Enhancing the Utility of Privacy-Preserving Cancer Classification using Synthetic Data

Richard Osuala[1,2,3], Daniel M. Lang[2,3], Anneliese Riess[2,3], Georgios Kaissis[2,3,4], Zuzanna Szafranowska[1], Grzegorz Skorupko[1], Oliver Diaz[1,5], Julia A. Schnabel[2,3,6], and Karim Lekadir[1,7]

[1] Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Spain
richard.osuala@ub.edu
[2] Helmholtz Center Munich, Munich, Germany
[3] Technical University of Munich, Munich, Germany
[4] Imperial College London, London, United Kingdom
[5] Computer Vision Center, Bellaterra, Spain
[6] Kings College London, London, UK
[7] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract. Deep learning holds immense promise for aiding radiologists in breast cancer detection. However, achieving optimal model performance is hampered by limitations in availability and sharing of data commonly associated to patient privacy concerns. Such concerns are further exacerbated, as traditional deep learning models can inadvertently leak sensitive training information. This work addresses these challenges exploring and quantifying the utility of privacy-preserving deep learning techniques, concretely, (i) differentially private stochastic gradient descent (DP-SGD) and (ii) fully synthetic training data generated by our proposed malignancy-conditioned generative adversarial network. We assess these methods via downstream malignancy classification of mammography masses using a transformer model. Our experimental results depict that synthetic data augmentation can improve privacy-utility tradeoffs in differentially private model training. Further, model pretraining on synthetic data achieves remarkable performance, which can be further increased with DP-SGD fine-tuning across all privacy guarantees. With this first in-depth exploration of privacy-preserving deep learning in breast imaging, we address current and emerging clinical privacy requirements and pave the way towards the adoption of private high-utility deep diagnostic models. Our reproducible codebase is publicly available at https://github.com/RichardObi/mammo_dp.

Keywords: Differential Privacy · Generative Models · Breast Imaging

1 Introduction

Breast cancer accounts for staggering estimates of 684.000 deaths and 2,26 million new cases worldwide per year [11]. Part of this burden could be reduced

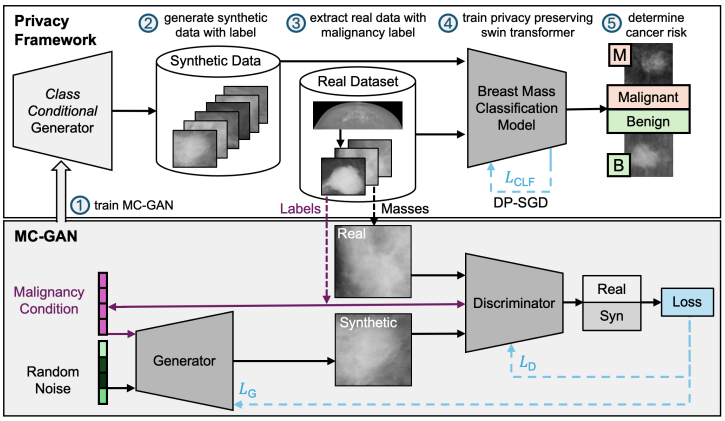

Fig. 1: Overview of our privacy-preserving deep learning pipeline and malignancy-conditioned generative adversarial network (MCGAN).

