Embedding-Based Federated Data Sharing via Differentially Private Conditional VAEs

Francesco Di Salvo* (✉), Hanh Huyen My Nguyen*, and Christian Ledig

xAILab Bamberg, University of Bamberg, Germany
francesco.di-salvo@uni-bamberg.de

Abstract. Deep Learning (DL) has revolutionized medical imaging, yet its adoption is constrained by data scarcity and privacy regulations, limiting access to diverse datasets. Federated Learning (FL) enables decentralized training but suffers from high communication costs and is often restricted to a single downstream task, reducing flexibility. We propose a data-sharing method via Differentially Private (DP) generative models. By adopting foundation models, we extract compact, informative embeddings, reducing redundancy and lowering computational overhead. Clients collaboratively train a Differentially Private Conditional Variational Autoencoder (DP-CVAE) to model a global, privacy-aware data distribution, supporting diverse downstream tasks. Our approach, validated across multiple feature extractors, enhances privacy, scalability, and efficiency, outperforming traditional FL classifiers while ensuring differential privacy. Additionally, DP-CVAE produces higher-fidelity embeddings than DP-CGAN while requiring 5× fewer parameters.

Keywords: Federated learning · Generative model · Differential privacy.

1 Introduction

Deep Neural Networks (DNNs) have driven remarkable advancements in medical imaging, yet their adoption in clinical practice remains constrained by limited data availability and stringent privacy requirements [26]. Medical datasets are siloed across institutions, and low-prevalence diseases further limit the availability of diverse, high-quality training data [18]. While collaborative data sharing could mitigate these challenges [25], strict privacy regulations (*e.g.*, HIPAA, GDPR) make centralized dataset aggregation infeasible. To address these constraints, Federated Learning (FL) [20] has emerged as a privacy-preserving alternative, allowing institutions to collaboratively train models without sharing raw data. A widely adopted strategy, FedAvg [20], aggregates model updates from participating clients to construct a global model. However, FL introduces several challenges. Communication overhead remains high, particularly when deploying deep architectures such as Vision Transformers (ViTs) [6], which significantly increases transmission costs. Furthermore, FL is typically restricted to a single downstream task (*e.g.,* classification, segmentation), limiting generalizability.

[*] Equal contribution. Code available at github.com/myng15/federated-dp-cvae.

To mitigate transmission costs, most FL research focuses on lightweight architectures, which often come at the expense of model performance and robustness. An alternative to model-sharing is data-sharing via privacy-preserving synthetic data generation, which reduces communication overhead while enabling broader downstream applications [12,14]. Several works have explored federated generative models for this purpose (see [8] for a comprehensive review). However, generating realistic, task-relevant synthetic data remains challenging, as it often requires a large number of diverse training samples to ensure fidelity to the original distribution. While Generative Adversarial Networks (GANs) [9] and diffusion models [10] achieve high-fidelity image synthesis, they exhibit notable limitations [12]. GANs suffer from mode collapse, producing low-diversity synthetic data, while diffusion models are computationally expensive and exhibit high latency, making them impractical for resource-constrained federated environments. In contrast, Variational Autoencoders (VAEs) and Conditional VAEs (CVAEs), despite producing lower-fidelity images, *e.g.,* blurred reconstructions, offer notable advantages. In fact, they avoid mode collapse while being more computationally efficient than GANs and diffusion models.

While VAEs and CVAEs have been explored in federated settings, prior work has primarily applied them to simpler generative tasks, such as MNIST-like datasets [24], sensor data [11], or joint training with a downstream classifier [2], limiting their adaptability. A recent work has demonstrated that generating synthetic feature embeddings using a CVAE preserved classification performance, comparable to real embeddings, while enhancing data privacy [5]. A key enabler of this approach is the use of foundation models [22], known to be robust to domain shifts [23]. Furthermore, these models produce compact and diagnostically relevant feature representations while reducing redundancy in raw images and enabling low-cost downstream learning. Training a CVAE on feature embeddings rather than raw images allows to better capture feature distributions, making it less susceptible to fidelity degradation. This motivates our extension of CVAE to a federated setting, where, as illustrated in Figure 1, clients collaboratively train a differentially-private global generative model. Unlike prior FL settings, our approach decouples generative modeling from task-specific constraints, allowing greater flexibility across applications. In summary, our contributions are:

