Medical Image Analysis 83 (2023) 102664

Contents lists available at ScienceDirect

Medical Image Analysis

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

Contrastive domain adaptation with consistency match for automated pneumonia diagnosis

Yangqin Feng [a], Zizhou Wang [a], Xinxing Xu [a,*], Yan Wang [a], Huazhu Fu [a], Shaohua Li [a], Liangli Zhen [a], Xiaofeng Lei [a], Yingnan Cui [a], Jordan Sim Zheng Ting [c], Yonghan Ting [c], Joey Tianyi Zhou [b], Yong Liu [a], Rick Siow Mong Goh [a], Cher Heng Tan [c,d]

[a] Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
[b] Centre for Frontier AI Research, Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
[c] Department of Diagnostic Radiology, Tan Tock Seng Hospital (TTSH), Singapore 308433, Singapore
[d] Lee Kong Chian School of Medicine, Singapore 308232, Singapore

ARTICLE INFO

Keywords:
Unsupervised domain adaptation
Contrastive learning
Consistency match
Chest X-ray screening
Automated disease diagnosis

ABSTRACT

Pneumonia can be difficult to diagnose since its symptoms are too variable, and the radiographic signs are often very similar to those seen in other illnesses such as a cold or influenza. Deep neural networks have shown promising performance in automated pneumonia diagnosis using chest X-ray radiography, allowing mass screening and early intervention to reduce the severe cases and death toll. However, they usually require many well-labelled chest X-ray images for training to achieve high diagnostic accuracy. To reduce the need for training data and annotation resources, we propose a novel method called Contrastive Domain Adaptation with Consistency Match (CDACM). It transfers the knowledge from different but relevant datasets to the unlabelled small-size target dataset and improves the semantic quality of the learnt representations. Specifically, we design a conditional domain adversarial network to exploit discriminative information conveyed in the predictions to mitigate the domain gap between the source and target datasets. Furthermore, due to the small scale of the target dataset, we construct a feature cloud for each target sample and leverage contrastive learning to extract more discriminative features. Lastly, we propose adaptive feature cloud expansion to push the decision boundary to a low-density area. Unlike most existing transfer learning methods that aim only to mitigate the domain gap, our method instead simultaneously considers the domain gap and the data deficiency problem of the target dataset. The conditional domain adaptation and the feature cloud generation of our method are learning jointly to extract discriminative features in an end-to-end manner. Besides, the adaptive feature cloud expansion improves the model's generalisation ability in the target domain. Extensive experiments on pneumonia and COVID-19 diagnosis tasks demonstrate that our method outperforms several state-of-the-art unsupervised domain adaptation approaches, which verifies the effectiveness of CDACM for automated pneumonia diagnosis using chest X-ray imaging.

1. Introduction

Pneumonia is a form of acute respiratory infection that inflames the air sacs in one or both lungs, and it can cause mild to life-threatening illnesses in people of all ages. According to the World Health Organization (World Health Organization, 2019), pneumonia is the single most significant infectious cause of death in children worldwide. Early diagnosis and intervention can significantly reduce the severe cases and death toll, especially for coronavirus disease-19 (COVID-19) pneumonia, which has a high-speed spread and requires mass screening during the pandemic. Chest X-ray screening is a commonly accessible radiological examination for diagnosing many lung diseases (Wang et al., 2017). For instance, as the global pandemic of coronavirus disease-19 (COVID-19) progresses, chest radiography (CXR) has been one of the most commonly utilised modalities for triage and primary clinical observations (Minaee et al., 2020). Automated computer-aided diagnosis from CXRs can improve the CXRs interpretation efficiency significantly (Lakhani and Sundaram, 2017).

Inspired by the great success of deep neural networks for computer vision tasks, such as face recognition, image classification, and object detection, data scientists from the healthcare domain have developed

* Corresponding author.
E-mail address: xuxinx@ihpc.a-star.edu.sg (X. Xu).

https://doi.org/10.1016/j.media.2022.102664
Received 16 December 2021; Received in revised form 2 September 2022; Accepted 13 October 2022
Available online 22 October 2022

