GradInvDiff: Stealing Medical Privacy in Federated Learning via Diffusion-Based Gradient Inversion

Zhiyuan Wang[1], Daisong Gan[2], Wenzhuo Fang[1],
Yuliang Zhu[2,3], and Kun Liu[1](✉)

[1] School of Automation, Beijing Institute of Technology, Beijing, China
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[3] University of Nottingham Ningbo China, Ningbo, China
kunliubit@bit.edu.cn

Abstract. Federated learning (FL) has become a crucial technique for medical imaging analysis, enabling multiple institutions to train machine learning models while preserving patient privacy collaboratively. However, recent research has uncovered the vulnerability of shared gradients in FL, which can be exploited through the gradient inversion attack (GIA) to reconstruct private medical images. While existing methods show promise in generic image tasks, their application to high-resolution medical images remains underexplored and ineffective due to data complexity. This paper introduces GradInvDiff, a novel GIA tailored for medical FL scenarios. Unlike traditional methods that rely solely on gradient guidance, our approach combines diffusion models with gradient matching optimization to iteratively refine the inference process. By replacing the standard random noise in the diffusion process with a direction derived from the difference between the optimized and original means, we inject a gradient-based condition into the noise to enhance image reconstruction quality. This method enables high-quality, pixel-level reconstruction of private medical images, even in the presence of large batch sizes or gradient noise. Our experiments demonstrate that GradInvDiff outperforms existing state-of-the-art gradient inversion methods and shows better accuracy and visibility when attacking medical FL models. We hope that this paper can raise public awareness of privacy leakage risks when using medical FL.

Keywords: Federated Learning · Gradient Inversion Attack · Diffusion Models.

1 Introduction

Federated learning (FL) has emerged as a pivotal paradigm for decentralized medical imaging analysis, enabling collaborative model training across insti-

Source code is available at https://github.com/R00TSEN1650/GradInvDiff.

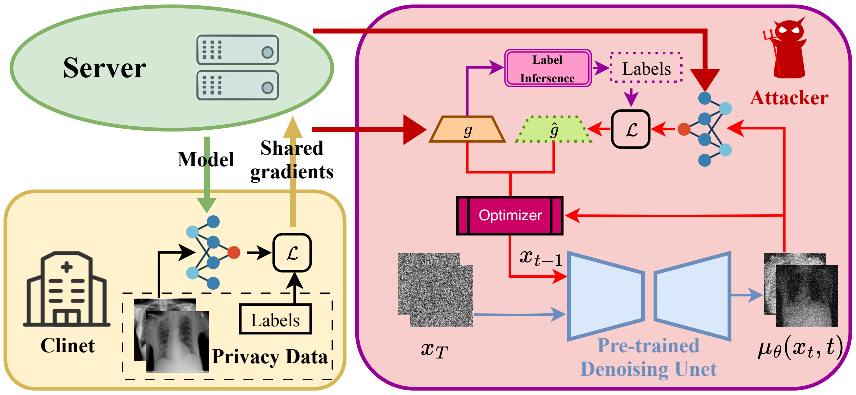


Fig. 1. Overview of the proposed Diffusion-based Gradient Inversion Architecture.

tutions while ostensibly preserving patient privacy through gradient exchange rather than raw data sharing [6,19,20]. Its critical impact is evident in applications like multi-institutional tumor segmentation, X-ray classification, and MRI reconstruction, which enhance diagnostic accuracy and protect patient data [8,10,21]. Nevertheless, this privacy promise is fundamentally challenged by the gradient inversion attack (GIA), where adversaries exploit shared gradients as "Trojan horses" to reconstruct private medical images [15,27]. While existing GIAs demonstrate risks in generic vision tasks, their ability to steal medical images with high clinical validity in medical FL - whether they can effectively reconstruct diagnostically meaningful information that preserves critical anatomical features and pathological characteristics - remains unverified.

