applied sciences

MDPI

Article

Local Differential Privacy Image Generation Using Flow-Based Deep Generative Models

Hisaichi Shibata [1,*], Shouhei Hanaoka [1], Yang Cao [2], Masatoshi Yoshikawa [3], Tomomi Takenaga [1], Yukihiro Nomura [4,5], Naoto Hayashi [5] and Osamu Abe [1]

1 Department of Radiology, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan
2 Graduate School of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo 060-0814, Japan
3 Faculty of Data Science, Osaka Seikei University, 1-3-7 Aikawa, Higashiyodogawa-ku, Osaka 533-0007, Japan
4 Center for Frontier Medical Engineering, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan
5 Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan
* Correspondence: sh@g.ecc.u-tokyo.ac.jp

Abstract: Diagnostic radiologists need artificial intelligence (AI) for medical imaging, but access to medical images required for training in AI has become increasingly restrictive. To release and use medical images, we need an algorithm that can simultaneously protect privacy and preserve pathologies in medical images. To address this, we introduce DP-GLOW, a hybrid that combines the local differential privacy (LDP) algorithm with GLOW, one of the flow-based deep generative models. By applying a GLOW model, we disentangle the pixelwise correlation of images, which makes it difficult to protect privacy with straightforward LDP algorithms for images. Specifically, we map images to the latent vector of the GLOW model, where each element follows an independent normal distribution. We then apply the Laplace mechanism to this latent vector to achieve ϵ-LDP, which is one of the LDP algorithms. Moreover, we applied DP-GLOW to chest X-ray images to generate LDP images while preserving pathologies. The ϵ-LDP-processed chest X-ray images obtained with DP-GLOW indicate that we have obtained a powerful tool for releasing and using medical images for training AI.

Keywords: differential privacy; deep generative models; medical images; privacy protection; database; image obfuscation

check for updates

Citation: Shibata, H.; Hanaoka, S.; Cao, Y.; Yoshikawa, M.; Takenaga, T.; Nomura, Y.; Hayashi, N.; Abe, O. Local Differential Privacy Image Generation Using Flow-Based Deep Generative Models. *Appl. Sci.* **2023**, *13*, 10132. https://doi.org/10.3390/app131810132

Academic Editors: Danila Germanese, Maria Antonietta Pascali and Lorenzo Faggioni

Received: 7 August 2023
Revised: 29 August 2023
Accepted: 6 September 2023
Published: 8 September 2023

1. Introduction

Diagnostic radiologists need artificial intelligence (AI) for medical imaging to reduce workloads and enhance productivity. However, access to medical images required for the training of AI has become increasingly restrictive owing to the increasing demands for the protection of personal information in each country. Moreover, data use agreements bind the purpose of use for even medical image datasets open to researchers upon request (e.g., [1]). Additionally, if someone specifies personal identities in medical images, irreversible leakage of personal information may occur.

Here, we assume a situation in which we anonymize test datasets when training datasets are open to the public. We further assume that probabilistic distributions of medical images for the training and test datasets are similar. For example, the training dataset can be a large-scale dataset already open to the public from a hospital, and the test dataset can be a dataset privately held by another hospital.

Differential privacy (DP) algorithms [2] have recently emerged as tools with a provable privacy protection guarantee and usefulness. We specifically focus on local DP (LDP) [3], which adds noise to each image so that we cannot specify any identity in an image. Because one can anonymize the image upstream of image processing using LDP algorithms, we can use the processed images without limiting the purpose of use. At the same time, LDP can retain valuable information in the original image, e.g., lung opacity suggesting pneumonia in chest X-ray (CXR) images.

Because CXR images are the most representative medical images, in this study, we adopt CXR images from the Radiological Society of North America (RSNA) dataset. However, we can extend the proposed method to natural images and medical images of arbitrary modality and dimensions.

To summarize, our contributions are as follows:

1. We adopt GLOW to disentangle image pixels and realize the ϵ-LDP algorithm for images (DP-GLOW).
2. We generate ϵ-LDP-processed CXR images.
3. We evaluate the usefulness of ϵ-LDP-processed CXR images using a pneumonia detection model.
4. We visually confirm how the proposed method obscures identities in medical images.

2. Related Works

Abadi et al. [4] proposed a differentially private stochastic gradient descent (DP-SGD) method, in which a controlled noise is added to the gradient of parameters and then clipped during the training of a deep model. Ziller et al. [5] proposed the training of a segmentation network for CXR images using a discriminative model trained with DP-SGD. Kossen et al. [6] proposed the generation of differentially private time-of-flight magnetic resonance angiography (TOF-MRA) images using generative adversarial networks (GANs) trained with DP-SGD. For DP-SGD, the theoretical guarantee that images generated using GANs trained with DP-SGD satisfy ϵ-LDP is not apparent.

Fan [7] adopted LDP (although it is not explicitly stated in the paper) for pixelized images. Image pixelization has the effect of reducing the global sensitivity of the LDP algorithm. Fan [8] proposed another LDP algorithm using pixelization and Gaussian blur. However, these methods can significantly degrade the quality of the original images. Croft et al. [9], Liu et al. [10], and Li and Clifton [11] almost simultaneously proposed another LDP algorithm for images. Liu et al. [10] showed a concrete implementation using GANs, whereas Croft et al. [9] showed an abstract formulation. The implementation of the LDP algorithm for images by Li and Clifton is similar to that by Liu et al., but Li and Clifton adopt clipping so that the generated LDP images are within the probabilistic distribution of training images. Finally, Croft et al. [12] experimented with an LDP algorithm for facial obfuscation. On the other hand, we propose an algorithm for local differential privacy using GLOW [13], one of the flow-based deep generative models. We adopt GLOW as it is a renowned and representative flow-based deep generative model. While it is challenging to map real images to latent space without degrading them in GANs, it is easily achievable with flow-based deep generative models.

In previous studies [7–12], LDP was not adopted for medical images. Therefore, previous studies did not evaluate the usefulness of LDP-processed images being able to contain valuable information for AI for medical imaging; in this study, we aim to evaluate the usefulness experimentally and quantitatively. As a metric of usefulness, we adopt the area under the curve (AUC) for pathology detection (we assume pneumonia detection in this study), which is essential in AI for medical imaging.

3. Materials and Methods

3.1. Dataset for CXR Images

We took CXR images from the RSNA Pneumonia Detection Challenge dataset [14]. This dataset comprises 30,000 frontal-view CXR images, with each image labeled as "Normal", "No Opacity/Not Normal", or "Opacity" by one to three board-certified radiologists. The Opacity group consists of images with suspicious opacities suggesting pneumonia, and the No Opacity/Not Normal group consists of images with abnormalities other than pneumonia.