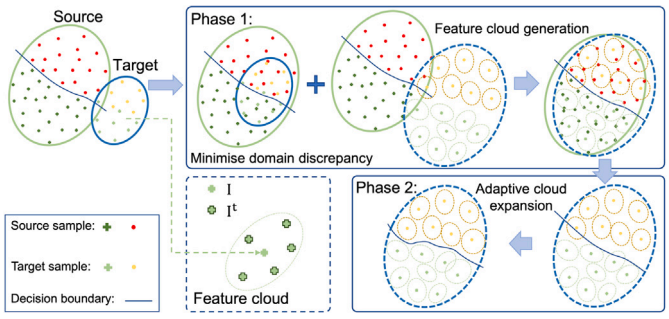


Fig. 1. The framework of Contrastive Domain Adaptation with Consistency Match (CDACM). The proposed CDACM contains two phases. Phase I aims to minimise domain discrepancy and generate a feature cloud for solving the target domain's data deficiency problem; In phase 2, it further adjusts the decision boundary of the target domain's classifier to improve the prediction accuracy by involving the adaptive feature cloud expansion.

some deep learning methods for automated disease diagnoses and achieved promising results (Lundervold and Lundervold, 2019; Yi et al., 2019; Schlemper et al., 2019; Gu et al., 2019). However, deep neural networks typically require a large amount of labelled data with diverse visual variations for training (Deng et al., 2009). It might not be practical for clinical applications since data from clinical practice presents heterogeneous acquisition conditions (e.g., different equipment), and the annotations are expensive for each new domain (e.g., a new dataset from the new equipment or another ethnic group). To reduce the annotation effort, transferring knowledge from label-rich source domains to the target domain is a feasible and widely-used solution (Saito et al., 2019), especially when samples in the target dataset are unlabelled. Nevertheless, dataset bias often reduces the model's generalisation to new data (Saenko et al., 2010). For example, when training the model on publicly available large-scale CXR datasets and then directly applying the model to perform the inference on the new dataset from clinical practice, the performance of the deep learning models degrade sharply. This phenomenon is known as domain shift (Sun et al., 2016).

In the scenarios where label information is unavailable for the target dataset, unsupervised domain adaption (UDA) methods (Cicek and Soatto, 2019; Kang et al., 2019; Tajbakhsh et al., 2020; Luo et al., 2020b; Zhang et al., 2020; Liu et al., 2019) have been developed to address the domain shift problem. They aim to transfer knowledge from a source dataset to a different (but related) target dataset (Redko et al., 2019). Among existing approaches, adversarial learning has been utilised to minimise the domain discrepancy between the source and target domains successfully (Tzeng et al., 2017; Ganin et al., 2016; Chadha and Andreopoulos, 2019; Long et al., 2018). For instance, Tzeng et al. proposed a two-player model named domain adversarial neural network (DANN) (Ganin et al., 2016), which contains a domain discriminator D and feature representation module F. The domain discriminator D aims to distinguish the input samples of the source domain from that of the target domain. Simultaneously, the feature representation module F is trained to confuse the domain discriminator D. They are learning to go against each other, and both become strong. Recent studies of Conditional Generative Adversarial Networks (CGANs) (Mirza and Osindero, 2014) show that different distributions can be matched better if we condition the generator and discriminator on relevant information, e.g., associated labels and affiliated modality. Although domain adaption can mitigate the gap between the source and target domains to some extent, automatic CXR interpretation is still a challenging task due to the extremely small-scale target dataset (Zhou et al., 2020).

To overcome the challenges mentioned above, we propose a novel method, called **Contrastive Domain Adaptation with Consistency Match** (CDACM), for automated CXRs interpretation. CDACM aims to learn domain-invariant features by minimising domain discrepancy and to improve the semantic quality of the learnt representations for target data with contrastive learning and consistency match. The framework of CDACM is shown in Fig. 1, from which we can see that it contains two phases. In Phase I, CDACM generates a feature cloud for each target sample upon the domain discrepancy minimisation to solve the data deficiency problem in the unlabelled target domain. The conditional domain adaptation and the feature cloud generation are optimised jointly to extract discriminative features in an end-to-end manner. In Phase II, the feature clouds expand adaptively to push the decision boundary to an area of low density, improving the learnt classifier's robustness in the target domain. Unlike most existing transfer learning methods that only aim to mitigate the domain gap, our method instead considers the domain gap and the data deficiency problem of the target domain simultaneously, making it more effective to transfer the knowledge of feature extraction from the source domain to the target domain. Besides, we design an adaptive feature cloud expansion mechanism to improve the model's generalisation ability in the target domain. The proposed CDACM is firstly applied to the child pneumonia diagnosis by leveraging a source dataset of adult CXRs. Also, CDACM is evaluated using a COVID-19 dataset we collected from the hospital as the target domain while using a large-scale publicly available dataset as the source domain. The experimental results demonstrate CDACM's effectiveness for pneumonia diagnosis from child CXRs and COVID-19 diagnosis in the clinical setting.

