Unsupervised contrastive unpaired image generation approach for improving tuberculosis screening using chest X-ray images

Daniel I. Morís [a,b], Joaquim de Moura [a,b,*], Jorge Novo [a,b], Marcos Ortega [a,b]

[a] Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, Xubias de Arriba, 84, A Coruña, 15006, Spain
[b] Centro de investigación CITIC, Universidade da Coruña, Campus de Elviña, s/n, A Coruña, 15071, Spain

A R T I C L E   I N F O

A B S T R A C T

Tuberculosis is an infectious disease that mainly affects the lung tissues. Therefore, chest X-ray imaging can be very useful to diagnose and to understand the evolution of the pathology. This image modality has a poorer quality in contrast with other techniques as the magnetic resonance or the computerized tomography, but chest X-ray is easier and cheaper to perform. Furthermore, data scarcity is challenging in the domain of biomedical imaging. In order to mitigate this problem, the use of Generative Adversarial Network models for image generation has proved to be a powerful approach to train the deep learning models with small datasets, representing an alternative to classic data augmentation strategies. In this work, we propose a fully automatic approach for the generation of novel synthetic chest X-ray images to mitigate the effect of data scarcity in order to improve the tuberculosis screening performance using 3 different publicly available representative datasets: Montgomery County, Shenzhen and TBX11K. Firstly, this approach trains image translation models with a large-sized dataset (TBX11K). Then, these models are used to generate the novel set of synthetic images using small-sized and medium-sized datasets (Montgomery County and Shenzhen, respectively). Finally, the novel set of generated images is added to the training set to improve the performance of an automatic tuberculosis screening. As a result, we obtained an 88.41% ± 5.27% of accuracy for the Montgomery County dataset and a 90.33% ± 1.41% for the Shenzhen dataset. These results demonstrate that the proposed method outperforms previous state-of-the-art approaches.

1. Introduction

Tuberculosis is an infectious disease caused by the bacillus *Mycobacterium tuberculosis* that can be transmitted through the air [1]. While the tuberculosis can damage several parts of the body, the main affectation is caused in the lungs. This disease is included within the 10 main causes of death worldwide. As reference, in 2019, around 10 million people were infected with tuberculosis causing, in addition, more than 1.4 million deaths [2].

Chest X-ray imaging has been widely used to diagnose and to assess the evolution of several pulmonary diseases such as pneumonia and, more recently, it has also proven its suitability in diagnosing the COVID-19 affectation. This image modality is used for those purposes because it offers several advantages, as it is cheap, non-invasive, and easy to capture, in contrast with more advanced imaging techniques as the computerized tomography (CT) [3] or the magnetic resonance (MRI) [4]. However, despite these advantages, the quality of the captures from chest X-ray devices is considerably poorer, which is reflected in a lower level of detail, an aspect that implies a more challenging diagnostic process for the clinicians.

Moreover, in the last years, biomedical imaging domains have been supported by computer-aided diagnosis (CAD) systems, that can help the clinicians to make decisions. As the diagnostic process is a tedious and error-prone task, automatic methods emerge as useful techniques to mitigate these problems. In that sense, deep learning strategies, very relevant in the current biomedical imaging discipline, have shown great potential in dealing with challenging problems obtaining satisfactory and robust results. In fact, as the tuberculosis is still a challenging disease in many countries, some works have addressed the problem of the automatic tuberculosis screening.

* Corresponding author.
  E-mail addresses: daniel.iglesias.moris@udc.es (D.I. Morís),
joaquim.demoura@udc.es (J. de Moura), jnovo@udc.es (J. Novo), mortega@udc.es
(M. Ortega).

As reference, we can find several remarkable contributions, such as the work from Pasa et al. [5] that proposes a novel and efficient deep network architecture that maintains a satisfactory effectiveness despite being simpler and faster than previous models of the state of the art. The contribution from Alfadhli et al. [6] uses classical computer vision techniques to extract features from images and Support Vector Machines (SVMs) to perform the classification. On the other hand, the work from Lopes et al. [7] takes advantage of pretrained models as feature extractors, using those features as an input for a Support Vector Machine. The work from Hwang et al. [8] proposes a deep convolutional neural network approach to tackle the tuberculosis screening, which is based on the well-known AlexNet architecture [9]. On the other hand, Jaeger et al. [10] propose a methodology that performs lung region segmentation followed by feature extraction using classical computer vision techniques based on color, texture, and shape characteristics that are finally used as input of an SVM model to perform a binary classification. Finally, the work from Ali et al. [11] proposes the application of a modular convolutional neural network architecture to solve problems in the scope of malaria detection, diabetic retinopathy detection and tuberculosis detection.