through earlier detection and timely treatment. Screening mammography is a cornerstone for early detection and further associated with a reduction in breast cancer mortality [21]. Recent literature emphasizes the potential of deep learning-based computer-aided diagnosis (CAD) [30,24,15,22], e.g., demonstrating that a symbiosis of deep learning models with radiologist assessment yields the highest breast cancer detection performances [21]. However, training deep learning models on patient data poses a risk of leakage of sensitive person-specific information during and after training [24], as models have the capacity to memorise sufficient information to allow for high-fidelity image reconstruction [3,13]. To avoid such leakage of private patient information, data needs to be protected during model training, in particular when the objective is to develop models to be used in clinical practice or shared among entities. Furthermore, international data protection regulations grant patients the right to request the removal of their information from data holders. For instance, point (b) of article 17(1) of the EU General Data Protection Regulation (GDPR) [9] stipulates that data subjects have a "right to be forgotten". Given, for instance, the proven possibility of reconstructing training data given a model's weights [3,13], these rights can extend to the removal of patient-specific information from already trained deep learning models [29]. However, it is known to be difficult to "reliably" and "provably" remove patient information — present in only one or few specific training data points — from already trained model weights [29]. A generic and verifiable alternative is given by the removal of a patient's data point from the training data and retraining of the respective model with the reminder of the dataset. This procedure is not only likely to have negative impacts on the performance of algorithms, but also emerges as a deterrence and risk for hospitals to adopt deep learning models, due to extensive economic, organisational, and environmental costs caused by retraining. Anticipating patient consent withdrawals, costly retraining can be avoided by demonstrating that deep learning model weights do

not include personally identifiable information (PII) about any specific patient. To this end, a powerful technique to ensure privacy during model training is given by Differentially Private Stochastic Gradient Descent (DP-SGD)[1], which quantifiably reduces the effect each single training sample can have on the resulting model weights. Furthermore, privacy-preservation can also be achieved by diagnostic models exclusively trained on synthetic data, which is not (unambiguously) attributable to any specific patient but rather contains anonymous samples representing the essence of the dataset [12,24]. The caveat of both DP-SGD and synthetic data strategies is, however, that they generally lead to a reduction in model performance, known as the privacy-utility trade-off. Investigating this trade-off in the realm of breast imaging, our core contributions are summarised as follows:

- We design and validate a transformer model, achieving promising performance as a backbone for privacy-preserving breast mass malignancy classification.
- We propose and validate a conditional generative adversarial network capable of differentiating between benign and malignant breast mass generation.
- We empirically quantify privacy-utility-tradeoffs in mass malignancy classification, assessing various differential privacy guarantees, and further combine and compare them with training on synthetic data.

2 Methods and Materials

Datasets and Preprocessing

We use the open-access Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) dataset [16], which consists of 891 scanned film mammography cases with segmented masses with biopsy-proven malignancy status. After extracting mass images from craniocaudal view (CC) and mediolateral oblique (MLO) views, we follow the predefined per-patient train-test split [16], allocating 1296 mass images for training and 402 (245 benign, 157 malignant) mass images to testing. We further divided this training set randomly per-patient into a training (1104 mass images, 525 malignant) and a validation set (192 mass images, 102 malignant). As external test set, we further adopt the publicly available BCDR cohort [19], which comprises 1010 patients, totalling 1493 lesions (639 masses) with contours and biopsy information from both digital mammograms (BCDR-DM) and film mammograms (BCDR-FM). Our final BCDR test set contains 1106 mass images extracted from CC and MLO views, 486 of which are malignant and 620 benign. To obtain mass patches from the mammograms, the lesion contour information is used to define bounding boxes, which enclose the mass. We then create a square patch around each bounding box with a minimum length and width of 128 pixels. Next, we increase the patch size by adding a margin of 60 pixels in each direction, before extracting the resulting patch, and resizing it to a pixel dimensions of 128x128 using inter-area interpolation. For classification, the mass patches are further resized to

224x224px maintaining image ratios, and stacked to 3 channels. Models were trained on either a single 8GB NVIDIA RTX 2080 Super or 48GB RTX A6000 GPU using PyTorch and opacus [31] for DP-SGD.