The novelty and main contributions of this work are summarised as follows:

- We propose a two-phase domain adaptation method for automated pneumonia diagnosis, which includes our designed contrastive and adaptive feature cloud expansion modules. It achieves effective domain adaptation when the target dataset is much smaller than the source dataset. The pneumonia diagnosis use case faces a data deficiency problem and can benefit from the proposed CDACM. Experimental results show that our proposed two-phase domain adaptation method can significantly improve pneumonia diagnosis performance compared with other peer methods and three junior radiologists.
- A contrastive learning-based strategy is presented to generate the feature cloud for each target sample to enforce target data representations dispersed in the feature space, thus improving the domain alignment when the target dataset is much more small-scale than the source dataset. A conditional adversarial domain adaption module and the contrastive learning module are trained jointly, making the two modules beneficial to each other along with the training process. To the best of our knowledge, this work is the first one that adopts contrastive learning for addressing data deficiency problem in domain adaptation.
- An adaptive feature cloud expansion mechanism is designed to improve the model's generalisation ability in the target domain. It expands the feature cloud only if the prediction results of existing samples in the feature cloud are consistent with high confidence. At the same time, the newly generated samples for the feature cloud are enforced to match the pseudo-label induced from the prediction results of the existing samples in the feature cloud.

2. Related work

The high prediction accuracy of deep neural networks usually comes at the cost of the massive amount of labelled data for training. In practice, however, it is expensive and hard to collect and annotate large-scale datasets for many real-world applications, especially for automated disease diagnosis, which requires experienced doctors to provide the labels. Transfer learning is one of the most effective ways to reduce the requirement of labelled data for the target task. A commonly used transfer learning strategy is to pre-train the model on a large-scale source dataset, e.g., ImageNet (Deng et al., 2009)

(which contains more than 2000 categories of more than 14 million images). Then, the model will be fine-tuned on the target dataset to achieve high prediction accuracy. For instance, Shin et al. evaluated the networks of AlexNet (Krizhevsky et al., 2012) and GoogLeNet (Szegedy et al., 2015) for the computer-aided detection (CADe) applications and proved that pre-training on ImageNet yields higher accuracy than training from scratch (Shin et al., 2016b). The effectiveness of transferring knowledge from natural images to medical images, such as CT and MRI images (Shin et al., 2016a), X-ray images (Ke et al., 2021), or neuroimaging data (Gupta et al., 2013), with pre-training also has been verified in the literature. However, these approaches have not considered different distributions of the two domains, which may cause the domain-shift problem (Tzeng et al., 2017).

Domain adaptation (DA) is one of the most effective ways of addressing the domain-shit problem (Patel et al., 2015). Considering label information availability, we can group it into three categories: supervised DA, semi-supervised DA, and UDA. This work lies in the last category. Pioneering UDA methods directly minimise the discrepancy between the source and target domains in the representation spaces of the deep learning models. For example, Long et al. proposed to minimise the maximum mean discrepancy (MMD) (Long et al., 2015) or the Joint MMD (Long et al., 2017), which has achieved impressive performance for image classification.

In recent years, adversarial learning has been applied to mitigate the domain gap between the two domains. For example, Hoffman et al. adopted the cycle-consistency strategy and proposed Cycle-Consistent Adversarial Domain Adaptation (CyCADA). It contains a reconstruction (cycle-consistency) loss and a semantic loss to preserve local structural information and enforce semantic consistency. There are also plenty of efforts on adversarial learning for unsupervised domain adaptation in the medical image analysis domain. The seminal works on these attempts investigate the brain lesion segmentation (Kamnitsas et al., 2017), cardiac structures segmentation (Dou et al., 2018), and multi-organ segmentation (Zhang et al., 2018). To further improve the performance, Chen et al. proposed Synergistic Image and Feature Adaptation (SIFA) (Chen et al., 2019, 2020), which provides a synergistic fusion of adaptations in both image and feature spaces for medical image segmentation.