Nevertheless, deep learning-based models need a considerable amount of data to be trained, which is scarce in many domains, as is the case of the biomedical imaging modalities. To mitigate the problem of data scarcity, several contributions have proposed the use of different strategies. As reference, some works have studied the application of incremental learning schemes [12] that consist of training with a subset of diseases that keeps growing on each incremental step but, at the same time, avoiding the forgetting problem, training with few amounts of images of the previous steps. These schemes have been properly adapted to medical imaging, in domains such as Optical Coherence Tomography (OCT) for eye fundus analysis [13] and chest X-ray imaging [14]. However, these schemes are unsuitable for scenarios where the number of classes is small as, for instance, in binary problems. For such cases, several authors have proposed different data augmentation techniques [15]. The most traditional data augmentation techniques consist on applying trivial transformations to the images such as random rotations, translations or pixel intensity changes to artificially increase the dimensionality of the original datasets. As trivial transformations could be insufficiently representative of the great variability of these domains, one of the alternatives proposed in the last years is the synthetic image generation, taking advantage of Generative Adversarial Network architectures (GANs) [16]. These models are able to generate realistic synthetic images representative of the domain.

There are several approaches of Generative Adversarial Networks. Particularly, in the field of image generation using image translation models, we can find several implementations, such as the Contrastive Unpaired Translation (CUT) [17]. Basically, these models are able to convert images from certain scenarios to other different scenarios without the necessity of paired data. Then, this emerges as a powerful approach for synthetic image generation. However, translation models can also be affected by data scarcity, as training with insufficient data can lead the models to perform useless transformations on images.

Some works have tackled the use of generative models in the context of chest X-ray imaging. As reference, the work from Malygina et al. [18] uses a specific GAN implementation for image translation, CycleGAN [19], to improve the pneumonia classification on chest X-ray images using the novel set of generated images and, therefore, avoid the effect of high-imbalanced datasets. More recently, the use of CycleGAN in this context has been also taken into account to improve the COVID-19 screening. As reference, the work from Morís et al. [20] proves the utility of generating synthetic chest X-ray images provided by portable devices in the con-

text of COVID-19, using the CycleGAN to perform a translation in 3 different scenarios. Later, the same authors proved in [21] that using the novel set of images generated by the CycleGAN improves the performance of the automatic COVID-19 screening. Moreover, the work from Bargshady et al. [22] also proposes the generation of synthetic chest X-ray images with the CycleGAN to improve the performance of a fine-tuned Inception V3 model to distinguish between COVID-19 and non-COVID-19.

In this work, we propose a fully automatic approach to generate useful and representative samples for the tuberculosis screening in chest X-ray image datasets. To do so, we take into account an image translation architecture to convert healthy samples to their hypothetical versions with tuberculosis affectation and vice versa. Then, the novel set of synthetic images is added to the original dataset to improve the performance of the screening model. To mitigate the problem of data scarcity in this restricted domain, we train the translation models using a large-sized dataset, the TBX11K, while the image generation is performed with small-sized and medium-sized datasets (Montgomery County and Shenzhen, respectively). To the best of our knowledge, this is the only work that addresses the adaption of the contrastive unpaired translation network for synthetic chest X-ray image generation to improve tuberculosis screening. Given the used imaging modality and the proposed methodology, our contribution is suitable for challenging scenarios as health emergency centers or screening programs.

The rest of the manuscript is structured as follows. Firstly, in Section 2, we describe the 3 public datasets that were used for the development of the work as well as the steps of the methodological proposal. In Section 3 we present the obtained results and their discussion. Finally, Section 4 states the main conclusions obtained in the same way as some possible lines of future work.

2. Materials and methods

The proposed methodology is divided into 2 different alternatives to prove both the performance of the model when training with only original images and when adding the novel set of generated images to the original datasets. In the first alternative, called baseline, the screening models are trained using only the original images from the input datasets. For the second alternative, we train 2 CUT models for image translation as a data augmentation approach, adding the novel set of generated images to the original dataset to improve the performance of the original screening model. An overview of the proposed methodology can be seen in Fig. 1 with both alternatives. As it can be observed, the first alternative (the baseline) has only one part, the automatic tuberculosis screening itself, while the second alternative (the data augmentation approach) has 2 different parts: the synthetic image generation followed by the automatic tuberculosis screening.

2.1. Chest X-ray images datasets

3 datasets were used for the development of this work. The first 2 are representative of a small and a medium-sized dataset, the Montgomery County [23] and the Shenzhen [23] datasets, respectively. Lastly, the third dataset, TBX11K, is representative of a large-sized dataset [24].

Montgomery County dataset (**available at** [25]). This dataset, often abbreviated as "Montgomery dataset", was acquired from the tuberculosis control program of the Department of Health and Human Services of Montgomery County in the United States having 138 subjects (58 normal and 80 pathological), being representative of a small dataset.

