

CX-DaGAN: Domain Adaptation for Pneumonia Diagnosis on a Small Chest X-ray Dataset

Karen Sanchez, *Student Member, IEEE*, Carlos Hinojosa, *Member, IEEE*,
Henry Arguello, *Senior Member, IEEE*, Denis Kouamé, *Senior Member, IEEE*, Olivier Meyrignac,
and Adrian Basarab, *Senior Member, IEEE*

Abstract—Recent advances in deep learning led to several algorithms for the accurate diagnosis of pneumonia from chest X-rays. However, these models require large training medical datasets, which are sparse, isolated, and generally private. Furthermore, these models in medical imaging are known to over-fit to a particular data domain source, i.e., these algorithms do not conserve the same accuracy when tested on a dataset from another medical center, mainly due to image distribution discrepancies. In this work, a domain adaptation and classification technique is proposed to overcome the over-fit challenges on a small dataset. This method uses a private-small dataset (target domain), a public-large labeled dataset from another medical center (source domain), and consists of three steps. First, it performs a data selection of the source domain's most representative images based on similarity constraints through principal component analysis subspaces. Second, the selected samples from the source domain are fit to the target distribution through an image to image translation based on a cycle-generative adversarial network. Finally, the target train dataset and the adapted images from the source dataset are used within a convolutional neural network to explore different settings to adjust the layers and perform the classification of the target test dataset. It is shown that fine-tuning a few specific layers together with the selected-adapted images increases the sorting accuracy while reducing the trainable parameters. The proposed approach achieved a notable increase in the target dataset's overall classification accuracy, reaching up to 97.78% compared to 90.03% by standard transfer learning.

Index Terms—Chest X-ray, deep learning, domain adaptation, generative adversarial network, pneumonia diagnosis.

I. INTRODUCTION

Pneumonia is a lung infection that affects 7% of the world population and is the leading cause of death worldwide in children under five years, with about 1.4 million deaths every year [1], [2]. Furthermore, every year are diagnosed two billion

Karen Sanchez, Carlos Hinojosa, and Henry Arguello are with the Universidad Industrial de Santander, Bucaramanga, 680002 Colombia, e-mail: (karen.sanchez2, carlos.hinojosa@saber.uis.edu.co, henarf@uis.edu.co).

Denis Kouamé is with the Université Paul Sabatier Toulouse 3, France.

Olivier Meyrignac is with the Bicêtre University Hospital, Assistance publique des hôpitaux de Paris, Biomaps, France.

Adrian Basarab is with the Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, France.

people with pneumonia. This respiratory disease can be treated with antibiotics or antivirals, and the success of the treatment strongly depends on its early detection [3].

Expert analysis of chest X-ray images is currently the most widely medical imaging technique used world-wide to diagnose pneumonia and a wide variety of diseases. The popularity of X-rays is due to their relatively low cost, low irradiation, and the easy access to the acquisition equipment. Consequently, an estimated of 2 billion X-ray images are acquired each year worldwide [4].

Nevertheless, the image quality of chest X-ray has some limitations, such as low contrast, overlapping organs and blurred boundaries, which seriously affect pneumonia detection [5], [6]. Especially in children, accurately diagnosing pneumonia from chest X-rays remains a time-consuming task even for experienced radiologists [7]. Another important problem is the lack of radiologists trained in low-income and developing countries, even more in rural areas.

Due to advances in computing of recent decades, computer-aided diagnosis (CAD) has emerged to support medical staff in decision-making [8], [9]. In this sense, multiple computational strategies such as generic, probabilistic, population-based, and surface learning models have been developed to help health professionals in the analysis of medical images, whether for disease prediction, diagnosis or treatment. Specifically, these strategies include Gaussian mixture models, conditional random fields, statistical atlases, logistic regression, nearest-neighbor methods, support vector machines, random forests, among others [10]–[13].

More recently, the use of deep learning (DL) models has revolutionized the field of medical imaging computing [14]–[16]. Due to its ability to process large amounts of data with high speed, and to extract complex characteristics not visible to the human eye, this area of research became the most appealing for the analysis of medical images in recent years [17]. Deep learning approaches include recurrent neural networks, convolutional neural networks (CNN), auto-encoders, and deep reinforcement learning techniques, among others [18]–[20]. Precisely, recent advances in the application of deep learning and computer vision in healthcare have allowed to attain radiologist-level performance for pneumonia diagnosis from chest X-ray images. For instance, authors in [21]–[23] achieved an overall accuracy of 98%, 98.43% and 99.41%, respectively, in the classification of a public pneumonia dataset

with 5,849 labeled images acquired in the Guangzhou Women and Children's Medical Center, China [24]. Despite these accurate results, deep learning classifiers are well-known to over-fit to a particular data domain source in medical imaging. That is, a deep learning model trained on a large dataset originating from one medical center does not conserve the same precision when tested on a dataset from another medical center. This challenge is mainly related to discrepancies in the appearance of the images, to differences in acquisition protocols and/or device technologies. Specifically, as we will show in this work, a model trained on the previously mentioned dataset with an overall accuracy of 98% only achieves 88% accuracy when classifying 173 non-public chest X-ray images acquired at a different clinical center. This represents a major challenge limiting the clinical applicability of such technologies.

In this context, domain adaptation (DA) has emerged as a transfer learning alternative to address the lack of massive amounts of labeled data and the difficulty of deep learning methods to obtain high performance when applied to small datasets different from the one used during the training phase [25], [26]. Precisely, DA is the area of machine learning that enables knowledge to be transferred from one source domain to a different but related target domain to increase the learning models' capability on the latter [27]. DA applications mainly focus on natural images [28], telecommunication problems [29], and, to a lesser extent, on medical applications [30].

It should be noted that other approaches related to the relatively new DA concept already exist. In particular, one may cite two large categories according to the modified domain: domain transformation (DT-DA) which translates images from one domain to another such that a specific task can be applied directly to the transformed images; and latent feature-space transformation (LFST-DA) which aligns the images from both domains in a hidden common feature space in which the task model is to be trained [31]. Both perspectives have recently been addressed separately in the literature to analyze lung X-rays using images from different demographic sources or disease states. In [26] and [32], the authors proposed DT-DA strategies for semi-supervised and unsupervised classification, respectively. Other DT-DA works for the study of lungs have been presented in the literature, based on feature transfer from one imaging modality to another (cross-modality). However, these works were mainly focused on image segmentation. For instance, in [33] and [34], DA was performed from computed tomography (CT) to X-rays. In [35], the authors used histopathological images to transfer characteristics from cytokeratins to PD-L1. Furthermore, a few unsupervised lung X-ray imaging works based on the LFST-DA approach have been studied, also focused on image segmentation [36], [37].

Considering the challenge discussed above on the scarcity of labeled medical data and the difficulty of deep learning models to generalize from one dataset to another, the main scientific question addressed here is how to improve the classification of a small dataset acquired in one center, taking advantage of information from a larger public dataset acquired in a different hospital?

This paper proposes a new DA technique to classify a small chest X-ray dataset taking advantage of a large pub-

TABLE I
SUMMARY OF MAIN NOTATIONS

	Variable notations
S	Source dataset (large)
T	Target dataset (small)
S_P, S_N	Pneumonia, normal classes data in S
T_P, T_N	Pneumonia, normal classes data in T
D	Number of image pixels in one image
\mathbf{y}	An image sample in $D \times 1$ vector form
d	Number of eigenvalues from T
U_P, U_N	Subspace of $T \in \mathbb{R}^{D \times d}$ for pneumonia and normal classes.
I_d	Identity matrix
E_P, E_N	The similarity score between an image and the pneumonia or normal class subspace.
n_1, n_2	Number of images from S_P, S_N

lic dataset acquired in another clinical center. The resulting algorithm is referred to as CX-DaGAN meaning Chest X-rays Domain adaptation with Generative Adversarial Network. The proposed technique is composed of three stages. First, a data selection from a source dataset through similarity constraints with the target dataset is performed. Then, a translation of the selected source images to the target domain with a generative adversarial network approach is processed. Finally, training of a CNN using both sets, target and translated sets, is performed in order to classify the target set. For testing, a sub-set from the target set is used. This work is a substantial improvement from the previous research published in [38].