We show the composition of the CXR image sets in Table 1. We randomly sampled CXR images for the three datasets. We then trained GLOW, one of the flow-based deep generative models (DGMs) using $\mathcal{S}^{\text{train}}_{\text{mixture}}$. The goal was to ensure that lung opacity, which suggests pneumonia, was not overly obfuscated for the DP-GLOW algorithm. For testing purposes, we used images from $\mathcal{S}^{\text{test}}_{\text{unknown}}$ to generate ϵ-LDP-CXR images with DP-GLOW. Subsequently, we detected pneumonia from these ϵ-LDP-CXR images. Additionally, we utilized $\mathcal{S}^{\text{train}}_{\text{normal}}$ to train a separate model specifically designed to detect pneumonia from CXR images (see Appendix A for details).

Table 1. Composition of datasets. $\mathcal{S}^{\text{train}}_{\text{normal}}$ and $\mathcal{S}^{\text{train}}_{\text{mixture}}$ share 6529 normal CXR images.

Set	Normal	Abnormal
$\mathcal{S}^{\text{train}}_{\text{normal}}$	7808	0
$\mathcal{S}^{\text{train}}_{\text{mixture}}$	6553	6631
$\mathcal{S}^{\text{test}}_{\text{unknown}}$	1358	13,863

3.2. Preliminary for GLOW

The probability distribution from which images arise is complex and difficult to handle. Thus, flow-based deep generative models (DGMs) [13,15,16] transform this probability distribution into more manageable distributions, e.g., the elementwise independent normal distribution, by requiring the neural network to be bijective. Specifically, using the change of variables formula, we obtain the following:

$$\log p(\boldsymbol{x}) = \log p(\boldsymbol{z}) + \log\left|\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\right|, \tag{1}$$

where $\boldsymbol{z}$ is a random variable vector that follows a Gaussian distribution independent for each element, and $\boldsymbol{x}$ is a random variable vector sampled from the probability distribution to which images belong. During this transformation process, the logarithm of the determinant appears with respect to the neural network. To efficiently compute this, the neural network is decomposed and represented as a product of functions where the determinant is easy to compute. Specifically, we define the relationship between $\boldsymbol{z}$ and $\boldsymbol{x}$:

$$\boldsymbol{z} = \boldsymbol{G}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}), \tag{2}$$

$$= g_K^{-1} \circ g_{K-1}^{-1} \cdots g_1^{-1}(\boldsymbol{x}), \tag{3}$$

where $\boldsymbol{G}_{\boldsymbol{\theta}}^{-1}$ is a trainable (parameters $\boldsymbol{\theta}$) invertible map between a probabilistic distribution of images and a tractable probabilistic distribution (elementwise independent normal distribution in our settings) and $g_K^{-1} \cdots g_1^{-1}$ are decomposed functions. Using Equation (1), the flow-based DGMs maximize the average logarithm likelihood of training images ($\mathcal{L}$) during training:

$$\mathcal{L} = \frac{1}{|\mathcal{D}|}\sum_{i}^{\mathcal{D}} \log p(\boldsymbol{x}_i), \tag{4}$$

where $\mathcal{D}$ represents image dataset, $\boldsymbol{x}_i$ indicates an image in the dataset, and $|\mathcal{D}|$ is number of images in the dataset. After the training, the flow-based DGMs can generate fake but realistic images using Equation (3) (sampling) and can explicitly compute the value of the probabilistic density function of images using Equation (1) (density estimation).

GLOW [13] is one of the flow-based DGMs. The deep network in GLOW recursively contains the actnorm, coupling, and permutation layers. The actnorm layer normalizes the data. The coupling layer contains deep convolutional neural networks while it guarantees the invertibility of the layer. The permutation layer ensures that processing in coupling layers affects all the elements of data and is implemented using 1×1 convolution. Moreover, GLOW adopts a multiscale architecture, which can contain multiple deep networks of different recursive levels. This approach can reduce memory requirements and computational costs without significant compromise to image quality. For a detailed understanding of GLOW, we refer the reader to [13].

3.3. Our Framework: DP-GLOW

We represent a gray-scale CXR image as a vector $\boldsymbol{x} \in \mathbf{R}^{H \times W}$, where H and W are the height and width of the image, respectively. We assume that the abovementioned GLOW has already been trained with many CXR images. Moreover, as a result of the training, we assume that we obtained an invertible map $\boldsymbol{G}_{\boldsymbol{\theta}}^{-1}$ between an elementwise independent normal distribution and a probabilistic distribution of CXR images.

Our objective is to generate another image $\tilde{\boldsymbol{x}}$ from $\boldsymbol{x}$, which satisfies the definition of ϵ-LDP:

$$\log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}) - \log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}') \quad \leq \quad \epsilon, \tag{5}$$

where $\epsilon(\geq 0)$ is the privacy budget, $\boldsymbol{x}'$ is an arbitrary image taken from a probabilistic distribution of CXR images, $p(\tilde{\boldsymbol{x}}|\boldsymbol{x})$ is the posterior probability to obtain $\tilde{\boldsymbol{x}}$ when we already obtained $\boldsymbol{x}$, and $p(\tilde{\boldsymbol{x}}|\boldsymbol{x}')$ is the posterior probability to obtain $\tilde{\boldsymbol{x}}$ when we already obtained $\boldsymbol{x}'$. Specifically, even if we simply add noise that follows a Laplace distribution to the image itself, the image satisfies local differential privacy. However, especially when the privacy budget is small, the intensity of the noise increases, and the class of the image (in this case, CXR images) is not retained. In such cases, the utility drastically decreases. To avoid this, in this study, we add noise following the Laplace distribution to the image mapped to the latent space and then revert the perturbed latent space vector back to the image space. To this end, we introduce a trained invertible vector function $\boldsymbol{G}_{\boldsymbol{\theta}}^{-1}$, which maps $\boldsymbol{x}$ into another vector $\boldsymbol{z}$ (a latent space vector), each element of which does not have a correlation with other elements. This vector function is, in general, dependent on the probabilistic distribution of images we adopt (e.g., CXR, head computed tomography, and mammography images). Therefore, we explicitly represent this dependence as parameters $\boldsymbol{\theta}$. We further define $\boldsymbol{z}' \equiv \boldsymbol{G}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}')$, and $\tilde{\boldsymbol{z}} \equiv \boldsymbol{G}_{\boldsymbol{\theta}}^{-1}(\tilde{\boldsymbol{x}})$. As we will prove later, we can add noise that follows the Laplace distribution to the latent space vector:

$$\tilde{z}_k \quad = \quad z_k + \mathcal{N}_k, \tag{6}$$

$$\mathcal{N}_k \quad \sim \quad \mathrm{Lap}\left(\mu_k = 0, \sigma_k = \frac{\Delta z_k}{\frac{\epsilon}{H \cdot W}}\right), \tag{7}$$