Cancer Classification Transformer Model

Given its reported high performance on classifying the presence of a lesion in mammography patches [30] and its shifted window mechanism, allowing to effectively attend to shapes of varying sizes, we adopt a swin transformer (Swin-T) [17] as cancer classification model, to distinguish between benign and malignant masses. We inititalize ImageNet-pretrained [6] network weights and, after following the Swin-T hyperparameter setup [17] (stride, window size), we adjust the last fully-connected layer of the swin transformer reinitializing it with two output nodes each one outputting the logits for one of our respective classes (i.e., malignant or benign). We only set the parameters of the adjusted fully-connected layer as trainable and apply a learning rate of 1e-5. A weight decay of 1e-8 is used following the fine-tuning experiment described in [17]. Furthermore, an adamw optimizer [20], label smoothing of 0.1, and a batch size of 128 are used. During training, random horizontal and vertical flips are applied as data augmentation and a cross entropy loss is backpropagated. Training for 300 epochs using a cross entropy loss function, the model from the epoch with the lowest area under the precision-recall curve (AUPRC) on the validation set is selected for testing.

Malignancy-Conditioned Generative Adversarial Network

Going beyond unconditional mass synthesis in the literature [30,2], we propose a malignancy conditioned generative adversarial network (MCGAN) to control the generation of either benign or malignant synthetic breast masses. In general, GANs consist of a generator (G) and a discriminator (D) network, which engage in a two-player zero-sum game, where G generates synthetic samples that D strives to distinguish from real ones [12]. We design G and D as deep convolutional neural networks [27] and, as shown in Fig. 1, integrate class-conditional information [23]. To this end, we extract the histopathology report's biopsy information for each mass from the metadata, and convert it into a discrete malignancy label. Then, we transform this label into a multi-dimensional embedding vector to either represent the (a) malignant or (b) benign class, before passing it through a fully-connected layer yielding a representation with the corresponding dimensionality to concatenate it to the generator input (100 dim noise vector) and to the discriminator input (128x128 input image). As D learns to associate class labels with patterns in the input images, it has to learn whether or not a given class corresponds to a given synthetic sample. Furthermore, as the discriminator loss is backpropagated into the generator, G is forced to synthesize samples corresponding to the provided class condition. This results in G learning a conditional distribution based on the value function

$$\min_G \max_D V(D,G) = \min_G \max_D [\mathbb{E}_{x\sim p_{\text{data}}}[\log D(x|y)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z|y)))]].$$

Optimizing the discriminator via binary cross-entropy [12], we define its loss in a class-conditional setup as

$$L_{D_{\mathrm{MCGAN}}} = -\mathbb{E}_{x\sim p_{\mathrm{data}}}[\log D(x|y)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z|y)))].$$

We train our MCGAN on the CBIS-DDSM training data, applying random horizontal (p=0.5) and vertical (p=0.5) flipping as well as random cropping with resizing, where the resize scale ranges from 0.9 to 1.1 and aspect ratio from 0.95 to 1.1. We further include one-sided label smoothing [27] in a range of [0.7, 1.2]. Following [2], we employ a discriminator convolutional kernel size of 6 and a generator kernel size of 4. We observe that this reduces checkerboard artefacts as D's field-of-view now requires G to create realistic transitions between the kernel-sized patches in the image. MCGAN is trained for 10k epochs with a batch size of 16. Based on the best quality-diversity tradeoff, we select the model from epoch 1.4k after qualitative visual assessment of generated samples .

Patient Privacy Preservation Framework

Privacy protection is an ethical norm and legal obligation, e.g. granting patients the right of their (retrospective) removal from databases [9]. Since (biomedical) deep learning models are vulnerable to information leakage, e.g. sensitive patient attributes [29,3,13], they can be affected by such (and future) regulations. However, privacy-preserving techniques can be integrated into deep learning frameworks and, to some extent, avoid compromising confidential data. For instance, (i) model training with DP-SGD [1] or (ii) training exclusively on synthetic data.

From a legal perspective, models trained on only synthetic data remain unaffected by patient consent withdrawal if "relatedness" between the data and the data subject cannot be established, or if "personal data has been rendered synthetic in such a manner that the data subject is no longer identifiable" [18] e.g., according to article 4(1) and recital 26 of the GDPR [9]. It is to be noted that in the "acceptable-risk" legal interpretation, a data subject's re-identification risk is reduced to an "acceptable" level rather than fully eradicated [18]. Hence, this interpretation enables approaches such as synthetic data and/or Differential Privacy (DP) model training to be used as legally compliant privacy preservation methods despite not guaranteeing a "zero-risk" of patient re-identification.