Existing GIAs are primarily categorized into three approaches. Optimization-based GIAs [12,29,31] iteratively optimize pixel values to match target gradients. While these methods are conceptually simple, they face challenges when applied to medical imaging due to the high dimensionality of such data, often leading to blurred reconstructions that fail to preserve diagnostically critical features. Analytic-based GIAs [4,11,30] require model modifications to access private data, limiting their applicability in the common "honest-but-curious" setting where model architecture is fixed. Generation-based GIAs leverage generative models as structural priors to constrain the solution space by learning from data distributions. These methods can be further classified based on their generative models. GAN-based GIAs [9,16,17] utilize adversarial training to achieve semantic consistency during reconstruction; however, despite their effectiveness in producing semantically similar images, they frequently introduce significant spatial misalignments in reconstructed anatomical structures, undermining their clinical utility [17]. Diffusion-based GIAs [13,24] leverage the prior knowledge of image generation diffusion models. DGGI [24] uses diffusion outputs only for initialization, lacking iterative gradient-guided refinement. Conversely, GGDM [13] integrates gradient loss directly into sampling, which can cause manifold deviation and degrade feature preservation [26].

To address the critical challenges of existing GIAs in medical FL, we propose GradInvDiff, a diffusion-based framework that integrates gradient matching optimization with medical imaging generation, as depicted in Fig. 1. Our method introduces two key innovations: (1) Adaptive Mean Optimization (AMO): A hybrid mean formulation, akin to classifier-free guidance, that dynamically adjusts the gradient conditioning strength to constrain the sampling path towards the data manifold. (2) Gradient-Aligned Noise Injection (GANI): A strategy that replaces standard stochastic noise by projecting it onto the gradient-matching direction, reducing sampling randomness and aligning the update with gradient information.

To the best of our knowledge, GradInvDiff is the first diffusion-based gradient inversion framework specifically designed for medical FL. Our main contributions are threefold: (1) **Diffusion-based Gradient Inversion Architecture**: We develop a novel framework that combines adaptive mean optimization with time-variant gradient-diffusion blending. This approach iteratively refines the diffusion sampling trajectory to better align with gradient-matching objectives, thereby improving the fidelity of reconstructed images. (2) **Gradient-Conditioned Sampling**: We propose a gradient-aligned noise projection mechanism that projects the added noise during the diffusion process onto the gradient residual subspace. This allows the reverse diffusion process to better preserve critical image details while ensuring gradient guidance is consistently integrated throughout the generation process. (3) **Clinical Validation**: Comprehensive evaluation across multiple imaging modalities demonstrates robust performance under practical FL settings, including large-batch training and gradient noise, with significant improvements in preserving pathological features compared to existing methods.

2 Method

2.1 Threat Model

Attacker's Knowledge and Capabilities In our threat model, the adversary operates as an honest-but-curious server with access to FL model parameters and shared gradients. While capable of storing and processing client updates, the adversary cannot modify the learning protocol or global parameters. The attack pipeline combines two key capabilities: robust label inference via RLU [5], which ensures effective label reconstruction under practical FL constraints, and anatomical prior integration through modality-specific diffusion models pre-trained on public medical imaging data. Fig. 1. illustrates the complete implementation workflow.

Attacker's Objective Given an FL model f_W with parameters W for medical image classification and batch-averaged gradients g computed from a private batch of images x and labels y, the attacker seeks to reconstruct private images $\hat{x}^* \in \mathbb{R}^{B\times H\times W\times C}$ by solving the following optimization problem:

$$\hat{x}^* = \arg\min_{\hat{x}} \mathcal{D}\left(\hat{g}, g\right), \tag{1}$$

where $\hat{g} = \frac{1}{B}\sum_{i=1}^{B} \nabla_W \mathcal{L}(f_W(\hat{x}_i), \hat{y}_i)$ represents the dummy gradients computed from the reconstructed images $\hat{x}$, with labels $\hat{y}$ inferred using RLU [5]. Following state-of-the-art methods [13], we adopt negative cosine similarity as the gradient matching loss $\mathcal{D}(\cdot,\cdot)$:

$$\mathcal{D}(g_1, g_2) = 1 - \frac{\langle g_1, g_2\rangle}{\|g_1\|\|g_2\|}. \tag{2}$$

2.2 Preliminary: Diffusion Model

Our generative framework leverages a Denoising Diffusion Probabilistic Model (DDPM) [14] trained on publicly available medical image datasets. The diffusion process decouples image generation into two distinct phases—deterministic denoising and stochastic noise injection—thus enabling high-fidelity synthesis through controlled corruption reversal.