II. METHODOLOGY

Overall, the proposed approach consists of three stages: first, it selects from the source dataset the images which are the most similar to the ones of the target domain; these images are chosen based on a similarity function that measures the subspace-projection error obtained by projecting the source data onto the target subspaces of each class: pathological and normal images. Second, it uses the selected source images as input for a Cycle-GAN to generate images in the target domain. This second step uses images from the train target set to discriminate between the real and generated images generated by the GAN. Finally, in the third stage, the translated images and the small target train set are used to feed a CNN pneumonia/normal classification network, and test it on the test set from the target dataset. The network parameters are further reduced by proposing a fine-tuning strategy. The overall proposed method is depicted in Fig. 1. The following sections provide more details of each stage (A, B, and C). Note that stages A and B are preprocessing steps before training a CNN network for pneumonia/normal classification of the X-ray images in step C.

Notation. Table I summarizes the main notations used in this paper.

A. Proposed Similarity-constrained Data Selection

This section introduces the step A of the proposed method shown in Fig. 1 and detailed in Fig. 2. Let us denote by S the set of medium to large labeled dataset (source), and by T the small labeled dataset (target). Throughout the paper, we refer

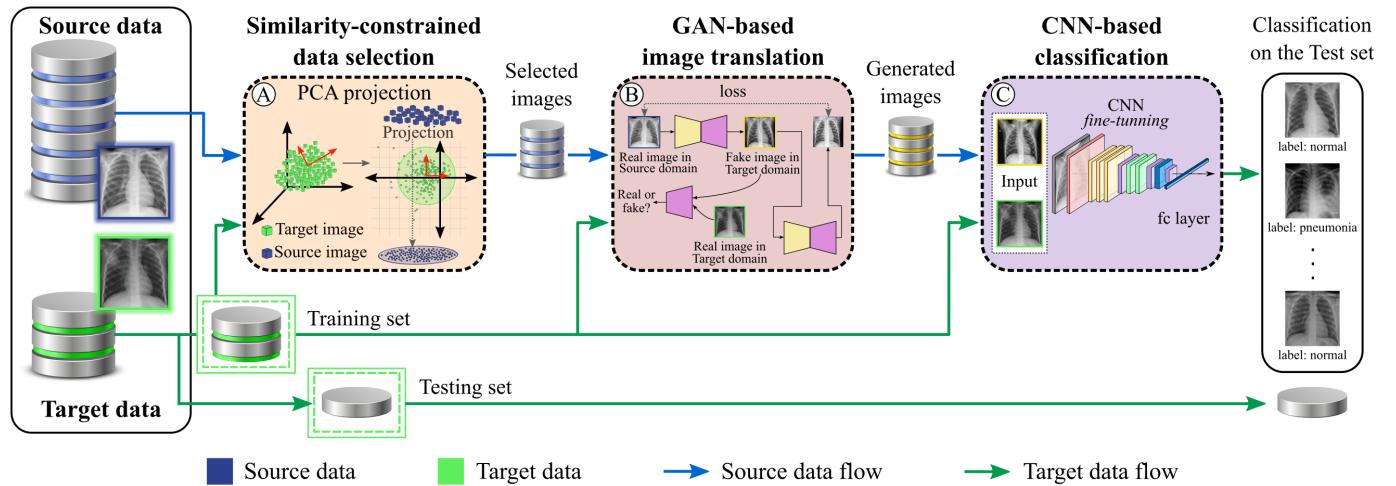


Fig. 1. Proposed domain adaptation and classification framework. First, the chest X-rays images from the source and training target domains are fed into step (A) for the similarity-constrained data selection process. In step (A), images from the source domain are selected using a similarity function that measures the subspace-projection error obtained by projecting the source data onto the training target data domain. Then, the selected source images and the train target set are used as input to the proposed GAN-based image-to-image translation (Step B). The output of step (B) consists of synthetic images generated from the GAN that follows the target image distribution. Finally, we fine-tune a pretrained CNN-based classification network for pneumonia diagnosis using the generated images in step (B) and the training target set as input. The performance of the proposed workflow is evaluated on the testing target set.

to “small” a dataset with less than 1,000 images, “medium” a dataset containing between 1,000 and 5,000 images, and “large” a dataset with more than 5,000 images. This choice is based on the number of images needed to train a X-ray-based pneumonia classifier from scratch with excellent (98%), good (95%) and insufficient (90%) accuracy [39]. Furthermore, let us denote by $S_P \subseteq S$, and $S_N \subseteq S$ the subsets of images labeled as pneumonia and normal, respectively, in the source domain. Similarly, $T_P \subseteq T$ and $T_N \subseteq T$ denote the subsets of images labeled as pneumonia and normal in the target domain, respectively. $\mathbf{y}_{S_P} \in \mathbb{R}^{D \times 1}$ represents an image sample from the S_P subset after reshaping it in a D -dimensional vector form, i.e., D corresponds to the total amount of pixels in the image. Similarly, $\mathbf{y}_{S_N} \in S_N$, $\mathbf{y}_{T_P} \in T_P$, and $\mathbf{y}_{T_N} \in T_N$ denote vectorized images from the corresponding subsets. First, every source and target data from T_P and T_N is normalized to have zero mean and unit standard deviation. Then, principal component analysis (PCA) is applied to select, for each domain, d eigenvectors corresponding to the d largest eigenvalues. These eigenvectors are used as bases of the subspace for each subset. Specifically, the matrices $U_P \in \mathbb{R}^{D \times d}$ and $U_N \in \mathbb{R}^{D \times d}$ are obtained, used as the subspaces. Note that U_P and U_N are semi-orthonormal, thus $U'_P U_P = I_d$ and $U'_N U_N = I_d$, where I_d is the identity matrix of size d^2 and $'$ denote the transpose of the matrix. Furthermore, two types of projections are performed: (1) project every image from each source class onto the target subspace of the same class, i.e., \mathbf{y}_{S_P} is projected onto U_P and \mathbf{y}_{S_N} onto U_N ; (2) project every image from each source class onto the target subspace of the opposite class, i.e., \mathbf{y}_{S_P} is projected onto U_N and \mathbf{y}_{S_N} onto U_P . Based on these projections, the following similarity functions considering the projection errors are defined:

$$E_P(\mathbf{y}) = \|U_P U'_P \mathbf{y} - \mathbf{y}\|_2, \quad (1)$$

$$E_N(\mathbf{y}) = \|U_N U'_N \mathbf{y} - \mathbf{y}\|_2, \quad (2)$$

where $\|\cdot\|_2$ stands for the ℓ_2 -norm. (1) and (2) are used to project all images in S_P and S_N and build four error vectors: $\mathbf{q}_1 \in \mathbb{R}^{n_1}$, $\mathbf{q}_2 \in \mathbb{R}^{n_2}$, $\mathbf{q}_3 \in \mathbb{R}^{n_1}$ and $\mathbf{q}_4 \in \mathbb{R}^{n_2}$. Specifically, the vector \mathbf{q}_1 is built as $\mathbf{q}_1 = \{E_P(\mathbf{y}_{S_P}^1), \dots, E_P(\mathbf{y}_{S_P}^{n_1})\}$ using all images (n_1) from S_P . Similarly, $\mathbf{q}_2 = \{E_N(\mathbf{y}_{S_N}^1), \dots, E_N(\mathbf{y}_{S_N}^{n_2})\}$ is formed using all images (n_2) from S_N ; $\mathbf{q}_3 = \{E_N(\mathbf{y}_{S_P}^1), \dots, E_N(\mathbf{y}_{S_P}^{n_1})\}$; and $\mathbf{q}_4 = \{E_P(\mathbf{y}_{S_N}^1), \dots, E_P(\mathbf{y}_{S_N}^{n_2})\}$.

The vectors \mathbf{q}_1 and \mathbf{q}_2 are sorted in ascending order, and vectors \mathbf{q}_3 and \mathbf{q}_4 in descending order. Finally, considering the first k values from each error vector, the corresponding k images from the source domain (S_N and S_P) are selected and used as input for the proposed cycle-GAN-based network shown in step (B) of Fig. 1. Note that in this document, “similarity” is named for the minimum mathematical difference between pixel values of two spatial sets. A human inspection of the images was not considered to establish the similarity. This way of selecting images from the source set is guided by the idea of choosing the images that are the most similar intra-class to those from the target set, and the most different inter-class between the two domains.