where μ_k and σ_k are, respectively, the expectation (a scalar) and scale (a scalar) of the Laplace distribution, ϵ is the user-defined privacy budget, and Δz_k is the sensitivity for the element k defined as:

$$\Delta z_k \quad := \quad \max_{z, z' : z \neq z'} |z_k - z'_k|, \tag{8}$$

where $\boldsymbol{z}$ and $\boldsymbol{z}'$ run latent vectors of all the training images. Moreover, we set a bound for each element of the latent space vector so that $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ are in the probabilistic distribution of CXR images:

$$z_k \quad \leftarrow \quad \mathrm{clip}\left(z_k, c_k - \frac{w_k}{2}, c_k + \frac{w_k}{2}\right), \tag{9}$$

$$\tilde{z}_k \quad \leftarrow \quad \mathrm{clip}\left(\tilde{z}_k, c_k - \frac{w_k}{2}, c_k + \frac{w_k}{2}\right), \tag{10}$$

$$w_k \quad = \quad \alpha \cdot (\max_z z_k - \min_z z_k), \tag{11}$$

$$c_k \quad = \quad \frac{(\max_z z_k + \min_z z_k)}{2}, \tag{12}$$

where min and max operators, respectively, find the minimum and maximum values over all the latent space vectors $\boldsymbol{z}$ for training images, the scalar function $\mathrm{clip}(x, a, b)$ bounds the value of x such that $a \leq x \leq b$, and the subscript k means the element number (runs all the elements of the latent space vector). We empirically set $\alpha = 0.4$ in this study. An overly large α can cause significant collapse of the input images, while an excessively small α tends to overly normalize them. Now we can generate a differentially private image $\tilde{\boldsymbol{x}}$ for the budget ϵ,

$$\tilde{\boldsymbol{x}} = \boldsymbol{G}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}). \tag{13}$$

We prepare the vector function $\boldsymbol{G}_{\boldsymbol{\theta}}^{-1}$ by training one of the flow-based DGMs, i.e., GLOW, with many CXR images prior to executing this locally differential private algorithm. We summarize our DP-GLOW algorithm to generate ϵ-LDP images below:

1. Train GLOW (maximize the average logarithm likelihood) with many CXR images to obtain $\boldsymbol{G}_{\boldsymbol{\theta}}^{-1}$.
2. Set the privacy budget ϵ.
3. Compute sensitivity from the training CXR images following Equation (8).
4. Compute image-dependent clipping parameters following Equations (11) and (12) from the training CXR images.
5. Map an image $\boldsymbol{x}$ from the test CXR images set onto the latent vector $\boldsymbol{z}$.
6. Clip $\boldsymbol{z}$ following Equation (9).
7. Add noise following Equation (6) to obtain $\tilde{\boldsymbol{z}}$.
8. Clip $\tilde{\boldsymbol{z}}$ following Equation (10).
9. Map the clipped latent vector $\tilde{\boldsymbol{z}}$ onto the CXR image space by $\boldsymbol{G}_{\boldsymbol{\theta}}$ to obtain a ϵ-LDP CXR image.

In Figure 1, we visually illustrate the DP-GLOW algorithm to clarify the step where we introduce noise using the Laplace mechanism.

Finally, we prove that our framework indeed satisfies ϵ-LDP. We can obtain the following equations using the change of variable formula:

$$\begin{aligned}
\log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}) &= \log p(\tilde{\boldsymbol{x}}, \boldsymbol{x}) - \log p(\boldsymbol{x}) && (14)\\
&= \log p(\tilde{\boldsymbol{z}}, \boldsymbol{z}) + \log\left|\det\left(\frac{\partial(\tilde{\boldsymbol{z}}, \boldsymbol{z})}{\partial(\tilde{\boldsymbol{x}}, \boldsymbol{x})}\right)\right| \\
&- \log p(\boldsymbol{z}) - \log\left|\det\left(\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\right)\right| && (15)\\
&= \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}) + \log\left|\det\frac{\partial \tilde{\boldsymbol{z}}}{\partial \tilde{\boldsymbol{x}}}\right| \\
&+ \log\left|\det\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\right| - \log\left|\det\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\right| && (16)\\
&= \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}) + \log\left|\det\frac{\partial \tilde{\boldsymbol{z}}}{\partial \tilde{\boldsymbol{x}}}\right|, && (17)
\end{aligned}$$

where $p(\boldsymbol{x})$ is the same as in Equation (1), and we used the following modification:

$$\begin{aligned}
\log\left|\det\left(\frac{\partial(\tilde{\boldsymbol{z}}, \boldsymbol{z})}{\partial(\tilde{\boldsymbol{x}}, \boldsymbol{x})}\right)\right| &= \log\begin{vmatrix} \frac{\partial \tilde{\boldsymbol{z}}}{\partial \tilde{\boldsymbol{x}}} & \frac{\partial \boldsymbol{z}}{\partial \tilde{\boldsymbol{x}}} \\ \frac{\partial \tilde{\boldsymbol{z}}}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} \end{vmatrix} && (18)\\
&= \log\begin{vmatrix} \frac{\partial \tilde{\boldsymbol{z}}}{\partial \tilde{\boldsymbol{x}}} & \frac{\partial \boldsymbol{z}}{\partial \tilde{\boldsymbol{x}}} \\ \boldsymbol{0} & \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} \end{vmatrix} && (19)\\
&= \log\left|\det\frac{\partial \tilde{\boldsymbol{z}}}{\partial \tilde{\boldsymbol{x}}}\right| + \log\left|\det\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\right|. && (20)
\end{aligned}$$