- We propose a lightweight federated generative model with differential privacy to address data scarcity and enable privacy-preserving data sharing in medical image analysis. Our approach decouples data-sharing from downstream tasks, enhancing generalizability and adaptability across applications.
- We empirically demonstrate that our federated generative approach and subsequent downstream training outperform traditional federated classifiers across multiple datasets, achieving higher balanced accuracy.
- We show that training a lightweight CVAE on feature embeddings achieves higher fidelity than GAN-based approaches while requiring approximately 5× fewer parameters, significantly improving computational efficiency.

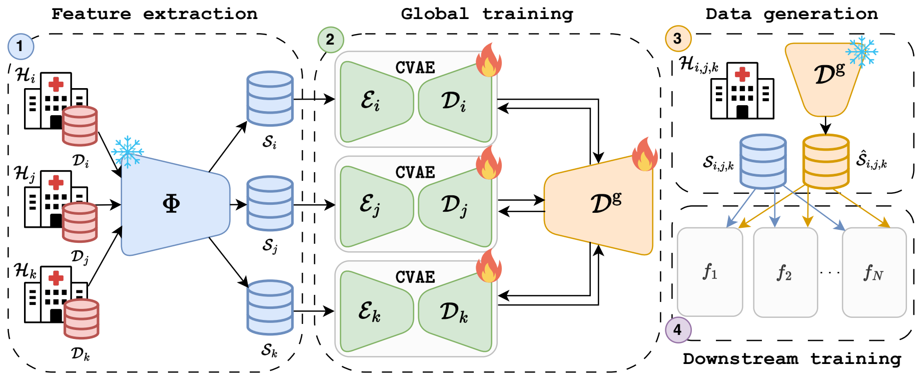

Fig. 1. Illustration of our proposed methodology. (1) Each client $\mathcal{H}$ encodes its image-based dataset $\mathcal{D}$ into an embedding-based dataset $\mathcal{S}$ using a large, pre-trained foundation model Φ, reducing data storage requirements and computational overhead. (2) Clients collaboratively train a lightweight DP-CVAE $(\mathcal{E}, \mathcal{D})$ and periodically share decoder weights, which are aggregated into a global decoder $\mathcal{D}^{\mathrm{g}}$. This shared decoder captures cross-client variation while preserving local data privacy. (3) Each client independently generates a synthetic dataset $\hat{\mathcal{S}}$ using the globally trained generative model, and (4) utilizes (real) local and (synthetic) global data for any downstream task f.

2 Method

2.1 Feature extraction

Traditional FL pipelines often use lightweight architectures [27,28] to mitigate communication overhead and computational demands. However, this may compromise robustness and generalization, which are essential in medical image analysis for capturing high-quality feature representations. To address these limitations, we utilize a (shared) large pre-trained foundation model to extract compact, informative feature embeddings while substantially reducing the inherent information redundancy of raw images. While any feature extractor can be used, in this work, we adopt DINOv2 Base [22] due to its state-of-the-art performance across different domains and tasks. Considering a federation of M clients, each client m holds a private image-based dataset $\mathcal{D}_m := \{(\mathbf{d}_i^m, y_i^m)\}_{i=1}^{n_m}$, where $\mathbf{d}_i^m$ represents the i-th raw image and $y_i^m \in \mathcal{Y}$ is its corresponding label. After applying feature extraction locally using a pre-trained foundation model, the raw images are transformed into feature embeddings, resulting in the feature-based dataset $\mathcal{S}_m := \{(\mathbf{x}_i^m, y_i^m)\}_{i=1}^{n_m}$ where $\mathbf{x}_i^m \in \mathcal{X}_m$ is the feature embedding of $\mathbf{d}_i^m$.