B. Proposed GAN-based Image-to-Image Translation

After selecting the most similar images from the source dataset (with respect to the target domain images) using the proposed subspace-based approach, a multi-domain and unpaired image-to-image (I2I) translation network is used to generate images following the target domain distribution. Specifically, Step B generates the same number of images that were selected from the source dataset by Step A. The proposed network is depicted in Fig. 1 (step B) and detailed in Fig. 3. Specifically, the I2I translation strategy [40] is adopted to map images from the two domains corresponding to the same class ($\mathbf{y}_{S_P} \rightleftharpoons \mathbf{y}_{T_P}$, $\mathbf{y}_{S_N} \rightleftharpoons \mathbf{y}_{T_N}$).

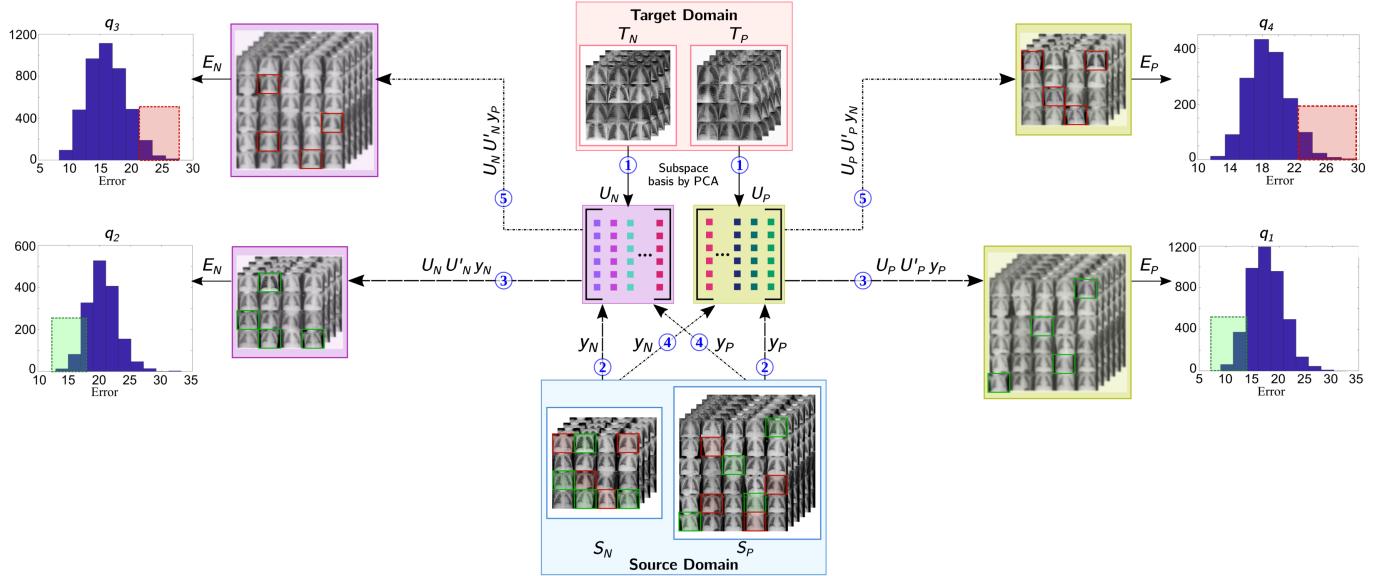


Fig. 2. Step A of Fig 1. Similarity-constrained data selection via subspace projection error. The circled blue numbers represent the order of the stages in the figure. First, we calculate a subspace basis for each target class using principal component analysis (PCA). Second, the images from each class in the source domain are projected onto the subspaces of T within their corresponding classes. Third, the results of these projections are obtained. Fourth, each source image is projected onto the opposite class subspace basis, and the results are obtained in the fifth lines. Sixth, for each case, we calculate the projection error between the projected and original corresponding image through the similarity functions shown in Eq. (1)-(2). Finally, we select the images with the lowest projection error when projecting the images onto subspaces of the same classes and the largest error when projecting the images onto subspaces of the opposite classes.

In this work, the Cycle-GAN [40] is adopted to learn two mappings: $S \rightarrow T$, and $T \rightarrow S$, with generators $G_{S \rightarrow T}(\mathbf{y}_S)$ and $G_{T \rightarrow S}(\mathbf{y}_T)$, so that discriminators D_T and D_S cannot distinguish between real and synthetic images generated by the generators. In a Cycle-GAN network, $G_{S \rightarrow T}$ and its discriminator D_T are used to define the adversarial learning objective loss as

$$\mathcal{L}_{\text{adv}}(G_{S \rightarrow T}, D_T) = \mathbb{E}_{y_s \sim \mathbf{y}_S} [\log(1 - D_T(G_{S \rightarrow T}(y_s))] + \mathbb{E}_{y_t \sim \mathbf{y}_T} [\log D_T(y_t)], \quad (3)$$

where \mathbb{E} denotes the expected value over the data instances specified in the subindex.

A similar adversarial loss can be designed for mapping $G_{T \rightarrow S}$ and its discriminator D_S as well, i.e., $\min_{G_{T \rightarrow S}} \max_{D_S} \mathcal{L}_{\text{adv}}(G_{T \rightarrow S}, D_S)$. To preserve sufficient low-level content information, we use the *cycle-consistency* loss [40] to force the reconstructed synthetic images y'_s and y'_t to resemble their inputs y_s and y_t :

$$\mathcal{L}_{\text{cyc}}(G_{S \rightarrow T}, G_{T \rightarrow S}) = \mathbb{E}_{y_s \sim \mathbf{y}_S} [\|y'_s - y_s\|_1] + \mathbb{E}_{y_t \sim \mathbf{y}_T} [\|y'_t - y_t\|_1], \quad (4)$$

where $y'_s = G_{T \rightarrow S}(G_{S \rightarrow T}(y_s))$, $y'_t = G_{S \rightarrow T}(G_{T \rightarrow S}(y_t))$, and $\|\cdot\|_1$ is the ℓ_1 -norm. The generative adversarial training with cycle-consistency enables synthesizing realistic-looking radiographs across domains. However, there is no guarantee that high-level semantics would be preserved during translation, thus decreasing the classification accuracy.

To improve the classification accuracy on the generated synthetic target images, a classification model \mathcal{F} was included in the GAN-based network to guide the training by considering

the classification loss. Specifically, the classification model \mathcal{F} is learned on the synthetic target data $\bar{T} = \{G_{S \rightarrow T}(\mathbf{y}_S), \bar{L}_S\}$, where \bar{L}_S represent the corresponding labels (Normal or Pneumonia) of the synthetic \bar{T} data. The binary cross-entropy loss was used to classify the two categories:

$$\mathcal{L}_{\text{cls}}(\mathcal{F}, \bar{T}) = -\mathbb{E}_{\bar{t} \sim \bar{T}} \sum_{c=1}^C \mathbb{1}_c \log (\sigma(\mathcal{F}^{(c)}(\bar{t}))), \quad (5)$$

where σ is the softmax function, $\mathbb{1}_c = 1$ if an input image \bar{t} belongs to class $c \in C = \{\text{Normal, Pneumonia}\}$, otherwise $\mathbb{1}_c = 0$. The final objective of our proposed GAN-based network for synthetic target images generation is the sum of adversarial learning losses, cycle consistency loss, and classification loss:

$$\mathcal{L} = \mathcal{L}_{\text{adv}}(G_{S \rightarrow T}, D_T) + \mathcal{L}_{\text{adv}}(G_{T \rightarrow S}, D_S) + \lambda \mathcal{L}_{\text{cyc}}(G_{S \rightarrow T}, G_{T \rightarrow S}) + \mathcal{L}_{\text{cls}}(\mathcal{F}, \bar{T}). \quad (6)$$

It is worth mentioning that, for ease of notation, the above equations were developed without distinguishing between the two classes. However, during implementation, four generators described the mappings from source/target images with pneumonia/normal to target/source images, respectively. Similarly, four discriminators were associated with each generator output. Also, the same classification network (\mathcal{F}) was used in the last step of the proposed framework shown in Fig. 1, which is described in the following section.