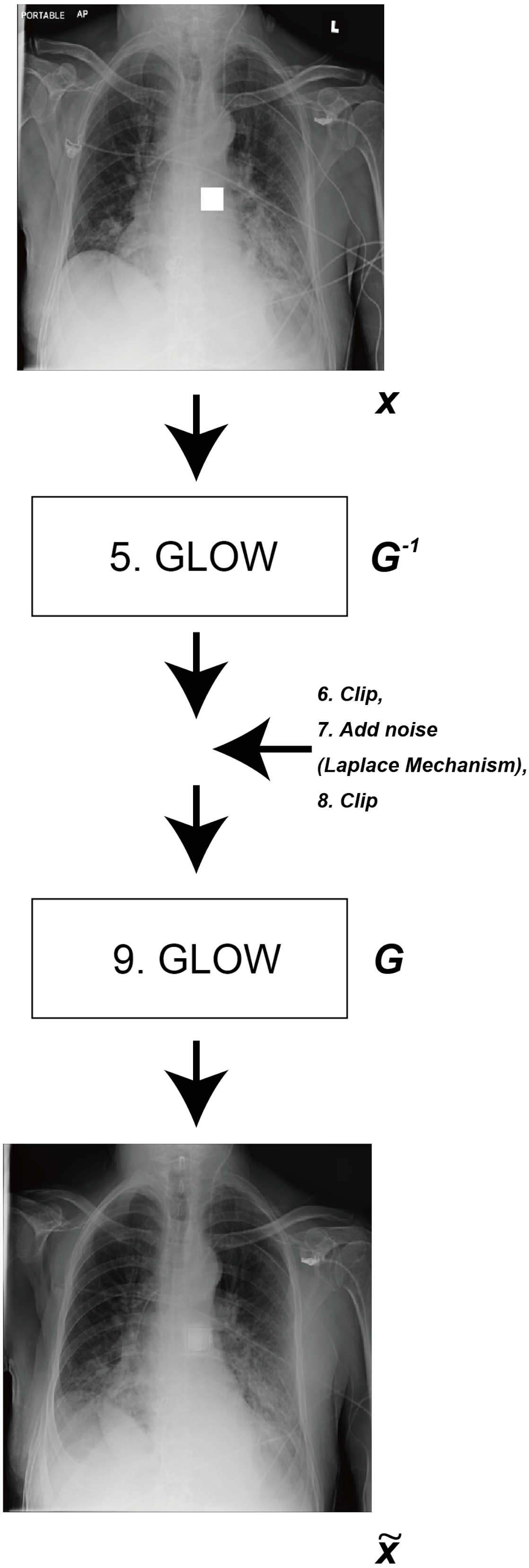


Figure 1. Overview of the DP-GLOW algorithm.

We have $\frac{\partial \tilde{z}}{\partial x} = \mathbf{0}$ because vector $\boldsymbol{z}$ is fixed and there is no correlation between probabilistic vector $\mathcal{N}$ and any image vector $\boldsymbol{x}$.

Therefore, we have:

$$\log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}) = \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}) + \log\left|\det \frac{\partial \tilde{\boldsymbol{z}}}{\partial \tilde{\boldsymbol{x}}}\right|, \tag{21}$$

and

$$\log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}') = \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}') + \log\left|\det \frac{\partial \tilde{\boldsymbol{z}}}{\partial \tilde{\boldsymbol{x}}}\right|. \tag{22}$$

Substituting those equations, we have:

$$\log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}) - \log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}') = \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}) - \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}'). \tag{23}$$

Combined with the Laplace mechanism, we can ensure:

$$\begin{aligned}
\log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}) - \log p(\tilde{\boldsymbol{x}}|\boldsymbol{x}') &= \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}) - \log p(\tilde{\boldsymbol{z}}|\boldsymbol{z}') && (24)\\
&= \sum_{k}^{H\cdot W} \log p(\tilde{z_k}|z_k) \\
&\quad - \sum_{k}^{H\cdot W} \log p(\tilde{z_k}|z'_k) && (25)\\
&= \sum_{k}^{H\cdot W} -\frac{|\tilde{z}_k - z_k|}{\frac{H\cdot W\cdot \Delta z_k}{\epsilon}} \\
&\quad + \frac{|\tilde{z}_k - z'_k|}{\frac{H\cdot W\cdot \Delta z_k}{\epsilon}} && (26)\\
&\leq \frac{\epsilon}{H\cdot W} \cdot \sum_{k}^{H\cdot W} \frac{|-z'_k + z_k|}{\Delta z_k} && (27)\\
&\leq \epsilon, && (28)
\end{aligned}$$

where k runs each element of vectors.

3.4. Hyperparameters

We show the hyperparameters for the training of CXR image sets $\mathcal{S}^{\text{train}}_{\text{normal}}$ and $\mathcal{S}^{\text{train}}_{\text{mixture}}$ in Table 2. The 'learn-top' option determines whether we train the means and variances of the latent space in GLOW. The minibatch size refers to the number of images trained simultaneously during batch learning. To train GLOW, we used Tensorflow 1.12.0 [17]. The versions of CUDA and cuDNN were 9.0 and 7.4, respectively. We carried out all processes in one computing node of the Reedbush-L supercomputer system, Rackable C1102-GP8, SGI, Mountain View, CA, USA, in the Information Technology Center, The University of Tokyo. The system consists of 64 computing nodes equipped with two Xeon E5-2695v4 processors, Intel, Santa Clara, CA, USA, 256 GB memory, and four GPUs (Tesla P100 SXM2 with 16 GB memory, NVIDIA, Santa Clara, CA, USA).

To obtain ϵ-LDP-CXR images and to detect pneumonia in CXR images, we used Tensorflow 1.15.5. The versions of CUDA and cuDNN were 9.0 and 8.1, respectively. We carried out all processes in one computing node of the Wisteria/B-DEC01 supercomputer system, PRIMERGY GX2570 M6, FUJITSU, Tokyo, Japan in the Information Technology Center, The University of Tokyo. The system (Wisteria-Aquarius) consists of 45 computing nodes equipped with two Xeon 8360Y processors, Intel, and eight GPUs (A100 with 40 GB memory, NVIDIA).

Table 2. Hyperparameters for GLOW.

Coupling Layer	Affine
Learn-top option	False
Flow permutation	1×1 convolution
Minibatch size	4
Number of training samples per epoch	50,000
Network levels	7
Depth per level	32
Image size (in pixel)	H512 $\times$ W512 $\times$ C1
Total epochs	200
Learning rate in steady state	10^{-3}

4. Results

4.1. ϵ-LDP-Processed CXR Images

In Figure 2, we show ϵ-LDP-processed CXR images of four clinical cases obtained with the image domain LDP, which directly imposes the Laplace mechanism on the input image, with different privacy budgets together with the original images. Figure 3 shows four ϵ-LDP-processed CXR images of clinical cases obtained with DP-GLOW and different privacy budgets together with the original images. In case 1 for DP-GLOW, there is decreased permeability in the bilateral hilar regions. Although this hilar opacity tends to be preserved with a larger privacy budget, the entire image is degraded when the privacy budget becomes $10^1 \cdot H \cdot W$. A similar tendency is observed in the images of all the four cases for DP-GLOW; for example, in case 4 with $\epsilon = 10^1 \cdot H \cdot W$, the lung opacity suggesting pneumonia in the right lower lung field is well preserved, while the entire image is degraded.