Shenzhen dataset (**available at** [26]). This dataset was retrieved from subjects that underwent routine care at the Shenzhen Hospital, China. This dataset has 662 samples, with 326 normal
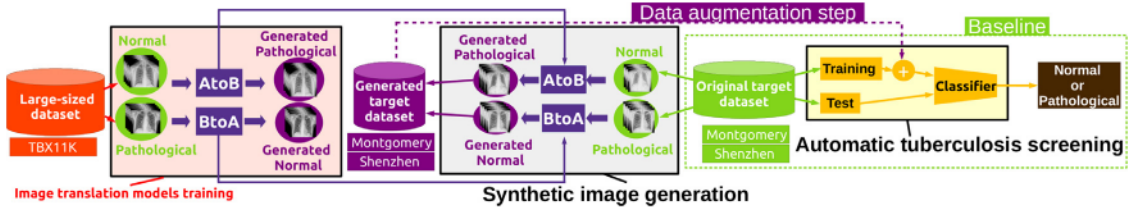
Fig. 1. Overall description of the methodology, showing the 2 alternatives: the baseline and the data augmentation approach. The dashed line represents the additional step necessary for the data augmentation approach.

cases and 336 with evidence of tuberculosis affectation, being representative of a medium-sized image dataset.

TBX11K dataset (available at [27]). TBX11K is a large-sized dataset of 11,200 samples (5,000 healthy cases, 5000 patients that present pathological affectation different from tuberculosis and 1200 that present different types of tuberculosis evidence). Due to the confusing notation of some samples, it being difficult to know their actual class, and in order to balance the dataset, only 1600 were randomly selected (with 800 healthy samples and 800 tuberculosis samples).

2.2. Baseline

In this alternative, the automatic tuberculosis screening is performed using only the original images of the small-sized and the medium-sized datasets (Montgomery and Shenzhen, respectively).

Network architecture and training details of the automatic tuberculosis screening. We adapted the architecture proposed by Pasa et al. [5] due to its demonstrated suitability for this problem to carry out this part of the methodology. This deep architecture is composed of 5 convolutional blocks that end up on a global average pooling layer, a fully connected layer and a softmax with 2 outputs. Moreover, all the images are resized to a resolution of 512×512 and the models are trained for 1750 epochs instead of the 150 epochs in the original work to ensure convergence. We use the Adam algorithm as the optimizer, with a learning rate of $\alpha = 8 \times 10^{-5}$, a value of $\beta_1 = 0.9$ and a value of $\beta_2 = 0.999$. As data augmentation, elastic deformations with a probability of 80% are performed, with the same parameters as in the reference work. To understand the behavior of the models, a 5-fold cross validation is performed with a mini-batch size of 4. To assess the effectiveness of the screening models, we considered the metrics of accuracy, F1-Score and AUC. To train the models, cross-entropy was used as the loss function, which is defined in Eq. (1):

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \cdot \log \hat{y}_i \qquad (1)$$

2.3. Data augmentation approach

For this part of the methodology, the synthetic image generation is performed adapting the Contrastive Unpaired Translation (CUT) paradigm [17] due to its suitability for image translation. Firstly, 2 different CUT models are trained with the large-sized dataset TBX11K to perform the translation between the classes Normal and Pathological and vice versa (therefore, each model performs the translation in only one direction). Once the translation models are trained, we perform the image generation using the Montgomery dataset and the Shenzhen dataset, respectively. After the novel sets of generated images are obtained, the automatic tuberculosis screening model is trained following the same details as stated in the baseline alternative but, in this case, the synthetic images are added to the original datasets to obtain the augmented datasets. In particular, these generated images are added to the

training set, while the validation and the test are performed with only original images.

Network architecture and training details of the image translation models. For the first part of the methodology, the CUT paradigm is adapted to our problem. In contrast with other image translation models, this paradigm only converts from one scenario to another, making it necessary to train 2 models for each dataset (i.e., one to convert from Normal to Pathological and another to do the opposite translation). Both models are trained during 200 epochs using the Adam algorithm [28] and all the input images are converted to grayscale and resized to a resolution of 512×512 instead of the original resolution of 256×256. Moreover, the mini-batch size is set to 1 and the learning rate is set to a constant value of $\alpha = 0.0002$.

The CUT model has a generator G and a discriminator D. G is an encoder-decoder that is based on a ResNet architecture with 6 residual blocks. It receives an original image a from class A as the input and returns its translated version $\hat{b}$ that presumably should belong to class B as the output. On its hand, D is a 70×70 Patch-GAN.

The CUT is trained using a loss composed of several elements. The first component is the adversarial loss that is computed using the output of the discriminator D, with the mean square error (MSE). Particularly, it is necessary for the discriminator to learn to classify the translated images as 1 and the original images as 0. Then, the adversarial loss can be expressed as stated in Eq. (2):

$$\mathcal{L}_{GAN}(G, D, A, B) = \mathbb{E}_{b \sim B} D(b)^2 + \mathbb{E}_{a \sim A} (D(G(a)) - 1)^2 \qquad (2)$$

The second component is the *PatchNCE* loss, a loss based on the idea of the contrastive learning [29], whose aim is to maximize the association between a patch from the original image, denoted as query, and its correspondent patch in the translated image, denoted as positive sample, which must be considerably greater than the association between that original patch and random patches different from the positive sample, which are denoted as negative samples. To do so, a number N is chosen to specify how many negative samples must be taken into account. To make the comparison between patches, some specific layers from the encoder of the generative model are selected, as after being trained for image translation, their weights can be used as features.