C. CNN-based Classification

The augmented training dataset obtained following the steps A and B detailed in the previous sections is used to feed a convolutional neural network (CNN) trained to perform the final

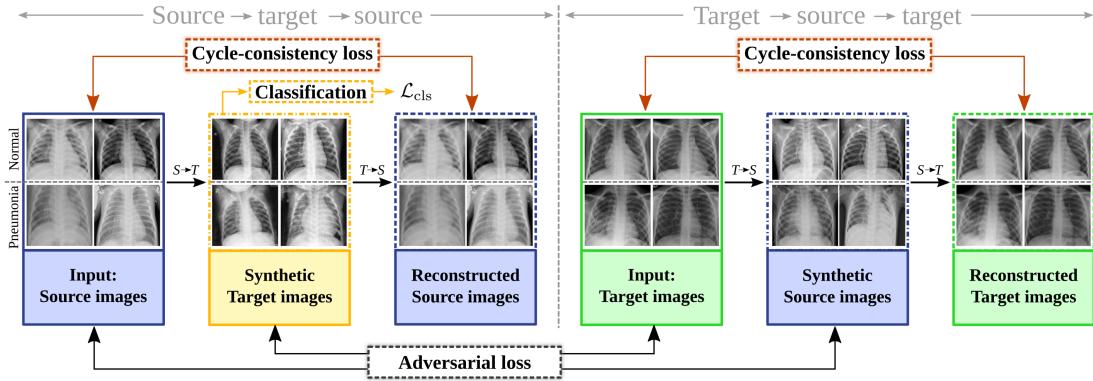


Fig. 3. Step B of Fig 1. Proposed GAN-based I2I architecture used to translate chest X-ray images from the source domain (Normal or Pneumonia) to its corresponding class in the target domain. The network also translates back the generated images to the source domain to maintain the cycle consistency. To ensure the generated synthetic images maintain the high-level semantics after the transformation and improve classification accuracy, a classification model was incorporated to guide the training by considering the classification loss. The training set of target domain was used to measure the adversarial loss during training.

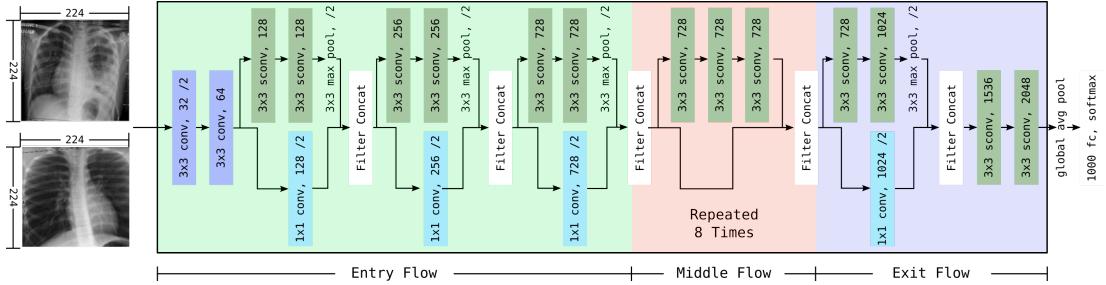


Fig. 4. Step C of Fig 1. It starts by using the pretrained weights of the Xception architecture on ImageNet and investigate different fine-tuning settings to achieve the highest accuracy while training fewer parameters. It uses the generated images obtained by following steps (A) and (B) of the proposed workflow as input to this architecture.

classification. In this work, the Xception CNN was adopted as backbone to extract features and used a fully connected layer at the end of the network to perform the classification. The Xception [41] is an extension of the Inception architecture which replaces the standard Inception modules with depthwise separable convolutions. Instead of partitioning input data into several compressed chunks, it maps the spatial correlations for each output channel separately, and then performs a 1×1 depthwise convolution to capture cross-channel correlation. This is essentially equivalent to an existing operation known as a “depthwise separable convolution”, which consists of a depthwise convolution (a spatial convolution performed independently for each channel) followed by a pointwise convolution (a 1×1 convolution across channels). The Xception architecture is shown in Fig. 4. In general, the network can be divided in three sections: the entry, middle and exit flow, where the middle flow is repeated eight times. Given the limited size of the training dataset, pre-trained weights from ImageNet dataset were used to initialize the network and fine-tuned the layers of the Xception to adapt them to the specific task of pneumonia detection. Section III-C investigates different fine-tuning settings to achieve high accuracy while training less parameters.

III. RESULTS

This section illustrates the efficiency of the proposed CX-DaGAN classification algorithm for normal and pneumonia

images on a small chest X-rays dataset. All simulations were implemented in Python with Tensorflow 2.3 and ran on an Nvidia Quadro RTX 6000 GPU with 24 GB of memory.

A. Datasets and Metrics

Datasets. The proposed CX-DaGAN algorithm was tested using two datasets for domain adaptation: a large source dataset (S) from which we extracted and transformed a selected number of images; second, a small target dataset (T) from which we performed the classification. In this work, a private dataset was used as T and a publicly available dataset as S .

Specifically, the “Chest X-ray Images (Pneumonia) dataset”¹ was used as S which consists of 5,849 labeled images acquired in the Guangzhou Women and Children’s Medical Center in China [24]. The 8-bit X-ray grayscale images are separated into 4,266 pneumonia (S_P) and 1,583 normal (S_N).

On the other hand, 573 chest X-ray images acquired at the Toulouse University Hospital in France were used as target dataset T . Each image of T was labeled as pneumonia or normal class by two expert radiologists. The dataset T is divided into 275 normal and 298 pneumonia images. In the following experiments, we split T in 400 images for

¹The dataset is available for free download at <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

training and the remaining 173 for testing, which corresponds to 69,8% and 30,2% of the data, respectively. To train and evaluate the proposed CX-DaGAN method, images with a fixed size of 224×224 pixels were considered.

Metrics. To quantitatively evaluate the performance of the proposed method, three metrics [42] were computed: Accuracy (ACC), F1 score (F1), and Area under the ROC curve (AUC).

B. Quantitative Classification Results

For the testing chest X-ray images, the CX-DaGAN algorithm was used to predict the probability of pneumonia. By comparing with the binary ground-truth labels, the overall accuracy of the proposed method was calculated in extensive simulations, as shown in Table III. In this table, each value corresponds to the average and standard deviation of 10 realizations of the proposed method, evaluated on 173 images from the target dataset. In this Table III, the number of images used for training throughout the entire framework was varied. Specifically, between columns and rows, the number of images from target and source, respectively, changes. Note that due to the small-size of the target dataset, the maximum number of target images used for training was 400 (see columns in Table III). On the other hand, the number of images from the source domain chosen by the similarity-constrained data selection step was simultaneously varied. It is worth noting that this method was designed for selecting the same amount of images from each class in S , i.e., after the similarity phase (A), the proposed method ensures 50% of normal (disease-free) and 50% of pneumonia images to feed the step B. Thus, considering the composition of the unbalanced public dataset used in these experiments, the maximum number of X-rays from S was 2400, which corresponds to 1200 images of each class (see rows in Table III).

In deep learning, specially for CNNs, it is well known that a greater number of labeled training samples leads to better classification [43], [39]. In general terms, Table III shows that the classification accuracy is lower using fewer target images for training the CX-DaGAN algorithm. Conversely, the best results are concentrated in the right area of the Table, where we used more images from the target dataset. However, it can also be observed vertically that using more images from source domain does not necessarily imply better accuracy. Instead, there is a central area with combinations of data that are particularly interesting.

In the highest case of average precision, the data showed that the pneumonia prediction accuracy obtained by the CX-DaGAN proposed algorithm is higher than 97% when training using 250 images from the target and 400 images from the source set. Similarly, high accuracy results can be achieved by training the algorithm with 400 images from the target and 200 images from the source. For a deeper analysis of results in Table III, Fig. 5 presents the 20 highest average classification results, without considering their standard deviation (STD), organized in descending order and graphed with their respective STD. Note that the best classification average value is 97,78. However, its corresponding STD is 0,7. On the other hand, the third highest average rating in this plot is 96,75. In this

case, the STD is lower (0,4), which can be interpreted as a statistically more stable and reliable result. Therefore, it is not possible to select only one combination of target/source data as the “best” since results can vary between the ranges defined by the STD. Please note that this table aims to provide results with different source/target data combinations to allow the user select the best combination according to the data availability. Figure 6 shows a histogram with all the error values of Table III to visualize their behavior. Note that the mean of the error values is 0,732 and the standard deviation of the error values is 0,231.