In the *forward* process, clean images are gradually degraded over T iterative steps by progressively adding Gaussian noise according to a predefined schedule $\{\beta_t\}_{t=1}^{T}$. This results in a sequence of images $\{x_t\}_{t=0}^{T}$ with increasing levels of noise:

$$x_t = \sqrt{1-\beta_t}\cdot x_{t-1} + \sqrt{\beta_t}\cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I). \tag{3}$$

Thus, the original image x_0 is systematically transformed into pure Gaussian noise x_T, allowing the model to learn an effective inversion of this degradation.

In the *reverse* process, starting from $x_T \sim \mathcal{N}(0, I)$, the model iteratively reconstructs the original image by refining the noise. The sampling process of the DDPM is illustrated in Fig. 2(a). At each timestep, it computes a denoised mean $\mu_\theta(x_t, t)$ conditioned on the current noisy image x_t, and combines it with a controlled noise component via $\Sigma_\theta(x_t, t)$:

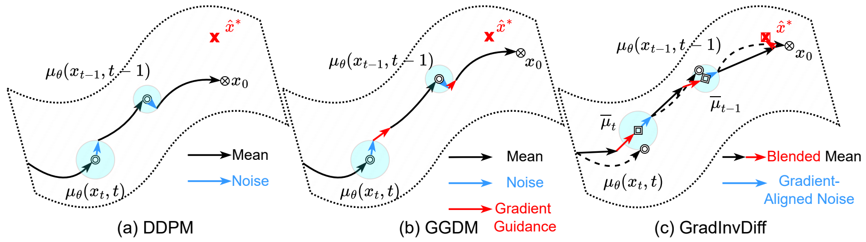
$$x_{t-1} = \mu_\theta(x_t, t) + \Sigma_\theta(x_t, t)\cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I). \tag{4}$$

This dual mechanism not only recovers the underlying image structure but also preserves critical anatomical details, making DDPMs particularly well-suited for high-fidelity medical image reconstruction.

2.3 Gradient Inversion with Diffusion Model

To effectively combine GIA and diffusion models, it is necessary to treat the gradient as a condition for the diffusion model. Existing methods like GGDM [13] adopt a classifier guidance approach [7] by computing virtual gradients through the FL model and labels, then incorporating gradient similarity guidance into the reverse process. As illustrated in Fig. 2(b), GGDM's sampling formula is:

$$x_{t-1} = \mu_\theta(x_t, t) + \Sigma_\theta(x_t, t)\left(\gamma \Sigma_\theta \nabla_{x_t} \frac{\langle \hat{g}, g\rangle}{\|\hat{g}\|\|g\|} + \epsilon\right), \quad \epsilon \sim \mathcal{N}(0, I), \tag{5}$$



(a) DDPM (b) GGDM (c) GradInvDiff

Fig. 2. Reverse diffusion sampling paths of DDPM, GGDM, and GradInvDiff

where γ serves as a constant guidance scale. However, this fixed-scale guidance can lead to manifold deviation [26] during the optimization process, as the gradient constraints pull the reconstruction away from the data distribution learned by the diffusion model. To overcome this limitation, we introduce a more flexible approach by dynamically adjusting the gradient conditioning through time-dependent parameters. Our framework incorporates two key mechanisms: Adaptive Mean Optimization and Gradient-Aligned Noise Injection, which better align the reconstruction process with the data manifold.