Then, the layers selected from the encoder are connected to a multilayer perceptron of 2 layers (denoted as H). The final result is a K-dimensional vector for each patch. To make the comparison among patches, the cosine similarity is computed on each pair of K-dimensional vectors. Therefore, to solve the problem, the similarity between the query patch and the positive patch is maximized while the similarity between the query patches and the random selected negative patches is minimized. This means that a (N+1)-classification problem is solved. With the cosine similarities results, the softmax function is used to normalize the outputs in the range [0, 1]. It is important to note that the distance among samples is scaled with a temperature value of $\tau = 0.07$. To compute the loss of a specific query patch v with its corresponding positive patch v^+ and the set of selected N random negative samples v^-,

the expression denoted in Eq. (3) is used:

$$l(v, v^+, v^-) =$$
$$-\log\left[\frac{\exp(v \cdot v^+/\tau)}{\exp(v \cdot v^+/\tau) + \sum_{n=1}^{N}\exp(v \cdot v_n^-/\tau)}\right] \quad (3)$$

This step produces a stack of features denoted as $\{z_l\}_L = \{H_l(G_{enc}^l(a))\}_L$ where l specifies the particular layer that is being referred and L refers to the total number of layers. Moreover, for each layer, there is a specific number of spatial locations S_l. Thus, the *PatchNCE* loss is defined as stated in Eq. (4):

$$\mathcal{L}_{PatchNCE}(G, H, A) = \mathbb{E}_{a \sim A} \sum_{l=1}^{L}\sum_{s=1}^{S_l} l(\hat{z}_l^s, z_l^s, z_l^{S\,s}) \quad (4)$$

Finally, to train the generative model, the adversarial loss and the PatchNCE loss are combined. Furthermore, the PatchNCE loss is also computed on images of class B $\mathcal{L}_{PatchNCE}(G, H, B)$ as a particular implementation of the identity loss that is used by several unpaired translation models as, for example, the CycleGAN. This prevents the appearance of unexpected changes made by the translation model. Moreover, 2 wt values are also added: λ_A and λ_B. These coefficients weight the contribution that both the PatchNCE loss and the identity loss make. Particularly, for the configuration used in this work, both values are 1.0. Joining all the components, the final expression to minimize is specified in Eq. (5):

$$\mathcal{L} = \mathcal{L}_{GAN}(G, D, A, B) + \lambda_A \mathcal{L}_{PatchNCE}(G, H, A) + \lambda_B \mathcal{L}_{PatchNCE}(G, H, B) \quad (5)$$

On the other hand, it is important to note that the discriminator model is only trained to minimize the adversarial loss previously stated in Eq. (2).

3. Results

To validate the proposal, we performed 2 different experiments for each target dataset (*i.e.*, Montgomery and Shenzhen), for a total of 4 experiments. The first experiment belongs to the baseline and, therefore, the automatic tuberculosis screening model is trained only with the original images. For the second experiment, the novel set of generated images by the CUT models (trained with the large-sized TBX11K dataset) is added to the training set to improve the performance of the screening process.

Regarding the training of the image translation models, Fig. 2 shows the losses when training the CUT models with the TBX11K dataset for both pathways, Normal to Pathological and Pathological to Normal, respectively. From that evolution, it can be concluded that the losses tend to achieve a convergence as the training progresses.

Moreover, the results of the baseline are shown in Table 1. There, it can be seen that the results are acceptable for both datasets. In the same way, it can be concluded that the results that were obtained for the Shenzhen dataset are slightly better in comparison with the Montgomery dataset, as expected. This is due to more severe data scarcity in the case of the Montgomery dataset, as this is the smallest one. In fact, it can be seen that this is noticeable not only in terms of effectiveness but also in terms

Table 1
Results obtained for the baseline with the 2 target datasets (Montgomery and Shenzhen).