2.2 Federated- and differentially-private generative model

CVAE A recent study [5] demonstrated that generating synthetic training sets at the embedding level, rather than raw images, introduces privacy preservation while observing only minimal performance degradation in downstream classification. Following this approach, we adopt a CVAE with a symmetric architecture, consisting of three linear layers for both the class-conditional encoder and decoder. The model is trained by minimizing a reconstruction loss (Mean Squared Error) while enforcing latent space normality through a Kullback-Leibler (KL) divergence loss with a standard normal prior.

Differential Privacy (DP) To enforce formal privacy guarantees, we integrate Differential Privacy [1,7], ensuring that the synthetic embeddings remain indistinguishable with respect to the presence or absence of any single data point in the original dataset. This is achieved by bounding the sensitivity of the generative model to individual samples, thereby preventing adversaries from reconstructing or inferring specific data points from the released synthetic dataset. DP-CVAE ensures that the posterior distribution of the generated embeddings remains statistically similar irrespective of whether a given sample is included in the training data. We enforce (ϵ, δ)-differential privacy by introducing calibrated noise into the generative process. Formally, a mechanism $\mathcal{M}$ applied to the private dataset $\mathcal{S}_m$ is (ϵ, δ)-differentially private if, for any two neighboring datasets $\mathcal{S}_m$ and $\mathcal{S}'_m$ differing by at most one sample, and for any possible output $\hat{\mathcal{S}}_m$:

$$\Pr[\mathcal{M}(\mathcal{S}_m) \in \hat{\mathcal{S}}_m] \leq e^{\epsilon} \Pr[\mathcal{M}(\mathcal{S}'_m) \in \hat{\mathcal{S}}_m] + \delta, \quad \epsilon > 0, \delta \in [0, 1) \tag{1}$$

where (ϵ, δ) is an upper bound on the privacy loss between before and after an individual was added to the dataset, giving us a formal privacy guarantee. To achieve this, we integrate DP-SGD [1] into the CVAE training process, where noise $\mathcal{N}(0, \sigma^2)$ is added to the *per-sample* gradients, ensuring that each model update is differentially private. Before adding noise, we clip gradients so that they have a maximum ℓ_2 norm of C to limit the influence of a single data point. Notably, each client performs DP locally on their CVAE before sending the local decoder to the server to update the global CVAE decoder. Any outputs of the resulting CVAE perturbed by the DP noise are guaranteed to protect an individual's data used in training according to the chosen ϵ. However, this noise comes at the cost of utility loss, which grows as ϵ decreases.

Model aggregation and data generation Each client maintains a personalized encoder, which adapts to its local data distribution, mapping samples into a shared latent space. Meanwhile, the decoders are jointly trained using a Federated Averaging (FedAvg) strategy, ensuring that the aggregated decoder learns a globally representative feature reconstruction function while remaining compact and efficient. Formally, at each communication round t, the server aggregates the decoder weights θ^m_{dec} from all M participating clients as follows:

$$\theta^{(t+1)}_{\text{dec}} = \sum_{m=1}^{M} w_m \theta^{m,(t)}_{\text{dec}}, \quad w_m = \frac{n^{\text{train}}_m}{\sum_{m=1}^{M} n^{\text{train}}_m} \tag{2}$$

where $\theta_{\text{dec}}^{m,(t)}$ represents the decoder parameters of client m at round t, and $\theta_{\text{dec}}^{(t+1)}$ represents the updated global decoder parameters distributed back to the clients. Once the federated training of the global decoder is completed, each client utilizes it to generate a synthetic dataset that approximates the global data distribution while ensuring privacy preservation. Given a target synthetic dataset size N, each client constructs its global dataset $\hat{\mathcal{S}}_m$ as follows:

$$\hat{\mathcal{S}}_m = \{(\hat{\mathbf{x}}_i, \hat{y}_i)\}_{i=1}^{N}, \qquad \hat{\mathbf{x}}_i = \text{Dec}_{\theta_{\text{dec}}^{(t+1)}}(\mathbf{z}_i \mid \hat{y}_i) \tag{3}$$

where $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is sampled from a standard normal Gaussian distribution, $\hat{y}_i \sim C$ is sampled from a selected class distribution, and $\text{Dec}_{\theta_{\text{dec}}^{(t+1)}}$ is the globally trained decoder obtained after FedAvg aggregation.

