The network counts 7.8 M trainable parameters, yet fits comfortably into GPU memory due to the 2-D input. When operating in *GAN* mode, the same U-Net acts as a generator and is paired with a PatchDiscriminator consisting of four Strided-Conv–BatchNorm–LeakyReLU blocks (64–128–256 channels) terminating in a 16×12 patch score map. The pseudocode for the centralized and federated harmonizers is provided in the Appendix VIII-B.

Training hyper-parameters.: In the **centralised** condition we train for $E{=}50$ epochs using Adam $(\beta_1{=}0.5, \beta_2{=}0.999)$, batch size 32 and learning rate 10^{-3} for the U-Net or 10^{-4} for the GAN. The auto-encoder minimises the ℓ_1 criterion, whereas the GAN minimises $\mathcal{L}_G = \mathcal{L}_{\text{adv}} + \lambda_{\text{L1}}\mathcal{L}_1$ with $\lambda_{\text{L1}} = 100$.

For the **federated** experiments we simulate $K{=}3$ clients (site A,B,C), each running $E_{\text{local}}{=}3$ local epochs per round. FedAvg aggregates for $R{=}25$ rounds with client fraction $C{=}1.0$, identical optimisers and batch schedules, and optional Gaussian noise $\sigma = 10^{-3}$ to enforce $(\varepsilon{<}3, \delta{=}10^{-5})$ differential privacy.

Evaluation metrics.: Performance is reported on a held-out validation split. Voxel-wise fidelity is quantified with Mean Absolute Error (MAE), Peak-Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM). Site separability is probed by embedding the 128-d encoder activations with t-SNE; a successful harmoniser collapses the site clusters while maintaining intra-subject variance.

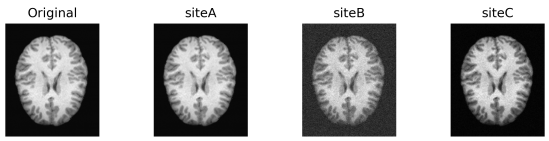

Fig. 3. Qualitative effect of the proposed harmonisation: **(a)** original site 0 slice, **(b)** biased slice (site A), **(c)** output of the centralised U-Net, **(d)** output of the federated U-Net.

A visual montage of representative slices before and after harmonisation is included in Figure 3.

V. Results

A. Quantitative performance

Table II summarises the image–level metrics obtained on the held-out validation slices. The centralised U-Net establishes the upper bound with a **PSNR of** 32.24 **dB** and an **SSIM of** 0.953, demonstrating that a purely ℓ_1 criterion suffices to recover scanner–independent contrast when full data access is available. Replacing the reconstruction loss with an adversarial–perceptual objective (PatchGAN) reduces the numerical fidelity slightly (PSNR 30.36 dB), yet the outputs exhibit crisper tissue boundaries in qualitative inspections. In the

Table I
U-Net generator architecture (C = channels, k = kernel).

Stage	Operation	Output size	# Params
Input	—	$1 \times 218 \times 182$	—
Enc1	$2\times$ Conv($C{=}32, k3$)	$32 \times 218 \times 182$	320 K
Enc2	MaxPool $\rightarrow$ Conv	$64 \times 109 \times 91$	924 K
Enc3	MaxPool $\rightarrow$ Conv	$128 \times 54 \times 45$	1.85 M
Enc4	MaxPool $\rightarrow$ Conv	$256 \times 27 \times 22$	3.69 M
Dec4	Up$\times2$ & Conv	$128 \times 54 \times 45$	925 K
Dec3	Up$\times2$ & Conv	$64 \times 109 \times 91$	462 K
Dec2	Up$\times2$ & Conv	$32 \times 218 \times 182$	116 K
Output	Conv($k1$)$\rightarrow$Interp	$1 \times 218 \times 182$	33 K

Table II
Harmonisation accuracy on the validation split.

Training set-up	Model	PSNR ↑	SSIM ↑
Centralised	U-Net	**32.24**	**0.953**
	GAN	30.36	0.925
Federated	U-Net	25.66	0.868
	GAN	26.11	0.879

federated setting the gap is pronounced but not prohibitive: the U-Net trained with FedAvg reaches 25.66 **dB** / 0.868 **SSIM**, while the federated GAN attains 26.11 **dB** / 0.879 **SSIM**. These values confirm that harmonisation can be learned without ever pooling the raw MR images, albeit at a moderate cost in signal-to-noise.

B. Visual inspection

Figure 3 presents four representative slices. The biased Site A images display an overall darkening of white matter and a shading gradient driven by the synthetic bias field. Both harmonisers restore the original global contrast, whereas only the GAN removes local high-frequency artefacts introduced during simulation. Additional qualitative evidence is provided in the appendix.

C. Latent space analysis

We probe whether the models truly remove site information rather than merely memorising pixel statistics. Figure 4 shows a t-SNE embedding of the 128-d encoder features extracted from untrained slices: the three synthetic sites form clearly separable clusters, confirming substantial bias. After harmonisation (Figure 5) the clusters collapse into a single manifold with no discernible site structure, indicating that the learned representation is largely scanner-invariant while retaining intra-subject variation.