Adaptive Mean Optimization Unlike GGDM, we optimize the predicted mean μ_θ directly through gradient matching. For each x_t, we perform K optimization steps with learning rate η to minimize the gradient matching loss $\mathcal{D}$ defined in Eq. 2. The optimized mean μ_t^* is selected as the iteration achieving minimal $\mathcal{D}$. Subsequently, we blend μ_t^* with the original μ_θ using a time-dependent schedule $\{\gamma_t\}_{t=1}^T$:

$$\overline{\mu} = \gamma_t \mu_t^* + (1 - \gamma_t)\mu_\theta(x_t, t). \tag{6}$$

The blending schedule prioritizes gradient guidance in high-frequency stages (near T) for rapid contour formation, while gradually reducing its influence in low-frequency stages (near 0) to preserve anatomical details. We use a linear decay to balance contour formation and detail preservation effectively. At high-frequency stages, the gradient influence is maximized ($\gamma_T = 1$), and it linearly decays to 0 in low-frequency stages, ensuring smooth transitions without unnecessary complexity.

Gradient-Aligned Noise Injection We introduce a noise projection mechanism to preserve gradient-matching information during stochastic sampling. The adjusted noise injection projects random noise onto the gradient residual subspace:

$$\epsilon_{\text{proj}} = \frac{\langle \epsilon, \Delta\mu \rangle}{\|\mu_t^* - \mu_\theta\|_2^2}(\mu_t^* - \mu_\theta), \quad \epsilon \sim \mathcal{N}(0, I). \tag{7}$$

This projection minimizes interference between gradient alignment and stochastic sampling. When noise is orthogonal to the gradient residual direction ($\Delta\mu$),

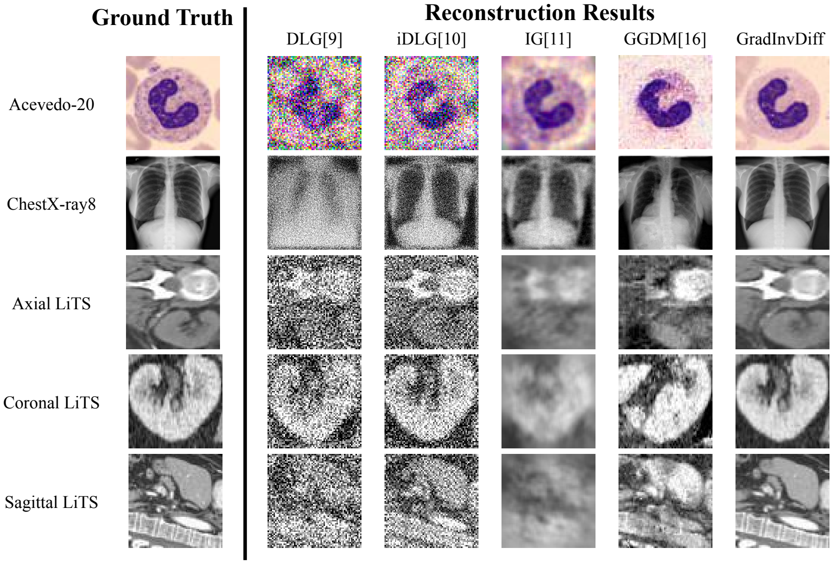


Fig. 3. Qualitative Comparisons of Image Reconstruction Across GIAs

the projected noise component becomes zero, effectively mitigating manifold deviation. As visualized in Fig. 2(c), this dual mechanism ensures gradient fidelity while maintaining the diffusion model's anatomical prior. The complete GradInvDiff sampling integrates both components:

$$x_{t-1} = \underbrace{\gamma_t \cdot \mu_t^* + (1-\gamma_t) \cdot \mu_\theta(x_t, t)}_{\text{Blended Mean}} + \underbrace{\Sigma_\theta(x_t, t) \cdot \epsilon_{\text{proj}}}_{\text{Gradient-Aligned Noise}} . \tag{8}$$

3 Experiments and Results

Implementation Details: We validate our method using two commonly used classification models in medical FL: LeNet7 and ResNet18, along with three medical image datasets: ChestX-ray8 [22] (resized to 224 × 224), Acevedo-20 [1] (resized to 64 × 64), and LiTS [2] (sliced into 64 × 64 2D images along sagittal, coronal, and axial planes), which have been preprocessed in the MedMNIST [25] dataset. The diffusion model is implemented using the iDDPM framework [18], with Gaussian initialization and an initial learning rate of 0.0001. For mean optimization, we configure 5 iterations with $\eta = 0.01$ using the Adam optimizer. All experiments are conducted on dual NVIDIA 3090 GPUs (48GB total memory).