Besides, by considering the condition of relevant information on the generator and discriminator, Conditional Domain Adversarial Networks (CDAN) (Long et al., 2018) is proposed to exploit discriminative information implicitly in the classifier predictions to improve the performance of adversarial adaptation. However, CDAN has not considered the semantic quality of the learnt representations in the target domain. Thomas et al. (Kang et al., 2019) developed a class-aware UDA method by estimating the label hypothesis of target samples via clustering. Thomas et al. claimed that these UDA methods are not as helpful as adapting to the test set directly and proposed the test-time unsupervised domain adaptation (TTUDA) (Varsavsky et al., 2020). In addition, some studies use a self-training strategy to mitigate the negative impact of noisy pseudo-labels generated for the target domain in the domain adaptation. For instance, Zhou et al. presented a confidence regularised self-training (CRST) (Zou et al., 2019) framework to treat pseudo-labels as continuous latent variables and jointly optimise them via alternating optimisation. These adversarial learning-based and self-trained-based methods have achieved promising performance for unsupervised domain adaptation.

However, these domain adaptation methods have not considered the data deficiency problem of the target dataset. To address the data deficiency problem of the target domain, some researchers have made some attempts in recent years. For instance, Ouyang et al. proposed a data-efficient method for multi-domain medical image segmentation (Ouyang et al., 2019) by combining a variational autoencoder-based feature prior matching and domain adversarial training to learn a shared domain-invariant latent space. It achieves excellent performance for medical image segmentation by transferring knowledge from the labelled source domain (3D MRI) to the unlabelled target domain (3D CT). Luo et al. proposed an Adversarial Style Mining approach (ASM) (Luo et al., 2020a) to search for new stylised samples to help the target model adapt to the almost unseen target domain. It assumes that the distribution gap between source and target domains consists of style differences. This assumption does not hold in automated pneumonia diagnosis.

To achieve effective domain adaption for automated pneumonia diagnosis in real-world deployment, we consider the data deficiency problem of the target dataset and the potential difference between the decision boundaries for the classifiers of the source and target domains. Our proposed CDACM addresses these two issues in two phases. Phase I aims to mitigate the domain gap using adversarial learning and simultaneously solve data deficiency problems with contrastive learning. Phase II is designed to adjust the decision boundary to improve the classification accuracy for the target domain.

3. Our proposed method

We denote the source domain as $D^s = \{(x_i^s, y_i^s)\}_{i=1}^{n^s}$ with n^s labelled CXRs drawn from the distribution P^s, and the target domain as $D^t = \{(x_i^t)\}_{i=1}^{n^t}$ with n^t unlabelled CXRs sampled from the distribution P^t, where the $P^s \neq P^t$. CDACM aims to build a deep learning model Θ to reduce the difference between the distributions P^s and P^t in the feature space, thus improving the semantic quality of the learnt representations for achieving high prediction accuracy in the target domain. In this work, we leverage two different types of transformations: weakly-augmented transformations (e.g., the horizontal flip of an image) and strongly-augmented transformations (e.g., rotation of the image with a large degree) to generate new samples for the target domain. The weakly-augmented dataset is represented as $\bar{D}^t = \{(\bar{x}_i^j)\}_{i=1}^{n^t}, j = 1, \ldots, \mu$, which contains a set of μ weakly-augmented samples $\bar{x}_i^1, \bar{x}_i^2, \ldots, \bar{x}_i^\mu$ for each target sample x_i^t. The strongly-augmented dataset is represented as $\tilde{D}^t = \{(\tilde{x}_i^k)\}_{i=1}^{n^t}, k = 1, \ldots, \nu$, which contains a set of ν strongly-augmented samples $\tilde{x}_i^1, \tilde{x}_i^2, \ldots, \tilde{x}_i^\nu$ for each target sample x_i^t. The weakly-augmented dataset is used for feature cloud construction, and the strongly-augmented dataset is used for adaptive feature cloud expansion.

3.1. Overview of CDACM

CDACM is a two-phase domain adaptation method, as shown in Fig. 1. Phase I conducts conditional domain adversarial learning to minimise the cross-domain discrepancy and contrastive training to generate the feature cloud for each target sample simultaneously. Thus, transferring the knowledge about the discriminative features of the target task from the source domain to the target domain with high quality. Phase II further adjusts the decision boundary to enforce it to be located in a low-density area, making CDACM generalise well to the target samples.