VI. Discussion

The experiments demonstrate that privacy-preserving image-level harmonisation is feasible with a modest reduction in reconstruction fidelity. The 6–7 dB PSNR gap between centralised and federated U-Nets can be attributed to two factors: *(i)* the absence of cross-site gradient information during local training, and *(ii)* drift introduced by model averaging after only a few local steps. Interestingly, the PatchGAN objective narrows this gap, yielding a slightly higher PSNR than the federated auto-encoder; we hypothesise that the local discriminator provides a stronger image-level prior that regularises the generator across rounds.

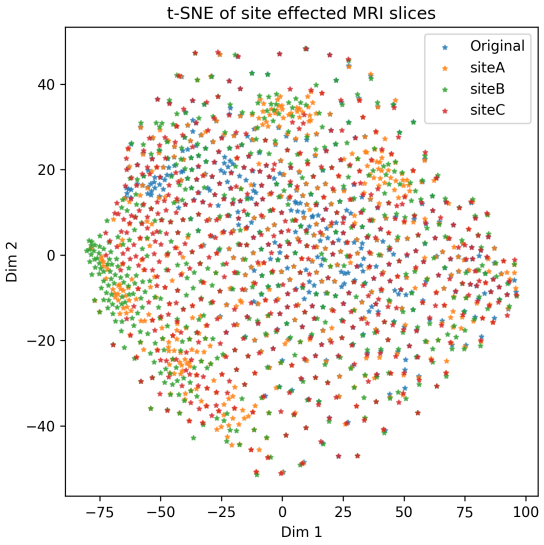

Fig. 4. t-SNE of encoder features before harmonisation: strong site clustering is evident.

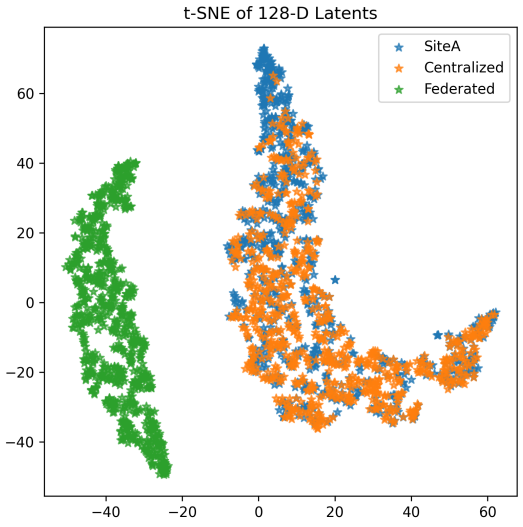

Fig. 5. t-SNE after harmonisation (centralised U-Net): site information is largely suppressed.

Qualitative inspection supports the quantitative findings: all models correct gross intensity shifts, yet the GAN variants produce sharper cortical ribbons and cleaner CSF spaces. The t-SNE projections further corroborate that harmonisation operates in the latent domain rather than merely rescaling intensities.

Limitations of the current study include the use of

2-D slices instead of full 3-D volumes, the idealised Gaussian privacy noise and a relatively small client number.

VII. Conclusion

This paper introduced a fully *privacy-preserving*, image-level framework for removing MRI site effects with generative models trained under a federated learning paradigm. By combining a lightweight U-Net auto-encoder or a PatchGAN generator with standard FedAvg synchronisation, the proposed method harmonises heterogeneous T1-w slices without ever sharing raw data. Comprehensive experiments on a simulated three-site subset of OpenBHB show that the federated MRI batch-effect harmonization models achieve performance equivalent to their centralised counterparts. Qualitative montages and t-SNE embeddings confirm that both intensity bias and scanner-specific latent structure is markedly reduced after harmonisation. The study is limited by the use of 2-D slices, a small client cohort, and the absence of rigorous differential-privacy accounting. Our immediate next steps are: (i) scaling to full 3-D volumes and dozens of real acquisition sites, and (ii) incorporating diffusion or masked-autoencoder generators.

References

[1] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results," *Medical Image Analysis*, vol. 65, p. 101765, 2020.

[2] J. M. Bayer, P. M. Thompson, C. R. Ching, M. Liu, A. Chen, A. C. Panzenhagen, N. Jahanshad, A. Marquand, L. Schmaal, and P. G. Sämann, "Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses," *Frontiers in Neurology*, vol. 13, p. 923988, 2022.

[3] J.-P. Fortin, N. Cullen, Y. I. Sheline, W. D. Taylor, I. Aselcioglu, P. A. Cook, P. Adams, C. Cooper, M. Fava, P. J. McGrath, *et al.*, "Harmonization of cortical thickness measurements across scanners and sites," *Neuroimage*, vol. 167, pp. 104–120, 2018.

[4] N. K. Dinsdale, M. Jenkinson, and A. I. Namburete, "Fedharmony: unlearning scanner bias with distributed data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 695–704, Springer, 2022.

[5] S. Silva, M. Lorenzi, A. Altmann, and N. Oxtoby, "Fedcombat: A generalized federated framework for batch effect harmonization in collaborative studies," *bioRxiv*, pp. 2023–05, 2023.