Baseline Methods and Evaluation Metrics: Our baseline methods include DLG [31], iDLG [29], IG [12], and GGDM [13], which also utilizes a pre-trained

Table 1. Experimental Results for GIAs on Different Networks

Method	Metric	Dataset and Network					
		Acevedo-20[1]		ChestX-ray8[22]		LiTS[2]	
		LeNet	ResNet	LeNet	ResNet	LeNet	ResNet
DLG[31]	PSNR	11.5	8.78	8.74	8.61	10.3	9.26
	SSIM	0.0641	0.192	0.0119	0.0214	0.0961	0.0787
	LPIPS	0.698	0.397	1.37	1.31	0.987	0.974
iDLG[29]	PSNR	11.0	11.2	10.9	8.84	12.94	12.1
	SSIM	0.0967	0.479	0.0246	0.0225	0.203	0.245
	LPIPS	0.658	0.216	1.32	1.29	0.809	0.678
IG[12]	PSNR	18.9	11.5	20.2	15.7	18.7	14.2
	SSIM	0.627	0.449	0.784	0.264	0.464	0.382
	LPIPS	0.148	0.315	0.583	0.797	0.373	0.655
GGDM[13]	PSNR	19.4	14.4	20.1	16.8	19.2	15.1
	SSIM	0.648	0.539	0.791	0.487	0.645	0.426
	LPIPS	0.128	0.238	0.445	0.589	0.270	0.599
GradInvDiff(Ours)	PSNR	20.7	17.8	21.3	16.4	23.6	20.7
	SSIM	0.633	0.578	0.716	0.471	0.816	0.711
	LPIPS	0.107	0.185	0.164	0.249	0.107	0.298

diffusion model similar to our approach. To assess the quality of the privacy information stolen via GIA, we adopted the following metrics: Peak Signal-to-Noise Ratio (PSNR ↑); Structural Similarity Index Measure (SSIM ↑) [28]; and Learned Perceptual Image Patch Similarity (LPIPS ↓) [23], which computes similarity between target and reconstructed images using a neural network.

Results and Discussion: Fig. 3. visually demonstrates the superiority of GradInvDiff over existing baselines. Optimization-based methods (DLG, iDLG, IG) produce reconstructions plagued by noise artifacts and structural blurring, a direct consequence of unconstrained pixel-space optimization. While GGDM achieves partial improvement through the use of a diffusion prior, it still introduces noticeable perceptual distortions and fails to recover diagnostically critical details. GradInvDiff overcomes these limitations via two synergistic mechanisms: (1) AMO progressively aligns medical structures with gradient constraints while preserving anatomical coherence through time-variant blending, and (2) GANI replaces stochastic noise with deterministic updates to preserve fine textures. Together, these innovations enable the recovery of diagnostically critical, high-frequency details that are consistently lost by other methods.

As shown in Table 1, GradInvDiff demonstrates superior performance in PSNR, SSIM, and LPIPS across the Acevedo-20, ChestX-ray8, and LiTS datasets. Specifically, the results for the LiTS dataset are derived by averaging the 2D results from the sagittal, coronal, and axial planes. GradInvDiff advantage is

Table 2. Ablation Study of GradInvDiff on the Acevedo-20 Data.

Method			LeNet			ResNet		
Ablation	Gradient Noise	Batchsize	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
-	×	1	20.7	0.633	0.107	17.8	0.578	0.185
w/o AMO	×	1	14.3	0.443	0.324	11.3	0.321	0.288
w/o GANI	×	1	16.5	0.513	0.227	12.6	0.472	0.256
-	✓	1	19.6	0.602	0.123	16.9	0.523	0.196
-	×	4	18.5	0.582	0.191	15.0	0.499	0.185
-	×	8	16.7	0.524	0.248	12.9	0.390	0.249

"w/o AMO": the variant without Adaptive Mean Optimization.
"w/o GANI": the variant without Gradient-Aligned Noise Injection.

particularly pronounced in the LPIPS metric, underscoring its enhanced capability for preserving perceptual details. This deliberate focus on visual fidelity naturally results in a slight compromise on pixel-level distortion metrics, aligning with the established perception-distortion trade-off [3].