The ten highest average accuracy results from Table III, their STD, and the corresponding target/source data combination for training are shown in Fig. 7. Note that the “x” and “y” axis correspond to the number of “target” and “source” images used for training, respectively. The size of each dot is associated with the accuracy value: larger dots indicate better precision, as conventions dictate; the color of the dot indicates the value of the standard deviation, where red corresponds to a higher STD and green to a lower one. The aim of this plot is to expose the most accurate options achieved by the proposed method, according to the required number of images for training, and their STD.

The results shown in Tables III, IV, and V were obtained using $d = 200$ in step A. Conversely, OA results of CX-DaGAN method using other values of $d = 60, 100, 140, 180$ for the step A are presented in Table II. Note that subspaces U_P and U_N have dimensions $\mathbb{R}^{D \times d}$ and consequently the upper bound of d in this example is 298 for U_P and 275 for U_N , corresponding to T_P and T_N , respectively. One may remark from the OA values in Table II that the choice of d has limited influence on the accuracy of CX-DaGAN method. This observation is sustained by the fact that 88% of the images selected from S within step A were the same for all d values evaluated. Accordingly, $d = 200$ eigenvectors ($\approx 70\%$) were chosen, to run the following simulations given that this value leads to the best balance between classification accuracy and algorithm performance.

TABLE II
CLASSIFICATION RESULTS IN TERMS OF OVERALL ACCURACY (OA) OF THE CX-DAGAN ALGORITHM FOR DIFFERENT VALUES OF d IN STEP A (AVERAGE OVER 10 REALIZATIONS).

d	60 (20,9%)	100 (34,9%)	140 (48,8%)	180 (62,8%)	200 (69,8%)
OA	$95,0 \pm 0,3$	$96,2 \pm 0,3$	$95,5 \pm 0,4$	$96,8 \pm 0,3$	$97,6 \pm 0,4$

C. Ablation Studies

Two ablation experiments were conducted to evaluate the configuration of the proposed training pipeline. The first ablation study (Ablation study 1) evaluates the influence of the fine-tune training in the step C of the proposed approach. The second ablation study (Ablation study 2) validates the importance of each step of the CX-DaGAN algorithm in the training procedure, evaluating the separate use of one or two of the three steps. Furthermore, in the Ablation study 2, the

TABLE III

QUANTITATIVE CLASSIFICATION RESULTS FOR DIFFERENT DATA COMBINATIONS (TARGET/SOURCE) IN THE TRAINING (AVERAGE OVER 10 REALIZATIONS).

Images ↓ Source	Target → %	(↑) Overall accuracy (%)								
		0 0%	50 12,5%	100 25%	150 37,5%	200 50%	250 62,5%	300 75%	350 87,5%	
0	0%	51,33 ± 1,2	74,92 ± 1,1	84,51 ± 1,2	89,75 ± 1,0	90,56 ± 0,8	92,04 ± 0,9	92,81 ± 0,8	94,69 ± 0,8	91,09 ± 0,9
50	0,9%	52,33 ± 1,0	80,39 ± 1,1	81,32 ± 0,8	88,48 ± 0,9	90,92 ± 0,9	91,08 ± 0,6	94,09 ± 0,6	94,75 ± 0,9	95,75 ± 0,7
100	1,7%	52,23 ± 1,1	78,85 ± 0,8	90,15 ± 1,2	92,81 ± 1,0	90,81 ± 0,8	95,21 ± 1,1	94,12 ± 0,4	94,90 ± 0,8	94,90 ± 0,6
150	2,6%	58,64 ± 0,9	83,15 ± 1,0	84,84 ± 0,8	86,71 ± 0,8	89,90 ± 1,0	93,89 ± 0,8	93,81 ± 0,9	95,58 ± 0,7	95,56 ± 0,9
200	3,4%	59,34 ± 1,1	80,42 ± 1,0	89,27 ± 0,9	91,01 ± 0,9	89,38 ± 0,8	94,68 ± 0,9	94,85 ± 0,7	95,62 ± 0,6	97,02 ± 0,8
250	4,3%	59,77 ± 0,9	86,23 ± 0,7	83,07 ± 0,8	84,34 ± 1,2	89,38 ± 1,0	91,82 ± 1,2	94,95 ± 0,6	95,75 ± 0,6	95,66 ± 0,7
300	5,1%	62,44 ± 1,0	81,19 ± 0,8	84,84 ± 0,7	87,38 ± 0,6	90,06 ± 0,7	93,41 ± 0,4	94,35 ± 0,6	94,89 ± 0,5	95,46 ± 0,5
350	6,0%	61,88 ± 1,0	80,61 ± 0,9	84,84 ± 0,6	89,37 ± 0,7	91,04 ± 0,6	91,06 ± 0,7	94,51 ± 0,4	95,01 ± 0,5	96,70 ± 0,6
400	6,8%	63,11 ± 1,2	80,51 ± 0,9	83,96 ± 1,0	86,91 ± 0,8	92,61 ± 0,5	97,78 ± 0,7	95,02 ± 0,4	96,75 ± 0,4	95,85 ± 0,3
600	10,3%	66,37 ± 0,9	76,11 ± 1,0	83,19 ± 0,6	85,84 ± 0,7	87,61 ± 0,6	84,07 ± 0,6	87,61 ± 0,5	90,27 ± 0,4	92,04 ± 0,4
1000	17,1%	67,26 ± 0,9	79,65 ± 0,6	87,61 ± 0,8	85,84 ± 0,5	87,61 ± 0,4	88,5 ± 0,7	90,27 ± 0,5	86,73 ± 0,5	93,81 ± 0,5
1600	27,4%	68,14 ± 0,8	75,22 ± 0,6	76,99 ± 0,7	80,53 ± 0,5	87,61 ± 0,4	84,96 ± 0,4	86,73 ± 0,4	92,04 ± 0,4	90,27 ± 0,4
2400	41,0%	70,80 ± 0,7	82,3 ± 0,6	83,19 ± 0,5	80,53 ± 0,6	80,53 ± 0,7	89,38 ± 0,5	86,73 ± 0,5	88,5 ± 0,4	91,15 ± 0,3

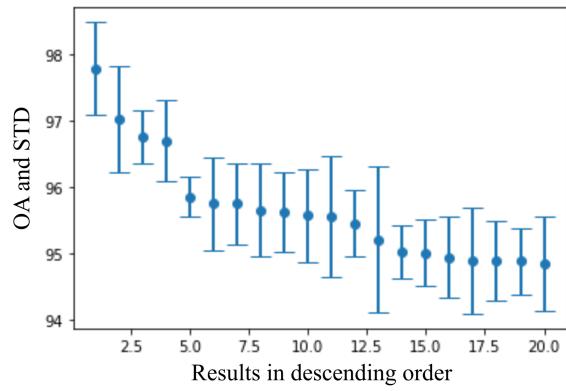


Fig. 5. Twenty highest overall accuracy results from Table III in descending order and their STD.

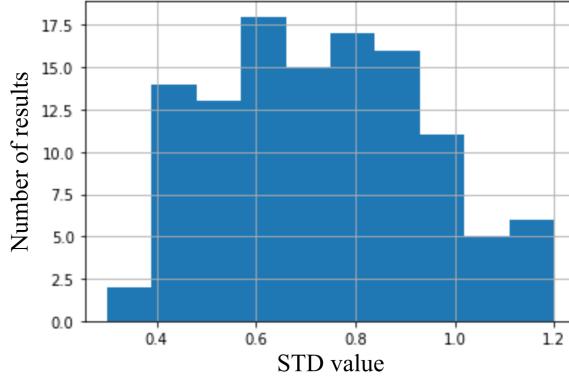


Fig. 6. Histogram of the error values from the results presented in Table III.

number of training images of the two sets, target and source, are updated simultaneously.

