On the Trade-off Between Privacy Protection and Data Utility for Chest X-ray Images

Truong Giang Vu
College of Engineering and Computer Science
VinUniversity
Hanoi, Vietnam
20giang.vt@vinuni.edu.vn

Nursultan Makhanov
Department of Computer Science
Nazarbayev University
Nur-Sultan, Republic of Kazakhstan
nursultan.makhanov@nu.edu.kz

Nguyen Anh Tu
Department of Computer Science
Nazarbayev University
Nur-Sultan, Republic of Kazakhstan
tu.nguyen@nu.edu.kz

Kok-Seng Wong
College of Engineering and Computer Science
VinUniversity
Hanoi, Vietnam
wong.ks@vinuni.edu.vn

2022 International Conference on Advanced Technologies for Communications (ATC) | 978-1-6654-5188-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ATC55345.2022.9942970

Abstract—**The rising advancement in deep learning (DL) techniques has enabled machine learning (ML) models to assist practitioners in performing medical tasks with high accuracy. However, it also poses privacy concerns regarding how such models will proceed with medical data containing protected patient health information. Therefore, some efforts have been made to anonymize medical data to preserve data privacy while keeping the model performance high enough to avoid wrong decisions in the medical field. Nevertheless, the adversary can develop an ML model to re-identify a patient's identity by matching an arbitrary chest X-ray image with a public or leaked image dataset with high accuracy. This paper aims to find a trade-off between our privacy protection method and data utility for medical images. Specifically, we propose a solution to anonymize chest X-ray images by directly adding noise to the images to prevent verification attacks and evaluate how well those images can maintain good performance in the lung disease classification task. Simulation results on real-world datasets show that the proposed solution achieved a good trade-off between privacy protection and data utility.**

Index Terms—**Data Privacy, Medical Image Classification, Differential Privacy, Data Utility**

I. Introduction

Chest X-ray is considered as the most common and accessible type of radiological examination procedure as it accounts for at least one third of all exams in a typical radiology department [1]. It enables scientists and medical professionals to observe the health condition of human lungs and diagnose our respiratory systems. In the time of Covid-19, X-ray examination has allowed medical personnel to identify the existence of Covid-19 on human lungs. For instance, several studies have shown some relationships between the lung abnormalities found on chest radiographs of Covid-19 positive patients and their disease severity [2], [3]. Moreover, for the ML/DL community, chest X-ray serves as a wealthy data source to build ML/DL models [4], [5] that can assist medical practitioners to diagnose patients.

To build ML/DL models, various public medical chest X-ray image datasets [5], [6] have been published online for academic or medical usage with personal information stripped off to protect patients' privacy. However, millions of medical images with sensitive patient information are exposed online, including X-rays, ultrasounds, and Computed Tomography (CT) scans. Moreover, as experimented in [7], publicly available medical chest X-ray datasets are not entirely anonymous. Unfortunately, privacy protection for these unstructured data is much more complex than that for structured data because data attributes of images are implicitly represented by sets of pixels covering irregular shapes and sizes [8].

This work is motivated by [7] where the authors raise the question of how publicly available X-ray images can remain anonymous. In this paper, we aim to decrease the patient verification performance by adding various type of statistical noises to the images and observe how well they can maintain their utility, i.e., for the disease classification task. We propose a method to anonymize chest X-ray images by modifying those images directly with noise. We conducted experiments to find the trade-off between the amount of noise to be added with the reduction in performance of the patient verification task, and the performance of the lung disease classification task.

The paper is structured as follow: In Section II, we briefly present background and previous work related to image classification task for X-ray images and differential privacy. Our proposed methodology is written in detail in Section III. In Section IV, we present our experiment setup, empirical results and evaluation in the patient verification task and X-ray disease classification task after applying our method. We discuss and summarize our findings in Section V and VI.

II. Background and Related Works

In this section, we provide the background on image classification approaches, differential privacy and related works.

A. Background

Deep learning has emerged to dominate the Computer Vision field since 2012 when Krizhevsky et al. introduced Deep Convolutional Neural Networks (CNNs) called AlexNet showing tremendous results in image classification task. It led to the development of deeper models such as VGG, GoogleNet, ResNet, SENet, etc. and becoming the state-of-the-art algorithms in the image classification problem. Superiority of finding such complex relationships in images was due to various convolution operations which use pooling and sampling strategies. Typical CNNs are composed of different

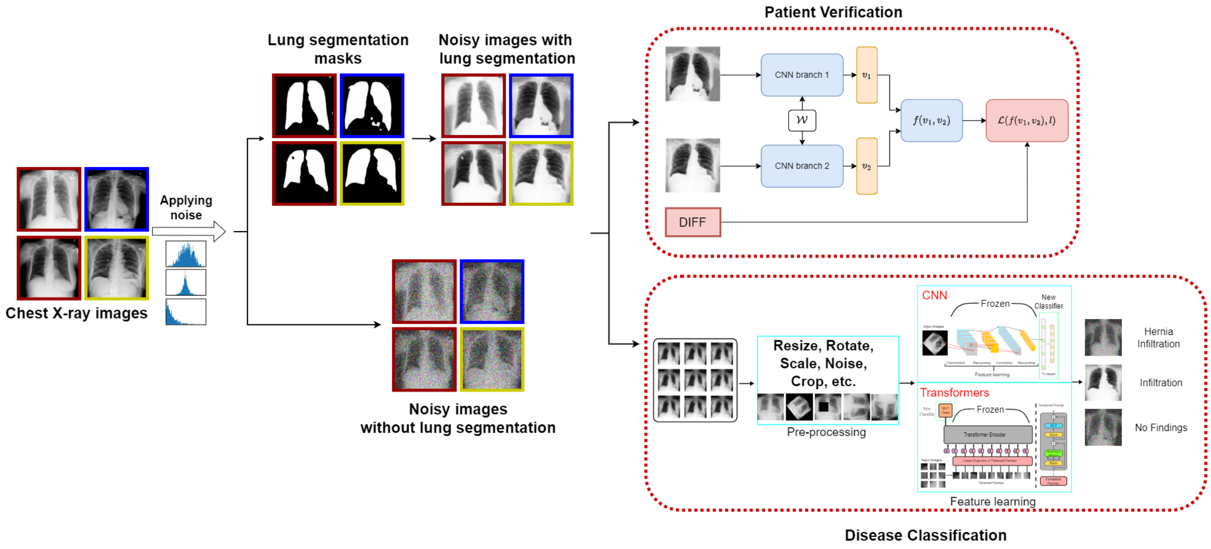


Fig. 1. Overview of the proposed framework. Chest X-ray images and their lung segmentation masks from the same patient have the same border color.

layers such as convolution, pooling, fully connected layers. Deep convolutional neural networks are the main choice when dealing with various image understanding tasks. However, CNNs are weak at encoding relative spatial information of certain features in the image. Recently, attention-based models [9] have shown superior results by addressing this encoding issue compared to many deep CNN-based models in Computer Vision tasks.

Differential privacy (DP) [10] is a strong notion of privacy that guarantees privacy protection in the presence of arbitrary auxiliary information. Intuitively, it aims to limit the information leakage from the output by making a small change on the inputs. It is a data-perturbation approach founded on the idea that systematically randomized modifications to a dataset or algorithm can reduce the information of a particular individual while preserving the statistical reasoning capabilities of the dataset [11]. DP offers a solution to protect against re-identification attacks such as linkage attacks while maintaining the dataset's utility. In the literature, DP has been applied in the input, output, or model updates during training.

One of DP's significant trade-offs lies in the perturbation process, i.e., data manipulation can possibly degrade the data. In the field of medical imaging, where the data is scarce and the results are sensitive to the human patients' outcomes, ill-formed data may be detrimental to the performance of algorithms [11]. This raises a question: how should we modify the data so that it can protect patients' privacy while still retaining good use for medical imaging tasks?

B. Related Works

Along with X-ray radiography, the recent development of machine learning techniques has allowed computers to detect abnormalities on chest X-ray images with high accuracy, aiming to assist medical doctors in diagnosing patients' conditions while avoiding human error and biases. For example, several pre-trained CNN models such as ResNet, VGG, XCeption, etc. have been utilized to extract features from chest X-ray images to detect pneumonia [4]. Notably, many researchers try to take advantage of these CNNs since they show huge potential. For instance, Wang et al. [6] proposed a large scale X-ray dataset to classify 8 chest pathologies. Further, it was modified with 6 more pathologies and Rajpurkar et al. [12] proposed ChexNet model to detect Pneumonia. Irvin et al. [5] proposed CheXpert datasets with 14 pathologies. Authors applied DenseNet121 architecture as a baseline and achieved pretty good results.

Medical images are an essential resource for diagnosing, monitoring, and treating diseases. There are several efforts to preserve the privacy protection of medical images while maintaining data utility in the literature. Ziller [13] implemented the Differential Privacy Stochastic Gradient Descent (DP-SGD) algorithm, which modifies user-supplied neural network architecture and adds noise to per-sample gradient to provide a formal privacy guarantee regardless of the dataset, learning task, and of model selection. Ziegler [14] incorporated Federated Learning with Rényi differential privacy and Gaussian noise mechanism to protect against data reconstruction attacks.

However, it is shown in [7] that chest X-ray images are vulnerable to linkage attacks. By retraining a pretrained ResNet-50 model [15], the attacker can detect whether two chest X-ray images belong to the same patient with high accuracy; and compare a given radiograph with public datasets to retrieve a list of chest X-ray images that are considered to be similar with. In case that the adversary has a chest X-ray image of a patient and has access to an arbitrary chest X-ray dataset, that patient is at risk of having his/her health conditions (over a period of time) exposed; since such a dataset usually contains the patient's history of chest radiography.

III. Proposed Methodology

As our work protects users' lung images and maintains utility, we divide our framework into two parts (as shown in Fig. 1). The first part (Patient Verification) considers the methodology for applying noise to the X-ray images and use those noisy images for patient verification task through Siamese network [16]. The second part (Disease Classification) considers image classification approaches to show the impact of noisy images on the classification performance.

A. Applying Noise on Chest X-ray Images

Our work involves the use of noise following one of the three statistical distribution: Gaussian, Exponential and Laplacian distributions.

Gaussian noise [17] is one of the most frequently occurring types in image processing as it happens under some reasonable circumstances in practice such as thermal, lighting, grain, etc. The density function of a value x in the univariate Gaussian noise $Q \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2 is:

$$p_Q(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\right) \tag{1}$$

The Exponential noise [17] is often used to model speckle in images. It is drawn from the exponential distribution $Q \sim Exp(\lambda)$, which has the following density function for $x > 0$ and variance $1/\lambda^2$:

$$p_Q(x) = \lambda \exp(-\lambda x) \tag{2}$$

In the experiment, the scale parameter $\beta = 1/\lambda$ will be used instead. The above formula is rewritten as below:

$$p_Q(x) = \frac{1}{\beta} \exp(-x/\beta) \tag{3}$$

Laplacian noise [18] is a type of noise that has the following density function: For a noise distribution $Q \sim Laplace(\mu, \lambda)$ having mean μ and variance $2\lambda^2$ $(\lambda > 0)$:

$$p_Q(x) = \frac{1}{2\lambda} \exp\Big(-\frac{|x-\mu|}{\lambda}\Big) \tag{4}$$

To apply noises on chest X-ray images, we have two options to apply the noise: either to the region outside the lung, or to the whole image. For the first option, we employed a lung segmentation model based on U-Net architecture from the RSNA Pneumonia Detection Challenge [19] to identify the lung area. We created a binary mask $\mathcal{M}$ over an image *img* to cancel the noise inside the lung. For the second option, the binary mask $\mathcal{M}$ is the whole image (with all values in the mask being 1).

The noise ε is generated as a tensor having the same shape as *img* and follows one of the statistical distributions: Gaussian, Laplace, and Exponential. The parameters for each noise distribution is listed in Table I. In the experiment, we denote noise distributions with $\lambda = 0$ or $\beta = 0$ as noise tensors having values 0 (no noise).

Afterwards, we created another noise ε_L by cancelling the noise ε in the masked area to avoid disturbing the actual lung, since we assume the quality of the inside-lung image is critical to the disease classification task. We applied an element-wise product $\odot$ of the binary lung mask tensor $\mathcal{M}$ and the noise tensor ε: Image points inside the mask have the value 0 in the $\mathcal{M}$, and therefore cancel the noise: $\varepsilon_L = \varepsilon \odot \mathcal{M}$. For an image *img*, it is converted from 8-bit format (value range is $[0-255]$) to floating-point format (with values in range $[0, 1]$). Then the noise ε_L is added to the image by a tensor addition operation, producing a "noisy image" img_N for *img*: $img_N = img + \varepsilon_L$. The image img_N is then converted back to the 8-byte format for model usage. Hereafter, we refer images that have noise by using this method as "noisy images", and "clean images" otherwise.

TABLE I
NOISE PARAMETERS

Distribution	Parameters
Gaussian	$\mu = 0$ $\sigma \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$
Exponential	$\beta \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$
Laplace	$\mu = 0$ $\lambda \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$

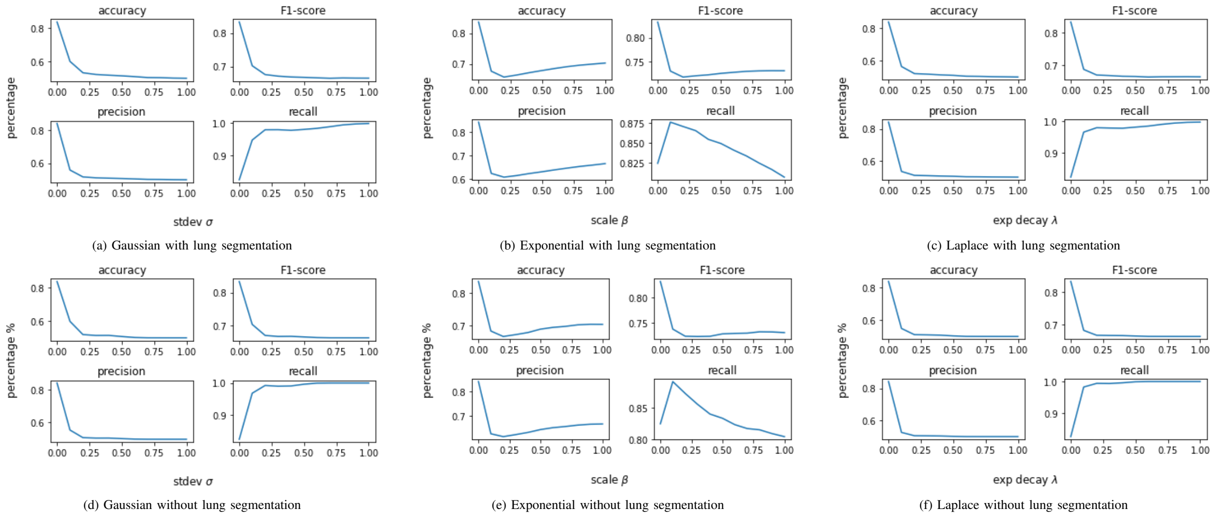
B. Siamese Network Architecture for Patient Verification

The Siamese Network Architecture is designed to learn the similarity metrics between two inputs [16]. It consists of two CNN branches sharing the same weights to calculate the two feature representation vectors of the input pair, a merge layer f to merge these results, and a loss layer $\mathcal{L}$ to calculate the loss value between the merged output and the label as illustrated in Fig. 1. This architecture is used for the Patient Verification task to match the features of a pair of chest X-ray images. We obtained the pretrained ResNet-50 model to serve as the backbone CNN branch for feature extraction of input images in the Patient Verification Model (PVM).

C. Transfer Learning for Disease Classification

Many research works have shown that deep features are useful for image classification. Extracting such features is more robust and effective than using hand-crafted image features. The technique based on the use of deep features is widely recognized as transfer learning, a research problem to store and apply knowledge or skills learned in several tasks to a target task. Specifically, the main goal is to obtain a robust feature representation for the target domain.

Disease classification part of Fig. 1 shows the logic behind testing the utilization of given X-ray images. We included pre-trained DL models as Vision Transformer (ViT), DenseNet121, and ResNet-50 and froze the intermediate layers. We kept the pre-trained weights of initial architectures to take advantage of existing models in order to reduce the training time. DL models were employed to process X-ray images and extract useful features of diseases. The extracted features were supplied into new network which acts as a classifier for 14 given diseases. All models share common hyper-parameters, i.e., the classifier consists of one linear layer followed by a sigmoid activation function. Such method is used to avoid re-training the whole model from scratch which requires a lot of computational power and time.



(a) Gaussian with lung segmentation

(b) Exponential with lung segmentation

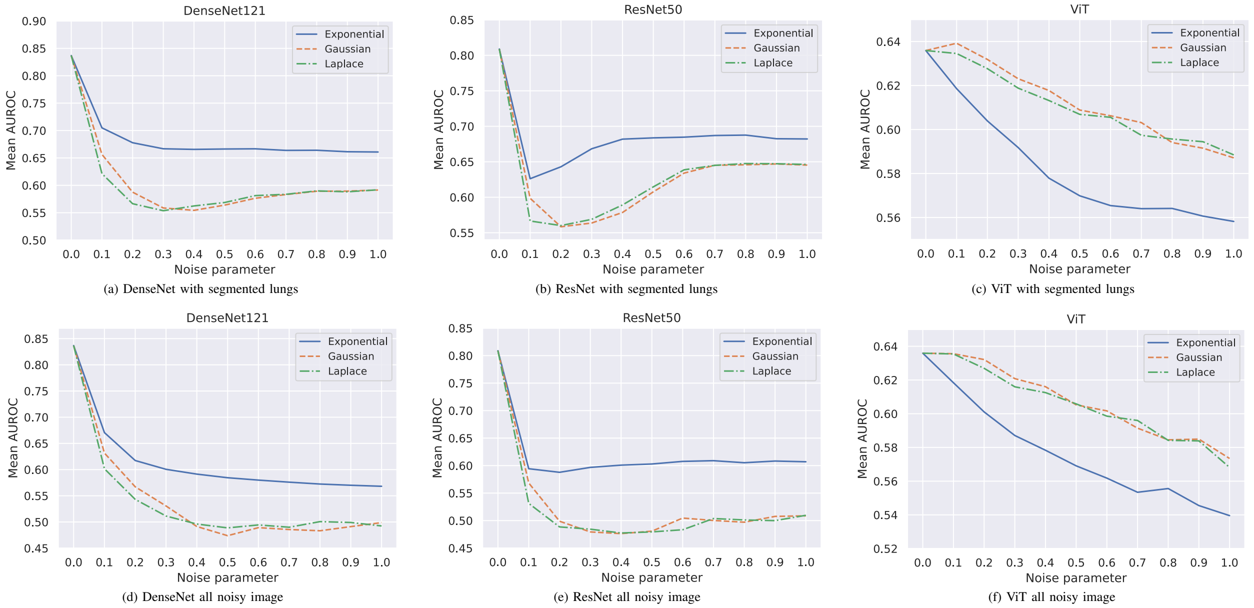
(c) Laplace with lung segmentation

(d) Gaussian without lung segmentation

(e) Exponential without lung segmentation

(f) Laplace without lung segmentation

Fig. 2. Verification performance for different setups (noise parameters, with/without lung segmentation)



(a) DenseNet with segmented lungs

(b) ResNet with segmented lungs

(c) ViT with segmented lungs

(d) DenseNet all noisy image

(e) ResNet all noisy image

(f) ViT all noisy image

Fig. 3. Mean AUROC scores for noisy images with/without lung segmentation

IV. Experimental Evaluation

Dataset and Augmentations: In an effort to show fair comparison in patient verification and disease classification tasks, we utilize the NIH Chest X-ray dataset [6], which is one of the largest publicly available chest X-ray datasets. It consists of over 112,000 deanonymized X-ray images from 30,805 unique patients with 14 pathologies. For classification task, we resized the images to 224 pixels, applied Random Horizontal Flip and Center Crop operations. Validation and Test images were resized to 256 pixels both for verification and classification tasks.

Patient Verification task: In this experiment, we reproduced the PVM from [7] to examine the performance of the verification task. The model takes two chest X-ray images and produces a boolean output: 0 if two images do not belong to the same person, and 1 otherwise. In the verification task, the following metrics were computed to evaluate the performance: accuracy, F1-score, precision, and recall. To conduct the patient verification task, we used the Kaggle service to train and test PVM as Python notebooks with GPU enabled in one run.

Disease Classification task: Learning rate was set to 0.001 with decay rate by 0.1 each round when learning does not happen. We also stopped the training process when validation shows no improvement in 3 consequent rounds. We set the batch size as 48 for DenseNet121 and ResNet-50, but we had to change the batch size to 16 on ViT model as it takes a lot of hardware resources. We set epoch numbers as 50 for all the models. To track the learning process we used Binary Cross Entropy (BCE) loss. BCE loss is a popular choice when working with multi-label classification task. Mean AUROC metric was used to test the performance of the models.

A. Patient Verification Performance

We used $400{,}000$ clean image pairs in the training phase, and $10{,}000$ fixed noisy image pairs with noise parameter values from Table I in the testing phase. While testing noisy images were repeatedly used for different noise parameters, there was no patient overlap between the training images and the testing images.

As shown in Fig. 2, there is a significant drop in accuracy, F1-score and precision metrics when noise is added to the images. In particular, accuracy and precision have an approximate 30% drop when we increase the parameter from 0 to 0.2 for Gaussian and Laplacian noise, and 20% for Exponential noise. F1-score experiences a 20% drop for Gaussian and Laplacian noise and over 10% for Exponential noise. However, for Gaussian and Laplacian noises, the recall metric gains substantially as the noise parameter grows. In Figs. 2(a), 2(c), 2(d) and 2(f), as the noise parameter reaches 1.0, the recall metric peaks at nearly 1.0. High recall signifies that the PVM has a high chance of correctly detecting two images belong to the same patient if it is true. It can also be seen that the Gaussian noise has similar behavior as the Laplacian does, due to the similarity between these two distributions: For the same μ, both are symmetrical about the mean μ. However, Gaussian distribution results in a bell-shaped curve, while Laplacian one tends to have a sharper peak towards the mean.

On the other hand, Exponential noise behaves differently from Gaussian and Laplacian. Figs. 2(b) and 2(e) show that the verification accuracy, F1-score and precision reach the lowest value of $0.6575, 0.7165, 0.6089$ respectively when $\beta = 0.2$. Meanwhile, also at $\beta = 0.2$, the recall is at its peak of 0.8705. For $\beta > 0.2$, all four metrics tends to change slowly but in different directions. Unlike Gaussian or Laplacian noise which continues to decrease in accuracy, F1-score, precision and increase in the recall metric, Exponential noise does the opposite as those three metrics gain and the recall reduces slightly. Moreover, for each type of noise, the PVM performance is similar whether to apply lung segmentation or not. All experiment settings displayed significant changes in performance metrics between testing without noise and with small noise parameter values; while for larger values, the changes were not as significant. We notice that since the PVM is trained on clean dataset, it also focuses on some region outside the lung; hence, it will behave poorly even with small amount of noise on the dataset. Therefore, the performance difference is obvious when testing between clean and noisy datasets, and is hardly noticeable among noisy datasets of various parameters.

B. Disease Classification Performance

In the first experiment, we trained all models on clean NIH dataset and we tested the results on 10,000 clean images. DenseNet121 model showed overall mean AUROC score of 0.83 which is close to ChexNet's [12] result 0.84. ResNet-50 model showed 0.81 mean AUROC score which is a good comparable result. However, ViT could show mean AUROC score of 0.64. This means that deep CNN's still learn better compared to Transformer architecture. Transformers need a lot of training images to learn by default. Nowadays Transformer based model which show state-of-the-art results in ImageNet dataset are pretrained on Google's private JFT-300 (300 mil.) and JFT-1B (1 bil.) images. We observed that the shortage of available medical data makes Transformer-based models weaker in performance compared to deep CNN models on clean X-ray images.

Another experiment we conducted is to train all models on clean dataset and test on noisy test data. All models show sudden drop when we apply a little noise except ViT. ViT's base prediction mean AUROC is 0.64 on clean X-ray images and when we apply noise parameter of 0.1 with or without lung segmentation algorithm, it shows 0.62-0.63 score for all types of noises (Figs. 3(c) and 3(f)). When we increase the noise parameter, mean AUROC scores fall gradually for images with clean segmented lungs and for noisy lung images. We can see that results with clean segmented lungs still higher than applying the noise to all parts of X-ray image. As our model trained on a clean data, we observed that all the models can see clean segmented lungs better than the noisy images with 100% coverage. Another finding we noticed is that applying exponential noise affects ViT model significantly. However, Gaussian and Laplacian noises show similar performance in both cases. Overall, we can see that ViT model is more robust to noises compared to deep CNN models' performance.

DenseNet121 and ResNet-50 showed the best performance on the NIH clean testing dataset (mean AUROC scores of 0.83 and 0.80, respectively). However, when we apply different types of noises to the images, we can see a sharp drop in the performance. For instance, when we segmented the lungs and applied Exponential noise (parameter 0.1), DenseNet's and ResNet's performance dropped by 0.12 and 0.175 scores, respectively. In Figs. 3(a) and 3(d), we observed that the increase of noise parameter caun AUROC core decreases moderately for Exponential noise. We noticed unusual behavior which can be noticed in Figs. 3(a) for DenseNet and 3(b) for ResNet models. When we increase the noise parameter of Gaussian and Laplacian noise more than 0.4, performance suddenly increases. Same situation applies for ResNet in Fig. 3(b) when we increase noise parameter from 0.3. Logical conclusion that we came up is that both DenseNet and ResNet models can distinguish better various diseases from clean segmented lungs even if we increase the noise parameter. When we generated lung segments and applied noise to the outside area of the lungs, in some cases, the pre-trained, U-Net segmentation algorithm could not identify lung areas. It led to the application of noises to the lung areas directly where the abnormal behavior was recorded.

In Figs. 3(d) and 3(e), we can see that the mean AUROC scores are impacted severely when the noise parameter is higher than 0.2. The AUROC score below 0.55 means that we cannot trust the results of the given models. Besides, Gaussian and Laplacian noises impacts the performance of CNN models more than Exponential in all cases. Overall, for CNN models the least impact was caused by Exponential noise for all cases and vice versa for ViT model.

V. Discussions

In this section, we summarize the findings for our experiments. Firstly, we can decrease the verification accuracy drastically by adding noise of small parameter values. We notice that since the PVM model focuses on some regions outside the lung, which we have made them noisy; therefore, the model performs poorly on our noisy dataset, and there are no significant changes when we increase the noise parameters.