Ablation Study As shown in Table 2, the ablation studies confirm the critical role of each proposed component. Removing Adaptive Mean Optimization (w/o AMO) leads to severe performance degradation across all metrics, underscoring its importance in gradient-guided alignment. Similarly, disabling Gradient-Aligned Noise Injection (w/o GANI) significantly reduces reconstruction fidelity, highlighting its necessity for preserving anatomical consistency. Under additive Gaussian gradient noise $\mathcal{N}(0, 0.01^2)$, GradInvDiff incurs only a marginal performance drop, demonstrating strong robustness to such perturbations. Furthermore, the method maintains stable reconstruction quality across increasing batch sizes ($B = 4$ or 8), proving practical applicability for real-world FL scenarios. These results validate the complementary nature of the two components and the framework's adaptability to common FL constraints.

4 Conclusion

In this paper, we present a novel gradient inversion attack method, GradInvDiff, designed for reconstructing high-resolution medical images in federated learning. By combining diffusion models with gradient-matching mechanisms, we significantly improve image reconstruction quality, overcoming challenges such as large batch sizes and gradient noise. Experimental results demonstrate that GradInvDiff outperforms existing methods in medical federated learning scenarios, enabling high-quality private image reconstruction. Through this research, we aim to raise awareness of the potential privacy leakage risks in medical federated learning and encourage more attention to security concerns in this field.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (no. U24A20264, 62273041), Open Projects of the Institute of Systems Science, Beiiing Wuzi University (BWUISS11).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Acevedo, A., et al.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in Brief **30**, 105474 (2020)
2. Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G. E. H., Chartrand, G., et al.: The liver tumor segmentation benchmark (LiTS). Medical Image Analysis **84**, 102680 (2023)
3. Blau Y, Michaeli T.: The perception-distortion tradeoff. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6228–6237. IEEE (2018) https://doi.org/10.1109/CVPR.2018.00652
4. Boenisch, F., Dziedzic, A., Schuster, R., Shamsabadi, A. S., Shumailov, I., Papernot, N.: When the Curious Abandon Honesty: Federated Learning Is Not Private. In: Proceedings of the 2023 IEEE 8th European Symposium on Security and Privacy, pp. 175–199. IEEE (2023) https://doi.org/10.1109/EuroSP57164.2023.00020
5. Chen, H., Vikalo, H.: Recovering labels from local updates in federated learning. In: Proceedings of the 41st International Conference on Machine Learning, pp. 7346–7372. PMLR (2024)
6. Dayan, I., Roth, H. R., et al.: Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine **27**(10), 1735–1743 (2021)
7. Dhariwal P, Nichol A.: Diffusion Models Beat GANs on Image Synthesis. In: Proceedings of the 35th International Conference on Neural Information Processing Systems, pp. 8780–8794. Curran Associates Inc. (2021)
8. Elmas, G., Dar, S. U. H., Korkmaz, Y., et al.: Federated learning of generative image priors for MRI reconstruction. IEEE Transactions on Medical Imaging **42**(7), 1996–2009 (2022)
9. Fang, H., Chen, B., Wang, X., Wang, Z., Xia, S. T.: GIFD: A generative gradient inversion method with feature domain optimization. In: Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision, pp. 4967—4976. IEEE (2023) https://doi.org/10.1109/ICCV51070.2023.00458.
10. Feki, I., Ammar, S., Kessentini, Y., et al.: Federated learning for COVID-19 screening from Chest X-ray images. Applied Soft Computing **106**, 107330 (2021)
11. Fowl, L. H., Geiping, J., Czaja, W., Goldblum, M., Goldstein, T.: Robbing the Fed: Directly Obtaining Private Data in Federated Learning with Modified Models. In: Proceedings of the 2022 International Conference on Learning Representations, pp. 1–15 OpenReview (2022)
12. Geiping, J., Bauermeister, H., Dröge, H., et al.: Inverting gradients - how easy is it to break privacy in federated learning? In: Proceedings of the 34th International Conference on Neural Information Processing Systems, pp. 16937–16947. Curran Associates Inc. (2020)
13. Gu, H., Zhang, X., Li, J., et al.: Federated learning vulnerabilities: Privacy attacks with denoising diffusion probabilistic models. In: Proceedings of the ACM Web Conference 2024, pp. 1149–1157. ACM (2024) https://doi.org/10.1145/3589334.3645514