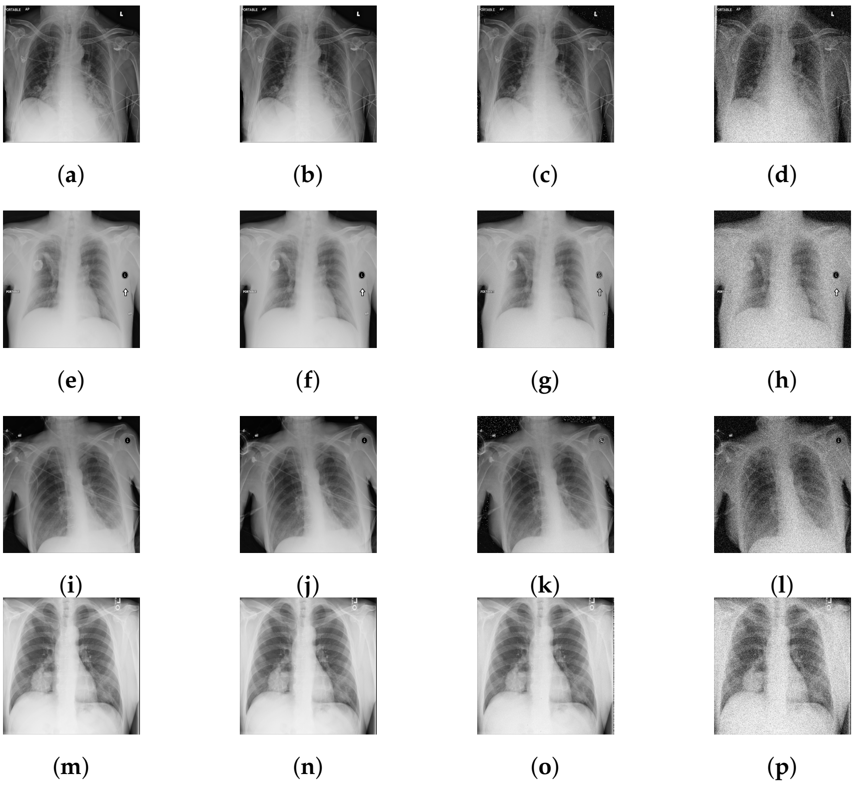


Figure 2. ϵ-LDP-processed CXR images (we applied the Laplace mechanism in the image domain). (**a**) Original, case 1; (**b**) $\epsilon = 10^3 \cdot H \cdot W$, case 1; (**c**) $\epsilon = 10^2 \cdot H \cdot W$, case 1; (**d**) $\epsilon = 10^1 \cdot H \cdot W$, case 1; (**e**) Original, case 2; (**f**) $\epsilon = 10^3 \cdot H \cdot W$, case 2; (**g**) $\epsilon = 10^2 \cdot H \cdot W$, case 2; (**h**) $\epsilon = 10^1 \cdot H \cdot W$, case 2; (**i**) Original, case 3; (**j**) $\epsilon = 10^3 \cdot H \cdot W$, case 3; (**k**) $\epsilon = 10^2 \cdot H \cdot W$, case 3; (**l**) $\epsilon = 10^1 \cdot H \cdot W$, case 3; (**m**) Original, case 4; (**n**) $\epsilon = 10^3 \cdot H \cdot W$, case 4; (**o**) $\epsilon = 10^2 \cdot H \cdot W$, case 4; (**p**) $\epsilon = 10^1 \cdot H \cdot W$, case 4.

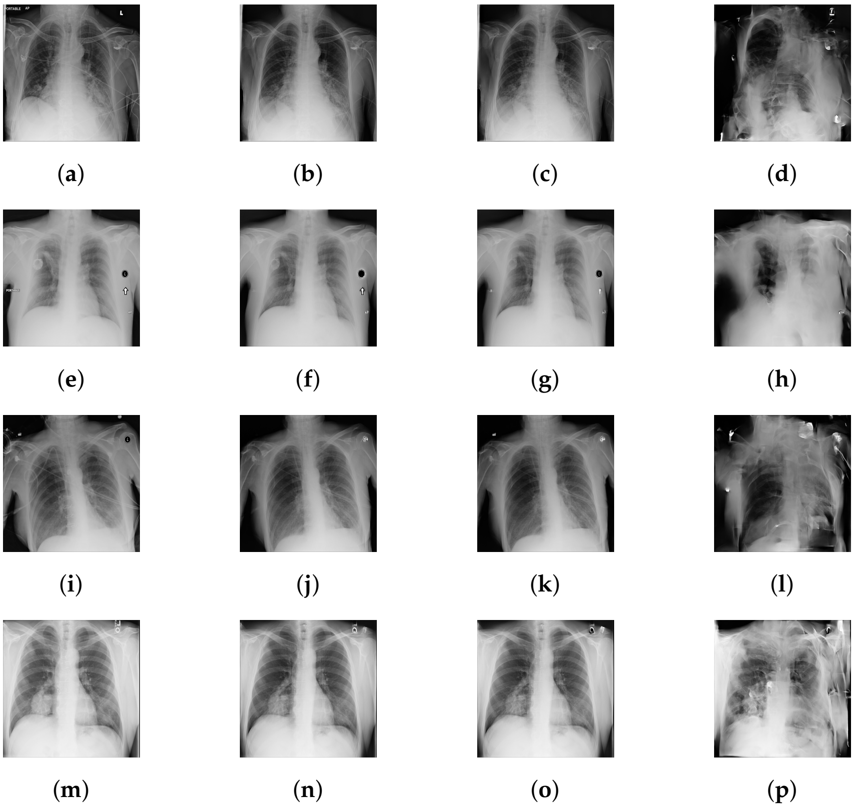
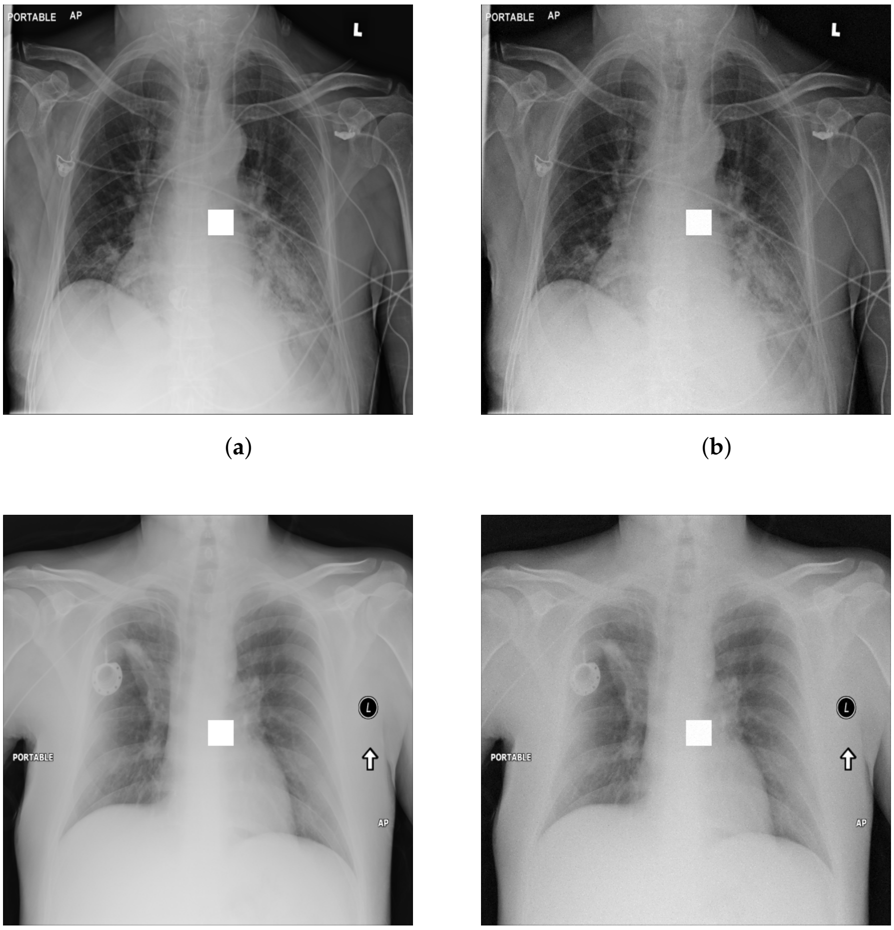