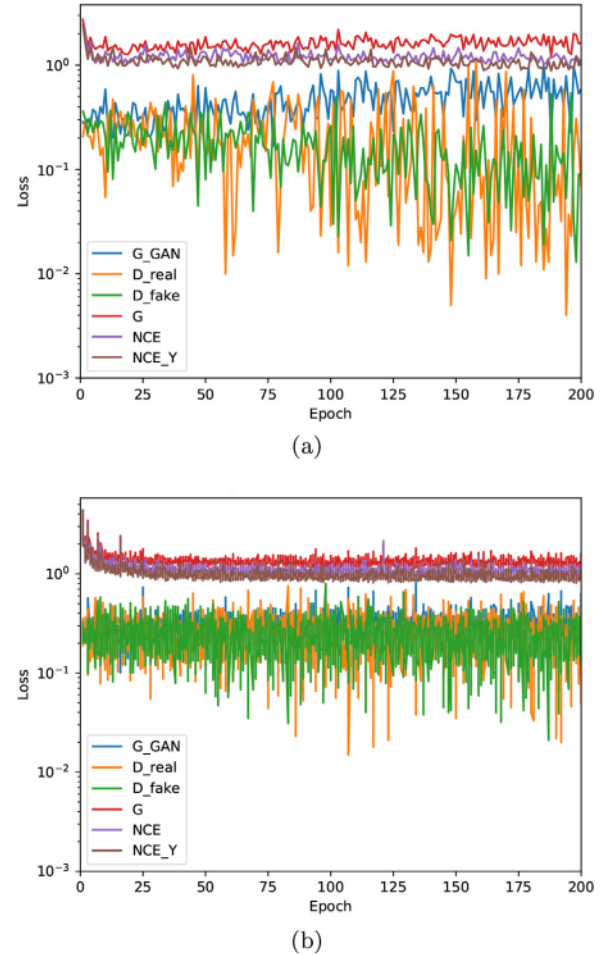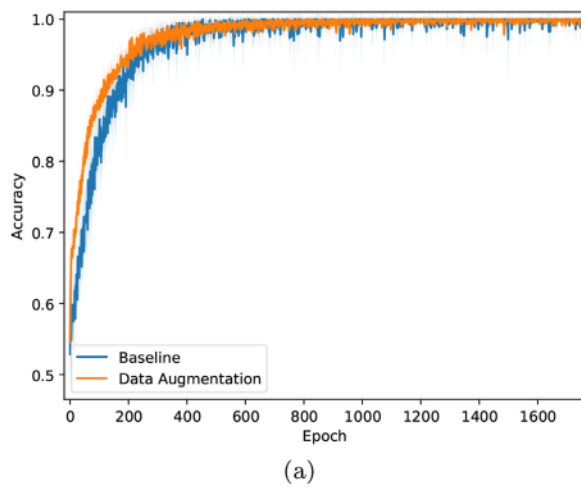|	Montgomery	Shenzhen
Accuracy	88.35% ± 6.28%	89.42% ± 2.30%
Recall	81.80% ± 12.04%	88.11% ± 3.42%
Specificity	91.53% ± 6.95%	90.60% ± 4.44%
Precision	86.35% ± 13.77%	90.90% ± 3.23%
F1-Score	82.89% ± 9.64%	89.41% ± 2.16%
AUC	0.8652 ± 0.0784	0.8846 ± 0.0379

(a)

(b)

Fig. 2. Training losses evolution for the translation models trained with the TBX11K dataset. (a) Model that translates from Normal to Pathological. (b) Model that translates from Pathological to Normal.

Table 2
Results obtained with the data augmentation using the target datasets (Montgomery and Shenzhen).

	Montgomery	Shenzhen
Accuracy	88.41% ± 5.27%	90.33% ± 1.41%
Recall	88.59% ± 9.88%	88.40% ± 2.51%
Specificity	89.38% ± 9.07%	92.21% ± 2.18%
Precision	86.24% ± 12.04%	92.28% ± 1.22%
F1-Score	86.32% ± 6.37%	90.27% ± 1.17%
AUC	0.8713 ± 0.0707	0.9088 ± 0.0229

of robustness. Particularly, in the case of the Montgomery dataset, the standard deviation values are always higher or equal to 6.28% and, in the case of the recall and the F1-Score, these values are even greater than 12.00%. On the other hand, the results of the Shenzhen dataset always have values of standard deviation that are equal or smaller than 4.44%.

In the case of the data augmentation approach, the results can be seen in Table 2. Once again, the results in the case of the Shenzhen dataset are more satisfactory than those obtained with the Montgomery dataset. In this case, it can also be seen in effectiveness as well as in robustness. In particular, the average values of the metrics are always lower than 90% in the case of the Montgomery dataset, while these are always higher than 90%, except in the case of the recall, where this value is an 88.40%.
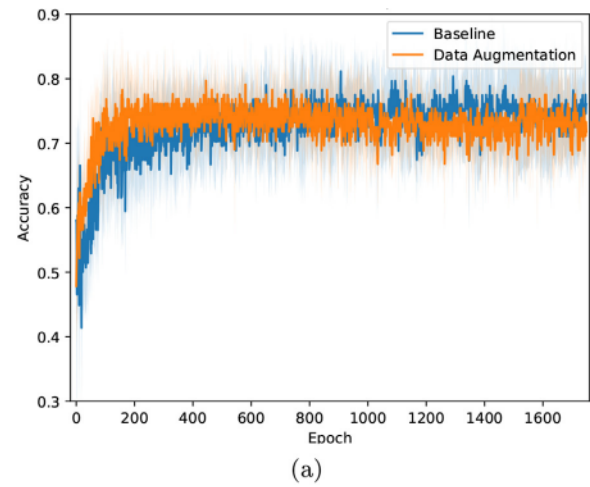
(a)
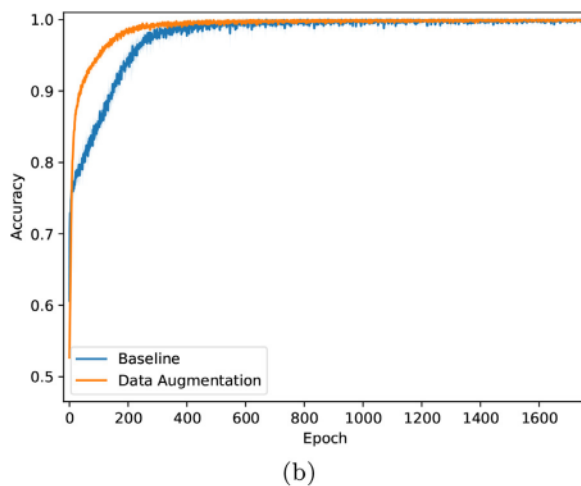
(b)

Fig. 3. Training accuracy evolution for the baseline and the data augmentation approach. The solid line shows the mean value of the folds, and the shadowed area shows the standard deviation. (a) Montgomery dataset. (b) Shenzhen dataset.