14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Proceedings of the 34th International Conference on Neural Information Processing Systems, pp. 574–585. Curran Associates Inc. (2020).
15. Huang, Y., Gupta, S., Song, Z., Li, K., Arora, S.: Evaluating gradient inversion attacks and defenses in federated learning. In: Proceedings of the 35th International Conference on Neural Information Processing Systems, pp. 7232–7241. Curran Associates Inc. (2021)
16. Jeon, J., Lee, K., Oh, S., Ok, J., et al.: Gradient inversion with generative image prior. In: Proceedings of the 35th International Conference on Neural Information Processing Systems, pp. 29898-–29908. Curran Associates Inc. (2021)
17. Li, Z., Zhang, J., Liu, L., Liu, J.: Auditing privacy defenses in federated learning via generative gradient leakage. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10132-–10142. IEEE (2022) https://doi.org/10.1109/CVPR52688.2022.00989.
18. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: Proceedings of the 38th International Conference on Machine Learning, pp. 8162–8171. PMLR (2021)
19. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., et al.: The future of digital health with federated learning. NPJ Digital Medicine **3**(1), 1–7 (2020)
20. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., et al.: Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports **10**(1), 1–12 (2020)
21. Tedeschini, B. C., Savazzi, S., Stoklasa, R., et al.: Decentralized federated learning for healthcare networks: A case study on tumor segmentation. IEEE Access **10**, 8693–8708 (2022)
22. Wang, X., et al.: ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3462—3471. IEEE (2017) https://doi.org/10.1109/CVPR.2017.369
23. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
24. Wu, L., Liu, Z., Pu, B., et al.: DGGI: Deep Generative Gradient Inversion with diffusion model. Information Fusion **113**, 102620 (2025)
25. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification. Scientific Data **10**(1), 41 (2023)
26. Yang, L., Ding, S., Cai, Y., Yu, J., Wang, J., Shi, Y.: Guidance with Spherical Gaussian Constraint for Conditional Diffusion. In: Proceedings of the 41st International Conference on Machine Learning, pp. 56071–56095. JMLR (2024)
27. Zhang, R., Guo, S., Wang, J., Xie, X., Tao, D.: A survey on gradient inversion: attacks, defenses and future directions. In: Proceedings of the 31st International Joint Conference on Artificial Intelligence, pp. 5678–5685. International Joint Conferences on Artificial Intelligence Organization (2022) https://doi.org/10.24963/ijcai.2022/791
28. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 586–595. IEEE (2018) https://doi.org/10.1109/CVPR.2018.00068

29. Zhao, B., Mopuri, K. R., Bilen, H.: IDLG: Improved deep leakage from gradients. arXiv preprint (2020) https://doi.org/10.48550/arXiv.2001.02610
30. Zhao, J. C., Sharma, A., Elkordy, A. R., Ezzeldin, Y. H., Avestimehr, S., Bagchi, S.: Loki: Large-scale Data Reconstruction Attack against Federated Learning through Model Manipulation. In: Proceedings of the 2024 IEEE Symposium on Security and Privacy, pp. 1287–1305. IEEE (2024) https://doi.org/10.1109/SP54263.2024.00030
31. Zhu, L., Liu, Z., Han, S.: Deep leakage from gradients. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 14774–14784. Curran Associates Inc. (2019)