3.2. Phase I of CDACM

The workflow of CDACM's Phase I is shown in Fig. 2, where we denote $f = F(x)$ as the feature representation and $g = G(x)$ as the classification prediction from the classifier. Inspired by conditional domain adversarial learning (Long et al., 2018), CDACM enables conditional adversarial domain adaptation over domain-specific feature representation f and classifier prediction g. Simultaneously, the feature representations f_i^{tj} of transformed images $\bar{x}_i^j$ for each target image x_i^t becomes a set of probable features scattered around itself (feature cloud). CDACM encourages the samples in the same feature cloud to be similar and those from different feature clouds to be dissimilar in the feature space. It is achieved by minimising a contrastive loss. The aforementioned two components of CDACM are trained jointly to achieve more effective domain adaptation since representations of the

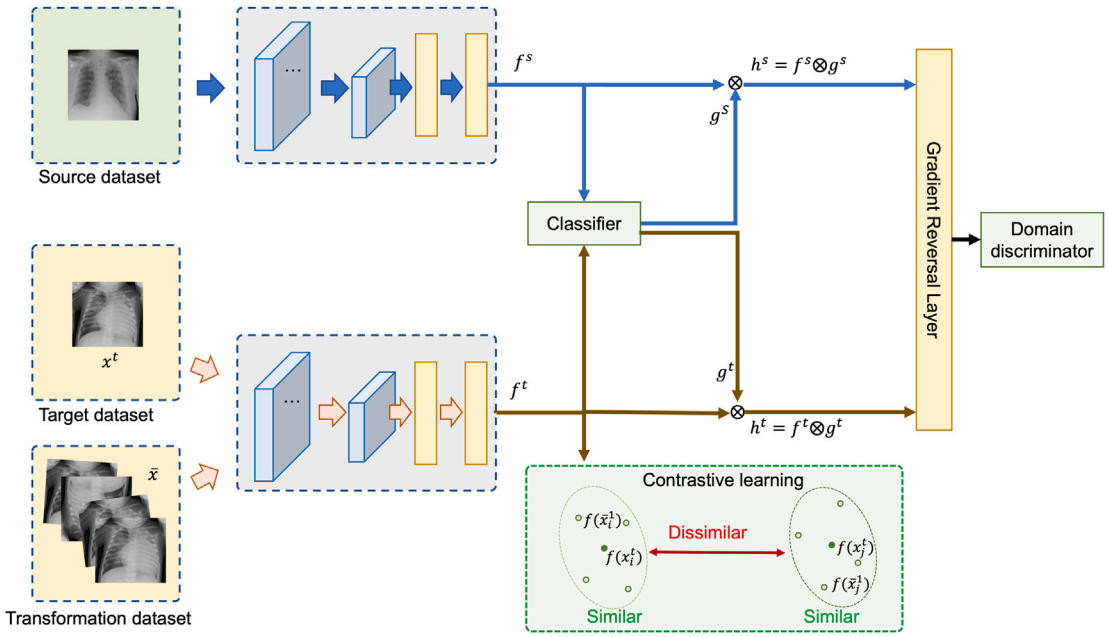


Fig. 2. The workflow of CDACM's Phase I. CDACM takes the source (D^s), the target (D^s) and the transformation (D^A) datasets as input to extract feature representations f^s, f^t, and f^A, respectively. Following Long et al. (2018), CDACM uses an entropy-based conditional domain adaptation loss. Different from Long et al. (2018), f^c is the centre of the feature cloud, and c^c denotes the average of the classifier outputs. CDACM encourages the representations of the ith target image to be similar to its transformed counterparts, while each feature cloud is dissimilar to others.

samples from the source and target domains tend to be mixed well in feature space along with the iterative optimisation of the training process.