(a)

(b)

Fig. 4. Test accuracy evolution for the baseline and the data augmentation approach. The solid line shows the mean value of the folds, while the shadowed area shows the standard deviation. (a) Montgomery dataset. (b) Shenzhen dataset.

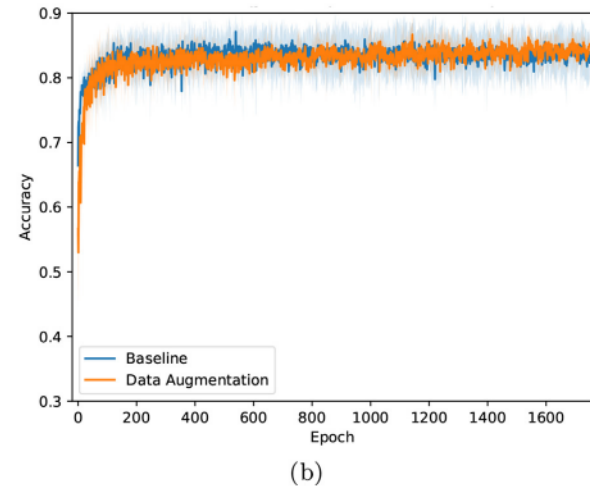### 3.1. Comparison between the baseline and the data augmentation approach

The training accuracy evolution comparison is depicted in Fig. 3. There, it can be seen that, for both cases and both datasets, the values achieve convergence as training progresses. The behavior is very similar in the case of the Montgomery dataset, even though the data augmentation approach tends to achieve higher values faster during the first 400 epochs, approximately. From that point onward, the behavior of both cases is very similar. In the case of the Shenzhen dataset, the discussion is alike, as data augmentation tends to achieve high values faster until reaching 600 epochs, approximately.

Regarding the test accuracy evolution comparison, the values are depicted in Fig. 4. In the case of the Montgomery dataset, the data augmentation approach has higher values of accuracy until 800 epochs, approximately, where the performance starts to experience a slight drop. In the case of the Shenzhen dataset, the performance is very similar between the data augmentation approach and the baseline, even though the first one is slightly higher from 800 epochs onward, approximately.

Putting the focus on the detailed results that are shown in Table 1 and 2, it can be seen that data augmentation implies an improvement of the baseline. In particular, for the Montgomery dataset, there is an improvement in accuracy from an average per-

formance of 88.35% to an 88.41%, with a standard deviation value that decreases from 6.28% to 5.27%. In terms of AUC, the average value increases from 0.8652 to 0.8713 and the standard deviation value drops from 0.0784 to 0.0707. Regarding recall, there is an important improvement in both the average value and in the standard deviation. In particular, recall improves from 81.80% to 88.59% in terms of the average value and from 12.04% to 9.88% in terms of standard deviation. In the same way, there is an important improvement for the F1-Score, which raises from 82.89% to 86.32% for the average value and drops from 9.64% to 6.37% for the standard deviation value. The precision value is very similar in both cases, while the specificity is worse, both in average and standard deviation terms. However, this drop is caused by the important improvement of the recall, which is representative of the unbalancing problem that the original dataset suffers from, but mitigated thanks to the proposed data augmentation approach.

On the other hand, in the case of the Shenzhen dataset, there is an improvement in terms of accuracy and AUC. In particular, the average value of accuracy improves from 89.42% to 90.33%, while the standard deviation improves from 2.30% to 1.41%. On the other hand, the AUC average value improves from 0.8846 to 0.9088 while the standard deviation drops from 0.0379 to 0.0229. The recall and the F1-Score remain very similar, with a slight improvement both for the average values and the standard deviation values. Finally, in terms of specificity and precision, there is also an improvement,

Table 3
Comparison between the accuracy of our data augmentation approach and the state-of-the-art methods.

Method	Montgomery	Shenzhen
Pasa et al. [5]	79.00%	84.40%
Alfadhli et al. [6]	79.10%	-
Lopes et al. [7]	82.60%	84.70%
Hwang et al. [8]	67.40%	83.70%
Jaeger et al. [10]	78.30%	84.10%
Sirshar et al. [14]	76.08%	76.73%
Ali et al. [11]	-	88.60%
Ours	**88.41%**	**90.33%**

with an average value that raises from 90.60% to 92.21% and a standard deviation that decreases from 4.44% to a 2.18% in the case of the specificity and an average value that improves from 90.90% to 92.28% alongside a standard deviation improvement from 3.23% to 1.22% in the case of precision.

Therefore, these results demonstrate that the data augmentation approach implies a performance improvement that leads the automatic screening system to have better effectiveness as well as better robustness, both demonstrated with the average values and with the standard deviation values, respectively. Similarly, the experiments that use the Montgomery dataset demonstrate that this methodology can help to mitigate the problem of unbalanced datasets, an aspect that is considerably accused in this particular case (80 normal cases against 58 pathological cases). It should be noted that, as data scarcity and the problem of dealing with unbalanced datasets are common issues in biomedical imaging, the extrapolation of this synthetic domain-related image generation approach to other scopes is extremely powerful.

3.2. Comparison with the state-of-the-art

The comparison between our method and the main state-of-the-art methods for the automatic tuberculosis screening can be seen in Table 3. In this comparative, we have selected a representative sample of the works that are most related to our contribution and that use at least one of the reference datasets also considered in this work (Montgomery County and Shenzhen datasets). Particularly, we take as reference the accuracy values that were obtained for the data augmentation approach. These results demonstrate that the method herein proposed beats the performance of the main state-of-the-art methods for automatic tuberculosis screening. A noticeable performance improvement can be seen when taking as reference the work of Lopes et al. [7] (the one with the best performance with the Montgomery County dataset from the state-of-the-art), as accuracy increases from 82.60% to 88.41% with the Montgomery dataset while it increases from 84.70% to 90.33% with the Shenzhen dataset. On the other hand, the work from Ali et al. [11], that achieves the best performance with the Shenzhen dataset from the state-of-the-art, is outperformed by almost 2% of accuracy as this value improves from 88.60% to 90.33%.

4. Conclusion

This work presents a fully automatic synthetic image generation approach to improve the tuberculosis screening on 2 different low dimensional chest X-ray images datasets. The methodology is divided into 2 different parts. Both of them were performed using different deep network architectures, where the first part conducts the synthetic image generation process, while the second part uses the novel set of generated images to improve the performance of the screening process. For the synthetic image generation, the data augmentation approach takes advantage of the

Contrastive Unpaired Translation (CUT) paradigm, a deep learning image translation model able to synthesize a novel set of relevant and useful chest X-ray images, representative of the variability of the domain, translating Normal samples to their Pathological version and vice versa. Particularly, in this methodology, we trained 2 CUT models to perform both image translation pathways using a large-sized dataset, as is the TBX11K, to mitigate the data scarcity effect. After that, we generated novel sets of synthetic images using the small-sized Montgomery and the medium-sized Shenzhen datasets, respectively. Finally, the novel sets of generated images are used to build the augmented datasets that, at the same time, are used to train the deep learning automatic screening model. To do so, the novel sets of synthetic images are only added to the training sets, while the validation and the test sets have only original images. Regarding the screening model, we adapted a previous proposal of a custom state-of-the-art deep network architecture specifically designed for this purpose.

To validate the proposal, we conducted 2 experiments for each target dataset (*i.e.*, Montgomery and Shenzhen). The first experiment trains the screening model with only original images, while the second experiment validates the result of adding the novel set of generated synthetic images. These results demonstrate that the data augmentation approach improves the performance of the automatic tuberculosis screening model for both datasets in terms of accuracy and can help to deal with the problem of training with considerably unbalanced datasets, as is the case of the Montgomery dataset. In the same way, the method herein proposed beats the performance of other main state-of-the art approaches for automatic tuberculosis screening using chest X-ray imaging by a wide margin. In particular, using the data augmentation approach, the system obtained $88.41\% \pm 5.27\%$ for accuracy using the Montgomery dataset and an accuracy of $90.33\% \pm 1.41\%$ in the case of the Shenzhen dataset. Overall, this contribution represents a powerful domain-related approach to mitigate the data scarcity problem by improving the performance of an automatic tuberculosis screening method with respect to other state-of-the-art proposals. Therefore, the proposal could be extrapolated to other biomedical imaging domains, as it is a common issue in these scopes.

Declaration of Competing Interest

None declared.

Acknowledgements

References

[1] D.N. McMurray, Disease model: pulmonary tuberculosis, Trends Mol Med 7 (3) (2001) 135–137, doi:10.1016/S1471-4914(00)01901-8.

[2] W.H. Organization, Global Tuberculosis Report 2020, 2020.

[3] S.L. Wellington, H.J. Vinegar, X-Ray computerized tomography, Journal of Petroleum Technology 39 (08) (1987) 885–898, doi:10.2118/16983-PA.

[4] R.W. Brown, Y.-C.N. Cheng, E.M. Haacke, M.R. Thompson, R. Venkatesan, Magnetic resonance imaging: physical principles and sequence design, John Wiley & Sons, 2014, doi:10.1002/9781118633953.

[5] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, D. Pfeiffer, Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization, Sci Rep 9 (1) (2019) 1–9, doi:10.1038/s41598-019-42557-4.

[6] F.H.O. Alfadhli, A.A. Mand, M.S. Sayeed, K.S. Sim, M. Al-Shabi, Classification of tuberculosis with surf spatial pyramid features, in: 2017 International Conference on Robotics, Automation and Sciences (ICORAS), IEEE, 2017, pp. 1–5, doi:10.1109/ICORAS.2017.8308044.