DP is a mathematical framework that allows practitioners to provide (worst-case scenario) theoretical privacy guarantees for an individual sharing their data to train a deep learning model. Consider two databases (e.g., containing image-label pairs), we call them adjacent if they differ in a single data point, i.e., one image is present in one database but not in the other. Then, a randomised mechanism $\mathcal{M}\colon \mathcal{D} \to \mathcal{R}$ with domain $\mathcal{D}$ and range $\mathcal{R}$ is said to satisfy (ε, δ)-differential privacy, if for any two adjacent databases $d, d' \in \mathcal{D}$ and for any subset of outputs $S \subseteq \mathcal{R}$, $\Pr[\mathcal{M}(d) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(d') \in S] + \delta$ holds. ε and δ bound a single data point's influence on a model's output (e.g. the models' weights or predictions). Thus, the smaller the value of these parameters, the higher the model's privacy and the harder it is for an attacker to retrieve information

about any training data point. DP-SGD [1] is the DP variant of the well-known SGD algorithm, and facilitates the training of a model under DP conditions. In particular, a model trained under (ε, δ)-DP is robust to post-processing, meaning only using its output for further computations also satisfies (ε, δ)-DP. Moreover, the choice of these parameters is application-dependent and normative [5] and varies strongly across real-world deployments [7]. In the case of mammography, multiple lesions of the same patient are available in the datasets, i.e. one from the CC view and one from the MLO view. Therefore, to preserve the privacy of one patient it is necessary to protect all their data points (i.e. all images). In such a case, DP group privacy is used to estimate a patient's DP privacy guarantee. However, for simplicity, in our subsequent experiments, we provide image-level privacy guarantees rather than per patient.

3 Experiments and Results

Experimental Setup		Metrics		
Dataset 1	Dataset 2	$\text{FID}_{\text{Img}} \downarrow$	$\text{FID}_{\text{Rad}} \downarrow$	FRD $\downarrow$
$\text{Syn}_{\text{MCGAN}}$	$\text{Real}_{\text{DDSM}}$	58.00±0.72	0.81±.013	18.12±1.01
$\text{Real}_{\text{DDSM}}$	$\text{Real}_{\text{DDSM}}$	29.25±0.82	0.31±.019	3.48±.352
$\text{Syn}_{\text{MCGAN}}$	$\text{Syn}_{\text{MCGAN}}$	20.90±0.16	0.32±.012	0.57±.094
$\text{Real}_{\text{DDSM}}$	$\text{Real}_{\text{BCDR}}$	156.43±14.3	3.88±.351	277.63±39.0

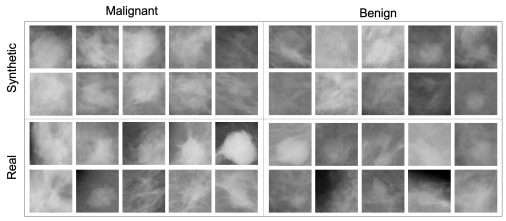

Fig. 2: Qualitative and quantitative synthesis results: Images are randomly selected malignant and benign real (CBIS-DDSM [16]) and MCGAN-generated masses. ImageNet [6] and RadImageNet [26,22] based FID [14] and FRD [25] scores are reported as mean ± standard deviation based on 3 subsets randomly sampled per patient ($N_{real} \approx 360$, $N_{syn} \approx 3240$). Row 4 indicates an BCDR-based[19] upper bound for comparison with synthetic data metrics in row 1.