Figure 3. ϵ-LDP-processed CXR images obtained with DP-GLOW. (**a**) Original, case 1; (**b**) $\epsilon = 10^3 \cdot H \cdot W$, case 1; (**c**) $\epsilon = 10^2 \cdot H \cdot W$, case 1; (**d**) $\epsilon = 10^1 \cdot H \cdot W$, case 1; (**e**) Original, case 2; (**f**) $\epsilon = 10^3 \cdot H \cdot W$, case 2; (**g**) $\epsilon = 10^2 \cdot H \cdot W$, case 2; (**h**) $\epsilon = 10^1 \cdot H \cdot W$,case 2; (**i**) Original, case 3; (**j**) $\epsilon = 10^3 \cdot H \cdot W$, case 3; (**k**) $\epsilon = 10^2 \cdot H \cdot W$, case 3; (**l**) $\epsilon = 10^1 \cdot H \cdot W$, case 3; (**m**) Original, case 4; (**n**) $\epsilon = 10^3 \cdot H \cdot W$, case 4; (**o**) $\epsilon = 10^2 \cdot H \cdot W$, case 4; (**p**) $\epsilon = 10^1 \cdot H \cdot W$, case 4.

4.2. Qualitative Assessment of LDP-Processed CXR Images

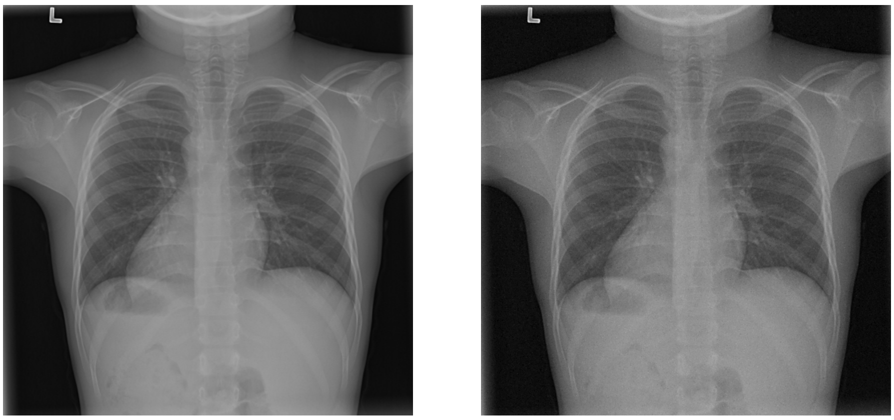
Here, we assume two possible privacy leakage scenarios. To CXR images, we intentionally add features that can lead to the re-identification of the subject appearing in a CXR image. The first feature is an artificial block marker. The second feature is a rare anatomical abnormality known as *situs inversus*, simulated by flipping a CXR image along the vertical axis. Figure 4a,c show CXR images with the artificial block marker. Figure 5a shows a flipped CXR image to represent a case of *situs inversus*. We applied DP-GLOW to these CXR images. In Figure 4b,d, the image domain LDP fails to obfuscate the artificial block marker with a moderate privacy budget. In contrast, in Figure 6b,d, DP-GLOW successfully obfuscated the artificial block marker with the moderate privacy budget. On the other hand, the anatomical shape of the chest and the abnormal opacity (hilar regions in the case 1) are preserved. In Figure 5b, we observed that the right edge of the heart does not become obfuscated with the image domain LDP. In contrast, in Figure 7b, we observed that the right edge of the heart becomes obfuscated and the heart appears at the center of the thoracic cage with DP-GLOW. However, DP-GLOW with this privacy budget is insufficient to almost completely erase the feature of *situs inversus*.



(**a**) (**b**)

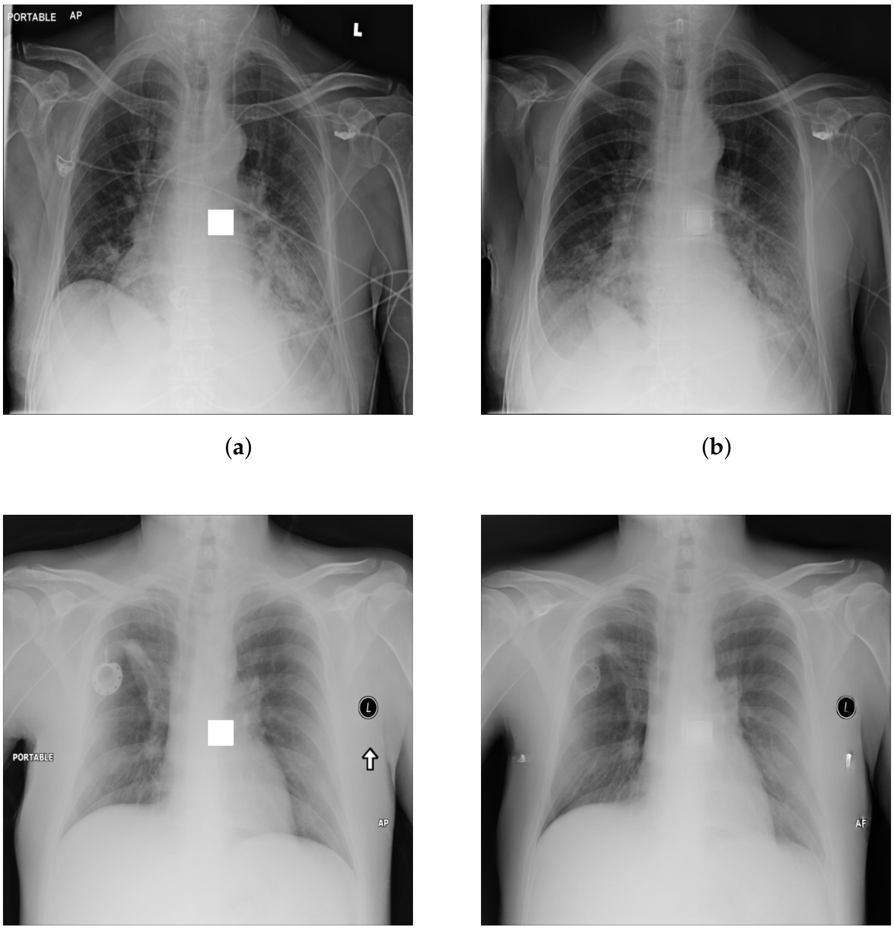
(**c**) (**d**)