1) **Ablation Study 1:** In this experiment, the fine-tune training on the Xception architecture shown in Section II-C was performed. The pre-trained weights from the ImageNet dataset were first loaded and then each block (entry, middle, and exit flow) fine-tuned while freezing the other layers. The network was trained using Adam optimizer with a learning rate of 0.0001, a batch size of 16, a dropout of 0.2 before the decision layer, and 100 epochs. The mean F1-score results when fine-

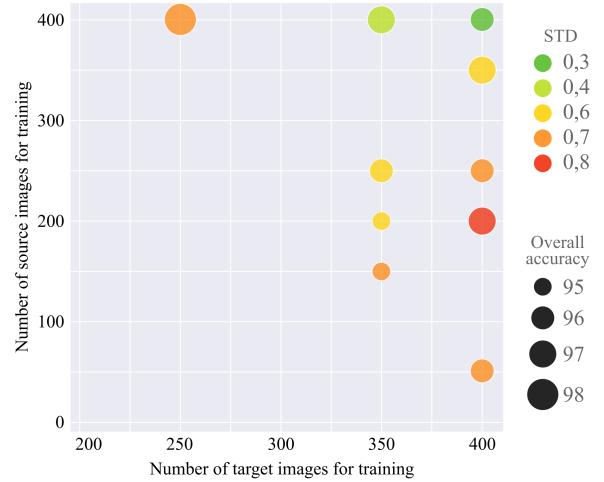


Fig. 7. Ten highest overall accuracy average results achieved with the proposed method (dot size), STD (dot color), and their corresponding data combinations of target/source in the training, presented in the "x" and "y" axis, respectively.

tuning different blocks of the Xception are shown in Fig. 8. One may observe that, fine-tuning the middle flow block of the Xception architecture leads to a very similar performance compared to fine-tuning all the network; hence, it is unnecessary to retrain all the Xception network. Consequently, in the following experiments, only the middle flow was fine-tuned as it provides the best results.

2) **Ablation Study 2:** For this experiment, combinations of number of images from target/source data were selected from Table III to perform an ablation study presented in Table II. Specifically, the accuracy results of the proposed method were calculated for seven target/source data combinations by eliminating one or two of the three main steps (A, B, and C) of our proposed CX-DaGAN algorithm.

The column *Step C*, which gives the worst results in Table IV, presents the classification of the test images of the small/target dataset through the Xception neural network. For this, the CNN was fine-tuned with the number of source and target images indicated in each row. In this case, source images used to complement the training dataset are randomly selected. Note

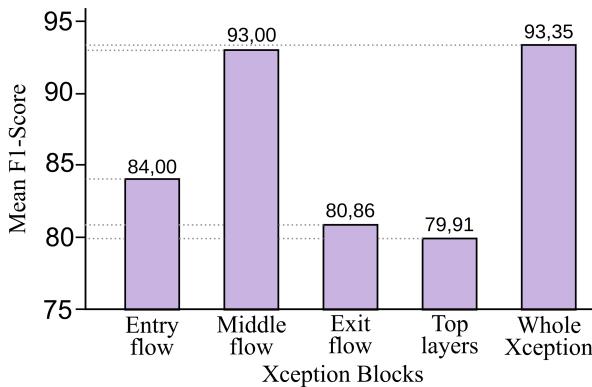


Fig. 8. Mean results when fine-tuning different blocks of Xception. Fine-tuning the middle flow block of the Xception architecture leads to a very similar performance compared to fine-tuning all the network.

that the most accurate result is obtained in this column with the largest number of images from each dataset. It should be noted that although 400 images from the target dataset are used in two cases, the highest precision is obtained when more images (350) are incorporated from the source dataset. Thus, combining a total of 750 tagged images for retraining, the accuracy was of 88.78%.

On the other hand, the results are better when we use two steps of our method. For instance, in “*Steps A + C*”, the similarity-constrained stage selects the source images to train the CNN. In this way, an increase of up to 7.91% in the average accuracy was achieved.

Then, the “*Steps B + C*” column shows the results of randomly selecting images from target and source, generating synthetic source images based on Cycle-GAN, and training the Xception. In this case, an increase of up to 4.73% compared to using only stage C was observed. Finally, the last row in Table IV shows the result using the three steps (A + B + C). It is evident that the simultaneous combination of all the steps allows a better performance of our proposed method, compared to using a part of it, thus proving the importance of each of these steps. It is worth mentioning that the results reported in the Tables III, IV, and V include the cross-validation technique, which is used to evaluate the results of statistical analysis and ensure that they are independent of the partition between training and test data. Note that, as indicated in the table titles, each result in Tables II and III included 10 realizations, and each result in Tables IV and V included 30 realizations.

D. Comparison Results

1) Other classification methods: In order to compare the performance of our proposed CX-DaGAN algorithm with other state-of-the-art methods, the above reference approaches were implemented and tested:

- TL:** Transfer learning with two chest X-rays datasets. This approach consists in considering a CNN previously trained with the ImageNet dataset, and retraining it with all available source images (5216 samples) + target images (400 samples). The resulting training network was used to classify the test target set.

TABLE IV
QUANTITATIVE CLASSIFICATION RESULTS OF THE ABLATION STUDY 1.
IMPORTANCE OF EACH STEP IN THE PERFORMANCE OF THE
CX-DAGAN ALGORITHM (AVERAGE OVER 30 REALIZATIONS).

target	source	Images for Train		(↑) Overall accuracy (%)		Proposed method: A + B + C
		Step C	Steps A + C	Steps B + C		
100	100	84.72±2.5	88.93±1.6	87.78±1.7	90.15 ± 1.2	
250	100	83.78±3.6	90.86±0.4	89.09±1.1	95.21 ± 1.1	
250	400	85.55±1.1	89.50±0.1	88.79±2.3	97.78 ± 1.5	
350	250	82.60±1.7	94.40±1.7	89.09±1.1	95.75 ± 0.6	
350	400	87.32±0.4	93.22±1.5	90.56±1.7	96.75 ± 1.0	
400	150	87.91±0.4	94.69±0.7	90.27±0.7	97.02 ± 1.8	
400	350	88.78±4.6	96.69±0.7	93.51±1.1	96.70 ± 1.4	

- NO-S:** No source images. In this experiment, a CNN pre-trained with the ImageNet dataset is re-trained with target images (400 samples), assuming no access to a second (source) X-ray dataset.
- RAND-S:** Random selection of source images. In this case, the CNN is retrained on a training dataset consisting in target images and randomly selected source images. This experiment aims the contrast of increasing the training dataset randomly compared to our source image selection method.
- SDASC:** Subspace-based Domain Adaptation using Similarity Constraints [38], a recent method of augmenting a target dataset with source images to improve classification results.

In all cases, the average over 30 realizations and the classification of 173 samples from the target dataset are reported. To ensure a fair comparison, all the methods used the same network backbone but with different optimization procedures. However, to broaden the comparison and to evaluate the consistency of our method, a discussion with other backbones is included using VGG-16, ResNet-50, and Xception networks. These results are reported in Table V.

TABLE V
QUANTITATIVE CLASSIFICATION RESULTS OF FIVE DATA-BASED METHODS (INCLUDING OUR PROPOSED METHOD) FOR THREE DIFFERENT CNNs. ALL METHODS INCLUDE 400 IMAGES FROM THE TARGET DOMAIN FOR TRAINING (AVERAGE OVER 30 REALIZATIONS).

Metric Method	S Images	VGG-16			ResNet-50			Xception		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
TL	5216	75.05	74.15	0.74	63.72	62.48	0.62	88.36	88.47	0.87
NO-S	0	85.84	85.83	0.85	90.27	90.26	0.90	90.03	90.03	0.88
RAND-S	100	85.84	85.84	0.85	87.61	87.61	0.88	89.52	88.19	0.90
	200	81.42	81.41	0.82	90.27	90.27	0.90	89.98	89.68	0.89
SDASC	100	84.96	84.96	0.84	91.15	91.15	0.91	93.25	92.54	0.92
	200	88.50	88.50	0.88	93.69	93.48	0.93	96.18	95.96	0.95
CX-DaGAN	100	85.25	86.22	0.86	92.04	92.03	0.92	94.90	92.03	0.92
	200	90.02	90.12	0.90	94.12	93.97	0.94	97.02	96.91	0.96

In general, all the methods shown in Table V perform better using the Xception network for the classification task, except for the NO-S method. NO-S method provides the best result by using only the target dataset. Overall, the results shown in Table V suggest that the classification accuracy is improved by adding data from another (source) dataset. However, the samples used to increase the size of the training database should be adequately selected. In particular, RAND-S method,

consisting in selecting randomly images from the source dataset, is shown to degrade the classifier's accuracy. On the other hand, the methods SDASC and CX-DaGAN achieve the best results due to careful data selection. Indeed, the proposed CX-DaGAN method presents a significant advantage in classification accuracy over the other algorithms in terms of ACC, F1, and AUC results. It is worth highlighting that SDASC and the proposed CX-DaGAN method were designed for binary classification. Therefore, a disadvantage of the proposed method, specifically for step A, which is based on data selection through equal and crossed classes projection, is that it can not be applied directly to a multi-class problem.

E. Visual Results

This section presents visual results of each stage of the proposed method.

1) Step A: First, images from the source dataset are selected based on error metrics that account for the similarity between these images and the target domain when projected on their subspaces. To get a deeper understanding of our CX-DaGAN, we visualized some of the images selected by the Step A. In Fig.9 the best and worst projected images are presented according to (1) and (2). Therefore, the figure is divided into four parts, one for each error metric. The two top rows depict the projection of each class into the same category of the target PCA subspaces. The first row presents original images from source domain, while the second row shows the result of the projections. Specifically, we show the images with the best (a and c) and the worst (b and d) projection error for each class, normal (left) and pneumonia (right). Note that the projected image is visually more distorted when the error is larger than when the projection error is smaller. On the other hand, the two bottom rows correspond to the projection of the source images on the target subspaces with opposite classes. In this case, the error numbers are similar to those in the upper part due to the remarkable similarity between all chest radiographs regardless of their pathology. However, our proposed method for cross-class projection involves, this time, selecting the source images with the highest error projection (b and d) as shown in the pink boxes.

2) Step B: The source samples selected in step A were used together with the training target images (400 images) to train the GAN-based image-to-image translation proposed in Section II-B. New synthetic images are generated within the target domain from the transformation of the previously chosen source images. In such a way, the number of source images feed the GAN network is equal to the number of output synthetic images. Fig. 10 shows four random source images selected by step A that entered the GAN, and in the row below, their respective transformations to the target domain.

3) Step C: Synthetic images and the target samples were used to train the Xception network with a particular proposed fine-tuning strategy. Fig. 11 depicts some examples of the classification results obtained when testing our method with step C on 173 images of the target domain. The correct predictions are presented with black labels, and erroneous predictions are shown with a red label.

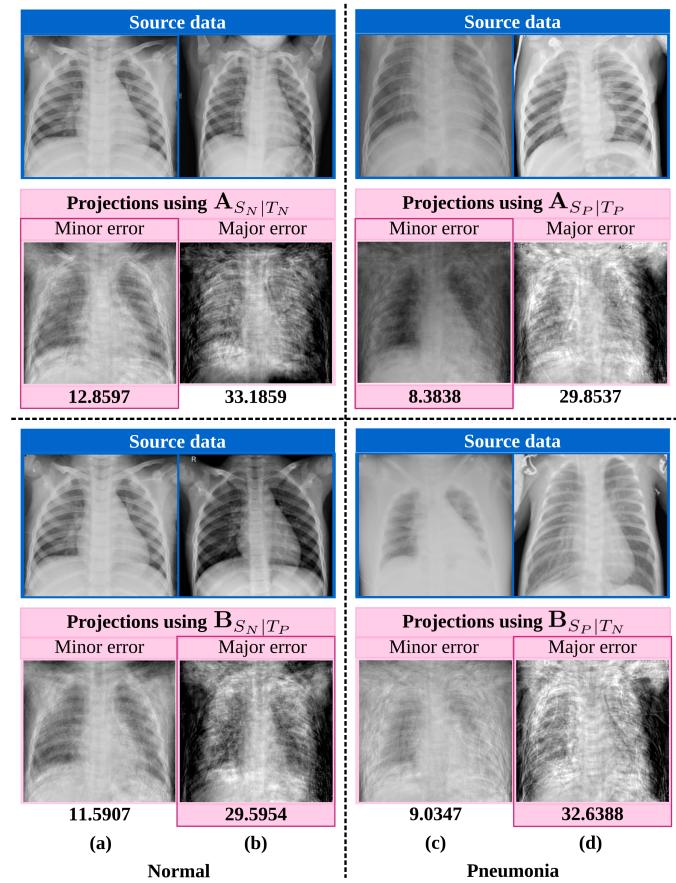


Fig. 9. Selected X-ray images from the source domain considering the projection error when projected onto target subspaces obtained by PCA. The selected images are highlighted in each case with pink color. The first two rows depict the projection of each class into the same category of the target PCA subspaces. The last two rows depict the projection of the source images on the target subspaces with opposite classes.

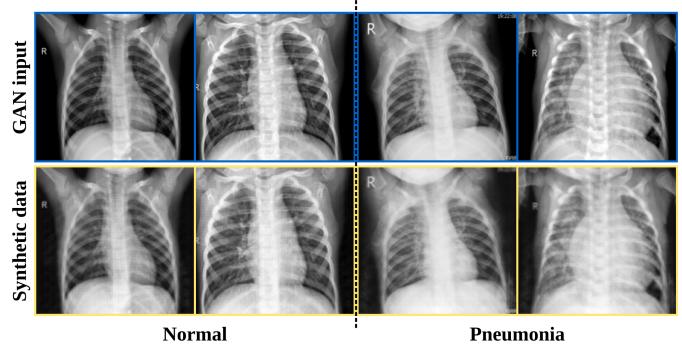


Fig. 10. Input (source domain) and output (synthetic data) examples of our GAN network.

IV. CONCLUSIONS

The main contribution of the proposed method is to take advantage of information from an extensive labeled public dataset to improve the classification accuracy of a small X-ray dataset acquired in a different hospital. Specifically, the main goal of the proposed approach is to select from a large dataset the images that best fit the small target dataset in the sense of their intra-class similarity and inter-class dissimilarity. In addition, a classification improvement is achieved by gen-

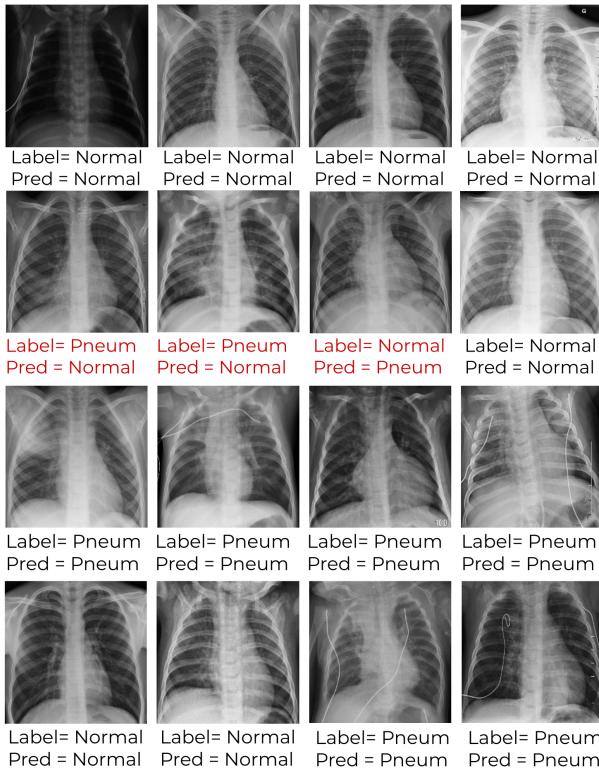


Fig. 11. Classification results of some chest X-ray images from the test subset. Correctly classified samples are shown with black labels and incorrect ones with red.

erating new images through a GAN network that follows the target data distribution. This paper introduced the CX-DaGAN algorithm, an original method to address the problem of chest X-ray pneumonia diagnosis on a small target dataset. To achieve this purpose, we propose to use information extracted from a larger and publicly available chest X-ray source dataset. Specifically, our proposed algorithm is a complete domain adaptation workflow which consists of three stages. First, we proposed a subspace-based domain adaptation method to select images from the large dataset (source domain). We then used the selected images and the train set of the small dataset (target domain) to train our proposed GAN-based image-to-image translation network. We finally used the synthetic images generated from GAN, which follow the target domain distribution, and the training set of the target dataset to fine-tune a pretrained CNN classification network to achieve the final classification accuracy. During the experiments, we observed that training on target data without performing our proposed domain adaptation workflow led to an overall accuracy of 88.36%. However, when we used our proposed workflow to augment the training set of target and carefully fine-tuning the Xception network, we achieved an overall accuracy of up to 97.78%. Future studies will consist in evaluating the performance of this new domain adaptation method for the classification of small datasets in other related medical tasks and involving other medical imaging modalities. Furthermore, it would be interesting to address the fundamental ideas behind the CX-DaGAN algorithm to extend its scopes to multi-class classification tasks. In this work, we considered as similarity

metric, rather the projection error using the ℓ_2 -norm than a human medical inspection. We will also investigate in our future studies the selection of the adequate images based on expert radiologists' decision, or using criteria specifically designed for medical images.