Mathematically, in Phase I of CDACM, we propose to minimise the following objective function:

$$\mathcal{L}_1(D^s, D^t, \bar{D}^t) = \mathcal{L}_C(\theta_f, \theta_c) + \lambda_1 \mathcal{L}_D(\theta_f, \theta_d) + \lambda_2 \mathcal{L}_{CL}(\theta_f), \tag{1}$$

where θ_f, θ_c, and θ_d are the weight parameters of the feature extractor, source classifier, and domain discriminator, respectively. $\mathcal{L}_C(\theta_f, \theta_c)$ is the classification loss of the source domain. We use Cross Entropy loss as the classification loss in this work. $\mathcal{L}_D(\theta_f, \theta_d)$ denotes the domain adaptation loss, and $\mathcal{L}_{CL}(\theta_f)$ denotes the contrastive loss. λ_1 and λ_2 are given scalars that trade-off the contributions of the three different terms. We will detail the definitions of $\mathcal{L}_{LC}(\theta_f)$ and $\mathcal{L}_D(\theta_f, \theta_d)$ in the following.

Triplet Margin Loss encourages relative distance constraint by minimising the distance between the anchor α and the positive sample β with the same identity and maximising the distance between the anchor α and the negative sample γ that with a different identify (Schroff et al., 2015). We construct all the possible triplets $(\alpha_i, \beta_i, \gamma_i), i = 1, \ldots, N$ by treating each sample from the target dataset as an anchor and the samples from the same feature cloud as the positive samples and that from the different feature cloud as the negative samples. The contrastive loss can be expressed as:

$$\mathcal{L}_{CL}(\theta_f) = \sum_{i=1}^{N} \left[\|f(\alpha_i) - f(\beta_i)\|_2^2 - \|f(\alpha_i) - f(\gamma_i)\|_2^2 + \tau \right], \tag{2}$$

where τ is a margin that is enforced between positive and negative pairs. Minimising $\mathcal{L}_{CL}(\theta_f)$ can enforce the feature representation of target image x_i^t to be similar to that of its transformed counterpart $\bar{x}_i^j$, and the feature representation of $\bar{x}_i^j$ to be dissimilar to that of other images x_j^t in the target domain. Note that we only use μ weakly-augmented samples to construct a feature cloud in Phase I, which aims to accelerate the convergence of the training optimisation.

The conditional domain adaptation loss of our method $\mathcal{L}_D(\theta_f, \theta_d)$ is defined as a entropy-based loss (Long et al., 2018):

$$\mathcal{L}_D(\theta_f, \theta_d) = log[\sigma(\phi(h_i^s))] + log[1 - \sigma(\phi(h_j^t))], \tag{3}$$

where ϕ is a multi-linear map, σ is the condition domain discriminator on classifier prediction g through joint variable h, and $h_i^s = f_i^s \bigotimes g_i^s$ and $h_j^t = f_j^t \bigotimes g_j^t$ are the joint variables of f and g of CXRs from the two domains respectively. f_i^s and g_i^s denote the feature representation and prediction of the ith image from the source domain, respectively. f_j^t and g_j^t denote the centre of the jth feature cloud and centre of the predictions of the jth feature cloud, respectively.

3.3. Phase II of CDACM

The workflow of CDACM's Phase II is shown in Fig. 3, from which we can see that CDACM expands the target sample's feature cloud if the weekly-augmented samples (which also are used in Phase I) in the feature cloud have consistent prediction results. Inspired by FixMatch (Sohn et al., 2020), which is originally proposed for semi-supervised learning, we generate strongly-augmented samples to expand the existing feature cloud if the weakly-augmented samples in the same target feature cloud have a consistency of high confidence. Furthermore, the strong-augmented samples are enforced to match the pseudo-label deduced from the weakly-augmented samples in the same feature cloud. Unlike FixMatch (Sohn et al., 2020) that only generates a single augmented sample and checks if the model produces a high-confidence prediction, our method generates a set of weakly-augmented samples and checks the consistency of the model's predictions on these samples, thus making CDACM more stable for the case where the target domain has insufficient data. During Phase II, the domain alignment module and the consistency match module are learning jointly to adjust the classifier's decision boundary without the labelled target samples.

Specifically, in Phase II of CDACM, we minimise the following objective function:

$$\mathcal{L}_2(D^t, \bar{D}^t, \tilde{D}^t) = \mathcal{L}_1(D^s, D^t, \bar{D}^t) + \lambda_3 \mathcal{L}_{CM}(\theta_f, \theta_c), \tag{4}$$

where λ_3 is a hyper-parameter to control the trade-off between the two terms. In the following, we will illustrate how to compute $\mathcal{L}_{CM}(\theta_f, \theta_c)$.