2.3 Downstream classification

In our setup, each client has access to both a local (private) dataset and a global dataset, the latter generated using the shared global decoder. This setup enables clients to train models tailored to their specific downstream tasks. For instance, clients can utilize the same synthetic data to train classifiers with different label granularities, model data distributions for anomaly detection, address out-of-distribution (OOD) detection problems, and more.

We consider image classification as the primary downstream task for a standardized evaluation. The availability of both local and global datasets enables *personalized* FL, in which each client trains a personalized model while still benefiting from other clients' knowledge. For example, kNN-Per [19] uses client-level local memorization to improve individual performance but still jointly learns the global representations, and FedRep [4] aims to learn a shared feature representation across clients. In contrast, we take advantage of the shared feature representations of a foundation model and train for each client one model on the local data and another on the global (synthetic) data for the downstream task (instead of for representation learning). The final classification prediction is obtained through a weighted interpolation between these two models, controlled by a tunable parameter λ_m, which is optimized based on the validation set to balance local specialization and global generalization. For a test sample $\mathbf{x}_{\text{test}}$, the interpolated probability distribution is computed as:

$$P_{\lambda_m}(y \mid \mathbf{x}_{\text{test}}) = \lambda_m P_{\text{local}}(y \mid \mathbf{x}_{\text{test}}) + (1 - \lambda_m) P_{\text{global}}(y \mid \mathbf{x}_{\text{test}}) \tag{4}$$

where $\lambda_m \in [0, 1]$ determines the trade-off between personalization and global knowledge. The final predicted class $\hat{y}_{\text{test}}$ is given by:

$$\hat{y}_{\text{test}} = \arg\max_{c \in \mathcal{Y}} P_{\lambda_m}(y \mid \mathbf{x}_{\text{test}}) \tag{5}$$

Intuitively, as $\lambda_m \to 1$, the model prioritizes local data, utilizing client-specific knowledge. Conversely, as $\lambda_m \to 0$, the model relies more on globally shared information, benefiting from knowledge aggregated across the federation.

3 Experimental results

3.1 Experimental settings

Datasets and metrics We evaluate the downstream classification performance of our method across two distinct classification settings. Following [17], we use the Abdominal CT dataset (Sagittal view) [29], utilizing the splits from [30]. It presents $25,211$ images across 11 classes. We distribute the data among 10 clients, simulating IID conditions, together with highly non-IID conditions using a Dirichlet distribution ($\alpha = 0.3$). Furthermore, as in [3], we use a subset of 4,600 images from Camelyon17-Wilds [13], a binary dataset consisting of histopathological images from five hospitals, treated as five individual clients. Each local dataset is further split into train–val–test (60:20:20). While Camelyon17-Wilds is class-balanced, the CT dataset is imbalanced, therefore, we report the overall mean and standard deviation across clients of accuracy and balanced accuracy.