Figure 4. Block obfuscation with the image domain LDP. (**a**) Original, case 1 with a block. (**b**) $\epsilon = 10^2 \cdot H \cdot W$, case 1. (**c**) Original, case 2 with a block. (**d**) $\epsilon = 10^2 \cdot H \cdot W$, case 2.



(**a**) (**b**)

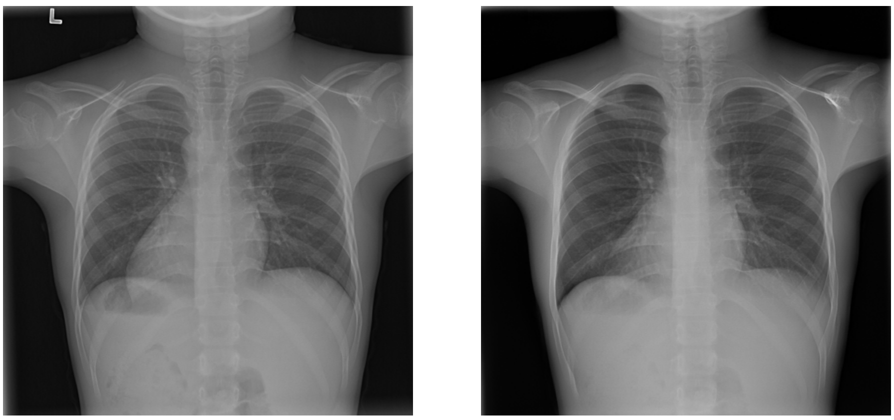
Figure 5. Flip obfuscation with the image domain LDP. (**a**) Original, case 5 (simulated *situs inversus*). (**b**) $\epsilon = 10^2 \cdot H \cdot W$, case 5.



(**a**) (**b**)

(**c**) (**d**)

Figure 6. Block obfuscation with DP-GLOW. (**a**) Original, case 1 with a block. (**b**) $\epsilon = 10^2 \cdot H \cdot W$, case 1. (**c**) Original, case 2 with a block. (**d**) $\epsilon = 10^2 \cdot H \cdot W$, case 2.



(**a**) (**b**)

Figure 7. Flip obfuscation with DP-GLOW. (**a**) Original, case 5 (simulated *situs inversus*). (**b**) $\epsilon = 10^2 \cdot H \cdot W$, case 5.

4.3. Pneumonia Detection in ϵ-LDP-Processed CXR Images

Table 3 shows the area under the curve (AUC) with different privacy budgets for ϵ-LDP-processed CXR images obtained with the image domain LDP and DP-GLOW. For details of AUC computation, see Appendix A.

Table 3. AUC for pneumonia detection (ID-LDP: Image Domain LDP).

ϵ	AUC (ID-LDP)	AUC (DP-GLOW)
∞ (without clip)	0.807	0.807
$10^3 \cdot H \cdot W$	0.813	0.679
$10^2 \cdot H \cdot W$	0.559	0.665
$10^1 \cdot H \cdot W$	0.643	0.539

5. Discussion

We showed an algorithm (DP-GLOW) for generating useful images for diagnosis and medical AI, while ϵ-LDP is guaranteed against any image in the training distribution. Furthermore, this is the first study to apply an ϵ-LDP algorithm against a medical image itself. Additionally, we validated the usefulness of the ϵ-LDP CXR images generated by AI for pneumonia detection: we showed the AUC as a function of the privacy budget. Finally, this is the first work to adopt flow-based DGMs for LDP processing.

For DP-GLOW, the AUCs for pneumonia detection significantly change from 0.539 to 0.807, while the privacy budget varies from $10^1 \cdot H \cdot W (= 2,621,440)$ to ∞. This means that this range of the privacy budget is indeed meaningful, whereas the privacy budget is very large compared with usual values of ϵ-LDP for scalar quantities. This finding implies that we must normalize the privacy budget so that we can consistently handle ϵ-LDP for vector quantities. To normalize the budget, we compute the privacy budgets per image pixel. To this end, we intentionally indicated the privacy budget to have a common factor $H \cdot W$. Therefore, the actual privacy budgets per image pixel in this study are from 10^1 to ∞, which are not much larger than commonly accepted privacy budgets.

Most of the approximate forms in CXR images are preserved and privacy is not protected with the image domain LDP. On the other hand, given a low privacy budget, DP-GLOW deforms the image so much that individuals cannot be identified. However, the AUCs for pneumonia detection are similar with the low privacy budget between DP-GLOW and the image domain LDP.

This study has several future directions. First, we adopted CXR images but one can adopt other kinds of image including nonmedical images. Second, we can readily extend this method to three-dimensional (3D) flow-based DGMs to generate ϵ-LDP 3D images, such as CT (Computed Tomography) and MR (Magnetic Resonance) images. Third, we can use other flow-based DGMs different from GLOW. Fourth, although we have formulated our LDP algorithm using flow-based deep generative models, we are now attempting to adapt it for denoising diffusion probabilistic models [18].

This study has a limitation—we must train GLOW with many medical images to generate ϵ-LDP-processed medical images using DP-GLOW. The training is very difficult when medical images of interest are not available. However, once the GLOW is trained with the medical images of interest, we can distribute the GLOW model to further release medical images of the same kind to the public using DP-GLOW. Empirically, we know that at least 1000 images are needed to train the GLOW model satisfactorily. However, the number of images required may vary if we use a different flow-based deep generative model. Additionally, due to memory limitations, it is expected that we cannot process large-sized 3D images. Furthermore, DP-GLOW cannot handle medical images with varying resolutions.