ACKNOWLEDGMENT

We thank Dr. Simon Freiss and Professor Nicolas Sans from the radiology department of Toulouse University Hospital for their help in collecting and anonymizing the images used in this study.

REFERENCES

- [1] M. Kolditz and S. Ewig, "Community-acquired pneumonia in adults," *Deutsches Ärzteblatt International*, vol. 114, no. 49, p. 838, 2017.
- [2] L. Liu, S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black, "Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals," *The Lancet*, vol. 388, no. 10063, pp. 3027–3035, 2016.
- [3] J. A. McCullers, "Effect of antiviral treatment on the outcome of secondary bacterial pneumonia after influenza," *Journal of Infectious Diseases*, vol. 190, no. 3, pp. 519–526, 2004.
- [4] H. Wu, P. Xie, H. Zhang, D. Li, and M. Cheng, "Predict pneumonia with chest x-ray images based on convolutional deep neural learning networks," *Journal of Intelligent & Fuzzy Systems*, no. Preprint, pp. 1–15, 2020.
- [5] M. I. Neuman, E. Y. Lee, S. Bixby, S. Diperna, J. Hellinger, R. Markowitz, S. Servaes, M. C. Monuteaux, and S. S. Shah, "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," *Journal of hospital medicine*, vol. 7, no. 4, pp. 294–298, 2012.
- [6] S. Andronikou, E. Lambert, J. Halton, L. Hilder, I. Crumley, M. D. Lytle, and C. Kosack, "Guidelines for the use of chest radiographs in community-acquired pneumonia in children and adolescents," *Pediatric Radiology*, vol. 47, no. 11, pp. 1405–1411, 2017.
- [7] M. Lavine, "The early clinical x-ray in the united states: patient experiences and public perceptions," *Journal of the history of medicine and allied sciences*, vol. 67, no. 4, pp. 587–625, 2012.
- [8] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics*, vol. 31, no. 4–5, pp. 198–211, 2007.
- [9] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, no. 1, pp. 1–13, 2016.
- [10] F. Riaz, S. Rehman, M. Ajmal, R. Hafiz, A. Hassan, N. R. Aljohani, R. Nawaz, R. Young, and M. Coimbra, "Gaussian mixture model based probabilistic modeling of images for medical image segmentation," *IEEE Access*, vol. 8, pp. 16 846–16 856, 2020.
- [11] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.
- [12] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.
- [13] T. T. Htay and S. S. Maung, "Early stage breast cancer detection system using glcm feature extraction and k-nearest neighbor (k-nn) on mammography image," in *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 2018, pp. 171–175.
- [14] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Medical physics*, vol. 47, no. 5, pp. e218–e227, 2020.
- [15] A. Anaya-Isaza, L. Mera-Jiménez, and M. Zequera-Díaz, "An overview of deep learning in medical imaging," *Informatics in Medicine Unlocked*, vol. 26, p. 100723, 2021.
- [16] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.

- [17] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [18] C. Calderón, K. Sanchez, S. Castillo, and H. Arguello, "Bilsk: A bilinear convolutional neural network approach for skin lesion classification," *Computer Methods and Programs in Biomedicine Update*, vol. 1, p. 100036, 2021.
- [19] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. I. Laurenson, "Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography," *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 3–13, 2022.
- [20] J. Escobar, K. Sanchez, C. Hinojosa, H. Arguello, and S. Castillo, "Accurate deep learning-based gastrointestinal disease classification via transfer learning strategy," in *2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, 2021, pp. 1–5.
- [21] T. Rahman, M. E. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray," *Applied Sciences*, vol. 10, no. 9, p. 3233, 2020.
- [22] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "Efficient pneumonia detection in chest x-ray images using deep transfer learning," *Diagnostics*, vol. 10, no. 6, p. 417, 2020.
- [23] M. Toğçaçar, B. Ergen, Z. Cömert, and F. Özyurt, "A deep feature learning model for pneumonia detection applying a combination of mrmr feature selection and machine learning models," *IRBM*, vol. 41, no. 4, pp. 212–222, 2020.
- [24] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (oct) and chest x-ray images for classification," *Mendeley data*, 2018.
- [25] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.
- [26] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation," in *2018 IEEE 15th International symposium on biomedical imaging ISBI*. IEEE, 2018, pp. 1038–1042.
- [27] Y. Zhang, S. Niu, Z. Qiu, Y. Wei, P. Zhao, J. Yao, J. Huang, Q. Wu, and M. Tan, "Covid-da: Deep domain adaptation from typical pneumonia to covid-19," *arXiv preprint arXiv:2005.01577*, 2020.
- [28] K. You, M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Universal domain adaptation," in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2019, pp. 2720–2729.
- [29] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 41–57, 2016.
- [30] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018.
- [31] A. Choudhary, L. Tong, Y. Zhu, and M. D. Wang, "Advancing medical imaging informatics by deep learning-based domain adaptation," *Yearbook of medical informatics*, vol. 29, no. 1, p. 129, 2020.
- [32] Y. Tang, Y. Tang, V. Sandfort, J. Xiao, and R. M. Summers, "Tuna-net: Task-oriented unsupervised adversarial network for disease recognition in cross-domain chest x-rays," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 431–440.
- [33] J. Jiang, Y.-C. Hu, N. Tyagi, P. Zhang, A. Rimner, G. S. Mageras, J. O. Deasy, and H. Veeraraghavan, "Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 777–785.
- [34] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 599–607.
- [35] A. Kapil, T. Wiestler, S. Lanzmich, A. Silva, K. Steele, M. Rebelatto, G. Schmidt, and N. Brieu, "Dasgan-joint domain adaptation and segmentation for the analysis of epithelial regions in histopathology pd-l1 images," *arXiv preprint arXiv:1906.11118*, 2019.
- [36] R. Venkataramani, H. Ravishankar, and S. Anamandra, "Towards continuous domain adaptation for medical imaging," in *2019 IEEE 16th International Symposium on Biomedical Imaging ISBI*. IEEE, 2019, pp. 443–446.
- [37] N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. Xing, "Unsupervised domain adaptation for automatic estimation of cardio-thoracic ratio," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 544–552.
- [38] K. Sanchez, C. Hinojosa, H. Arguello, S. Freiss, N. Sans, D. Kouamé, O. Meyrignac, and A. Basarab, "Subspace-based domain adaptation using similarity constraints for pneumonia diagnosis within a small chest x-ray image dataset," in *IEEE International Symposium on Biomedical Imaging ISBI*, 2021.
- [39] J. G. A. Barbedo, "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification," *Computers and electronics in agriculture*, vol. 153, pp. 46–53, 2018.
- [40] J.-Y. Zhu, T. Park, P. Isola, and 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.
- [41] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [42] M. Hossin and M. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, p. 1, 2015.
- [43] A. Diker, Z. Cömert, E. Avcı, M. Toğçaçar, and B. Ergen, "A novel application based on spectrogram and convolutional neural network for ecg classification," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*. IEEE, 2019, pp. 1–6.