Implementation details Both the federated and locally trained classifiers are implemented as single-layer linear models, a standard approach for evaluating foundation model embeddings [22]. Both the FedAvg classifier and our DP-CVAE are trained for 50 communication rounds, with 5 local epochs per round, using the SGD optimizer with a learning rate of $1{\times}10^{-3}$. We apply DP to our CVAE and, for comparison, to a Conditional GAN (CGAN), using the OPACUS library [31], with $(\epsilon, \delta) = (1.0, 1 \times 10^{-4})$ and a clipping norm of 1.5. This choice, with $\epsilon \leq 1$ and $\delta \ll 1/n$, where n is the average number of training samples per client in our experiments, ensures meaningful privacy guarantees while maintaining data utility [15,21]. The interpolation parameter λ_m (*c.f.* Equation 4) is automatically selected from $\{0.0, 0.1, \ldots, 1.0\}$ based on validation set performance. Lastly, our downstream classifiers are trained separately on local and global data, using Adam optimizer with a learning rate of $1{\times}10^{-3}$ for 100 epochs. For comparison, we evaluate FedAvg and its enhanced version, FedProx [16], which addresses non-IID data by adding a regularization term that penalizes large deviations from the global model. As a direct competitor to our approach, we include “FedLambda”, our adapted version of kNN-Per [19], where the local kNN component is replaced with a local linear classifier and the global classifier is trained through FedAvg.

3.2 Downstream classification

Table 1 presents the classification results across different experimental settings. Overall, federated data-sharing schemes using generative models outperform federated classifiers, with DP-CVAE achieving the highest performance in most cases. Notably, incorporating the local predictions under personalized FL settings leads to a substantial improvement over standard FedAvg, highlighting the benefits of a personalized adaptation. Furthermore, it is important to note that these federated classifiers (including FedLambda) are not differentially private, whereas our approach enforces differential privacy guarantees, which may have a slight impact on performance. In the heterogeneous CT setting ($\alpha = 0.3$), our

Table 1. Accuracy (ACC) and balanced accuracy (BACC), averaged across clients over three seed runs, between federated classifiers and our proposed data-sharing and classification method. We highlight in **bold** the top two methods.

	CT (IID)		CT ($\alpha = 0.3$)		Camelyon
	ACC	BACC	ACC	BACC	ACC
FedAvg	73.33±1.14	67.00±1.15	64.85±5.88	**58.74±2.93**	90.62±2.34
FedProx	73.26±1.19	66.95±1.17	64.76±6.13	58.66±5.84	90.65±2.01
FedLambda	77.21±0.80	71.32±0.91	81.03±3.83	**59.08±2.54**	92.54±1.03
DP-CGAN	**77.60±1.36**	**71.94±1.18**	**88.84±1.99**	57.49±2.41	**93.95±0.98**
DP-CVAE	**77.54±0.76**	**71.84±0.82**	**88.94±1.35**	57.58±3.33	**94.49±1.28**

method achieves comparable or slightly lower balanced accuracy than federated classifiers, yet it substantially outperforms them in terms of accuracy. The most notable performance gap between our generative-based approach and federated classifiers is observed on Camelyon17, where each client has an average of solely 500 train samples. This underscores that our generative method remains highly effective even under limited-data conditions. Furthermore, while DP-CGAN and DP-CVAE achieve comparable performance overall, DP-CVAE notably outperforms DP-CGAN on Camelyon17, suggesting that CVAE is less data-hungry, making it particularly suited for limited-data scenarios. Lastly, the parameter λ_m, which weighs the local model predictions per client, yields the best accuracy in the range $0.4 - 0.7$. This reflects a balanced mix of local and global data, thereby empirically confirming the high utility of the synthetic data.

3.3 Ablation studies

Choice of the generative model This ablation study compares DP-CVAE and DP-CGAN in their computational efficiency and ability to preserve the original data distribution. We compute the average Wasserstein distance ($\mathcal{W}$) between each client's real dataset and the synthetic dataset they generate using the globally trained DP-CGAN or DP-CVAE decoder. As shown in Figure 2, DP-CVAE consistently achieves a lower $\mathcal{W}$, indicating better fidelity. Notably, DP-CVAE requires 5× fewer parameters than DP-CGAN, enhancing efficiency and reducing latency, making it more suitable for FL.

Generalizability and robustness Utilizing CT (IID) as a reference, we demonstrate in Figure 3 that our approach generalizes effectively across different (Small) backbones, including DINOv2, DINO, and ViT, maintaining consistent classification performance. Additionally, as the number of clients increases, leading to fewer samples per client, it exhibits a minimal performance drop. This suggests that our approach is data-efficient and remains robust across varied feature representations and dataset sizes.