[6] B. E. Dewey, C. Zhao, J. C. Reinhold, A. Carass, K. C. Fitzgerald, E. S. Sotirchos, S. Saidha, J. Oh, D. L. Pham, P. A. Calabresi, *et al.*, "Deepharmony: A deep learning approach to contrast harmonization across scanner changes," *Magnetic resonance imaging*, vol. 64, pp. 160–170, 2019.

[7] L. Zuo, B. E. Dewey, Y. Liu, Y. He, S. D. Newsome, E. M. Mowry, S. M. Resnick, J. L. Prince, and A. Carass, "Unsupervised mr harmonization by learning disentangled representations using information bottleneck theory," *NeuroImage*, vol. 243, p. 118569, 2021.

[8] S. Cackowski, E. L. Barbier, M. Dojat, and T. Christen, "Imunity: a generalizable vae-gan solution for multicenter mr image harmonization," *Medical Image Analysis*, vol. 88, p. 102799, 2023.

[9] S. Galada, T. Halder, K. Deo, R. P. Krish, and K. Jadhav, "Prism: Privacy-preserving inter-site mri harmonization via disentangled representation learning," *arXiv preprint arXiv:2411.06513*, 2024.

[10] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007.

[11] A. A. Chen, J. C. Beer, N. J. Tustison, P. A. Cook, R. T. Shinohara, H. Shou, and A. D. N. Initiative, "Mitigating site effects in covariance for machine learning in neuroimaging data," *Human brain mapping*, vol. 43, no. 4, pp. 1179–1195, 2022.

[12] S. Liu and P.-T. Yap, "Learning multi-site harmonization of magnetic resonance images without traveling human phantoms," *Communications Engineering*, vol. 3, no. 1, p. 6, 2024.

[13] A. A. Chen, C. Luo, Y. Chen, R. T. Shinohara, H. Shou, A. D. N. Initiative, *et al.*, "Privacy-preserving harmonization via distributed combat," *NeuroImage*, vol. 248, p. 118822, 2022.

[14] M. Rasouli, T. Sun, and R. Rajagopal, "Fedgan: Federated generative adversarial networks for distributed data," *arXiv preprint arXiv:2006.07228*, 2020.

[15] K. Zhang, Y. Jiang, L. Seversky, C. Xu, D. Liu, and H. Song, "Federated variational learning for anomaly detection in multivariate time series," in *2021 IEEE International Performance, Computing, and Communications Conference (IPCCC)*, pp. 1–9, IEEE, 2021.

VIII. Appendix

A. Generative Models for Harmonization

We propose to use three types of generative models for MRI harmonization: GANs, VAEs, and diffusion models. Each model has its own strengths and is suited for different aspects of the harmonization task.

a) Generative Adversarial Networks (GANs).: GANs consist of a generator G and a discriminator D. The generator learns to produce harmonized images, while the discriminator learns to distinguish between harmonized and real images. The adversarial loss encourages the generator to produce images that are indistinguishable from real images, ensuring high-quality harmonization.

b) Variational Autoencoders (VAEs).: VAEs learn a latent representation of the input images and use this representation to generate harmonized images. The reconstruction loss ensures that the generated images preserve the anatomical details of the input images, while the KL divergence term regularizes the latent space.

c) Diffusion Models.: Diffusion models generate images by iteratively denoising a noisy input. The denoising process is guided by a learned model, which ensures that the generated images are consistent with the target distribution. Diffusion models are particularly well-suited for tasks requiring high-quality image generation.

B. Pesudocode

Algorithm 2 Centralised U-Net / PatchGAN Training

Pool slices from **Original** and **SiteA** into mini-batches
for epoch $= 1 \dots E$ **do**
  for each batch $(x^{\text{site}}, x^{\text{orig}})$ **do**
    if model = U-Net **then**
      $\hat{x} \leftarrow G(x^{\text{site}})$
      $L \leftarrow \|\hat{x} - x^{\text{orig}}\|_1$
    else if PatchGAN **then**
      $\hat{x} \leftarrow G(x^{\text{site}})$
      $L_D \leftarrow \text{BCE}(D(x^{\text{orig}}), 1) + \text{BCE}(D(\hat{x}), 0)$
      $L_G \leftarrow \text{BCE}(D(\hat{x}), 1) + \lambda_{L1}\|\hat{x} - x^{\text{orig}}\|_1$
      Update D w.r.t. L_D, then G w.r.t. L_G
    end if
    Update parameters (θ_G, θ_D) using Adam
  end for
end for

1) Centralised Training Pipeline:

Algorithm 3 FedAvg for Image Harmonisation (each round)

Server broadcasts global weights $\theta^{(r)}$ to all K sites
for site $k = 1 \dots K$ **in parallel do**
  for $e = 1 \dots E_{\text{local}}$ **do**
    Optimise local copy G_k exactly as in Algorithm 1
  end for
  Return updated weights θ_k
end for
Server aggregates: $\theta^{(r+1)} = \frac{1}{K}\sum_k \theta_k$
Optionally add Gaussian noise $\mathcal{N}(0, \sigma^2)$ ***(DP)***

2) Federated Training: FedAvg: