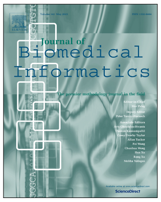
Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin



PRIMIS: Privacy-Preserving Medical Image Sharing via Deep Sparsifying Transform Learning with Obfuscation[⋆]

Isaac Shiri[a,b], Behrooz Razeghi[c,d], Sohrab Ferdowsi[e], Yazdan Salimi[a], Deniz Gündüz[f], Douglas Teodoro[e], Slava Voloshynovskiy[c,*], Habib Zaidi[a,g,h,i,**]

[a] *Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland*
[b] *Department of Cardiology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland*
[c] *Department of Computer Science, University of Geneva, Switzerland*
[d] *Idiap Research Institute, Switzerland*
[e] *Department of Radiology and Medical Informatics, University of Geneva, Switzerland*
[f] *Department of Electrical and Electronic Engineering, Imperial College London, UK*
[g] *Department of Nuclear Medicine and Molecular Imaging, University of Groningen, The Netherlands*
[h] *Department of Nuclear Medicine, University of Southern Denmark, Denmark*
[i] *University Research and Innovation Center, Óbuda University, Budapest, Hungary*

ARTICLE INFO

Article history:

Privacy
Sparse Coding
Obfuscation
Medical Image Sharing
Representation Learning

ABSTRACT

Objective: The primary objective of our study is to address the challenge of confidentially sharing medical images across different centers. This is often a critical necessity in both clinical and research environments, yet restrictions typically exist due to privacy concerns. Our aim is to design a privacy-preserving data-sharing mechanism that allows medical images to be stored as encoded and obfuscated representations in the public domain without revealing any useful or recoverable content from the images. In tandem, we aim to provide authorized users with compact private keys that could be used to reconstruct the corresponding images. **Method:** Our approach involves utilizing a neural auto-encoder. The convolutional filter outputs are passed through sparsifying transformations to produce multiple compact codes. Each code is responsible for reconstructing different attributes of the image. The key privacy-preserving element in this process is obfuscation through the use of specific pseudo-random noise. When applied to the codes, it becomes computationally infeasible for an attacker to guess the correct representation for all the codes, thereby preserving the privacy of the images. **Results:** The proposed framework was implemented and evaluated using chest X-ray images for different medical image analysis tasks, including classification, segmentation, and texture analysis. Additionally, we thoroughly assessed the robustness of our method against various attacks using both supervised and unsupervised algorithms. **Conclusion:** This study provides a novel, optimized, and privacy-assured data-sharing mechanism for medical images, enabling multi-party sharing in a secure manner. While we have demonstrated its effectiveness with chest X-ray images, the mechanism can be utilized in other medical images modalities as well.

[⋆] I. Shiri, B. Razeghi and S. Ferdowsi contributed equally to this work.
[⋆⋆] Implementation codes available at: https://github.com/sssohrab/PRIMIS.

[*] Corresponding author, svolos@unige.ch (Slava Voloshynovskiy)
[**] Corresponding author, habib.zaidi@hcuge.ch (Habib Zaidi)

1. Introduction

A prime challenge in data science is finding a balance between the need to extract useful features from data and the need to protect the privacy of individuals and proprietary information behind the data [1, 2, 3]. In diagnostic imaging, large volumes of data are created routinely in hospitals, which are often necessary to be shared with various departments or centers as part of clinical practice or research [1, 2, 3]. However, this can raise concerns regarding the sensitive and private information of patients being exposed and/or patients' privacy being violated [1, 2, 3]. To address the above-mentioned issues, data governance frameworks must carefully consider appropriate data handling and sharing approaches that ensure and minimize risks of sensitive information leakage [1, 2, 3]. This requires a careful balance between the need for data representation and the need to safeguard data privacy [1, 3].

With the widespread availability of imaging and the need for automated algorithms to analyze these images, machine learning (ML) has been successfully applied for different tasks [4]. To ensure the development of reliable and effective ML models, it is crucial to have access to large and diverse sets of images acquired under various settings considering different devices and environments, as well as different acquisition and reconstruction protocols [5]. However, sharing medical images is highly restricted owing to legal and ethical concerns and strict privacy regulations [6, 7, 8, 9]. Most medical images are only available within a single department or hospital, as it is difficult to completely de-identify personal information from medical images and guarantee patients' privacy [10]. Moreover, techniques like three-dimensional rendering of tomographic images such as MRI, CT and PET may potentially reveal the patient's identity. Therefore, it is essential to carefully consider and evaluate the ethical and legal concerns of sharing medical images.

There are various strategies to design ML algorithms that take into account data privacy concerns [10, 6, 7, 8, 9]. For example, in distributed learning, local models are developed based only on individual center's data, and shared with a central server to create a generalized model that works with different datasets [11]. By iterating between local training and model fusion at the central server, on the other hand, fully decentralized learning involves distributing computation across different parties without the use of a central server [12]. In cases where data owners, such as healthcare centers, need to share their data with trusted parties, multiparty data-sharing mechanisms may be necessary [10]. Several techniques, such as cryptographic [13], differential privacy [14], generative adversarial [15], and embedding [16], have been developed to protect data privacy in these scenarios. Conventional image-sharing algorithms, such as obfuscation techniques (pixelization and blurring) and encryption, have been outperformed by new techniques and approaches [17, 18, 19, 20]. Sparse coding with ambiguation (SCA) is a recent technique that allows for the sharing of compressed and obfuscated images with authorized parties [17, 18, 19, 20]. These parties can then grant access and regenerate the original images. SCA has been shown to be a promising solution for preserving privacy in data release techniques for various applications, including identification [17], near neighbor search [18, 19], and image sharing [20].

Privacy-preserving data release mechanisms are methods or technologies designed to protect individuals' privacy when their personal data are collected, stored, or shared [21, 22]. These mechanisms can be applied in various contexts. One possible example of a privacy-assuring data release mechanism is the use of data anonymization techniques, which aim at removing or obscuring personal information from datasets [23]. This can be done in various ways, such as *de-identification*, which removes or replaces critical information with random values or codes; *generalization*, which involves replacing specific values with more general categories; and *suppression*, which involves removing data altogether [24]. However, it is important to note that data anonymization is not a perfect solution for privacy preservation, as it may not always be possible to remove personal sensitive information from datasets completely [25, 26]. There is also the risk of re-identification, where third parties are able to link the anonymized data back to a specific individual through other sources of information [27]. Another example of a privacy-preserving data release mechanism is the use of encryption, which is a method of encoding data so that only individuals with valid decryption keys can access (decode) it [28, 29]. Encrypting personal data can help protect it from unauthorized access or disclosure, but it is not a guaranteed solution. For example, the encrypted data may become vulnerable to unauthorized access if the decryption keys are lost or stolen or if there is a deficiency in the designed encryption algorithm itself [30].

These technical measures are required by various legal and policy frameworks to help ensure personal data privacy [31, 32]. Data protection laws and regulations, such as the General Data Protection Regulation (GDPR) [31] in Europe, and the Health Insurance Portability and Accountability Act (HIPAA) [32] in the US, establish rules for the collection, use, and disclosure of personal data and provide individuals with the right to access, correct, or delete their data [31, 32]. These laws and regulations also include provisions for data breach notification, which require organizations to notify individuals and authorities when their personal data have been compromised [31, 32]. However, the effectiveness of these frameworks can vary depending on the jurisdiction in which they are applied [33, 34].

Medical imaging uses various cutting-edge privacy protection technologies to safeguard patients' privacy and their sensitive (private) information. *Technical approaches* and *legal and policy measures* are the two broad categories into which these techniques can be split. Technical approaches employ various techniques and methods to protect the confidentiality of medical images and data. Regulations for the gathering, use, and disclosure of personal health information are set forth by legal and policy measures.

Examples of technical privacy-preservation techniques applied in medical imaging include:

Anonymization. This includes removing or obfuscating personal sensitive information (attributes), such as name, gender, age, weight, address, and other identifying information from medical images, and data [35, 36, 37, 38]. Anonymization can be achieved using various methods, such as removing this information, perturbation, which involves adding random noise to the data to make it difficult to identify individuals, or replacing personal information with random codes or values [39]. However, even with de-identification and anonymization in medical images, the images may contain certain protected information. For example, some identifiable information could appear in ultrasound radiographic and mammographic images due to additional objects linked to the patient's information during image acquisition [39]. Moreover, 3D facial reconstruction in tomographic images, such as PET, CT and MRI could reveal the patient's identity in head and neck region imaging.

Encryption. Encryption is a widely used privacy-enhancing mechanism in medical imaging [40]. It requires the use of complex mathematical algorithms to encrypt medical images and data so that only those with the necessary decryption key can access them [41]. This technology protects patient privacy by preventing unauthorized access or disclosure of sensitive medical data. One type of encryption that has received particular attention in medical imaging is *homomorphic encryption* [42, 43]. This technique enables one to perform calculations on encrypted data without first decrypting it [44, 42]. This is useful in situations where medical images or data need to be shared or analyzed by multiple parties, as it allows collaboration without the risk of exposing sensitive information [45, 46]. However, it is important to note that homomorphic encryption can be computationally intensive and may not be suitable for all situations [45, 46]. In general, encryption is useful for protecting medical information from unauthorized access or disclosure. Keys must be secure, and encryption must be strong enough to prevent unauthorized decryption to protect patient privacy effectively [47].

Access Controls. This includes implementing policies and procedures to control who has access to medical images and data. Access control can include measures, such as authentication (verifying a person's identity before granting access to data) and authorization (only authorized personnel have access to data) [48, 49].

Image Deformation in Medical Imaging. Medical image deformation can be used to protect patient privacy by deforming identifiable features (such as facial features in tomographic images) to prevent the image from being traced to a specific individual [50]. It should be considered that while image deformation can help protect the privacy of patients in medical imaging, it can also reduce the diagnostic value of images when retrieving the original image from deformation.

Federated Learning. Federated learning is a learning approach to collaboratively train a machine learning model without sharing data across centers [6, 7, 8, 9]. This can be particularly useful in the context of medical imaging, where sensitive patient data needs to be protected while still allowing for the development and improvement of machine learning models [6, 7, 8, 9]. Several state-of-the-art privacy-preserving techniques based on federated learning have been developed for use in medical imaging [51, 52, 53, 9, 8, 54, 6, 7]. However, this approach fulfills privacy during ML algorithmic developments and is not an option for scenarios involving data sharing.

Generative Adversarial-based Learning. Generative adversarial learning methods have been used to generate and synthesize medical images that could be shared and made available in the public domain [55, 56]. The generated images could be used for different medical image analysis tasks but are not suited for clinical practice.

The following legal and policy measures can also be considered to protect patient privacy in medical imaging.

- **Data Protection Laws and Regulations:** These establish rules and guidelines for the collection, use, and disclosure of personal health information and may grant individuals the right to access, correct, or delete their personal information [31, 32].

- **Consent:** This includes obtaining the individual's consent before **(i)** collecting, **(ii)** using or **(iii)** disclosing the individual's health information [31, 32]. Depending on the individual or the situation, consent may be implied or expressed, as in the case of a written agreement [31, 32].

This study aims to design a practical and efficient privacy-preserving data-driven medical image-sharing mechanism for outsourcing medical images. For this purpose, we build upon our previous work [20] on SCA privacy-assuring mechanism, which is a generalized randomization technique that combines sparse lossy coding with ambiguity, allowing for a trade-off between utility and privacy in a principled manner [20]. We extended our previous work [20] by evaluating the SCA privacy-assuring mechanism on larger, more diverse medical images across various downstream imaging tasks. We also introduced a new architectural innovation for use in convolutional neural networks (CNNs) to improve the practicality and scalability of the SCA mechanism.

Our main contributions are:

- Introducing a novel unsupervised privacy-preserving data release mechanism for protecting patients' private information when sharing medical images.

- Using the S-sparsity inducing non-linearity along with other standard non-linearities in a CNN and showing that it does not slow down the training.

- From the deep learning perspective, we designed a fully convolutional neural network whose architecture allows compact bottlenecked codes, can produce sparse codes, and is highly efficient and seamless to train.

- Evaluating the performance of our approach on medical imaging tasks, including classification, segmentation, and texture analysis using reconstructed images.

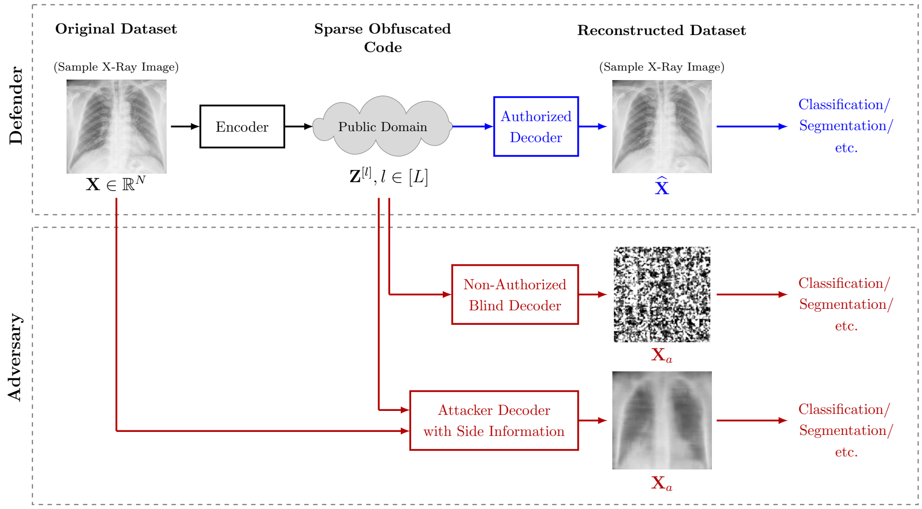

Defender
Original Dataset
(Sample X-Ray Image)
$\mathbf{X} \in \mathbb{R}^N$
Encoder
Sparse Obfuscated Code
Public Domain
$\mathbf{Z}^{[l]}, l \in [L]$
Authorized Decoder
Reconstructed Dataset
(Sample X-Ray Image)
$\widehat{\mathbf{X}}$
Classification/ Segmentation/ etc.
Adversary
Non-Authorized Blind Decoder
$\mathbf{X}_a$
Classification/ Segmentation/ etc.
Attacker Decoder with Side Information
$\mathbf{X}_a$
Classification/ Segmentation/ etc.

Fig. 1. General block diagram of the proposed Privacy-Preserving Medical Image Sharing (PRIMIS) Mechanism: A Defender with full access to the dataset releases the sparsely obfuscated code dataset to the public domain. Authorized users (highlighted in blue) employ the shared support to de-obfuscate (purify) the code and recover the original image data for applications such as segmentation or classification. Conversely, adversaries (highlighted in red), whether having access only to the public code or partial/full access to the original dataset, might also attempt to reconstruct the original data for different downstream purposes.

Problem	Medical images need confidential sharing across centers, but privacy concerns create challenges.
What is Already Known	SCA privacy-assuring mechanism, a generalized randomization technique, blends sparse lossy coding with ambiguity for a privacy-utility trade-off. Previous research has focused on this but faced scalability and diversity challenges with medical images.
What This Paper Adds	This paper introduces an unsupervised privacy-preserving mechanism specifically for medical images, integrates S-sparsity inducing non-linearity in CNNs without compromising training speed, and introduces an optimized CNN architecture for compact, sparse codes while offering comprehensive evaluations on multiple downstream medical imaging tasks.

2. Methodology

2.1. Problem Formulation

Envision a tripartite data disclosure situation that involves **(i)** a data owner, **(ii)** data users (authorized clients), and **(iii)** service provider(s)/server(s) [19, 20]. The data owner releases certain form of medical images they hold to 'honest but curious' server(s) [19, 20]. The data owner seeks to: **(i)** safeguard primary dataset from server-side examination; **(ii)** provide a specified utility to authorized users; **(iii)** safeguard the primary dataset from unauthorized entities [19, 20]. We explicitly define our measure of utility and privacy as the capability of reconstruction for authorized and unauthorized parties, respectively [19, 20]. Moreover, to adhere to Kerckhoffs's [57] principle in cryptography, we operate under the assumption that the data-release mechanism is publicly disclosed [19, 20, 57].

To encourage the network to learn somehow semantically disentangled representations, we use grouped convolutions (i.e., by grouping the convolutional filters) with CNNs [58]. Moreover, as far as the sparsification of the representations is concerned, this serves an essential practical purpose [20]. Since convolutional filters cannot be sparsified directly, as they are largely correlated, we need a fully connected linearity to diversify the activities before sparsification [20]. However, imposing sparsity on a single but large code requires a very large fully connected layer that increases the risk of over-training [20]. By independently applying smaller linear layers on top of convolutional filters, we reduce the computational cost of matrix multiplications and significantly avoid over-training [20].

2.2. Notations

In this manuscript, the superscript $(\cdot)^T$ indicates the transpose operation, while $(\cdot)^\dagger$ denotes the pseudo-inverse. Boldface lower-case letters (e.g., $\mathbf{x}$) represent vectors, while boldface upper-case letters (e.g., $\mathbf{X}$) signify matrices [19, 20]. we use identical notation for both a random vector $\mathbf{x}$ and its actual realization. The distinction between them should be evident from the context [19, 20]. The i-th element of the vector $\mathbf{x}$ is represented by x_i, while $\mathbf{x}_j$ signifies the j-th column of matrix $\mathbf{X}$ [19, 20]. In addition, we use the notation $[N]$ for the set $\{1, 2, ..., N\}$ and $\mathbb{N}^0$ for the set of non-negative integers [19, 20].

2.3. Sparse Data Representation Models

Sparse data representation has become popular thanks to its ability to significantly lower communication, storage, and computation costs [20, 59, 60]. Feature extraction, clustering, classification, and reconstruction are just a few of the signal processing applications where it has been extensively used [20, 61, 62, 63]. There are three main models for sparse representations, as follows.

Synthesis Model: Synthesis-based sparse representation model assumes that a data sample $\mathbf{x}_i \in \mathbb{R}^N$ is approximated by a linear combination $\mathbf{y}_i \in \mathbb{R}^M$ (referred to as sparse data representation) of a small number of columns (atoms) from a dictionary $\mathbf{D} \in \mathbb{R}^{N \times M}$ [20, 64], as $\mathbf{x}_i = \mathbf{D}\mathbf{y}_i + \mathbf{v}_i$, where $\|\mathbf{y}_i\|_0 \ll M$, $\mathbf{v}_i \in \mathbb{R}^N$ denotes the approximation error in the *original data domain*.

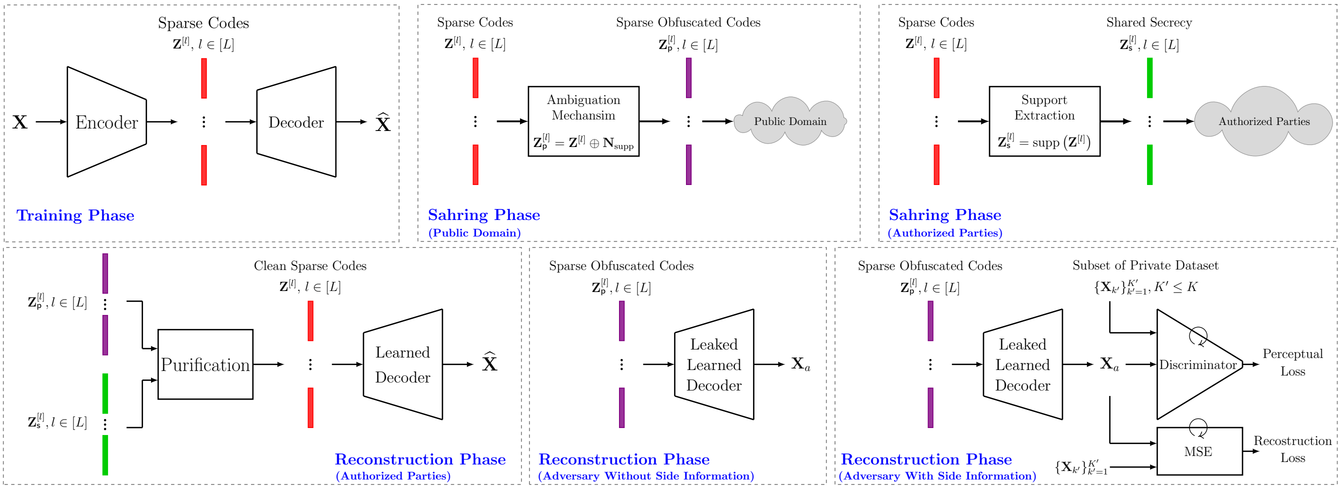

Sparse Codes
$\mathbf{Z}^{[l]}, l \in [L]$
$\mathbf{X}$
Encoder
Decoder
$\widehat{\mathbf{X}}$
Training Phase
Sparse Codes
$\mathbf{Z}^{[l]}, l \in [L]$
Sparse Obfuscated Codes
$\mathbf{Z}_{\mathsf{p}}^{[l]}, l \in [L]$
Ambiguation Mechansim
$\mathbf{Z}_{\mathsf{p}}^{[l]} = \mathbf{Z}^{[l]} \oplus \mathbf{N}_{\text{supp}}$
Public Domain
Sahring Phase
(Public Domain)
Sparse Codes
$\mathbf{Z}^{[l]}, l \in [L]$
Shared Secrecy
$\mathbf{Z}_{\mathsf{s}}^{[l]}, l \in [L]$
Support Extraction
$\mathbf{Z}_{\mathsf{s}}^{[l]} = \text{supp}\left(\mathbf{Z}^{[l]}\right)$
Authorized Parties
Sahring Phase
(Authorized Parties)
$\mathbf{Z}_{\mathsf{p}}^{[l]}, l \in [L]$
$\mathbf{Z}_{\mathsf{s}}^{[l]}, l \in [L]$
Purification
Clean Sparse Codes
$\mathbf{Z}^{[l]}, l \in [L]$
Learned Decoder
$\widehat{\mathbf{X}}$
Reconstruction Phase
(Authorized Parties)
Sparse Obfuscated Codes
$\mathbf{Z}_{\mathsf{p}}^{[l]}, l \in [L]$
Leaked Learned Decoder
$\mathbf{X}_a$
Reconstruction Phase
(Adversary Without Side Information)
Sparse Obfuscated Codes
$\mathbf{Z}_{\mathsf{p}}^{[l]}, l \in [L]$
Subset of Private Dataset
$\{\mathbf{X}_{k'}\}_{k'=1}^{K'}, K' \leq K$
Leaked Learned Decoder
$\mathbf{X}_a$
Discriminator
Perceptual Loss
MSE
Recostruction Loss
$\{\mathbf{X}_{k'}\}_{k'=1}^{K'}$
Reconstruction Phase
(Adversary With Side Information)

Fig. 2. Operational setups of the proposed PRIMIS framework utilizing SCA mechanism.

Analysis Model: Analysis model uses a dictionary $\mathbf{\Omega} \in \mathbb{R}^{M \times N}$ with $M > N$ to analyze the data sample $\mathbf{x}_i \in \mathbb{R}^N$. Given a data sample $\mathbf{x}_i \in \mathbb{R}^N$ and dictionary $\mathbf{\Omega} \in \mathbb{R}^{M \times N}$ it assumes the representation $\mathbf{y}_i = \mathbf{\Omega}\mathbf{x}_i$ is sparse, i.e., $\|\mathbf{y}_i\|_0 \ll M$.

Transform Model: The sparsifying transform model [20, 65] assumes that the data sample $\mathbf{x}_i$ is approximately sparsifiable using a linear transform $\mathbf{A} \in \mathbb{R}^{M \times N}$, i.e., $\mathbf{A}\mathbf{x}_i = \mathbf{y}_i + \mathbf{z}_i$, where $\mathbf{y}_i \in \mathbb{R}^M$ is sparse, i.e., $\|\mathbf{y}_i\|_0 \ll M$, and $\mathbf{z}_i$ is the representation error of the data sample $\mathbf{x}_i$ in the *transform domain*.

2.4. Sparse Coding with Ambiguation Mechanism

Given a data sample $\mathbf{x}_i \in \mathbb{R}^N$ and two integer parameters $0 \leq S_x \leq N$, $0 \leq S_n \leq M - S_x$, the SCA privacy-preserving data release mechanism $\mathsf{SCA} : \mathbb{R}^N \times \mathbb{N}^0 \times \mathbb{N}^0 \to \mathbb{R}^M$ is defined as follows [17, 18, 66, 19, 20]:

$$\mathsf{SCA}\left(\mathbf{x}_i, S_x, S_n\right) \triangleq f\left(\mathbf{x}_i\right) \oplus \mathbf{n}_{\text{supp}}, \tag{1}$$

where $f : \mathbb{R}^N \to \mathbb{R}^M$ is a nonlinear sparsifying transform, $\mathbf{n}_{\text{supp}}$ is (pseudo) random noise vector which is added to the orthogonal complement of the sparse representation $f(\mathbf{x}_i)$.

2.5. Bottlenecked Auto-Encoders Model

We present an end-to-end (distributed) optimization of a nonlinear image compression scheme with a privacy guarantee inspired by the SCA mechanism [20]. Our model can be interpreted as a 'deep sparsifying transform learning' model with *layered successive* encoding. Fig. 1 shows the high-level schematic of our proposed framework. Fig. 2 illustrates the operational setups of the three main phases of our framework which can be described as follows.

Training Phase. Given a collection of image instances $\{\mathbf{X}_k\}_{k=1}^{K} \in \mathbb{R}^N$, a bottlenecked auto-encoder, comprising of L independent encoders (see Fig. 3), denoted by $f^{[1]}(\cdot), \cdots, f^{[L]}(\cdot)$, is trained where the input sample $\mathbf{X}$ is encoded to L new sparse codes as $\mathbf{Z}^{[l]} = f^{[l]}(\mathbf{X}), \forall l \in [L]$, with $\mathbf{Z}^{[l]} \in \mathbb{R}^M$ [20]. The original domain image is reconstructed as $\widehat{\mathbf{X}}^{[l]} = g^{[l]}\left(\mathbf{z}^{[l]}\right), \forall l \in [L]$, where $g^{[l]} : \mathbb{R}^M \to \mathbb{R}^N, \forall l \in [L]$, are L independent decoders, each one is paired with the corresponding $f^{[l]}, \forall l \in [L]$ [20]. The encoding process is designed to ensure that the codes are S_x-sparse, i.e., $\text{card}\left(\text{supp}\left(\mathbf{Z}^{[l]}\right)\right) = S_x, \forall\, l$, where $\text{supp}\left(\mathbf{Z}^{[l]}\right)$ represents the index set of nonzero element of $\mathbf{Z}^{[l]}$ and $\text{card}\left(\cdot\right)$ denotes cardinality of the set [20]. The encoding rate (bit rate) and reconstruction fidelity (distortion), i.e., rate-distortion trade-off, of our auto-encoding mechanism is controlled by the sparsity level S_x and the latent dimension M [20].

Sharing Phase. Considering the sparse representation $\mathbf{Z}^{[l]} = \mathbf{z}^{[l]}$ with sparsity level S_x and taking into account the obfuscation level $S_n \geq 0$, the SCA privacy-preserving mechanism (1) adds S_n random noise components to the orthogonal complement of $\mathbf{z}^{[l]}$, i.e., non-informative components, with the same statistics (mean and variance) as generated sparse representation to safeguard the indistinguishably in the statistical properties [20]. Hence, we have:

$$\mathbf{Z}_{\mathsf{p}}^{[l]} = \mathbf{Z}^{[l]} \oplus \mathbf{N}_{\text{supp}}. \tag{2}$$

The public code $\mathbf{Z}_{\mathsf{p}}^{[l]}$, is S_{p}-sparse, since $\|\mathbf{Z}_{\mathsf{p}}^{[l]}\|_0 = S_x + S_n = S_{\mathsf{p}}$. The ambiguated sparse representations $\mathbf{Z}_{\mathsf{p}}^{[l]}, \forall\, l$ are then shared to the public domain [20]. The support of the sparse clean code $\mathbf{Z}^{[l]}$, denoted by $\mathbf{Z}_{\mathsf{s}}^{[l]} = \text{supp}\left(\mathbf{Z}^{[l]}\right)$, is considered as *shared secrecy* [17, 18, 19, 20], which is shared with the authorized parties (centers). This secure part can even be encrypted with low complexity [20].

Reconstruction Phase. Considering the public obfuscated sparse representations $\mathbf{Z}_{\mathsf{p}}^{[l]}, \forall l \in [L]$, and private support information $\mathbf{Z}_{\mathsf{s}}^{[l]}, \forall l \in [L]$, the authorized user (data center) has the ability to 'purify' the obfuscated codes. This is achieved by removing the nosiy components of $\mathbf{Z}_{\mathsf{p}}^{[l]}$ using $\mathbf{Z}_{\mathsf{s}}^{[l]}$, or equivalently, by decoding on the support intersection of these codes. Therefore, the authorized user can reconstruct the original image data as $\widehat{\mathbf{X}}^{[l]} = g^{[l]}\left(\mathbf{Z}_{\mathsf{p}}^{[l]} | \mathbf{Z}_{\mathsf{s}}^{[l]}\right)$ [20]. We consider two hypotheses to support secrecy [17, 20, 67]:

- $\mathcal{H}_1$: The authorized support $\mathbf{Z}_{\mathsf{s}}$.
- $\mathcal{H}_0$: The unauthorized support generated and claimed by an adversary.

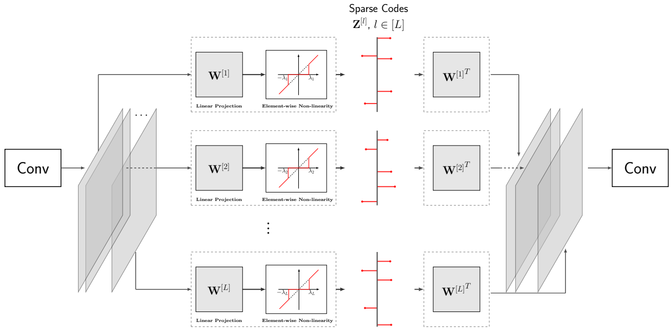


Fig. 3. Sparse Code-Map Generation using Grouped Linear Blocks: In this neural network architecture, each convolutional feature map on the encoder side is individually processed through a dedicated fully-connected linear layer followed by an element-wise nonlinearity to induce sparsity. Correspondingly, the decoder employs tied connections to accurately reconstruct these sparsified convolutional feature maps. This specialized block of grouped linear layers is strategically integrated at the middle of the network.

We consider two adversarial strategies in our study. The first strategy involves an unauthorized reconstruction where the adversary does not have access to a subset of the original dataset. However, they do have access to the trained decoder belonging to the data owner (defender), which was leaked or stolen. In the second strategy, it is assumed that the adversary has access to a subset of the original image samples, as well as the trained encoder-decoder. This means that, in this scenario, the adversary possesses pairs of images along with their corresponding sparse, obfuscated codes and the reconstructed images.

2.6. Connection with Modern Data Compression Schemes

Data compression refers to the method of shrinking the size of a file or data stream to save storage space or speed up the transmission of data over a network [68]. Classical data compression schemes often rely on a 'transform coding' technique [69, 70], which involves transforming the data into a different representation that is more amenable to compression. This is typically done using a mathematical transformation, such as a discrete cosine transform or a wavelet transform, which converts the data from the original domain (e.g., time or space) into a new domain, where it can be more easily compressed [69]. The transformed data is then quantized, which involves dividing it into a finite number of levels or bins and encoded using a lossless compression algorithm, such as Huffman coding [71]. The resulting code is then transmitted or stored and can be recovered at the destination by reversing the lossless compression and transform operations [72, 73]. Transform coding is a widely used technique in data compression because it can effectively remove redundancies and regularities in the data, making it easier to compress [69]. More recent data compression techniques often use auto-encoders and generative models [74, 75, 76, 77]. An auto-encoder is designed to learn how to encode data in a low-dimensional bottleneck representation, replacing the typically linear transform code with a learned nonlinear transform. When the data has intricate structures and patterns that may not be easily projected into a sparse domain, autoencoders can still be able to learn an efficient representation on a low dimensional manifold. The latent representation is still quantized and undergoes entropy coding, similar to classical compression techniques. However, the distribution used for entropy coding can also be learned directly form data to minimize rate loss. Generative models can further be used as the decoder, particularly to improve the perceived quality of the reconstructed image [78, 79].

2.7. Connection with the CLUB Model

The complexity-leakage-utility bottleneck (CLUB) [80] model is a generalization of the sufficient statistic methods that allow a model to smoothly trade-off the maximality of the informativeness of the bottleneck variable ($\mathbf{Z}$) for the utility task at hand ($\mathbf{U}$), against the compressiveness of the bottleneck variable ($\mathbf{Z}$) from data ($\mathbf{X}$), while limiting statistical inference about a sensitive random object $\mathbf{S}$ that depends on $\mathbf{X}$ and is possibly depended on $\mathbf{U}$. Considering the Markov chain $(\mathbf{U}, \mathbf{S}) \to \mathbf{X} \to \mathbf{Z}$ and denoting the mutual information between $\mathbf{X}$ and $\mathbf{Z}$ by $\mathrm{I}(\mathbf{X}; \mathbf{Z})$, this trade-off can be formulated by CLUB functional as follows:

$$\mathsf{CLUB}\left(R^{\mathrm{u}}, R^{\mathrm{s}}, P_{\mathbf{U},\mathbf{S},\mathbf{X}}\right) := \inf_{\substack{P_{\mathbf{Z}|\mathbf{X}}: \\ (\mathbf{U},\mathbf{S}) -\circ- \mathbf{X} -\circ- \mathbf{Z}}} \mathrm{I}(\mathbf{X}; \mathbf{Z}) \quad \text{s.t.} \quad \mathrm{I}(\mathbf{U}; \mathbf{Z}) \geq R^{\mathrm{u}}, \mathrm{I}(\mathbf{S}; \mathbf{Z}) \leq R^{\mathrm{s}}. \tag{3}$$

Setting $R^{\mathrm{s}} \geq \mathrm{H}(\mathbf{S})$ in (3), the CLUB model reduces to the information bottleneck (IB) principle [80, 81], while setting $\mathbf{U} \equiv \mathbf{X}$ and $R^{\mathrm{z}} \geq \mathrm{H}(\mathbf{X})$ in (3), the CLUB model reduces to the privacy funnel (PF) model [80, 82].

In the privacy-preserving image-sharing framework, our utility variable is $\mathbf{U} \equiv \mathbf{X}$, i.e., our goal is to reconstruct the original data $\mathbf{X}$ using the bottleneck variable $\mathbf{Z}$, with minimum information loss [80]. This allows for further downstream tasks whether using the encoded compressed representations $\mathbf{Z}$ or the reconstructed original domain data $\widehat{\mathbf{X}}$ [80]. This scenario is referred to as unsupervised CLUB in [80]. In this scenario, our objective is to obtain a compact information-preserving representation $\mathbf{Z}$ of original data $\mathbf{X}$, which can also be used for various utility tasks at the authorized parties in an unsupervised fashion [80]. As studied in [80], the information utility part of deep variational CLUB (DVCLUB) Lagrangian functional can be decomposed into two terms (i) reconstruction fidelity, and (ii) distribution discrepancy loss. In this research, we use a number of different reconstruction fidelity measures, such as mean absolute error (MAE), mean error (ME), mean squared error (MSE), root mean square error (RMSE), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR).

2.8. Bottleneck Auto-Encoder Architecture

We use a similar network as in [20], whose design is motivated by three important principles. Firstly, as is fundamental to the structure of our proposed privacy solution, the network should provide compact codes through bottlenecked structures [20]. This rules out many popular designs like the family of U-Net architectures, where skip connections break the bottleneck constraint [20]. Secondly, the compact codes should be

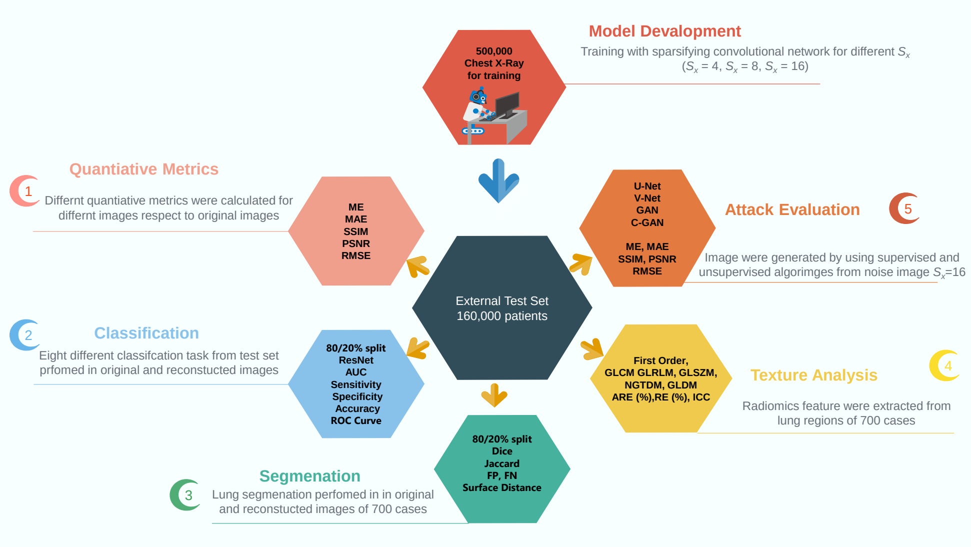

Model Devalopment
Training with sparsifying convolutional network for different S_x
(S_x = 4, S_x = 8, S_x = 16)
500,000
Chest X-Ray
for training
1
Quantiative Metrics
Differnt quantiative metrics were calculated for
differnt images respect to original images
ME
MAE
SSIM
PSNR
RMSE
U-Net
V-Net
GAN
C-GAN
ME, MAE
SSIM, PSNR
RMSE
Attack Evaluation
5
Image were generated by using supervised and
unsupervised algorimges from noise image S_x=16
External Test Set
160,000 patients
2
Classification
Eight different classifcation task from test set
prfomed in original and reconstucted images
80/20% split
ResNet
AUC
Sensitivity
Specificity
Accuracy
ROC Curve
First Order,
GLCM GLRLM, GLSZM,
NGTDM, GLDM
ARE (%),RE (%), ICC
Texture Analysis
4
Radiomics feature were extracted from
lung regions of 700 cases
80/20% split
Dice
Jaccard
FP, FN
Surface Distance
3
Segmenation
Lung segmenation perfomed in in original
and reconstucted images of 700 cases

Fig. 4. The schematic flowchart of the current study including model development, quantitative metrics, classification, segmentation, texture analysis, and attack evaluation.

sparse [20]. While this is not a usual constraint within modern deep learning networks, we found that the “top-S” operator implemented within deep learning (DL) frameworks works better than hard- or soft-thresholding [20]. Thirdly, the network design should fit a typical DL practice (e.g., smooth backpropagation, avoiding over-fitting) [20]. Similarly to ReLU, the top-S operator (equivalent to a adaptive-threshold hard-thresholding function) has discontinuities only at the threshold points and does not hinder back-propagation to run smoothly [20]. However, on the one hand, applying them directly on convolutional layer outputs does not provide sparsity at diverse locations, and hence, limits the coding efficiency [20]. On the other hand, using a dense matrix multiplication after the convolutional outputs would require a huge matrix to be trained, which can easily overfit [20]. Therefore, a matrix multiplication with a block-sparse structure is used after convolutional filters to avoid inter-mixing values from different convolutional filters [20]. This would diversify the sparsity pattern between different images and avoid the huge number of training parameters [20]. While these principles were common in the work of [20], current work improves the efficiency of this latter stage by implementing this group-sparse matrix multiplication as an equivalent convolution operation that could accept arbitrary image sizes as the input and has less number of parameters. Therefore, our design is fully convolutional, as opposed to the one proposed in [20].

2.9. Framework, Dataset, Training, Evaluation Approaches

Fig. 4 provides a summary of this study. We focused on chest X-ray images, and datasets were gathered from different data sources[83, 84, 85, 86, 87, 88]. All networks were trained in a 2D manner with an Adam optimization with a learning rate starting with 0.001 and a weight decay of 0.0001 in 500,000 images using a patch size of 64 by 64. Mean squared error loss was used for network training. We evaluated the proposed model in different scenarios, including quantitative analysis and classification, segmentation, and texture analysis tasks. In addition, we performed attacks on the random output of the model using different supervised and unsupervised deep neural networks.

2.9.1. Quantitative Image Level Analysis

A qualitative evaluation of the proposed method was performed on 160,000 external test samples of chest X-ray images. To this end, the predicted images were compared with reference original images. The quality of predicted images was assessed using voxel-wise ME, voxel-wise MAE, and voxel-wise RMSE. Moreover, the SSIM and peak PSNR were used as quantitative measures of the quality of the predicted chest X-ray images.

2.9.2. Image Classification Task

For classification, we used a subset of external validation set for different classification tasks, including:

Task 1: Three class classification of normal (2,020 patients), bacterial pneumonia (2,300 patients), and viral (2,270 patients) pneumonia.

Task 2: Classification of bacterial pneumonia (2,020 patients) against viral pneumonia (2,070 patients).

Task 3: Classification of viral COVID-19 pneumonia (610 patients) against viral pneumonia (610 patients).

Task 4: Classification of viral COVID-19 pneumonia (610 patients) against viral pneumonia + bacterial pneumonia (630 patients).

Task 5: Classification of normal cases (3,630 patients) against viral pneumonia + bacterial pneumonia (4,290 patients).

Task 6: Classification of normal (2,300 patients) cases against viral pneumonia(2,270 patients).

Task 7: Classification of normal (2,020 patients) cases against bacterial pneumonia (2,300 patients).

Task 8: Classification of normal (610 patients) cases against viral COVID-19 pneumonia (610 patients).

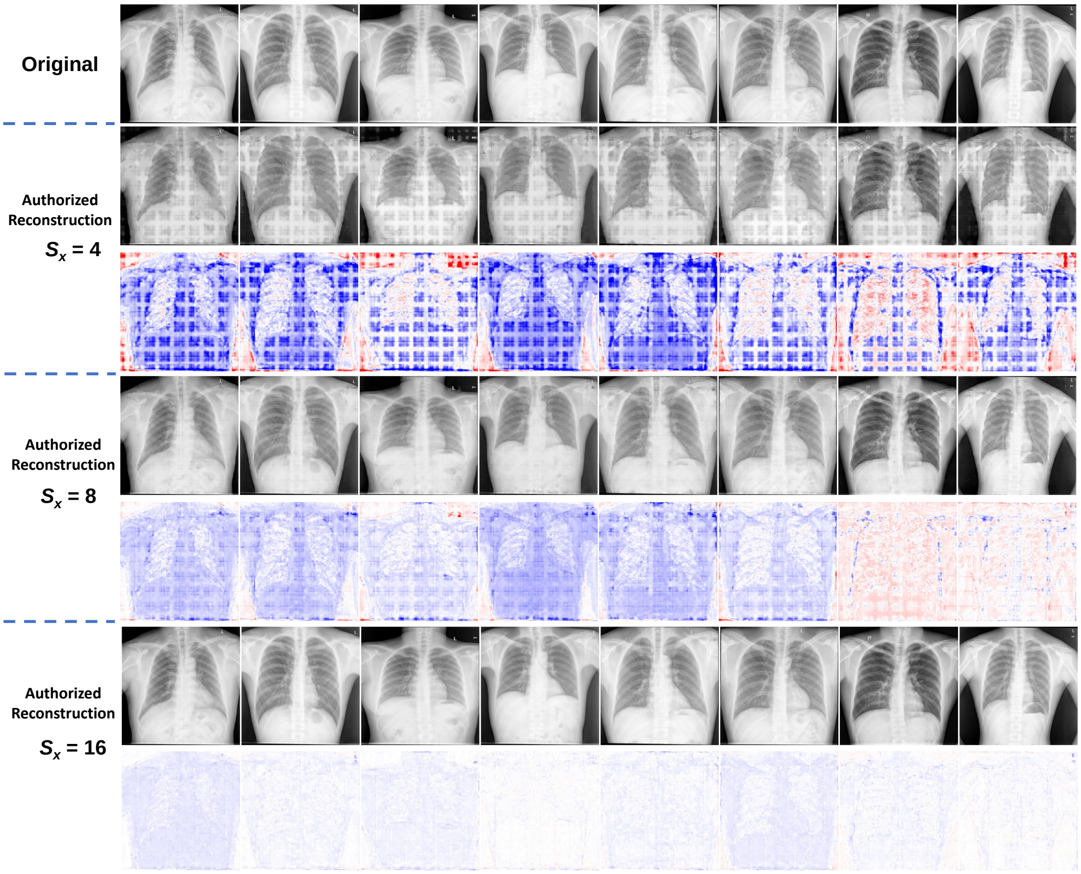

Original
Authorized Reconstruction
$S_x = 4$
Authorized Reconstruction
$S_x = 8$
Authorized Reconstruction
$S_x = 16$

Fig. 5. Comparison of the original images vs (i) authorized reconstructed images for sparsity levels $S_x = 4, 8, 16$, and (ii) their corresponding difference bias maps in different patients. Patchy structures could be seen in these images as a result of the training patch size in images.

Table 1. Summary of quantitative parameters (mean±SD and CI95%) for image domain fidelity measures: MAE, ME, RMSE, PSNR, and SSIM; setting sparsity levels $S_x = 16, 8, 4$.

Parameter	Sparsity Level	MAE	ME	RMSE	PSNR	SSIM
Mean±Sd	$S_x = 4$	9.35 ± 8.19	9.35 ± 8.19	37.63 ± 9.22	16.86 ± 2.03	0.48 ± 0.05
	$S_x = 8$	1.07 ± 0.27	1.07 ± 0.27	5.06 ± 1.29	34.29 ± 2.06	0.95 ± 0.01
	$S_x = 16$	0.71 ± 0.2	0.71 ± 0.2	3.08 ± 0.58	38.48 ± 1.56	0.98 ± 0.01
CI95%	$S_x = 4$	9.31 to 9.39	9.31 to 9.39	37.58 to 37.67	16.85 to 16.87	0.48 to 0.48
	$S_x = 8$	1.07 to 1.07	1.07 to 1.07	5.05 to 5.07	34.28 to 34.3	0.95 to 0.95
	$S_x = 16$	0.71 to 0.71	0.71 to 0.71	3.08 to 3.09	38.48 to 38.49	0.98 to 0.98

For each task, data were split into train/validation (70/10%) and test sets (20%), and all quantitative analyses were performed and reported on unseen test sets (there is no overlap between different sets). Training was performed on each set of images separately and tested on different images.

2.9.3. *Image Segmentation Task*

Image segmentation was performed on 700 image samples with whole lung segmentation as a subset of the external validation set. Data were split into train/validation (70/10%) and test set (20%), and quantitative analyses performed and reported on unseen test set. We implemented U-Net architecture for the core of the segmentation task. Training was performed on each set of images separately and tested on different images. Different evaluation metrics, including quantitative segmentation metrics, were implemented to evaluate the segmentation performance on original and predicted images. Standard image segmentation metrics, including the Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), false-negative rate, false-positive rate, mean and standard deviation (SD) of surface distance, and Hausdorff distance, were used for assessment.

2.9.4. *Image Texture Task*

In this task, we used 700 patient images available with whole lung segmentation. Prior to feature extraction, the image voxel was resized to an isotropic pixel size of 1 × 1 mm^2, and the intensity was discretized into 64 bins. All radiomics feature extraction was performed using PyRadiomics [89] Python library, including intensity (n = 18), second-order texture features, such as gray level co-occurrence matrix (GLCM, $n = 24$), higher-order features, namely gray level size zone matrix (GLSZM, $n = 16$), neighboring gray tone difference matrix (NGTDM, $n = 5$), gray level run length matrix (GLRLM, $n = 16$), and gray level dependence matrix (GLDM, $n = 14$). All extracted radiomics features are compliant with the Image biomarker standardization initiative (IBSI) guidelines [89, 90]. We calculated percent relative error (RE) and percent absolute relative error (ARE) with respect to the original image for each predicted image. Intraclass correlation (ICC) tests were performed for radiomics feature reproducibility in different approaches with respect to the original image. We classified radiomic features based on the ICC value into 4 groups: Poor reproducibility (ICC ≤ 0.40), fair reproducibility (0.40 < ICC ≤ 0.60), good reproducibility (0.60 <ICC ≤ 0.75), and excellent reproducibility (0.75 < ICC ≤ 1.00).

2.9.5. Adversarial Attack Analysis

For attack analysis, we used the obfuscated image as input and tried to infer the original image using different supervised and unsupervised networks. To this end, image-to-image translation, and supervised algorithms, including U-Net, V-Net, and GAN, were used in the supervised approach, whereas Cycle GAN was implemented in the unsupervised approach. In this evaluation, we used 160,000 external test samples, where the data were split into train/validation (70/10%) and test set (20%). Quantitative analyses were performed and reported on unseen test sets. A batch size of 30, an Adam optimizer, a learning rate of 0.001, an L2-norm loss, and a weight decay of 0.0001 were used in these networks.

3. Results

3.1. Qualitative Image Analysis

For visual comparison, Fig. 5 shows some examples of the original image and predicted images for different S_x's and their corresponding bias maps with respect to the original images. These figures show that the predicted images generated by different S_x are in good agreement with original images despite the variability in structures and textures. When $S_x = 4$, the patch structure affects the image texture; while with $S_x = 16$ reconstructed images were almost identical to the original ones, achieving the lowest differences.

3.2. Quantitative Image Level Analysis

Table 1 presents the reconstruction error for different sparsity levels ($S_x = 4$, $S_x = 8$ and $S_x = 16$) with respect to the original image using the 160,000 external test set (Supplemental Fig. 1 presents the results as a box plot for better visualization). For all the metrics compared, the results show that the lowest error was achieved with $S_x = 16$, followed by $S_x = 8$. In terms of ME, $S_x = 16$ and $S_x = 8$ achieved, 0.71±0.2 (CI95%: 0.71-0.71) and 1.07±0.27 (CI95%: 1.07-1.07), respectively. However, a ME of 9.35±8.9 (CI95%: 9.31-9.39) was achieved for $S_x = 4$ images. $S_x = 16$ images, generated the highest SSIM and PSNR (38.48±1.56 (CI95%: 38.48-34.3) and 0.98±0.01 (CI95%: 0.98-0.98) respectively).

Fig. 6 shows the joint histogram analysis displaying the correlation between the original and different predicted images. The results show that the images obtained with the $S_x = 16$ had the highest correlation with $R^2 = 0.9999$ followed by $S_x = 8$ with $R^2 = 0.9991$, and the lowest correlation is achieved when $S_x = 4$ with $R^2 = 0.7306$.

3.3. Image Classification

Table 2 presents the image classification results for models trained and tested with different S_x values. The results show that models trained with $S_x = 16$, $S_x = 8$, and original images yielded almost the same performance for different tasks, and performance did not change drastically when testing on these three (i.e., trained on original and tested on $S_x = 16$, $S_x = 8$). This illustrates that the important features for classification are preserved in these images. Models trained on $S_x = 4$ and tested on $S_x = 4$ revealed the same performance compared to other models; however, testing with $S_x = 16$, $S_x = 8$, and original images depicted low performance. For models trained with $S_x = 16$, $S_x = 8$, and original images, the lowest performance was achieved in the test data set generated by $S_x = 4$. In addition,

These tasks were performed on subsets of external validation sets. Fig 7 represents the ROC curve for comparison of different Tasks (1-8). The model was trained on original images and tested on different images, including original and predicted images. We also provided trained and tested models using different image sets. The ROC curves were presented in supplemental Figs 2-9 for Tasks 1-8, respectively. Supplemental Tables 1-3 summarize the accuracy, sensitivity, and specificity metrics for training and testing on different sets for different tasks, respectively.

3.4. Image Texture Analysis

Fig. 8 represents the heat map of ARE, RE, and ICC metrics of radiomic features extracted from lung segmentation in different reconstructed images with respect to the original images. As shown in this figure, most features showed RE less than 10 % and ICC higher than 0.75 in $S_x = 16$ and $S_x = 8$, demonstrating the excellent recovery of subtle textures of images.

3.5. Image Segmentation Analysis

Table 3 (CI95% presented in Supplemental Table 4) provides a summary of the quantitative analysis of segmentation metrics for different training and test sets. As presented in this table, training and testing on a different set of images provide quantitative metrics that are in good agreement. The lowest dice score (0.92±0.05 was achieved when training on $S_x = 16$ and testing on $S_x = 4$. There were no statistically significant differences between the original, $S_x = 8$ and $S_x = 16$ images when training and testing these three image sets.

For visual comparison, Fig. 9 depicts some examples of segmentation when training and testing a different set of images, including original and predicted images from a subset of external test sets. As shown in this figure, the segmentations provided by different training and test sets are in good agreement with manual segmentation in different patients.

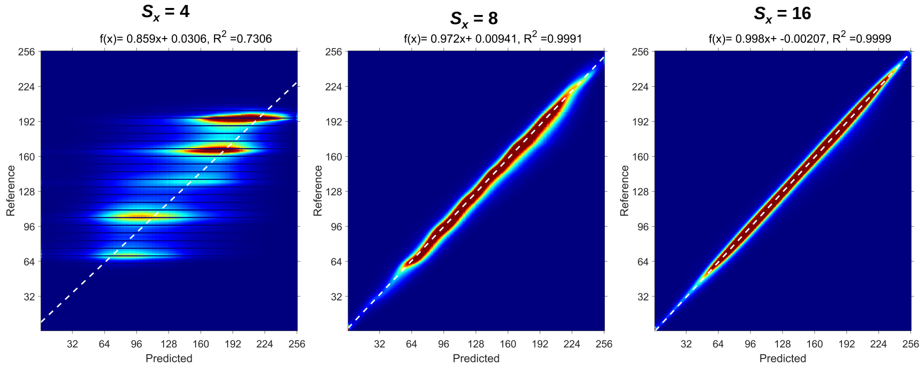


Fig. 6. Joint histogram analysis displaying the correlation between different reconstructed images and original images. The plot shows that $S_x = 16$ images had the highest correlation with R^2 of 0.9999 followed by $S_x = 8$ with R^2 of 0.9991 and the lowest correlation achieved by $S_x = 4$ images with R^2 of 0.7306.

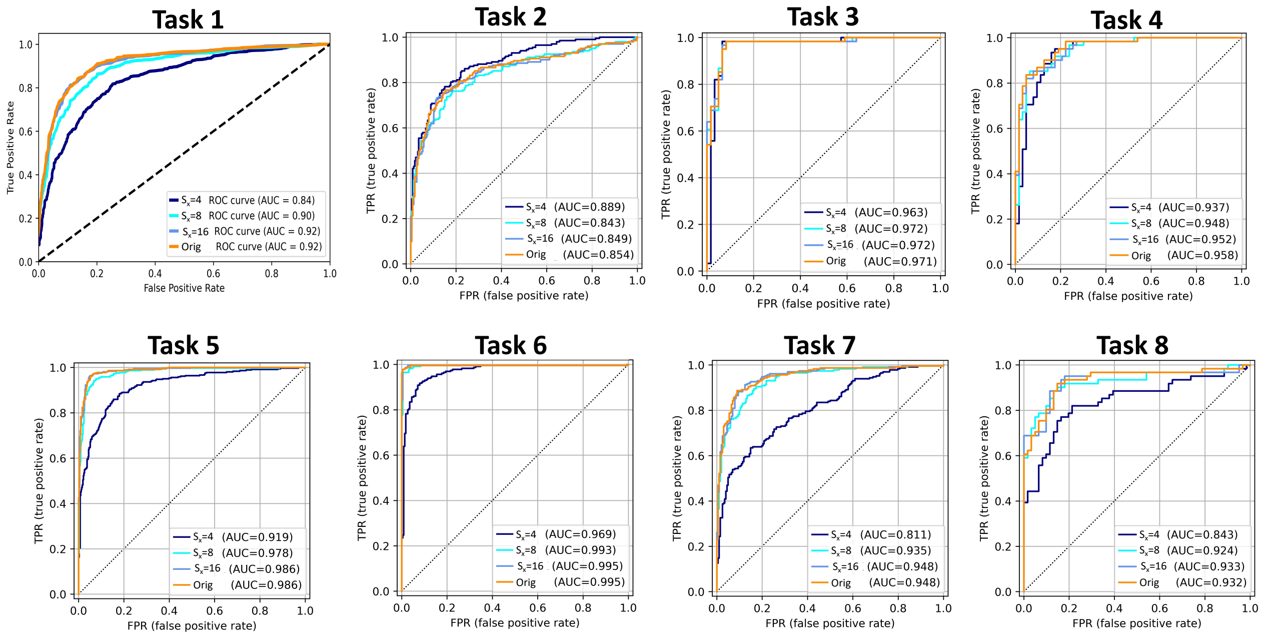


Fig. 7. ROC curve comparison for different training and test sets in different Tasks 1-8. The training was performed on the original images, whereas the tests were performed on different images for each task. Task 1: Three class classification of normal, bacterial, and viral pneumonia; Task 2: Classification of bacterial pneumonia against viral pneumonia; Task 3: Classification of viral COVID-19 pneumonia against viral pneumonia; Task 4: Classification of viral COVID-19 pneumonia against viral pneumonia + bacterial pneumonia, Task 5: Classification of normal cases against viral pneumonia +bacterial pneumonia, Task 6: classification of normal cases against viral pneumonia, Task 7: classification of normal cases against bacterial pneumonia, Task 8: classification of normal cases against viral COVID-19 pneumonia.

3.6. Attack Analysis

Table 4 shows the attacks outcomes for four networks (U-Net, V-Net, GAN, and C-GAN) when applied to noisy images with $S_x = 16$. As seen in Table 3, none of these methods successfully recovered the original images from the noisy structures. The supervised GAN achieved the highest SSIM (0.54+0.07 (CI95%: 0.54-55))). Fig. 10 shows the comparison of the outputs of these networks when applied to attack analysis using different supervised and unsupervised networks. As shown in the figure, none of the networks was able to recover the original images from the noisy ones effectively.

4. Discussion

Our research focused on using the SCA privacy-assuring mechanism for privacy-preserving medical image sharing. The SCA mechanism is a generalization of randomization techniques that allow for a trade-off between utility and privacy in a principled manner [19, 20, 18]. To improve the practicality and scalability of the SCA mechanism for use in CNNs, we proposed two architectural innovations: multiple code-maps using fully-connected groups on convolutional filters and the S-sparsity non-linearity in CNNs [19, 20, 18]. Additionally, we connected our framework to modern data compression techniques and the CLUB [80] model to further enhance its effectiveness and efficiency.

Table 2. AUC of different classification tasks for different training and test sets, Task 1: Three class classification of normal, bacterial, and viral pneumonia, Task 2: Classification of bacterial pneumonia against viral pneumonia, Task 3: Classification of viral COVID-19 pneumonia against viral pneumonia, Task 4: Classification of viral COVID-19 pneumonia against viral pneumonia + bacterial pneumonia, Task 5: Classification of normal cases against viral pneumonia +bacterial pneumonia, Task 6: classification of normal cases against viral pneumonia, Task 7: classification of normal cases against bacterial pneumonia, Task 8: classification of normal cases against viral COVID-19 pneumonia.

Train	Test	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
$S_x = 4$	$S_x = 4$	0.93	0.85	0.98	0.94	0.99	0.99	0.96	0.93
	$S_x = 8$	0.86	0.78	0.98	0.92	0.98	0.99	0.92	0.86
	$S_x = 16$	0.83	0.75	0.97	0.88	0.98	0.99	0.9	0.83
	Original	0.83	0.73	0.97	0.89	0.98	0.99	0.88	0.82
$S_x = 8$	$S_x = 4$	0.86	0.85	0.99	0.95	0.96	0.99	0.91	0.97
	$S_x = 8$	0.91	0.85	0.99	0.95	0.99	0.99	0.96	0.98
	$S_x = 16$	0.89	0.87	0.99	0.94	0.98	0.99	0.96	0.98
	Original	0.88	0.86	0.99	0.95	0.98	0.99	0.95	0.97
$S_x = 16$	$S_x = 4$	0.85	0.8	0.98	0.9	0.94	0.99	0.89	0.97
	$S_x = 8$	0.91	0.79	0.98	0.92	0.98	0.99	0.97	0.97
	$S_x = 16$	0.93	0.82	0.98	0.92	0.99	0.99	0.98	0.98
	Original	0.93	0.82	0.98	0.93	0.99	0.99	0.97	0.97
Original	$S_x = 4$	0.84	0.89	0.96	0.94	0.92	0.97	0.81	0.84
	$S_x = 8$	0.9	0.84	0.97	0.95	0.98	0.99	0.94	0.92
	$S_x = 16$	0.92	0.85	0.97	0.95	0.99	0.99	0.95	0.93
	Original	0.92	0.85	0.97	0.96	0.99	0.99	0.95	0.93

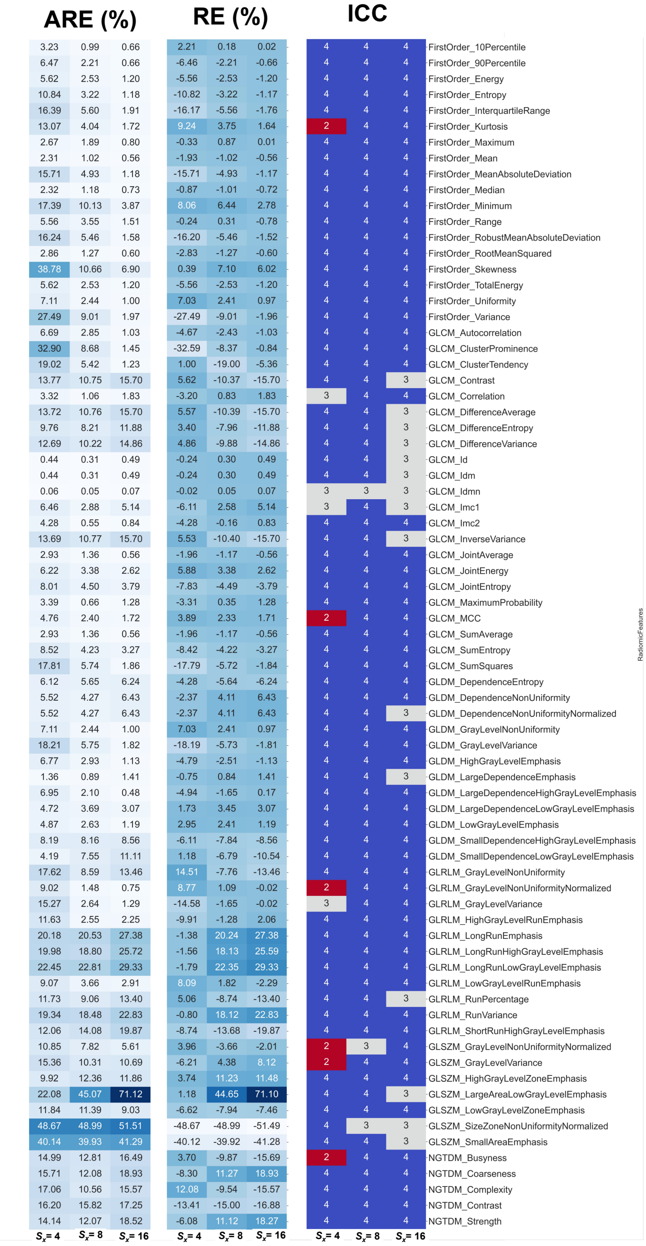
Table 3. Segmentation results for different training and test sets.

Train	**Test**	**Dice**	**Jaccard**	**False Negative**	**False Positive**	**Mean Surface Distance**	**Std Surface Distance**
$S_x = 4$	$S_x = 4$	0.94 ± 0.04	0.88 ± 0.06	0.1 ± 0.05	0.02 ± 0.05	0.14 ± 0.09	1.21 ± 1.02
	$S_x = 8$	0.94 ± 0.04	0.88 ± 0.06	0.1 ± 0.05	0.02 ± 0.05	0.14 ± 0.09	1.21 ± 1.02
	$S_x = 16$	0.93 ± 0.04	0.88 ± 0.06	0.1 ± 0.05	0.02 ± 0.05	0.14 ± 0.09	1.23 ± 1.01
	Original	0.93 ± 0.04	0.88 ± 0.06	0.1 ± 0.05	0.02 ± 0.05	0.15 ± 0.10	1.30 ± 1.09
$S_x = 8$	$S_x = 4$	0.93 ± 0.03	0.88 ± 0.05	0.1 ± 0.05	0.03 ± 0.03	0.13 ± 0.07	1.20 ± 0.80
	$S_x = 8$	0.94 ± 0.03	0.89 ± 0.05	0.1 ± 0.05	0.01 ± 0.02	0.12 ± 0.06	1.02 ± 0.80
	$S_x = 16$	0.94 ± 0.03	0.89 ± 0.05	0.1 ± 0.05	0.01 ± 0.01	0.12 ± 0.06	1.02 ± 0.79
	Original	0.94 ± 0.03	0.89 ± 0.05	0.1 ± 0.05	0.01 ± 0.01	0.12 ± 0.06	1.02 ± 0.79
$S_x = 16$	$S_x = 4$	0.92 ± 0.05	0.85 ± 0.08	0.08 ± 0.08	0.08 ± 0.06	0.19 ± 0.18	1.82 ± 1.66
	$S_x = 8$	0.94 ± 0.04	0.89 ± 0.06	0.09 ± 0.06	0.02 ± 0.03	0.12 ± 0.08	1.10 ± 0.93
	$S_x = 16$	0.94 ± 0.03	0.89 ± 0.05	0.09 ± 0.05	0.02 ± 0.03	0.12 ± 0.08	1.07 ± 0.89
	Original	0.94 ± 0.03	0.9 ± 0.05	0.08 ± 0.05	0.02 ± 0.03	0.11 ± 0.08	1.06 ± 0.89
Original	$S_x = 4$	0.93 ± 0.07	0.87 ± 0.09	0.08 ± 0.09	0.06 ± 0.04	0.17 ± 0.39	1.59 ± 2.63
	$S_x = 8$	0.94 ± 0.03	0.90 ± 0.06	0.09 ± 0.05	0.02 ± 0.03	0.11 ± 0.07	0.99 ± 0.80
	$S_x = 16$	0.94 ± 0.04	0.90 ± 0.06	0.09 ± 0.05	0.02 ± 0.04	0.11 ± 0.08	1.04 ± 0.94
	Original	0.94 ± 0.03	0.90 ± 0.06	0.09 ± 0.05	0.02 ± 0.04	0.11 ± 0.08	1.03 ± 0.92

Table 4. Summary of quantitative metrics (mean±sd, and CI95) for image domain parameters of different attacks. MAE: Mean Absolute Error, ME: Mean Error, RMSE: Root Mean Square Error, PSNR: peak signal-to-noise ratio, SSIM: Structural Similarity Index, U-Net: Supervised 2D U-Net, V-Net: Supervised 2D V-Net, GAN: Supervised 2D GAN, C-GAN: Unsupervised 2D Cycle GAN.

Parameter	Networks	MAE	ME	RMSE	PSNR	SSIM
Mean±Sd	U-Net	22.93 ± 4.67	-0.63 ± 11.47	29.5 ± 5.45	18.90 ± 1.49	0.51 ± 0.06
	V-Net	23.60 ± 4.51	-4.57 ± 11.47	30.27 ± 5.26	18.67 ± 1.45	0.50 ± 0.06
	GAN	21.67 ± 5.70	4.68 ± 11.33	27.95 ± 6.68	19.44 ± 1.83	0.54 ± 0.07
	C-GAN	31.13 ± 6.38	5.36 ± 13.82	40.33 ± 7.12	16.16 ± 1.32	0.36 ± 0.04
CI95%	U-Net	22.88 to 22.98	-0.76 to -0.51	29.44 to 29.56	18.88 to 18.92	0.51 to 0.51
	V-Net	23.55 to 23.65	-4.69 to -4.44	30.21 to 30.33	18.65 to 18.68	0.50 to 0.50
	GAN	21.60 to 21.73	4.55 to 4.80	27.88 to 28.03	19.42 to 19.46	0.54 to 0.55
	C-GAN	31.06 to 31.2	5.21 to 5.51	40.25 to 40.41	16.15 to 16.18	0.36 to 0.36

In our three-party data release scenario, the data owner shares representations of their medical image data with an ‘honest but curious’ server [19, 20, 18]. The data owner’s goal is to (i) protect original images from server-side analysis, (ii) provide a predetermined level of utility for their authorized clients, and (iii) protect original images from unauthorized parties or potential adversaries [19, 20, 18]. We define the measure of utility as the capability of reconstruction for authorized parties, and



	ARE (%)			RE (%)			ICC		
	$S_x = 4$	$S_x = 8$	$S_x = 16$	$S_x = 4$	$S_x = 8$	$S_x = 16$	$S_x = 4$	$S_x = 8$	$S_x = 16$
FirstOrder_10Percentile	3.23	0.99	0.66	2.21	0.18	0.02	4	4	4
FirstOrder_90Percentile	6.47	2.21	0.66	-6.46	-2.21	-0.66	4	4	4
FirstOrder_Energy	5.62	2.53	1.20	-5.56	-2.53	-1.20	4	4	4
FirstOrder_Entropy	10.84	3.22	1.18	-10.82	-3.22	-1.17	4	4	4
FirstOrder_InterquartileRange	16.39	5.60	1.91	-16.17	-5.56	-1.76	4	4	4
FirstOrder_Kurtosis	13.07	4.04	1.72	9.24	3.75	1.64	2	4	4
FirstOrder_Maximum	2.67	1.89	0.80	-0.33	0.87	0.01	4	4	4
FirstOrder_Mean	2.31	1.02	0.56	-1.93	-1.02	-0.56	4	4	4
FirstOrder_MeanAbsoluteDeviation	15.71	4.93	1.18	-15.71	-4.93	-1.17	4	4	4
FirstOrder_Median	2.32	1.18	0.73	-0.87	-1.01	-0.72	4	4	4
FirstOrder_Minimum	17.39	10.13	3.87	8.06	6.44	2.78	4	4	4
FirstOrder_Range	5.56	3.55	1.51	-0.24	0.31	-0.78	4	4	4
FirstOrder_RobustMeanAbsoluteDeviation	16.24	5.46	1.58	-16.20	-5.46	-1.52	4	4	4
FirstOrder_RootMeanSquared	2.86	1.27	0.60	-2.83	-1.27	-0.60	4	4	4
FirstOrder_Skewness	38.78	10.66	6.90	0.39	7.10	6.02	4	4	4
FirstOrder_TotalEnergy	5.62	2.53	1.20	-5.56	-2.53	-1.20	4	4	4
FirstOrder_Uniformity	7.11	2.44	1.00	7.03	2.41	0.97	4	4	4
FirstOrder_Variance	27.49	9.01	1.97	-27.49	-9.01	-1.96	4	4	4
GLCM_Autocorrelation	6.69	2.85	1.03	-4.67	-2.43	-1.03	4	4	4
GLCM_ClusterProminence	32.90	8.68	1.45	-32.59	-8.37	-0.84	4	4	4
GLCM_ClusterTendency	19.02	5.42	1.23	1.00	-19.00	-5.36	4	4	4
GLCM_Contrast	13.77	10.75	15.70	5.62	-10.37	-15.70	4	4	3
GLCM_Correlation	3.32	1.06	1.83	-3.20	0.83	1.83	3	4	4
GLCM_DifferenceAverage	13.72	10.76	15.70	5.57	-10.39	-15.70	4	4	3
GLCM_DifferenceEntropy	9.76	8.21	11.88	3.40	-7.96	-11.88	4	4	3
GLCM_DifferenceVariance	12.69	10.22	14.86	4.86	-9.88	-14.86	4	4	3
GLCM_Id	0.44	0.31	0.49	-0.24	0.30	0.49	4	4	3
GLCM_Idm	0.44	0.31	0.49	-0.24	0.30	0.49	4	4	3
GLCM_Idmn	0.06	0.05	0.07	-0.02	0.05	0.07	3	3	3
GLCM_Imc1	6.46	2.88	5.14	-6.11	2.58	5.14	3	4	3
GLCM_Imc2	4.28	0.55	0.84	-4.28	-0.16	0.83	4	4	4
GLCM_InverseVariance	13.69	10.77	15.70	5.53	-10.40	-15.70	4	4	3
GLCM_JointAverage	2.93	1.36	0.56	-1.96	-1.17	-0.56	4	4	4
GLCM_JointEnergy	6.22	3.38	2.62	5.88	3.38	2.62	4	4	4
GLCM_JointEntropy	8.01	4.50	3.79	-7.83	-4.49	-3.79	4	4	4
GLCM_MaximumProbability	3.39	0.66	1.28	-3.31	0.35	1.28	4	4	4
GLCM_MCC	4.76	2.40	1.72	3.89	2.33	1.71	2	4	4
GLCM_SumAverage	2.93	1.36	0.56	-1.96	-1.17	-0.56	4	4	4
GLCM_SumEntropy	8.52	4.23	3.27	-8.42	-4.22	-3.27	4	4	4
GLCM_SumSquares	17.81	5.74	1.86	-17.79	-5.72	-1.84	4	4	4
GLDM_DependenceEntropy	6.12	5.65	6.24	-4.28	-5.64	-6.24	4	4	4
GLDM_DependenceNonUniformity	5.52	4.27	6.43	-2.37	4.11	6.43	4	4	4
GLDM_DependenceNonUniformityNormalized	5.52	4.27	6.43	-2.37	4.11	6.43	4	4	3
GLDM_GrayLevelNonUniformity	7.11	2.44	1.00	7.03	2.41	0.97	4	4	4
GLDM_GrayLevelVariance	18.21	5.75	1.82	-18.19	-5.73	-1.81	4	4	4
GLDM_HighGrayLevelEmphasis	6.77	2.93	1.13	-4.79	-2.51	-1.13	4	4	4
GLDM_LargeDependenceEmphasis	1.36	0.89	1.41	-0.75	0.84	1.41	4	4	3
GLDM_LargeDependenceHighGrayLevelEmphasis	6.95	2.10	0.48	-4.94	-1.65	0.17	4	4	4
GLDM_LargeDependenceLowGrayLevelEmphasis	4.72	3.69	3.07	1.73	3.45	3.07	4	4	4
GLDM_LowGrayLevelEmphasis	4.87	2.63	1.19	2.95	2.41	1.19	4	4	4
GLDM_SmallDependenceHighGrayLevelEmphasis	8.19	8.16	8.56	-6.11	-7.84	-8.56	4	4	4
GLDM_SmallDependenceLowGrayLevelEmphasis	4.19	7.55	11.11	1.18	-6.79	-10.54	4	4	4
GLRLM_GrayLevelNonUniformity	17.62	8.59	13.46	14.51	-7.76	-13.46	4	4	4
GLRLM_GrayLevelNonUniformityNormalized	9.02	1.48	0.75	8.77	1.09	-0.02	2	4	4
GLRLM_GrayLevelVariance	15.27	2.64	1.29	-14.58	-1.65	-0.02	3	4	4
GLRLM_HighGrayLevelRunEmphasis	11.63	2.55	2.25	-9.91	-1.28	2.06	4	4	4
GLRLM_LongRunEmphasis	20.18	20.53	27.38	-1.38	20.24	27.38	4	4	4
GLRLM_LongRunHighGrayLevelEmphasis	19.98	18.80	25.72	-1.56	18.13	25.59	4	4	4
GLRLM_LongRunLowGrayLevelEmphasis	22.45	22.81	29.33	-1.79	22.35	29.33	4	4	4
GLRLM_LowGrayLevelRunEmphasis	9.07	3.66	2.91	8.09	1.82	-2.29	4	4	4
GLRLM_RunPercentage	11.73	9.06	13.40	5.06	-8.74	-13.40	4	4	3
GLRLM_RunVariance	19.34	18.48	22.83	-0.80	18.12	22.83	4	4	4
GLRLM_ShortRunHighGrayLevelEmphasis	12.06	14.08	19.87	-8.74	-13.68	-19.87	4	4	4
GLSZM_GrayLevelNonUniformityNormalized	10.85	7.82	5.61	3.96	-3.66	-2.01	2	3	4
GLSZM_GrayLevelVariance	15.36	10.31	10.69	-6.21	4.38	8.12	2	4	4
GLSZM_HighGrayLevelZoneEmphasis	9.92	12.36	11.86	3.74	11.23	11.48	4	4	4
GLSZM_LargeAreaLowGrayLevelEmphasis	22.08	45.07	71.12	1.18	44.65	71.10	4	4	3
GLSZM_LowGrayLevelZoneEmphasis	11.84	11.39	9.03	-6.62	-7.94	-7.46	4	4	4
GLSZM_SizeZoneNonUniformityNormalized	48.67	48.99	51.51	-48.67	-48.99	-51.49	4	3	3
GLSZM_SmallAreaEmphasis	40.14	39.93	41.29	-40.12	-39.92	-41.28	4	4	3
NGTDM_Busyness	14.99	12.81	16.49	3.70	-9.87	-15.69	2	4	4
NGTDM_Coarseness	15.71	12.08	18.93	-8.30	11.27	18.93	4	4	4
NGTDM_Complexity	17.06	10.56	15.57	12.08	-9.54	-15.57	4	4	4
NGTDM_Contrast	16.20	15.82	17.25	-13.41	-15.00	-16.88	4	4	4
NGTDM_Strength	14.14	12.07	18.52	-6.08	11.12	18.27	4	4	4

Fig. 8. Heat map of absolute relative error (ARE), relative error (RE), and ICC test of radiomic features extracted from lung segmentation in different reconstructed images with respect to original images.

the measure of privacy as the capability of reconstruction for unauthorized parties [19, 20, 18].

In order to follow Kerckhoffs' Principle in cryptography [57], we assume that the data release mechanism is publicly known [19, 20, 18]. In our evaluation, we found that the proposed method preserves important image content, and the different developed models achieve original image-level performance for various tasks, such as image classification, image segmentation, and texture analysis. Moreover, we discovered that various learning approaches, such as supervised and unsupervised image-to-image conversion, were unable to recover the images during attack analysis effectively.

There have been several studies on privacy-preserving techniques for medical images. In [91], a client-server system based on adversarial learning was proposed to obfuscate patient images to protect brain MR images' privacy. The system consists of encoders to remove patients' identity features, discriminators to identify patients from the encoded images and medical image analysis networks for image segmentation. However, the method presented in [50] does not encode the segmentation map, which could potentially reveal patient information, such as 3D renderings of the segmentation. In [50], a client-server system was proposed to preserve patient identity in brain MR images through the use of pseudo-random non-linear deformations on MR images, resulting in proxy images. A deep neural network was trained in an adversarial manner, with the flow-field generator, generating pseudo-random deformations to remove structural information.

Chen et al. [92] proposed a combination of encryption and digital watermark technology for the privacy-preserving sharing of medical images. The authoritative diagnosis results and image hash are integrated into QR code images using the discrete cosine transform (DCT) and inverse DCT (IDCT) algorithms and presented on watermark images [92]. This method was evaluated on chest X-rays and effectively preserved privacy against various attacks. In a more recent study, [3] used a GAN to create a medical image dataset that overcomes data-sharing barriers. The goal was to generate synthetic patient data with similar properties to the original images but without personal information [3]. The method was evaluated on chest X-ray and CT images and was found to produce high-quality generated images [3]. Popescu et al. [93] proposed the use of a variational auto-encoder combined with random non-bijective pixel intensity mapping, called obfuscation, for angiographic images. They claimed to be able to ensure privacy without allowing the recovery of coronary vessels using AI attacks.

The SCA privacy mechanism has several advantages for sharing privacy-preserving medical images. In the current research, we developed a system that is immune to various cyber-attacks and can be implemented on large datasets. This approach allows compact images to be shared with any party, even publicly available, and authorized clients can restore the images without loss of information or ambiguity. Moreover, our method allows the sharing of real data instead of synthetic data. In addition, the SCA approach can be integrated into any learning method, such as centralized, decentralized, distributed, and blended learning methods.

As for the real clinical implementation of the proposed solution in a practical setting, the training phase has similar considerations as any typical modern DL application. This comes with a strong emphasis on the security of patient data during training, as is required in hospital settings. More importantly, the key encryption and sharing phase should be done with care and through secure communication and storage protocols, as the security of the whole system relies on keeping the keys secure. Since image sharing is typically done in multi-party settings, e.g., between patients, clinics, imaging centers, and hospitals, asymmetric key encryption solutions, such as RSA [94], could be used to encrypt the keys based on the recipient's public key. Fortunately, the size of plaintext keys, i.e., the correct

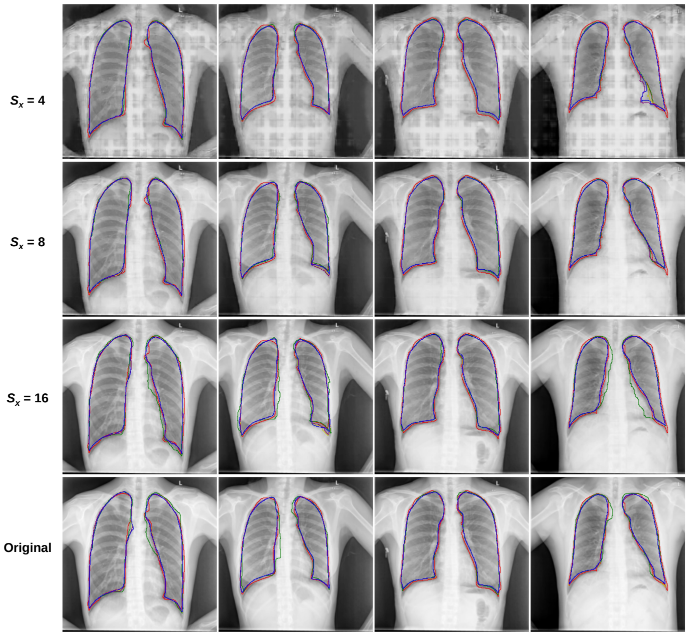


Fig. 9. Comparison of different image segmentation of the original images, reconstructed ($S_x = 16$, $S_x = 8$ and $S_x = 4$) images. Red: Ground Truth, Green: $S_x = 4$, Blue: $S_x = 8$, Yellow: $S_x = 16$, purple : Original.

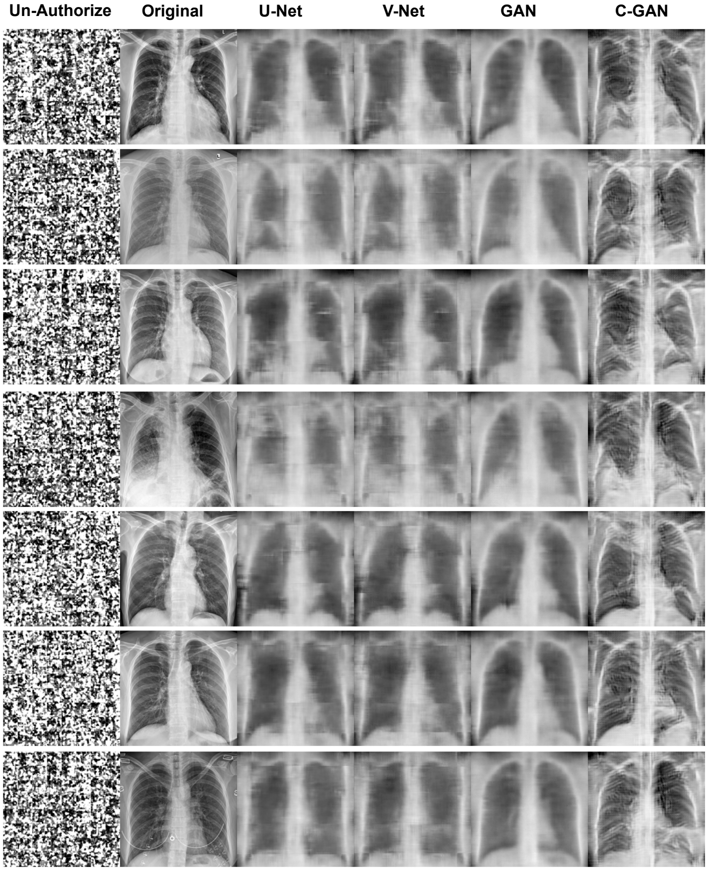


Fig. 10. Comparison of different network outputs for different attack analyses using different supervised (U-Net, V-Net, and GAN) and unsupervised (C-GAN) networks. Noisy images of $S_x = 16$ were set as input, and the neural networks attempted to reconstruct the original images.

position of non-zero elements in the sparse codes, is typically very small, as we showed in our experiments, and hence, the RSA or its variants may directly be applied. Finally, for the last phase, the decryption and decoding could be done in the user's computer, as the proposed decoder is relatively small and fast inference on consumer CPUs is possible in seconds for relatively large images.

The current study also has limitations. One limitation is that the proposed method only uses chest X-rays for evaluation. Other imaging modalities, such as CT, MRI, and PET, may contain more sensitive personal information that can be extracted from the images, such as the reconstruction of faces from 3D renderings [95]. However, the current study demonstrates the feasibility of preserving patient privacy in medical images, and further studies are needed to evaluate its performance on other imaging modalities. Furthermore, it is important to consider the privacy-preserving limitations of SCA methods while preserving important features relevant to radiological imaging tasks such as classification, segmentation, and texture analysis. The proposed methodology has shown promise in preserving privacy in 2D chest X-ray images. There is potential to expand it to other imaging modalities, such as ultrasound, and tomographic images including PET, CT, and MRI.

Our evaluation of generated images primarily relied on global image metrics, such as quantitative image-based metrics and radiomic features. However, due to the inherent limitations of the encoder-decoder architecture, smaller structures might be overlooked. Therefore, future studies should include a qualitative analysis by physicians to assess the impact of these limitations on small structures and clinical decision-making.

In addition to these areas of research, the SCA mechanism could be used in other situations, such as data sharing between several data owners for developing ML algorithms and any image-sharing protocol among authorized parties. The suggested approach could also be used to address privacy concerns during decentralized, distributed, and federated learning [6, 7, 8, 9]. As data collection and statistical analysis become increasingly prevalent in the healthcare industry, it is important to develop robust privacy-preserving techniques that can be widely adopted and used to ensure the protection of patient privacy. The SCA mechanism provides a promising robust solution to these challenges.

5. Conclusion

We developed a privacy-preserving medical image-sharing system that is resistant to different types of network attacks. Our approach leverages the SCA mechanism [18, 20], a generalization of randomization techniques that allows for a tradeoff between utility and privacy in a principled manner. We also introduced two architectural innovations for use in CNNs to improve the practicality and scalability of the SCA mechanism: multiple code maps using fully connected groups on convolutional filters and the S-sparsity non-linearity in CNNs. The results demonstrate the promising potential of the proposed method in private medical image sharing. While the current study only evaluated the proposed method on chest X-ray images, future research could explore its performance on other imaging modalities and in different privacy scenarios. The SCA mechanism offers a promising solution to the challenges of preserving privacy while still allowing for accurate and useful analysis.

6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. Acknowledgements

This work was supported by the Swiss National Science Foundation under grant 320030-176052 and the Private Foundation of Geneva University Hospitals under Grant RC-06–01.

References

[1] T. Carvalho, N. Moniz, P. Faria, L. Antunes, Survey on privacy-preserving techniques for data publishing (2022). `arXiv:2201.08120`.

[2] G. A. Kaissis, M. R. Makowski, D. Rückert, R. F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nature Machine Intelligence 2 (6) (2020) 305–311.

[3] A. DuMont Schütte, J. Hetzel, S. Gatidis, T. Hepp, B. Dietz, S. Bauer, P. Schwab, Overcoming barriers to data sharing with medical image generation: a comprehensive evaluation, NPJ digital medicine 4 (2021) 1–14.

[4] E. Topol, Deep medicine: how artificial intelligence can make healthcare human again, Hachette UK, 2019.

[5] B. Saboury, T. Bradshaw, R. Boellaard, I. Buvat, J. Dutta, M. Hatt, A. K. Jha, Q. Li, C. Liu, H. McMeekin, et al., Artificial intelligence in nuclear medicine: Opportunities, challenges, and responsibilities toward a trustworthy ecosystem, Journal of Nuclear Medicine (2022).

[6] I. Shiri, A. V. Sadr, M. Amini, Y. Salimi, A. Sanaat, A. Akhavanallaf, B. Razeghi, S. Ferdowsi, A. Saberi, H. Arabi, et al., Decentralized distributed multi-institutional pet image segmentation using a federated deep learning framework, Clinical Nuclear Medicine 47 (7) (2022) 606–617.

[7] I. Shiri, A. Vafaei Sadr, A. Akhavan, Y. Salimi, A. Sanaat, M. Amini, B. Razeghi, A. Saberi, H. Arabi, S. Ferdowsi, et al., Decentralized collaborative multi-institutional pet attenuation and scatter correction using federated deep learning, European Journal of Nuclear Medicine and Molecular Imaging 50 (4) (2023) 1034–1050.

[8] I. Shiri, Y. Salimi, M. Maghsudi, E. Jenabi, S. Harsini, B. Razeghi, S. Mostafaei, G. Hajianfar, A. Sanaat, E. Jafari, et al., Differential privacy preserved federated transfer learning for multi-institutional 68ga-pet image artefact detection and disentanglement, European journal of nuclear medicine and molecular imaging (2023) 1–14.

[9] I. Shiri, B. Razeghi, A. V. Sadr, M. Amini, Y. Salimi, S. Ferdowsi, P. Boor, D. Gündüz, S. Voloshynovskiy, H. Zaidi, Multi-institutional pet/ct image segmentation using federated deep transformer learning, Computer Methods and Programs in Biomedicine 240 (2023) 107706.

[10] C.-R. Shyu, K. T. Putra, H.-C. Chen, Y.-Y. Tsai, K. T. Hossain, W. Jiang, Z.-Y. Shae, A systematic review of federated learning in the healthcare area: From the perspective of data properties and applications, Applied Sciences 11 (23) (2021) 11191.

[11] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, H. V. Poor, Distributed learning in wireless networks: Recent progress and future challenges, IEEE Journal on Selected Areas in Communications 39 (12) (2021) 3579–3605.

[12] H. Kasyap, S. Tripathy, Privacy-preserving decentralized learning framework for healthcare system, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 17 (2s) (2021) 1–24.

[13] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, R. Sirdey, Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain, IEEE Signal Processing Magazine 30 (2) (2013) 108–117.

[14] C. Dwork, Differential privacy: A survey of results, in: International conference on theory and applications of models of computation, Springer, 2008, pp. 1–19.

[15] C. Huang, P. Kairouz, X. Chen, L. Sankar, R. Rajagopal, Context-aware generative adversarial privacy, Entropy 19 (12) (2017) 656.

[16] M. Gheisari, T. Furon, L. Amsaleg, B. Razeghi, S. Voloshynovskiy, Aggregation and embedding for group membership verification, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 2592–2596.

[17] B. Razeghi, S. Voloshynovskiy, D. Kostadinov, O. Taran, Privacy preserving identification using sparse approximation with ambiguization, in: IEEE Workshop on Information Forensics and Security (WIFS), 2017.

[18] B. Razeghi, S. Voloshynovskiy, Privacy-preserving outsourced media search using secure sparse ternary codes, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.

[19] B. Razeghi, S. Ferdowsi, D. Kostadinov, F. P. Calmon, S. Voloshynovskiy, Privacy-preserving near neighbor search via sparse coding with ambiguation, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 2635–2639.

[20] S. Ferdowsi, B. Razeghi, T. Holotyak, F. P. Calmon, S. Voloshynovskiy, Privacy-preserving image sharing via sparsifying layers on convolutional groups, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 2797–2801.

[21] M. E. Gursoy, A. Inan, M. E. Nergiz, Y. Saygin, Privacy-preserving learning analytics: challenges and techniques, IEEE Transactions on Learning technologies 10 (1) (2016) 68–81.

[22] M. Al-Rubaie, J. M. Chang, Privacy-preserving machine learning: Threats and solutions, IEEE Security & Privacy 17 (2) (2019) 49–58.

[23] S. Kahn, V. Sheshadri, Medical record privacy and security in a digital environment, IT professional 10 (2) (2008) 46–52.

[24] F. Cao, H. K. Huang, X. Zhou, Medical image security in a hipaa mandated pacs environment, Computerized medical imaging and graphics 27 (2-3) (2003) 185–196.

[25] Y. A. A. S. Aldeen, M. Salleh, M. A. Razzaque, A comprehensive review on privacy preserving data mining, SpringerPlus 4 (2015) 1–36.

[26] P. Goswami, S. Madan, Privacy preserving data publishing and data anonymization approaches: A review, in: International Conference on Computing, Communication and Automation, IEEE, 2017, pp. 139–142.

[27] M. G. Hansson, H. Lochmüller, O. Riess, F. Schaefer, M. Orth, Y. Rubinstein, C. Molster, H. Dawkins, D. Taruscio, M. Posada, et al., The risk of re-identification versus the need to identify individuals in rare disease research, European Journal of Human Genetics 24 (11) (2016) 1553–1558.

[28] M. M. P. Mr, C. A. Dhote, D. H. S. Mr, Homomorphic encryption for security of cloud data, Procedia Computer Science 79 (2016) 175–181.

[29] P. Jain, M. Gyanchandani, N. Khare, Big data privacy: a technological perspective and review, Journal of Big Data 3 (2016) 1–25.

[30] J. K. Liu, K. Liang, W. Susilo, J. Liu, Y. Xiang, Two-factor data security protection mechanism for cloud storage system, IEEE Transactions on Computers 65 (6) (2015) 1992–2004.

[31] European Commission, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) (2016).

[32] G. J. Annas, Hipaa regulations: a new era of medical-record privacy?, New England Journal of Medicine 348 (2003) 1486.

[33] D. J. Solove, P. M. Schwartz, Information privacy law, Aspen Publishing, 2020.

[34] M. B. Forcier, H. Gallois, S. Mullan, Y. Joly, Integrating artificial intelligence into health care through data access: can the gdpr act as a beacon for policymakers?, Journal of Law and the Biosciences 6 (1) (2019) 317.

[35] W. Newhauser, T. Jones, S. Swerdloff, W. Newhauser, M. Cilia, R. Carver, A. Halloran, R. Zhang, Anonymization of dicom electronic medical records for radiation therapy, Computers in biology and medicine 53 (2014) 134–140.

[36] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, M. Michalski, Medical image synthesis for data augmentation and anonymization using generative adversarial networks, in: International workshop on simulation and synthesis in medical imaging, Springer, 2018, pp. 1–11.

[37] T. Kossen, P. Subramaniam, V. I. Madai, A. Hennemuth, K. Hildebrand, A. Hilbert, J. Sobesky, M. Livne, I. Galinovic, A. A. Khalil, et al., Synthesizing anonymized and labeled tof-mra patches for brain vessel segmentation using generative adversarial networks, Computers in biology and medicine 131 (2021) 104254.

[38] G. Li, R. Togo, T. Ogawa, M. Haseyama, Compressed gastric image generation based on soft-label dataset distillation for medical data sharing, Computer Methods and Programs in Biomedicine 227 (2022) 107189.

[39] L. Hadjiiski, K. Cha, H.-P. Chan, K. Drukker, L. Morra, J. J. Näppi, B. Sahiner, H. Yoshida, Q. Chen, T. M. Deserno, et al., Aapm task group report 273: Recommendations on best practices for ai and machine learning for computer-aided diagnosis in medical imaging, Medical Physics.

[40] W. Cao, Y. Zhou, C. P. Chen, L. Xia, Medical image encryption using edge maps, Signal Processing 132 (2017) 96–109.

[41] A. Banu S, R. Amirtharajan, A robust medical image encryption in dual domain: chaos-dna-iwt combined approach, Medical & biological engineering & computing 58 (2020) 1445–1458.

[42] A. Wood, K. Najarian, D. Kahrobaei, Homomorphic encryption for machine learning in medicine and bioinformatics, ACM Computing Surveys (CSUR) 53 (4) (2020) 1–35.

[43] R. L. Rivest, L. Adleman, M. L. Dertouzos, et al., On data banks and privacy homomorphisms, Foundations of secure computation (1978).

[44] C. Zhang, S. Li, J. Xia, W. Wang, F. Yan, Y. Liu, {BatchCrypt}: Efficient homomorphic encryption for {Cross-Silo} federated learning, in: USENIX Annual Technical Conference, 2020, pp. 493–506.

[45] G. Kaissis, A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, I. Lima, J. Mancuso, F. Jungmann, M.-M. Steinborn, et al., End-to-end privacy preserving deep learning on multi-institutional medical imaging, Nature Machine Intelligence 3 (6) (2021) 473–484.

[46] R. L. Lagendijk, Z. Erkin, M. Barni, Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation, IEEE Signal Processing Magazine 30 (1) (2012).

[47] M. Naehrig, K. Lauter, V. Vaikuntanathan, Can homomorphic encryption be practical?, in: Proceedings of the 3rd ACM workshop on Cloud computing security workshop, 2011, pp. 113–124.

[48] M. Peleg, D. Beimel, D. Dori, Y. Denekamp, Situation-based access control: Privacy management via modeling of patient data access scenarios, Journal of Biomedical Informatics 41 (6) (2008) 1028–1040.

[49] M. A. Habib, M. Ahmad, S. Jabbar, S. Khalid, J. Chaudhry, K. Saleem, J. J. Rodrigues, M. S. Khalil, Security and privacy based access control model for internet of connected vehicles, Future Generation Computer Systems 97 (2019) 687–696.

[50] B. N. Kim, J. Dolz, C. Desrosiers, P.-M. Jodoin, Privacy preserving for medical image analysis via non-linear deformation proxy, arXiv preprint arXiv:2011.12835 (2020).

[51] M. Malekzadeh, B. Hasircioglu, N. Mital, K. Katarya, M. E. Ozfatura, D. Gündüz, Dopamine: Differentially private federated learning on medical data, in: AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI), 2021.

[52] D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, W.-J. Hwang, Federated learning for smart healthcare: A survey, ACM Computing Surveys (CSUR) 55 (3) (2022) 1–37.

[53] I. Shiri, E. Showkatian, R. Mohammadi, B. Razeghi, S. Bagheri, G. Hajianfar, Y. Salimi, M. Amini, M. Ghelich Oghli, S. Ferdowsi, S. Voloshynovskiy, H. Zaidi, Collaborative multi-institutional prostate lesion segmentation from mr images using deep federated learning framework, in: IEEE Nuclear Science Symposium, Medical Imaging Conference, 2022.

[54] I. Shiri, M. Amini, Y. Salimi, A. Sanaat, A. Saberi, B. Razeghi, S. Ferdowsi, A. V. Sadr, S. Voloshynovskiy, D. Gündüz, A. Rahmim, H. Zaidi, Multi-institutional pet/ct image segmentation using a decentralized federated deep transformer learning algorithm, Journal of Nuclear Medicine 63 (supplement 2) (2022) 3348–3348.

[55] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, Medical image analysis 58 (2019) 101552.

[56] Y. Chen, X.-H. Yang, Z. Wei, A. A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, Q. Guan, Generative adversarial networks in medical image augmentation: a review, Computers in Biology and Medicine (2022).

[57] A. Kerckhoffs, La cryptographic militaire, Journal des sciences militaires (1883) 5–38.

[58] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C. Burges, L. Bottou, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 25, Curran Associates, Inc., 2012.

[59] C. Renggli, S. Ashkboos, M. Aghagolzadeh, D. Alistarh, T. Hoefler, Sparcml: High-performance sparse communication for machine learning, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2019, pp. 1–15.

[60] Z. Qin, J. Fan, Y. Liu, Y. Gao, G. Y. Li, Sparse representation for wireless communications: A compressive sensing approach, IEEE Signal Processing Magazine 35 (3) (2018) 40–58.

[61] Y. Oktar, M. Turkan, A review of sparsity-based clustering methods, Signal processing 148 (2018) 20–30.

[62] K. Huang, S. Aviyente, Sparse representation for signal classification, Advances in neural information processing systems 19 (2006).

[63] S. Ravishankar, J. C. Ye, J. A. Fessler, Image reconstruction: From sparsity to data-adaptive methods and machine learning, Proceedings of the IEEE 108 (1) (2019) 86–109.

[64] D. Kostadinov, S. Voloshynovskiy, S. Ferdowsi, Learning overcomplete and sparsifying transform with approximate and exact closed form solutions, in: European Workshop on Visual Information, IEEE, 2018.

[65] S. Ravishankar, Y. Bresler, Learning sparsifying transforms, IEEE Transactions on Signal Processing 61 (5) (2012) 1072–1086.

[66] S. Rezaeifar, B. Razeghi, O. Taran, T. Holotyak, S. Voloshynovskiy, Reconstruction of privacy-sensitive data from protected templates, in: 2019 IEEE International Conference on Image Processing (ICIP), IEEE, 2019, pp. 1163–1167.

[67] B. Razeghi, S. Rezaeifar, S. Ferdowsi, T. Holotyak, S. Voloshynovskiy, Compressed data sharing based on information bottleneck model, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 3009–3013.

[68] K. Sayood, Introduction to data compression, Morgan Kaufmann, 2017.

[69] V. K. Goyal, Theoretical foundations of transform coding, IEEE Signal Processing Magazine 18 (5) (2001) 9–21.

[70] M. Gregorová, M. Desaules, A. Kalousis, Learned transform compression with optimized entropy encoding, in: ICLR Neural Compression Workshop, 2021.

[71] D. A. Huffman, A method for the construction of minimum-redundancy codes, Proceedings of the IRE 40 (9) (1952) 1098–1101.

[72] R. Clarke, Transform coding of images, Academic Press Professional, Inc., 1985.

[73] H. S. Malvar, D. H. Staelin, The lot: Transform coding without blocking effects, IEEE Transactions on Acoustics, Speech, and Signal Processing 37 (4) (1989) 553–559.

[74] G. Toderici, S. M. O'Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, R. Sukthankar, Variable rate image compression with recurrent neural networks, in: International Conference on Learning Representations (ICLR), 2016.

[75] K. Gregor, F. Besse, D. Jimenez Rezende, I. Danihelka, D. Wierstra, Towards conceptual compression, Advances In Neural Information Processing Systems 29 (2016).

[76] J. Ballé, V. Laparra, E. P. Simoncelli, End-to-end optimized image compression, in: International Conference on Learning Representations (ICLR), 2017.

[77] L. Theis, W. Shi, A. Cunningham, F. Huszár, Lossy image compression with compressive autoencoders, in: International Conference on Learning Representations (ICLR), 2017.

[78] L. Wu, K. Huang, H. Shen, A gan-based tunable image compression system, in: IEEE/CVF winter conference on applications of computer vision, 2020, pp. 2334–2342.

[79] E. Hoogeboom, E. Agustsson, F. Mentzer, L. Versari, G. Toderici, L. Theis, High-fidelity image compression with score-based generative models, arXiv preprint arXiv:2305.18231 (2023).

[80] B. Razeghi, F. P. Calmon, D. Gunduz, S. Voloshynovskiy, Bottlenecks CLUB: Unifying information-theoretic trade-offs among complexity, leakage, and utility, IEEE Transactions on Information Forensics and Security 18 (2023) 2060–2075.

[81] N. Tishby, F. C. Pereira, W. Bialek, The information bottleneck method, in: IEEE Allerton, 2000.

[82] A. Makhdoumi, S. Salamatian, N. Fawaz, M. Médard, From the information bottleneck to the privacy funnel, in: IEEE Information Theory Workshop (ITW), 2014, pp. 501–505.

[83] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: IEEE conference on computer vision and pattern recognition, 2017.

[84] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, cell 172 (5) (2018) 1122–1131.

[85] H. Q. Nguyen, K. Lam, L. T. Le, H. H. Pham, D. Q. Tran, D. B. Nguyen, D. D. Le, C. M. Pham, H. T. Tong, D. H. Dinh, et al., Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations, Scientific Data 9 (1) (2022) 429.

[86] A. Bustos, A. Pertusa, J.-M. Salinas, M. De La Iglesia-Vaya, Padchest: A large chest x-ray image dataset with multi-label annotated reports, Medical image analysis 66 (2020) 101797.

[87] 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: AAAI conference on artificial intelligence, Vol. 33, 2019.

[88] L. Wang, Z. Q. Lin, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, Scientific reports 10 (1) (2020) 19549.

[89] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, H. J. Aerts, Computational radiomics system to decode the radiographic phenotype, Cancer research 77 (21) (2017) e104–e107.

[90] A. Depeursinge, V. Andrearczyk, P. Whybra, J. van Griethuysen, H. Müller, R. Schaer, M. Vallières, A. Zwanenburg, Standardised convolutional filtering for radiomics, arXiv preprint arXiv:2006.05470 (2020).

[91] B. N. Kim, J. Dolz, P.-M. Jodoin, C. Desrosiers, Privacy-net: An adversarial approach for identity-obfuscated segmentation of medical images, IEEE Transactions on Medical Imaging 40 (7) (2021) 1737–1749.

[92] L. Chen, W. Bai, Z. Yao, A secure and privacy-preserving watermark based medical image sharing method, Chinese Journal of Electronics 29 (5) (2020) 819–825.

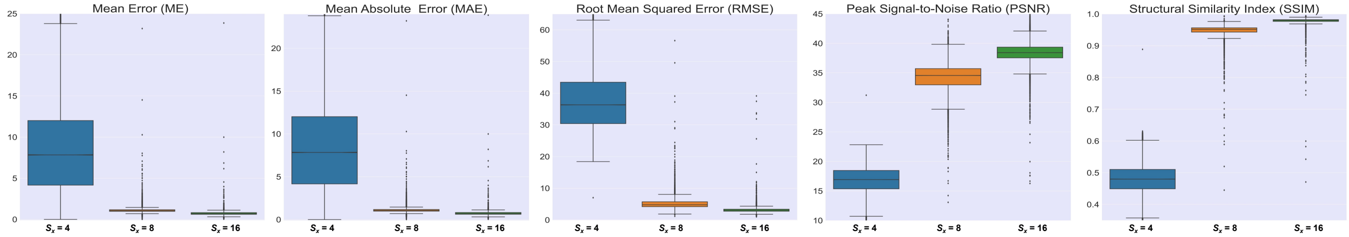
[93] A. B. Popescu, I. A. Taca, A. Vizitiu, C. I. Nita, C. Suciu, L. M. Itu, A. Scafa-Udriste, Obfuscation algorithm for privacy-preserving deep learning-based medical image analysis, Applied Sciences 12 (8) (2022).

[94] R. L. Rivest, A. Shamir, L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Communications of the ACM 21 (2) (1978) 120–126.

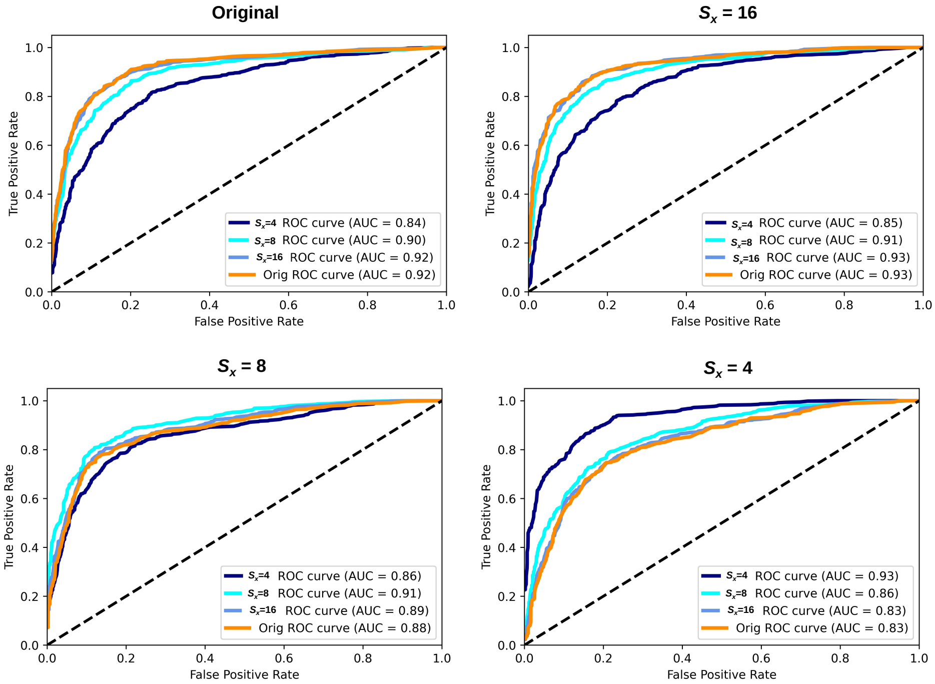
[95] J. C. Mazura, K. Juluru, J. J. Chen, T. A. Morgan, M. John, E. L. Siegel, Facial recognition software success rates for the identification of 3d surface reconstructed facial images: implications for patient privacy and security, Journal of digital imaging 25 (2012) 347–351.

Supplemental Results

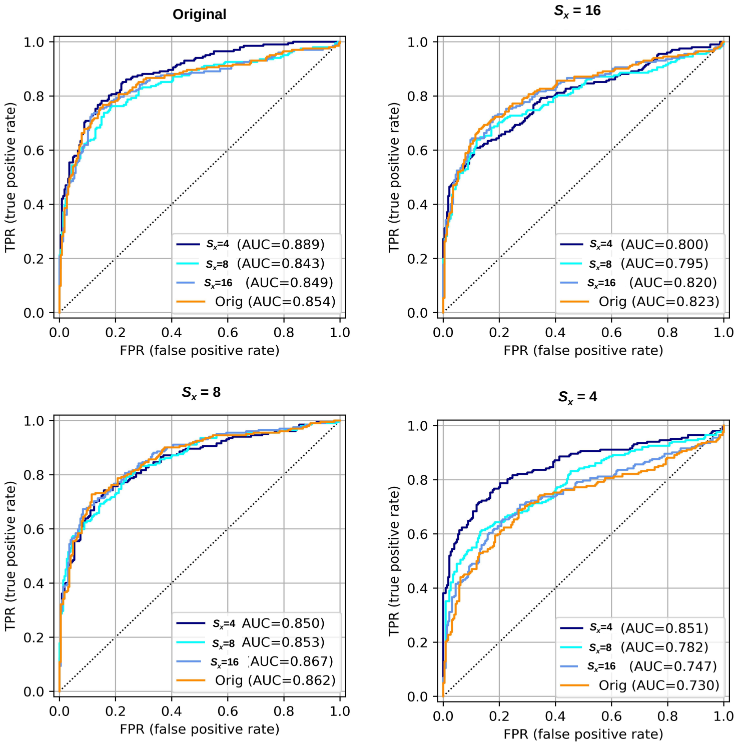
(Supplemental Figures)



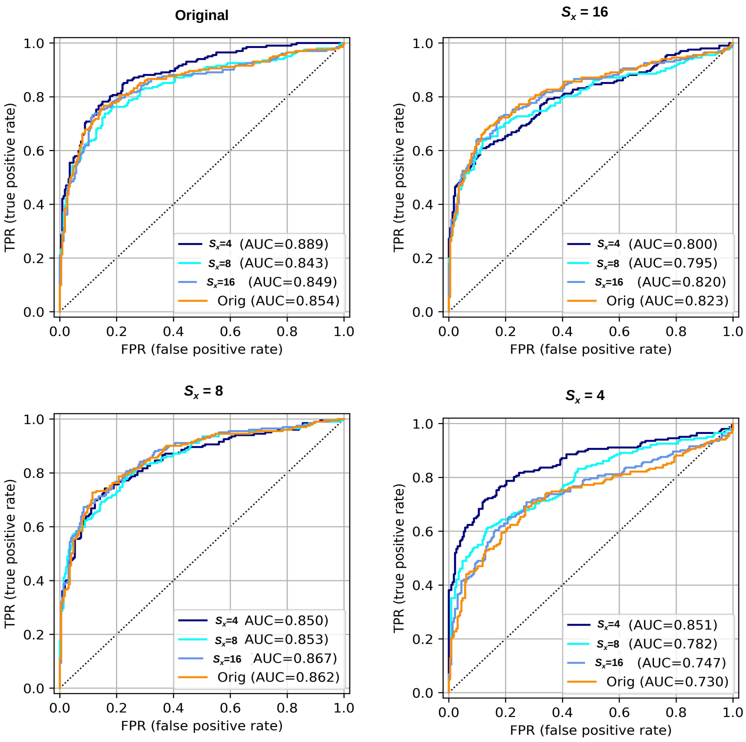
Supplemental Fig 1. Comparison of fidelity measures including mean error (ME), mean absolute error (MAE), peak signal to noise ratio (PSNR), structural similarity index (SSIM), and root-mean-square error (RMSE) as evaluated by an authorized party. The results were obtained by comparing reconstructed images with sparsity levels S_x of 4, 8, and 16 to the original image. The performance of each measure is evaluated in terms of its ability to reflect the quality of the reconstructed images accurately. The box plots display the intra-quartile range (IQR), minimum (Q1–1.5×IQR), first quartile (Q1), median, third quartile (Q3), maximum (Q3+1.5×IQR) and outliers.



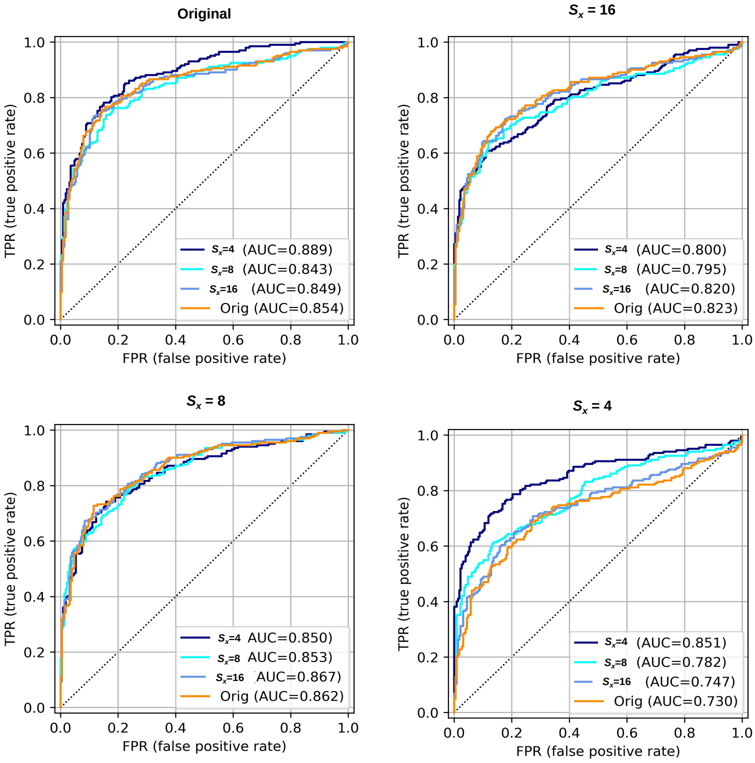
Supplemental Fig 2. Comparison of receiver operating characteristic (ROC) curves for different training and test sets in a three-class classification Task 3 for normal, bacterial, and viral pneumonia. The model was trained on a set of original images and then tested on a separate set of original images as well as reconstructed images with sparsity levels of S_x = 4, 8, and 16. The ROC curves represent the performance of the model in terms of true positive rate and false positive rate when classifying the different types of pneumonia.



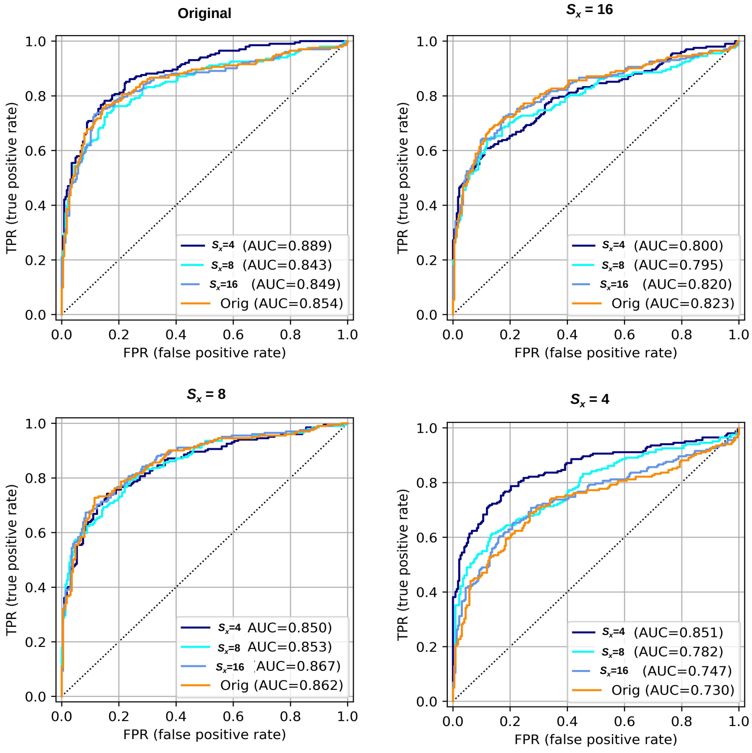
Supplemental Fig 3. ROC cure comparison for different training and test sets in Task 2 classification: classification of bacterial pneumonia against viral pneumonia. Training and testing were on different images, i.e, model built on original images as training sets and then tested on 20% of the test set of original images, and reconstructed images with sparsity levels S_x of 4, 8, and 16.



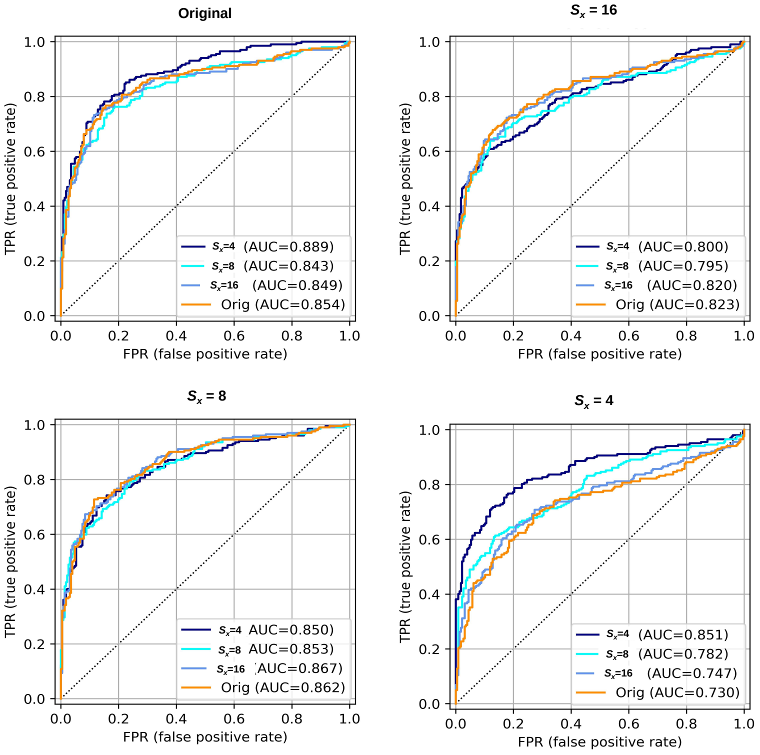
Supplemental Fig 4. ROC cure comparison for different training and test sets in Task 3 classification: classification of viral COVID-19 pneumonia against viral pneumonia. Training and testing were on different images, i.e, model built on original images as training sets and then tested on 20% of the test set of original images, and reconstructed images with sparsity levels S_x of 4, 8, and 16.



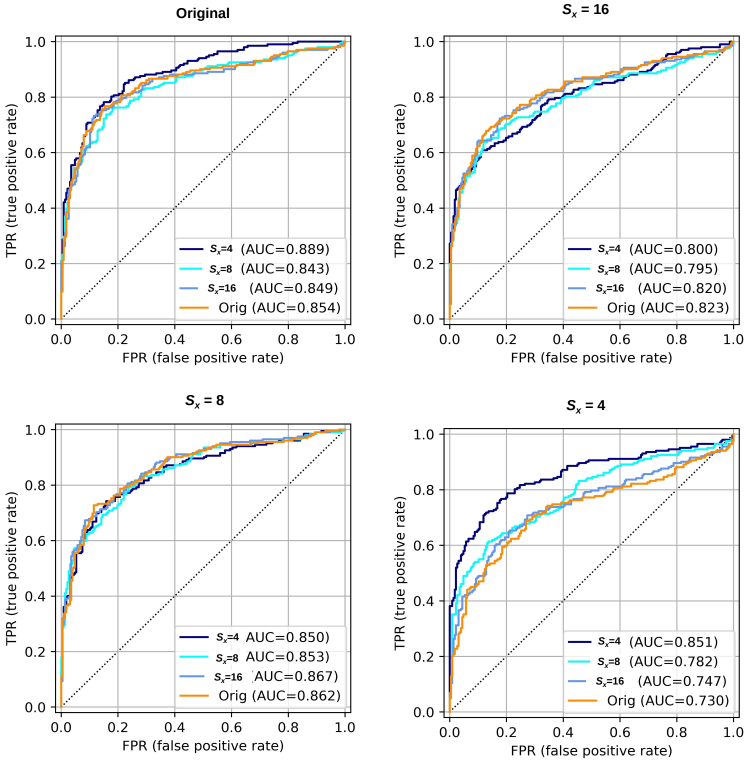
Supplemental Fig 5. ROC cure comparison for different training and test sets in Task 4 classification: classification of viral COVID-19 pneumonia against viral pneumonia + bacterial pneumonia. Training and testing were on different images, i.e, model built on original images as training sets and then tested on 20% of the test set of original images, and reconstructed images with sparsity levels S_x of 4, 8, and 16.



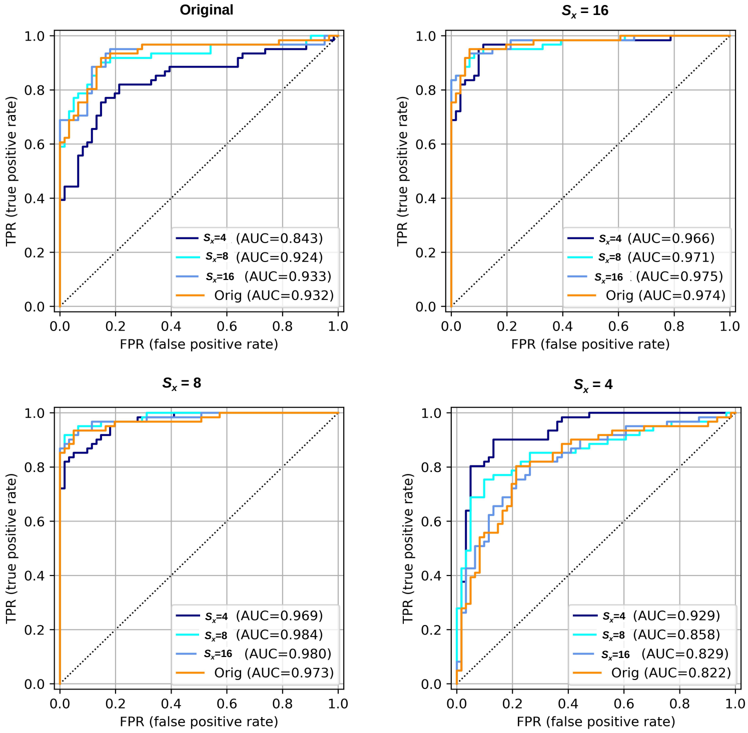
Supplemental Fig 6. ROC cure comparison for different training and test sets in Task 5 classification: Classification of normal cases against viral pneumonia +bacterial pneumonia. Training and testing were on different images, i.e, model built on original images as training sets and then tested on 20% of the test set of original images, and reconstructed images with sparsity levels S_x of 4, 8, and 16.



Supplemental Fig 7. ROC cure comparison for different training and test sets in Task 6 classification: classification of normal cases against viral pneumonia. Training and testing were on different images, i.e, model built on original images as training sets and then tested on 20% of the test set of original images, and reconstructed images with sparsity levels S_x of 4, 8, and 16.



Supplemental Fig 8. ROC cure comparison for different training and test sets in Task 7 classification: classification of normal cases against bacterial pneumonia. Training and testing were on different images, i.e, model built on original images as training sets and then tested on 20% of the test set of original images, and reconstructed images with sparsity levels S_x of 4, 8, and 16.



Supplemental Fig 9. ROC cure comparison for different training and test sets in Task 8 classification: classification of normal cases against viral COVID-19 pneumonia.Training and testing were on different images, i.e, model built on original images as training sets and then tested on 20% of the test set of original images, and reconstructed images with sparsity levels S_x of 4, 8, and 16.

(Supplemental Tables)

Supplemental Table 1. Accuracy of different classification tasks for different training and test sets, Task 1: Three class classification of normal, bacterial, and viral pneumonia, Task 2: Classification of bacterial pneumonia against viral pneumonia, Task 3: Classification of viral COVID-19 pneumonia against viral pneumonia, Task 4: Classification of viral COVID-19 pneumonia against viral pneumonia + bacterial pneumonia, Task 5: Classification of normal cases against viral pneumonia +bacterial pneumonia, Task 6: classification of normal cases against viral pneumonia, Task 7: classification of normal cases against bacterial pneumonia, Task 8: classification of normal cases against viral COVID-19 pneumonia.

Train	Test	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
$S_x = 4$	$S_x = 4$	0.80	0.80	0.93	0.90	0.96	0.98	0.92	0.89
	$S_x = 8$	0.68	0.75	0.93	0.85	0.94	0.98	0.86	0.83
	$S_x = 16$	0.66	0.73	0.93	0.82	0.94	0.98	0.83	0.78
	Original	0.66	0.71	0.93	0.85	0.94	0.97	0.82	0.80
$S_x = 8$	$S_x = 4$	0.73	0.79	0.99	0.9	0.91	0.96	0.84	0.90
	$S_x = 8$	0.8	0.78	0.98	0.90	0.95	0.98	0.93	0.95
	$S_x = 16$	0.75	0.80	0.98	0.88	0.94	0.99	0.91	0.93
	Original	0.75	0.81	0.98	0.90	0.94	0.99	0.90	0.94
$S_x = 16$	$S_x = 4$	0.66	0.76	0.94	0.87	0.87	0.96	0.83	0.93
	$S_x = 8$	0.78	0.77	0.95	0.86	0.95	0.99	0.91	0.93
	$S_x = 16$	0.80	0.78	0.97	0.86	0.96	0.99	0.93	0.93
	Original	0.82	0.78	0.96	0.87	0.96	0.99	0.92	0.94
Original	$S_x = 4$	0.67	0.82	0.96	0.89	0.85	0.92	0.74	0.80
	$S_x = 8$	0.76	0.79	0.95	0.9	0.94	0.98	0.87	0.88
	$S_x = 16$	0.78	0.81	0.95	0.89	0.96	0.99	0.90	0.89
	Original	0.79	0.81	0.95	0.90	0.95	0.99	0.90	0.89

Supplemental Table 2. Sensitivity of different classification tasks for different training and test sets, Task 1: Three class classification of normal, bacterial, and viral pneumonia, Task 2: Classification of bacterial pneumonia against viral pneumonia, Task 3: Classification of viral COVID-19 pneumonia against viral pneumonia, Task 4: Classification of viral COVID-19 pneumonia against viral pneumonia + bacterial pneumonia, Task 5: Classification of normal cases against viral pneumonia +bacterial pneumonia, Task 6: classification of normal cases against viral pneumonia, Task 7: classification of normal cases against bacterial pneumonia, Task 8: classification of normal cases against viral COVID-19 pneumonia.

Train	Test	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
$S_x = 4$	$S_x = 4$	0.79	0.71	0.93	0.92	0.97	0.97	0.9	0.90
	$S_x = 8$	0.66	0.61	0.92	0.82	0.97	0.97	0.80	0.75
	$S_x = 16$	0.64	0.60	0.98	0.9	0.96	0.97	0.84	0.82
	Original	0.64	0.53	0.98	0.77	0.95	0.96	0.84	0.80
$S_x = 8$	$S_x = 4$	0.71	0.74	0.98	0.84	0.93	0.94	0.82	0.82
	$S_x = 8$	0.79	0.69	0.97	0.80	0.96	0.99	0.95	0.92
	$S_x = 16$	0.74	0.67	0.98	0.80	0.97	0.98	0.93	0.92
	Original	0.74	0.73	0.98	0.84	0.92	0.99	0.88	0.93
$S_x = 16$	$S_x = 4$	0.65	0.6	0.93	0.9	0.88	0.96	0.77	0.95
	$S_x = 8$	0.78	0.64	0.93	0.87	0.95	0.99	0.94	0.92
	$S_x = 16$	0.79	0.64	0.97	0.75	0.97	0.98	0.93	0.92
	Original	0.81	0.68	0.97	0.82	0.99	0.98	0.89	0.95
Original	$S_x = 4$	0.65	0.77	0.98	0.93	0.86	0.92	0.85	0.75
	$S_x = 8$	0.75	0.74	0.97	0.84	0.94	0.97	0.84	0.89
	$S_x = 16$	0.77	0.73	0.97	0.82	0.96	0.98	0.88	0.93
	Original	0.78	0.77	0.98	0.84	0.96	0.98	0.91	0.92

Supplemental Table 3. Specificity of different classification tasks for different training and test sets, Task 1: Three class classification of normal, bacterial, and viral pneumonia, Task 2: Classification of bacterial pneumonia against viral pneumonia, Task 3: Classification of viral COVID-19 pneumonia against viral pneumonia, Task 4: Classification of viral COVID-19 pneumonia against viral pneumonia + bacterial pneumonia, Task 5: Classification of normal cases against viral pneumonia +bacterial pneumonia, Task 6: classification of normal cases against viral pneumonia, Task 7: classification of normal cases against bacterial pneumonia, Task 8: classification of normal cases against viral COVID-19 pneumonia.

Train	Test	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
$S_x = 4$	$S_x = 4$	0.90	0.88	0.93	0.87	0.95	0.99	0.93	0.87
	$S_x = 8$	0.84	0.86	0.93	0.87	0.91	0.99	0.90	0.90
	$S_x = 16$	0.83	0.84	0.87	0.75	0.93	0.99	0.83	0.74
	Original	0.83	0.87	0.87	0.92	0.93	0.99	0.81	0.79
$S_x = 8$	$S_x = 4$	0.86	0.84	0.99	0.95	0.9	0.98	0.87	0.98
	$S_x = 8$	0.90	0.85	0.99	0.98	0.95	0.97	0.92	0.98
	$S_x = 16$	0.88	0.92	0.98	0.95	0.92	0.99	0.90	0.95
	Original	0.87	0.89	0.98	0.95	0.95	0.99	0.93	0.95
$S_x = 16$	$S_x = 4$	0.83	0.89	0.95	0.84	0.87	0.96	0.87	0.90
	$S_x = 8$	0.89	0.88	0.97	0.86	0.94	0.99	0.89	0.93
	$S_x = 16$	0.90	0.90	0.97	0.97	0.95	0.99	0.92	0.95
	Original	0.91	0.87	0.95	0.92	0.93	0.99	0.94	0.93
Original	$S_x = 4$	0.83	0.86	0.93	0.84	0.85	0.92	0.63	0.85
	$S_x = 8$	0.88	0.84	0.93	0.95	0.93	0.99	0.90	0.87
	$S_x = 16$	0.89	0.89	0.93	0.95	0.95	0.99	0.91	0.84
	Original	0.90	0.85	0.92	0.95	0.95	0.99	0.89	0.85

Supplemental Table 4. CI95% Segmentation parameters for different training and test sets.

Train	Test	Dice	Jaccard	False Negative	False Positive	Mean Surface Distance	Std Surface Distance
$S_x = 4$	$S_x = 4$	0.93 to 0.94	0.87 to 0.89	0.10 to 0.11	0.01 to 0.03	0.12 to 0.15	1.07 to 1.35
	$S_x = 8$	0.93 to 0.94	0.87 to 0.89	0.10 to 0.11	0.01 to 0.03	0.12 to 0.15	1.07 to 1.35
	$S_x = 16$	0.93 to 0.94	0.87 to 0.89	0.10 to 0.11	0.01 to 0.03	0.13 to 0.15	1.09 to 1.37
	Original	0.93 to 0.94	0.87 to 0.88	0.10 to 0.11	0.02 to 0.03	0.13 to 0.16	1.15 to 1.46
$S_x = 8$	$S_x = 4$	0.93 to 0.94	0.87 to 0.89	0.09 to 0.11	0.02 to 0.03	0.12 to 0.14	1.09 to 1.32
	$S_x = 8$	0.94 to 0.95	0.88 to 0.90	0.09 to 0.11	0.01 to 0.01	0.11 to 0.13	0.91 to 1.13
	$S_x = 16$	0.94 to 0.95	0.88 to 0.90	0.09 to 0.11	0.01 to 0.01	0.11 to 0.13	0.91 to 1.13
	Original	0.94 to 0.95	0.88 to 0.90	0.09 to 0.11	0.01 to 0.01	0.11 to 0.13	0.91 to 1.12
$S_x = 16$	$S_x = 4$	0.91 to 0.92	0.84 to 0.86	0.07 to 0.09	0.07 to 0.09	0.16 to 0.21	1.59 to 2.05
	$S_x = 8$	0.94 to 0.95	0.88 to 0.90	0.08 to 0.10	0.02 to 0.03	0.11 to 0.13	0.97 to 1.23
	$S_x = 16$	0.94 to 0.95	0.89 to 0.90	0.08 to 0.10	0.02 to 0.03	0.11 to 0.13	0.95 to 1.20
	Orig	0.94 to 0.95	0.89 to 0.90	0.08 to 0.09	0.02 to 0.03	0.10 to 0.13	0.93 to 1.18
Original	$S_x = 4$	0.92 to 0.94	0.86 to 0.88	0.07 to 0.09	0.05 to 0.06	0.12 to 0.23	1.23 to 1.96
	$S_x = 8$	0.94 to 0.95	0.89 to 0.90	0.08 to 0.09	0.02 to 0.02	0.10 to 0.12	0.88 to 1.10
	$S_x = 16$	0.94 to 0.95	0.89 to 0.90	0.08 to 0.09	0.02 to 0.03	0.10 to 0.12	0.91 to 1.17
	Original	0.94 to 0.95	0.89 to 0.90	0.08 to 0.09	0.02 to 0.03	0.10 to 0.12	0.91 to 1.16