[7] U. Lopes, J. Valiati, Pre-trained convolutional neural networks as feature extractors for tuberculosis detection, Comput. Biol. Med. 89 (2017) 135–143, doi:10.1016/j.compbiomed.2017.08.001.

[8] S. Hwang, H.-E. Kim, J. Jeong, H.-J. Kim, A novel approach for tuberculosis screening based on deep convolutional neural networks, in: Medical Imaging 2016: Computer-aided Diagnosis, Volume 9785, International Society for Optics and Photonics, 2016, p. 97852W, doi:10.1117/12.2216198.

[9] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, volume 25, Curran Associates, Inc., 2012, doi:10.1145/3065386.

[10] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R.K. Singh, S. Antani, et al., Automatic tuberculosis screening using chest radiographs, IEEE Trans Med Imaging 33 (2) (2013) 233–245, doi:10.1109/TMI.2013.2284099.

[11] R. Ali, R. Hardie, B. Narayanan, T. Kebede, Imnets: deep learning using an incremental modular network synthesis approach for medical imaging applications, Applied Sciences 12 (2022) 5500, doi:10.3390/app12115500.

[12] L.T. Nguyen-Meidine, M. Kiran, M. Pedersoli, J. Dolz, L.-A. Blais-Morin, E. Granger, Incremental multi-target domain adaptation for object detection with efficient domain transfer, 2021. doi:10.48550/ARXIV.2104.06476.

[13] T. Hassan, B. Hassan, M.U. Akram, S. Hashmi, A.H. Taguri, N. Werghi, Incremental cross-domain adaptation for robust retinopathy screening via bayesian deep learning (2021). doi:10.48550/ARXIV.2110.09319.

[14] M. Sirshar, T. Hassan, M.U. Akram, S.A. Khan, An incremental learning approach to automatically recognize pulmonary diseases from the multi-vendor chest radiographs, Comput. Biol. Med. 134 (2021) 104435, doi:10.1016/j.compbiomed.2021.104435.

[15] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J Big Data 6 (1) (2019) 1–48, doi:10.1186/s40537-019-0197-0.

[16] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A.A. Bharath, Generative adversarial networks: an overview, IEEE Signal Process Mag 35 (1) (2018) 53–65, doi:10.1109/MSP.2017.2765202.

[17] T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image-to-image translation, in: European Conference on Computer Vision, Springer, 2020, pp. 319–345, doi:10.48550/arXiv.2007.15651.

[18] T. Malygina, E. Ericheva, I. Drokin, Gans' N lungs: improving pneumonia prediction, arXiv (2019), doi:10.48550/arXiv.1908.00433.

[19] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232, doi:10.48550/arXiv.1703.10593.

[20] D.I. Morís, J. de Moura, J. Novo, M. Ortega, Cycle generative adversarial network approaches to produce novel portable chest X-rays images for covid-19 diagnosis, in: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 1060–1064, doi:10.1109/ICASSP39728.2021.9414031.

[21] D.I. Morís, J. de Moura, J. Novo, M. Ortega, Data augmentation approaches using cycle-consistent adversarial networks for improving covid-19 screening in portable chest X-ray images, Expert Syst Appl 185 (2021) 115681, doi:10.1016/j.eswa.2021.115681.

[22] G. Bargshady, X. Zhou, P.D. Barua, R. Gururajan, Y. Li, U.R. Acharya, Application of CycleGAN and transfer learning techniques for automated detection of covid-19 using x-ray images, Pattern Recognit Lett 153 (2022) 67–74, doi:10.1016/j.patrec.2021.11.020.

[23] S. Jaeger, S. Candemir, S. Antani, Y.-X.J. Wáng, P.-X. Lu, G. Thoma, Two public chest X-ray datasets for computer-aided screening of pulmonary diseases, Quant Imaging Med Surg 4 (6) (2014) 475, doi:10.3978/j.issn.2223-4292.2014.11.20.

[24] Y. Liu, Y.-H. Wu, Y. Ban, H. Wang, M.-M. Cheng, Rethinking computer-aided tuberculosis diagnosis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2646–2655, doi:10.1109/CVPR42600.2020.00272.

[25] Montgomery County X-ray Set, (available at https://academictorrents.com/details/ac786f74878a5775c81d490b23842fd4736bfe33).

[26] Shenzhen Hospital X-ray Set, (available at https://academictorrents.com/details/462728e890bd37c05e9439c885df7afc36209cc8).

[27] TBX11K Dataset, (available at https://drive.google.com/file/d/1r-oNYTPiPCOUzSjChjCIYTdkjBTugqxR/view).

[28] I.K.M. Jais, A.R. Ismail, S.Q. Nisa, Adam optimization algorithm for wide and deep neural network, Knowledge Engineering and Data Science 2 (1) (2019) 41–46, doi:10.17977/um018v2i12019p41-46.

[29] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607, doi:10.48550/arXiv.2002.05709.