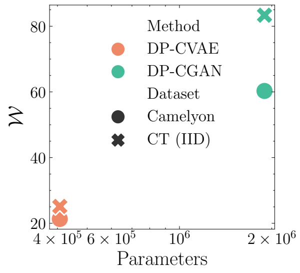

Fig. 2. Reconstruction fidelity, measured by the Wasserstein distance ($\mathcal{W} \downarrow$) between real and synthetic samples, comparing DP-CVAE and DP-CGAN. Results are analyzed in relation to the number of model parameters.

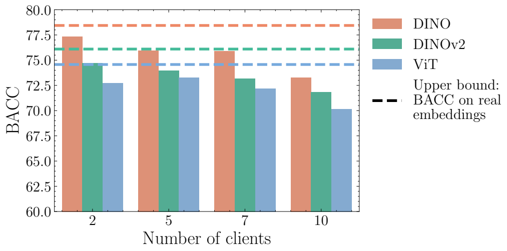

Fig. 3. Generalizability of our method across different backbones and increasing numbers of clients (*i.e.*, fewer samples per client), utilizing CT (IID). The dashed lines represent the BACC obtained when training a linear classifier on real image embeddings with the original train–val–test split.

4 Discussion

Limitations and future work Although generative models trained on embedding spaces are more efficient and less data-demanding, they remain susceptible to data and label imbalance, a well-known challenge in federated learning. This limitation can be addressed by incorporating techniques from long-tailed learning [32], enhancing robustness across imbalanced distributions. Furthermore, our current generative model samples embeddings with a fixed unit variance. Introducing learned or class-specific variance parameters could improve the quality and expressiveness of synthetic embeddings, leading to better downstream performance. Finally, conditioning the generative model on additional confounders (*e.g.*, domain-specific attributes) could further enhance data diversity and mitigate inherent biases in the training distribution, improving generalization and fairness in real-world applications.

Conclusions This work introduces a federated learning method that shifts from traditional (downstream) model-sharing to privacy-preserving data-sharing, utilizing DP-CVAEs trained on foundation model embedding spaces. Unlike conventional FL approaches that are constrained to a single downstream task, our method enables flexible and adaptive synthetic data generation, allowing clients to tailor their datasets for diverse applications. Through comprehensive experiments on medical imaging datasets, we demonstrated that our approach outperforms federated classifiers, achieving substantially higher accuracy. Additionally, we showed that training a lightweight CVAE on feature embeddings preserves data fidelity more effectively than GANs, while requiring significantly fewer parameters. These findings position our approach as a promising alternative for enabling secure, flexible, and high-performance FL in medical image analysis.

Acknowledgments. This study was funded through the Hightech Agenda Bayern (HTA) of the Free State of Bavaria, Germany.