In addition, the ViT model showed great performance when noises were added to the images. If we train the ViT model with considerable amount of training data, we might get better performance which is robust to the noise. This can be attributed to the transformer architecture's attention module which sees the whole picture while convolution operation in CNN's sees only neighboring pixels. During the training phase of disease classification task, we set the epoch number as 50, but normally the training accuracy converged after 30 epochs. This behavior is due to the decay rate which dropped the learning rate by 0.1 each round when the learning ability was stable for 3 epochs.

Another observation we found is that in some cases pretrained U-Net model could not segment the lung areas because of the low X-ray image quality. Some of the X-ray images consist too much white are on lungs which makes U-Net model harder to understand the image and segment properly the lung areas. This fact greatly impacted the classification accuracy which led to low performance or even strange behaviour of the ResNet model (Figs. 3(b) and 3(e)). Further research needs to be done in finding better lung segmentation model.

VI. Conclusion

In this paper, we analyzed the trade-off between patient verification and classification accuracy of DL models when we applied Gaussian, Exponential, and Laplacian noises to X-ray images. Specifically, we proposed to segment lungs and add noises to the outside area. Unlike other solutions in the literature, we added the noise directly into the images that helps to protect the data privacy from the source. All the experiments were conducted on the real-world NIH dataset. Experimental results showed that applying a small noise parameter can significantly reduce the verification accuracy, but slightly hurt the disease classification performance compared to clean images. Thus, a solution that can enhance data privacy and preserve ML/DL performance is detrimental to one of them. Consensus towards data protection and data utility require further research in emerging cryptographic solutions.

VII. Acknowledgments

This work was supported by Faculty Development Competitive Research Grant Program (No. 11022021FD2925) at Nazarbayev University.

References

[1] B. Van Ginneken, B. Ter Haar Romeny, and M. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1228–1241, 2001.

[2] J. P. Kanne, H. Bai, A. Bernheim, M. Chung, L. B. Haramati, D. F. Kallmes, B. P. Little, G. Rubin, and N. Sverzellati, "Covid-19 imaging: what we know now and what remains unknown," *Radiology*, vol. 299, no. 3, pp. E262–E279, 2021.

[3] N. Makhanov, N. A. Tu, and K.-S. Wong, "A survey on deep learning advances and emerging issues in pneumonia and covid19 prediction," in *2022 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 96–103, IEEE, 2022.

[4] S. L. K. Yee and W. J. K. Raymond, "Pneumonia diagnosis using chest x-ray images and machine learning," in *proceedings of the 2020 10th international conference on biomedical engineering and technology*, pp. 101–105, 2020.

[5] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 590–597, 2019.

[6] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[7] K. Packhäuser, S. Gündel, N. Münster, C. Syben, V. Christlein, and A. Maier, "Is medical chest x-ray data anonymous?," *arXiv preprint arXiv:2103.08562*, 2021.

[8] B. Liu, M. Ding, H. Xue, T. Zhu, D. Ye, L. Song, and W. Zhou, "Dp-image: differential privacy for image data in feature space," *arXiv preprint arXiv:2103.07073*, 2021.

[9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[10] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*, pp. 265–284, Springer, 2006.

[11] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.

[12] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[13] A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, and G. Kaissis, "Medical imaging deep learning with differential privacy," *Scientific Reports*, vol. 11, no. 1, pp. 1–8, 2021.

[14] J. Ziegler, B. Pfitzner, H. Schulz, A. Saalbach, and B. Arnrich, "Defending against reconstruction attacks through differentially private federated learning for classification of heterogeneous chest x-ray data," *arXiv preprint arXiv:2205.03168*, 2022.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[16] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *2016 23rd international conference on pattern recognition (ICPR)*, pp. 378–383, IEEE, 2016.

[17] C. Boncelet, "Image noise models," in *The essential guide to image processing*, pp. 143–167, Elsevier, 2009.

[18] S. Kotz, T. J. Kozubowski, and K. Podgorski, "The laplace distribution and generalizations: A revisit with applications to communications," *Economics, Engineering, and Finance*, vol. 183, 2001.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.