Synthetic Data Evaluation Qualitatively assessing the synthetic images in Fig. 2, it is not readily possible to distinguish synthetic from real masses in terms of image fidelity or diversity. We note the absence of clear visual indicators to distinguish between malignant and benign images for both real and synthetic images. This is in line with the difficulty of determining the malignancy of a mammographic lesion shown by high clinical error rates and inter-observer variability [8]. However, results for training our malignancy classification model on only synthetic data (see *Syn* and *SynPre* in Table 1) show that the synthetic data captures the conditional distribution effectively generating either malignant or benign masses. Both, vanilla ImageNet-based Fréchet Inception Distance (FID) [14,6] and radiology domain-specific RadImageNet-based FID [26,22], concur that the synthetic data (FID_{Img}=58±.72) is substantially closer to the real CBIS-DDSM [16] distribution compared to BCDR [19] (FID_{Img}=156.43±1.43). This

Table 1: Results for within-domain (CBIS-DDSM [16]) and out-of-domain (BCDR [19]) breast cancer malignancy classification masses extracted from mammograms. *Syn* indicates 3k synthetic images being part of the fine-tuning training data, while *SynPre* represents pretraining all trainable model params with those 3k synthetic images (without DP guarantee), before fine-tuning the last two layers on real data with DP guarantee (*RealFT*). AUROC and AUPRC are reported as mean ± std based on 3 random seed runs. Best results in bold.

Experimental Setup			CBIS-DDSM [16]		BCDR [19]	
Model	ε	δ	AUROC ↑	AUPRC ↑	AUROC ↑	AUPRC ↑
$\text{SwinT}_{\text{Real}}$	∞	∞	0.778±.001	0.85±.001	0.695±.002	0.726±.003
$\text{SwinT}_{\text{Syn}}$	∞	∞	0.597±.011	0.696±.011	0.566±.064	0.602±.048
$\text{SwinT}_{\text{SynPre}}$	∞	∞	0.639±.016	0.733±.001	0.622±.032	0.660±.017
$\text{SwinT}_{\text{Real}}$	1	$1e^{-4}$	0.525±.043	0.640±.030	0.487±.020	0.549±.020
$\text{SwinT}_{\text{Real+Syn}}$	1	$1e^{-4}$	**0.553±.040**	**0.665±.025**	**0.521±.023**	**0.573±.024**
$\text{SwinT}_{\text{SynPre+RealFT}}$	$\infty\|1$	$\infty\|1e^{-4}$	0.661±.018	0.741±.007	0.637±.026	0.67±0013
$\text{SwinT}_{\text{Real}}$	6	$1e^{-4}$	0.572±.031	0.679±.019	0.532±.031	0.579±.029
$\text{SwinT}_{\text{Real+Syn}}$	6	$1e^{-4}$	**0.617±.013**	**0.708±.015**	**0.609±.027**	**0.647±.024**
$\text{SwinT}_{\text{SynPre+RealFT}}$	$\infty\|6$	$\infty\|1e^{-4}$	0.677±.014	0.752±.009	0.647±.022	0.679±.009
$\text{SwinT}_{\text{Real}}$	12	$1e^{-4}$	0.596±.023	0.702±.013	0.559±.033	0.600±.030
$\text{SwinT}_{\text{Real+Syn}}$	12	$1e^{-4}$	**0.624±.010**	**0.704±.012**	**0.625±.020**	**0.663±.012**
$\text{SwinT}_{\text{SynPre+RealFT}}$	$\infty\|12$	$\infty\|1e^{-4}$	0.688±.012	0.758±.011	0.654±.019	0.685±.007
$\text{SwinT}_{\text{Real}}$	20	$1e^{-4}$	0.611±.018	**0.715±.012**	0.581±.028	0.618±.026
$\text{SwinT}_{\text{Real+Syn}}$	20	$1e^{-4}$	**0.630±.003**	0.699±.008	**0.641±.018**	**0.685±.012**
$\text{SwinT}_{\text{SynPre+RealFT}}$	$\infty\|20$	$\infty\|1e^{-4}$	0.697±.012	0.763±.012	0.659±.017	0.689±.006
$\text{SwinT}_{\text{Real}}$	60	$1e^{-4}$	0.622±.014	**0.721±.110**	0.605±.019	0.640±.017
$\text{SwinT}_{\text{Real+Syn}}$	60	$1e^{-4}$	**0.629±.002**	0.694±.005	**0.650±.013**	**0.696±.007**
$\text{SwinT}_{\text{SynPre+RealFT}}$	$\infty\|60$	$\infty\|1e^{-4}$	0.712±.013	0.776±.013	0.671±.014	0.697±.004

is even more pronounced when comparing the variation of extracted radiomics features for CBIS-DDSM to synthetic (FRD=18.12) and BCDR (FRD=277.63) images using the Fréchet Radiomics Distance (FRD) [25]. While this indicates desirable synthetic data fidelity, we also observe good diversity. The latter is shown by comparing subsets of the same datasets with each other, where the variation within the synthetic data (e.g., FID_{Rad}=0.32±.12) closely resembles the variation within the real CBIS-DDSM dataset (e.g., FID_{Rad}=0.31±.19). Notwithstanding less variation in radiomics imaging biomarkers within the synthetic data (FRD_{Syn}=0.57 vs. FRD_{Real}=3.48), this overall points to a valid coverage of the distribution and an absence of mode collapse.

Mass Malignancy Classification As shown in Table 1, we conduct experiments with and without formal privacy guarantees. For scenarios where a formal privacy guarantee is not strictly required and, thus, synthetic data suffices as

privacy mechanism, we compare the results of training SwinT on synthetic data (*Syn*) and on real data (*Real*) with DP-SGD. *Kaissis et al.* [15] defined $\varepsilon = 6$ as suitable privacy budget for their medical imaging dataset. Compared to DP-SGD with $\varepsilon = 6$, synthetic data achieves better AUPRCs for within-domain tests on CBIS-DDSM ($\text{SwinT}_{\text{Syn}}$=0.696 vs $\text{SwinT}_{\text{Real}(\varepsilon=6)}$=0.679) and is on par for out-of-domain (ood) tests on BCDR ($\text{SwinT}_{\text{Syn}}$=0.602 vs $\text{SwinT}_{\text{Real}(\varepsilon=6)}$=0.600). However, training all SwinT layers using synthetic data (*SynPre*), achieves substantially better performance only approximated by DP results for $\varepsilon = 60$ for within-domain ($\text{SwinT}_{\text{SynPre}}$=0.733 vs $\text{SwinT}_{\text{Real}(\varepsilon=60)}$=0.721) and ood ($\text{SwinT}_{\text{SynPre}}$ =0.66 vs $\text{SwinT}_{\text{Real}(\varepsilon=60)}$=0.64) tests. Further fine-tuning $\text{SwinT}_{\text{SynPre}}$ on real data using DP-SGD results in additional improvement across all privacy parameters for within-domain and ood testing. For instance, training $\text{SwinT}_{\text{SynPre+RealFT}}$ with $\varepsilon = 1$ results in an AUPRC of 0.74 and 0.67 for CBIS-DDSM and BCDR, respectively. To assess scenarios where a formal guarantee is required, we further compare DP-SGD training of SwinT on real data (*Real*) with DP-SGD training on a mix of real and synthetic data (*Real+Syn*). To this end, our experiments show that such synthetic data augmentation can improve the privacy-utility tradeoff. This is exemplified by $\text{SwinT}_{\text{Real+Syn}(\varepsilon=6)}$ accomplishing an AUPRC of 0.708 within-domain and 0.647 ood, while $\text{SwinT}_{\text{Real}(\varepsilon=6)}$ achieved 0.679 and 0.579, respectively. We further observe the trend that stricter privacy budgets (i.e., smaller ε) can be associated with more added performance of synthetic data as additional classification model training data.

4 Discussion and Conclusion

We introduce a privacy preservation framework based on differential privacy (DP) and synthetic data and apply it to the diagnostic task of classifying the malignancy of breast masses extracted from screening mammograms. We further propose, train, and evaluate a malignancy-conditioned generative adversarial network to generate a dataset of benign and malignant synthetic breast masses. Next, we train a swin transformer model on mass malignancy classification and assess, compare and combine training under DP and training on synthetic data. This analysis revealed that when training with DP, synthetic data augmentation can notably improve classification performance for within-domain and out-of-domain test cases. Apart from that, we show, across privacy mechanisms and across domains, that the performance of models pretrained on synthetic data can be further improved by DP fine-tuning on real data.

This finding is particularly important considering that synthetic data, if not directly attributable to any specific patient, can become a valid, legally compliant alternative to strict DP guarantees in clinical practice. Consequently, it is to be further investigated where and when deterministic mechanisms without formal DP guarantees can suffice to shield against different privacy attacks [4]. In particular, we motivate future work to analyse the extent to which the inherent properties of synthetic data generation algorithms can provide empirical protection against attacks. In this regard, a comparison of generation algorithms such

as GANs [12] and denoising probabilistic diffusion models (DDPMs) [28] can provide insights towards further improving tconditional mass synthesis, while also enabling to quantify and compare the extent and effect of training data memorization in these models. A methodological alternative to our approach is to assess privacy-utility tradeoffs when training the generative model itself using DP-SGD [10,24], resulting in formal privacy guarantees of the generated synthetic datasets. Thus, a further avenue to explore then lies within the question whether randomness inherent in randomised data synthesis algorithms (e.g., based on the noise in DDPMs or GANs) can be used to amplify the privacy of the DP versions of such synthesis algorithms, thereby potentially further enhancing privacy-utility tradeoffs. To this end, our study constitutes a crucial first step leading towards the clinical adoption of diagnostic deep learning models, enabling practical privacy-utility tradeoffs all while anticipating respective legal obligations and clinical requirements.

Acknowledgments. This study has received funding from the European Union's Horizon research and innovation programme under grant agreement No 952103 (EuCanImage) and No 101057699 (RadioVal). It was further partially supported by the project FUTURE-ES (PID2021-126724OB-I00) from the Ministry of Science and Innovation of Spain. RO acknowledges a research stay grant from the Helmholtz Information and Data Science Academy (HIDA). GK received support from the German Federal Ministry of Education and Research and the Bavarian State Ministry for Science and the Arts under the Munich Centre for Machine Learning (MCML), from the German Ministry of Education and Research and the Medical Informatics Initiative as part of the PrivateAIM Project, from the Bavarian Collaborative Research Project PRIPREKI of the Free State of Bavaria Funding Programme "Artificial Intelligence – Data Science", and from the German Academic Exchange Service (DAAD) under the Kondrad Zuse School of Excellence for Reliable AI (RelAI).

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. pp. 308–318 (2016)
2. Alyafi, B., Diaz, O., Marti, R.: DCGANs for realistic breast mass augmentation in x-ray mammography. In: Medical Imaging 2020: Computer-Aided Diagnosis. vol. 11314, p. 1131420. International Society for Optics and Photonics (2020)
3. Balle, B., Cherubin, G., Hayes, J.: Reconstructing training data with informed adversaries. In: 2022 IEEE Symposium on Security and Privacy (SP). pp. 1138–1156. IEEE (2022)
4. Cohen, A., Nissim, K.: Towards formalizing the gdpr's notion of singling out. Proceedings of the National Academy of Sciences **117**(15), 8344–8352 (2020)

5. De, S., Berrada, L., Hayes, J., Smith, S.L., Balle, B.: Unlocking high-accuracy differentially private image classification through scale. arXiv preprint arXiv:2204.13650 (2022)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
7. Dwork, C., Kohli, N., Mulligan, D.: Differential privacy in practice: Expose your epsilons! Journal of Privacy and Confidentiality **9**(2) (2019)
8. Ekpo, E.U., Alakhras, M., Brennan, P.: Errors in mammography cannot be solved through technology alone. Asian Pacific journal of cancer prevention: APJCP **19**(2), 291 (2018)
9. European Parliament and Council of European Union: General Data Protection Regulation (GDPR), REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Online at https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679/ (2018)
10. Ghalebikesabi, S., Berrada, L., Gowal, S., Ktena, I., Stanforth, R., Hayes, J., De, S., Smith, S.L., Wiles, O., Balle, B.: Differentially private diffusion models generate useful synthetic images. arXiv preprint arXiv:2302.13861 (2023)
11. Global Cancer Observatory: The global cancer observatory (gco) is an interactive web-based platform presenting global cancer statistics to inform cancer control and research. https://gco.iarc.fr/ (2023), accessed on 2023-01-17
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
13. Haim, N., Vardi, G., Yehudai, G., Shamir, O., Irani, M.: Reconstructing training data from trained neural networks. arXiv preprint arXiv:2206.07758 (2022)
14. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500 (2017)
15. Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima Jr, I., Mancuso, J., Jungmann, F., Steinborn, M.M., et al.: End-to-end privacy preserving deep learning on multi-institutional medical imaging. Nature Machine Intelligence **3**(6), 473–484 (2021)
16. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific data **4**(1), 1–9 (2017)
17. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
18. López, C.A.F.: On the legal nature of synthetic data. In: NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research (2022)
19. Lopez, M.G., Posada, N., Moura, D.C., Pollán, R.R., Valiente, J.M.F., Ortega, C.S., Solar, M., Diaz-Herrero, G., Ramos, I., Loureiro, J., et al.: BCDR: a breast cancer digital repository. In: 15th International conference on experimental mechanics. vol. 1215 (2012)
20. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
21. McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., et al.: International evaluation of an ai system for breast cancer screening. Nature **577**(7788), 89–94 (2020)

22. Mei, X., Liu, Z., Robson, P.M., Marinelli, B., Huang, M., Doshi, A., Jacobi, A., Cao, C., Link, K.E., Yang, T., et al.: RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning. Radiology: Artificial Intelligence p. e210315 (2022)
23. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
24. Osuala, R., Kushibar, K., Garrucho, L., Linardos, A., Szafranowska, Z., Klein, S., Glocker, B., Diaz, O., Lekadir, K.: Data synthesis and adversarial networks: A review and meta-analysis in cancer imaging. Medical Image Analysis **84**, 102704 (2023)
25. Osuala, R., Lang, D., Verma, P., Joshi, S., Tsirikoglou, A., Skorupko, G., Kushibar, K., Garrucho, L., Pinaya, W.H., Diaz, O., et al.: Towards learning contrast kinetics with multi-condition latent diffusion models. arXiv preprint arXiv:2403.13890 (2024)
26. Osuala, R., Skorupko, G., Lazrak, N., Garrucho, L., García, E., Joshi, S., Jouide, S., Rutherford, M., Prior, F., Kushibar, K., et al.: medigan: a python library of pretrained generative models for medical image synthesis. Journal of Medical Imaging **10**(6), 061403 (2023)
27. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
28. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. pp. 2256–2265. PMLR (2015)
29. Su, R., Liu, X., Tsaftaris, S.A.: Why patient data cannot be easily forgotten? In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VIII. pp. 632–641. Springer (2022)
30. Szafranowska, Z., Osuala, R., Breier, B., Kushibar, K., Lekadir, K., Diaz, O.: Sharing generative models instead of private data: a simulation study on mammography patch classification. In: 16th International Workshop on Breast Imaging (IWBI2022). vol. 12286, pp. 169–177. SPIE (2022)
31. Yousefpour, A., Shilov, I., Sablayrolles, A., Testuggine, D., Prasad, K., Malek, M., Mironov, I.: Opacus: User-friendly differential privacy library in pytorch. arXiv preprint arXiv:2109.12298 (2021)