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

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (2016)
2. Chen, B., Li, H., Guo, L., Wang, L.: Label-wise distribution adaptive federated learning on non-iid data. In: 2023 IEEE Wireless Communications and Networking Conference (WCNC). pp. 1–6 (2023)
3. Chen, M., Jiang, M., Dou, Q., Wang, Z., Li, X.: Fedsoup: improving generalization and personalization in federated learning via selective model interpolation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 318–328. Springer (2023)
4. Collins, L., Hassani, H., Mokhtari, A., Shakkottai, S.: Exploiting shared representations for personalized federated learning. In: International conference on machine learning. pp. 2089–2099. PMLR (2021)
5. Di Salvo, F., Tafler, D., Doerrich, S., Ledig, C.: Privacy-preserving datasets by capturing feature distributions with conditional vaes. The 35th British Machine Vision Conference (2024)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
7. Dwork, C.: Differential privacy. In: Proceedings of the 33rd International Colloquium on Automata, Languages and Programming. pp. 1–12. Springer Berlin Heidelberg (2006)
8. Gargary, A.V., De Cristofaro, E.: A systematic review of federated generative models. arXiv preprint arXiv:2405.16682 (2024)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
11. Kaspour, S., Yassine, A.: Variational auto-encoder model and federated approach for non-intrusive load monitoring in smart homes. In: 2023 IEEE Symposium on Computers and Communications (ISCC). pp. 1110–1115 (2023)
12. Koetzier, L.R., Wu, J., Mastrodicasa, D., Lutz, A., Chung, M., Koszek, W.A., Pratap, J., Chaudhari, A.S., Rajpurkar, P., Lungren, M.P., et al.: Generating synthetic data for medical imaging. Radiology **312**(3), e232471 (2024)
13. Koh, P.W., Sagawa, S., Marklund, H., Xie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R.L., Gao, I., et al.: Wilds: A benchmark of in-the-wild distribution shifts. In: International conference on machine learning (2021)
14. Ktena, I., Wiles, O., Albuquerque, I., Rebuffi, S.A., Tanno, R., Roy, A.G., Azizi, S., Belgrave, D., Kohli, P., Cemgil, T., et al.: Generative models improve fairness of medical classifiers under distribution shifts. Nature Medicine **30**(4) (2024)

15. Lange, L., Schneider, M., Christen, P., Rahm, E.: Privacy in practice: Private covid-19 detection in x-ray images (extended version). arXiv:2211.11434 (2022)
16. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
17. Li, X., Zhang, W., Yu, Y., Zheng, W.S., Zhang, T., Wang, R.: Sift: A serial framework with textual guidance for federated learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2024)
18. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
19. Marfoq, O., Neglia, G., Vidal, R., Kameni, L.: Personalized federated learning through local memorization. In: International Conference on Machine Learning. pp. 15070–15092. PMLR (2022)
20. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)
21. Nasr, M., Song, S., Thakurta, A., Papernot, N., Carlini, N.: Adversary instantiation: Lower bounds for differentially private machine learning. In: 2021 IEEE Symposium on Security and Privacy (S&P). IEEE (2021)
22. Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., HAZIZA, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: DINOv2: Learning robust visual features without supervision. Transactions on Machine Learning Research (2024)
23. Paul, S., Chen, P.Y.: Vision transformers are robust learners. In: Proceedings of the AAAI conference on Artificial Intelligence. vol. 36, pp. 2071–2081 (2022)
24. Pfitzner, B., Arnrich, B.: Dpd-fvae: Synthetic data generation using federated variational autoencoders with differentially-private decoder. arXiv:2211.11591 (2022)
25. Shilo, S., Rossman, H., Segal, E.: Axes of a revolution: challenges and promises of big data in healthcare. Nature medicine **26**(1), 29–38 (2020)
26. Stacke, K., Eilertsen, G., Unger, J., Lundström, C.: Measuring domain shift for deep learning in histopathology. IEEE journal of biomedical and health informatics **25**(2), 325–336 (2020)
27. Wu, N., Yu, L., Yang, X., Cheng, K.T., Yan, Z.: Fediic: Towards robust federated learning for class-imbalanced medical image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2023)
28. Xia, Y., Ma, B., Dou, Q., Xia, Y.: Enhancing federated learning performance fairness via collaboration graph-based reinforcement learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 263–272. Springer (2024)
29. Xu, X., Zhou, F., Liu, B., Fu, D., Bai, X.: Efficient multiple organ localization in ct image using 3d region proposal network. IEEE transactions on medical imaging **38**(8), 1885–1898 (2019)
30. 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)
31. Yousefpour, A., Shilov, I., Sablayrolles, A., Testuggine, D., Prasad, K., Malek, M., Nguyen, J., Ghosh, S., Bharadwaj, A., Zhao, J., Cormode, G., Mironov, I.: Opacus: User-friendly differential privacy library in PyTorch. arXiv:2109.12298 (2021)

32. Zhang, Y., Kang, B., Hooi, B., Yan, S., Feng, J.: Deep long-tailed learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)