6. Conclusions

We proposed DP-GLOW, the ϵ-LDP algorithm for images built upon the flow-based DGMs, which can simultaneously ensure provable privacy protection and usefulness, e.g., the preservation of pathologies, with a controllable privacy budget. The ϵ-LDP-processed CXR images obtained with DP-GLOW indicate that we have obtained a powerful tool to release and use medical images for training AI. Furthermore, DP-GLOW could benefit other areas such as images obtained from surveillance cameras and those uploaded to social networking services.

Author Contributions: Conceptualization, H.S. and S.H.; methodology, H.S. and S.H.; software, H.S.; validation, H.S.; formal analysis, H.S. and S.H.; investigation, H.S.; data curation, H.S.; writing—original draft preparation, H.S.; writing—review and editing, H.S., S.H., Y.C., M.Y., T.T., Y.N., N.H. and O.A.; visualization, H.S.; supervision, N.H. and O.A.; project administration, S.H.; funding acquisition, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JST, CREST Grant Number JPMJCR21M2, Japan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: This study handled only a dataset (the RSNA Pneumonia Detection Challenge dataset) open to the public.

Data Availability Statement: The RSNA Pneumonia Detection Challenge dataset is open to the public [14].

Acknowledgments: Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital is sponsored by HIMEDIC Inc. and Siemens Healthcare K.K.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area Under the Curve
CXR	Chest X-Rays
DGM	Deep Generative Model
DP	Differential Privacy
GPU	Graphical Processing Unit
LDP	Local Differential Privacy
SGD	Stochastic Gradient Descent

Appendix A. Pneumonia Detection with GLOW

Using the density estimation obtained by two trained flow-based DGMs and Bayes' theorem, Shibata et al. [19] proposed the computation of the logarithm posterior $\log p(C_n|\boldsymbol{x}_i)$, where C_n is a classification label that an image is a normal case, as follows:

$$\log p(C_n|\boldsymbol{x}_i) = \log p(\boldsymbol{x}_i|C_n) - \log p(\boldsymbol{x}_i) + \log p(C_n), \tag{A1}$$

where $\log p(\boldsymbol{x}_i|C_n)$ is a conditional likelihood that we can estimate with a flow-based DGM trained with images of normal cases ($\mathcal{S}^{\text{train}}_{\text{normal}}$), $\log p(\boldsymbol{x}_i)$ is a likelihood that we can estimate with the other flow-based DGM trained with images of normal and abnormal cases ($\mathcal{S}^{\text{train}}_{\text{mixture}}$), and $\log p(C_n)$ is a constant, which we can safely neglect when we draw the receiver operating characteristic curve (ROC curve) and thus when computing the area under the curve (AUC). We adopt the logarithm posterior to detect lung opacity suggesting pneumonia and other abnormalities from CXR images. We share the model trained with $\mathcal{S}^{\text{train}}_{\text{mixture}}$ between DP-GLOW for generating ϵ-LDP CXR images and pneumonia detection with GLOW.

References

1. Alzheimer's Disease Neuroimaging Initiative. Available online: https://adni.loni.usc.edu/ (accessed on 1 November 2022).
2. Dwork, C. Differential privacy: A survey of results. In Proceedings of the International Conference on Theory and Applications of Models Of Computation, Xi'an, China, 25–29 April 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19.
3. Erlingsson, Ú.; Pihur, V.; Korolova, A. Rappor: Randomized aggregatable privacy-preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, 3–7 November 2014; pp. 1054–1067.

4. 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, Vienna, Austria, 24–28 October 2016; pp. 308–318.
5. Ziller, A.; Usynin, D.; Braren, R.; Makowski, M.; Rueckert, D.; Kaissis, G. Medical imaging deep learning with differential privacy. *Sci. Rep.* **2021**, *11*, 13524. [CrossRef] [PubMed]
6. Kossen, T.; Hirzel, M.A.; Madai, V.I.; Boenisch, F.; Hennemuth, A.; Hildebrand, K.; Pokutta, S.; Sharma, K.; Hilbert, A.; Sobesky, J.; et al. Toward Sharing Brain Images: Differentially Private TOF-MRA Images with Segmentation Labels Using Generative Adversarial Networks. *Front. Artif. Intell.* **2022**, *5*, 85. [CrossRef] [PubMed]
7. Fan, L. Image pixelization with differential privacy. In Proceedings of the IFIP Annual Conference on Data and Applications Security and Privacy, Bergamo, Italy, 16–18 July 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 148–162.
8. Fan, L. Differential privacy for image publication. In Proceedings of the Theory and Practice of Differential Privacy (TPDP) Workshop, London, UK, 11 November 2019; Volume 1, p. 6.
9. Croft, W.L.; Sack, J.R.; Shi, W. Obfuscation of images via differential privacy: From facial images to general images. *Netw. Appl.* **2021**, *14*, 1705–1733. [CrossRef]
10. Liu, B.; Ding, M.; Xue, H.; Zhu, T.; Ye, D.; Song, L.; Zhou, W. DP-Image: Differential Privacy for Image Data in Feature Space. *arXiv* **2021**, arXiv:2103.07073.
11. Li, T.; Clifton, C. Differentially private imaging via latent space manipulation. *arXiv* **2021**, arXiv:2103.05472.
12. Croft, W.L.; Sack, J.R.; Shi, W. Differentially private facial obfuscation via generative adversarial networks. *Future Gener. Comput. Syst.* **2022**, *129*, 358–379. [CrossRef]
13. Kingma, D.P.; Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 10236–10245.
14. Shih, G.; Wu, C.C.; Halabi, S.S.; Kohli, M.D.; Prevedello, L.M.; Cook, T.S.; Sharma, A.; Amorosa, J.K.; Arteaga, V.; Galperin-Aizenberg, M.; et al. Augmenting the national institutes of health chest radiograph dataset with expert annotations of possible pneumonia. *Radiol. Artif. Intell.* **2019**, *1*, e180041. [CrossRef] [PubMed]
15. Dinh, L.; Krueger, D.; Bengio, Y. Nice: Non-linear independent components estimation. *arXiv* **2014**, arXiv:1410.8516.
16. Dinh, L.; Sohl-Dickstein, J.; Bengio, S. Density estimation using real nvp. *arXiv* **2016**, arXiv:1605.08803.
17. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467.
18. Shibata, H.; Hanaoka, S.; Nakao, T.; Kikuchi, T.; Nakamura, Y.; Nomura, Y.; Yoshikawa, T.; Abe, O. Practical Medical Image Generation with Provable Privacy Protection based on Denoising Diffusion Probabilistic Models for High-resolution Volumetric Images. *TechRxiv* **2023**. [CrossRef]
19. Shibata, H.; Hanaoka, S.; Nomura, Y.; Nakao, T.; Sato, I.; Sato, D.; Hayashi, N.; Abe, O. Versatile anomaly detection method for medical images with semi-supervised flow-based generative models. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 2261–2267. [CrossRef] [PubMed]