To compute the consistency match loss, we first compute the model's output class distributions for the given μ weakly-augmented samples. For the feature cloud of x_i^t, the class distribution of $\bar{x}_i^j$ is $p_i^j = g(\bar{x}_i^j)$, which denotes a vector of the probabilities of belonging to C different

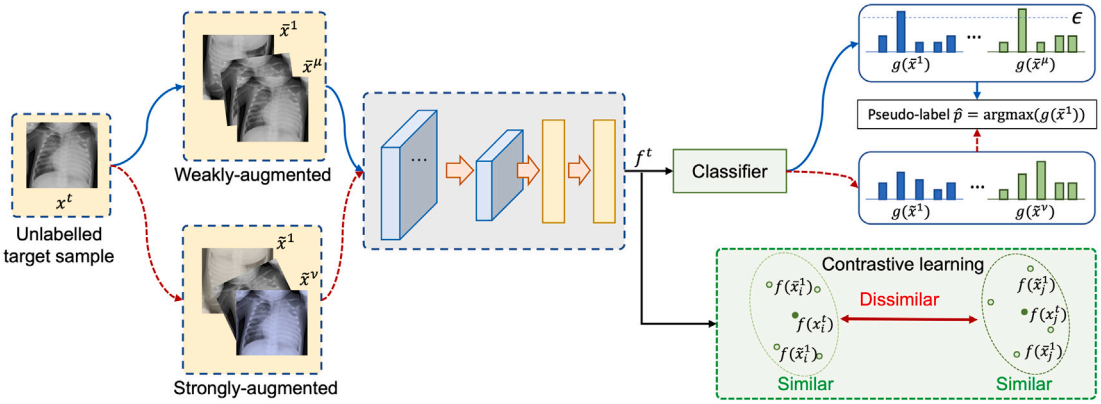


Fig. 3. The workflow of CDACM's Phase II. The weakly-augmented samples are existing samples in the feature cloud, and the strongly-augmented samples are new samples for feature cloud expansion. A feature cloud expansion is conducted based on if there is a consistency of the high confidence predicted by the model on existing samples in the feature cloud. The consistency is measured if all of the existing samples have a probability of one common class above a threshold ϵ, and it generates a pseudo-label for the feature cloud. Then, the model's predictions for new samples in the feature cloud (i.e., strongly-augmented samples) are enforced to match the pseudo-label. At the same time, contrastive learning is conducted on the feature clouds.

classes. Then, we expand the feature cloud for x_i^t only if the following conditions meet:

$$\begin{aligned} &\min(\max(p_i^1), \max(p_i^2), \dots, \max(p_i^\mu)) \geq \epsilon; \\ &\text{argmax}(p_i^1) = \text{argmax}(p_i^2) = \cdots = \text{argmax}(p_i^\mu), \end{aligned} \tag{5}$$

and set the expansion flag $\delta_i = 1$, otherwise $\delta_i = 0$. The pseudo-label for the feature cloud of x_i^t is $\hat{p}_i = \text{argmax}(p_i^1)$.

Next, we can compute the consistency match loss on the strongly-augmented samples $\tilde{x}_i^1, \tilde{x}_i^2, \dots, \tilde{x}_i^\nu$ as follows:

$$\mathcal{L}_{CM}(\theta_f, \theta_c) = \sum_{i=1}^{n^t} \left[\delta_i \sum_{k=1}^{\nu} \psi(\hat{p}_i, q_i^k) \right], \tag{6}$$

where $q_i^k = g(\tilde{x}_i^k)$ denotes the final class distribution and $\psi(\cdot)$ denotes the cross-entropy loss.

Note that our method does not include the strongly-augmented data of all target samples for model training since some target samples' strongly-augmented data may hurt the deep model's learning. Thus, our method includes the strongly-augmented data for learning only if the model has consistent prediction results on all the weakly-augmented samples for the target sample.

Regarding the contrastive loss, it computes the loss on both $\bar{D}^t$ and $\tilde{D}^t$ in Phase II, which is different from that in Phase I computing it on $\bar{D}^t$ only.

4. Experimental study

4.1. Dataset

We evaluate CDACM's effectiveness on four datasets for two classification tasks. All datasets are split into three subsets, and the statistical information is summarised in Table 1. The details of the tasks and datasets are as follows:

Binary classification: We use the Radiological Society of North America (RSNA) pneumonia detection challenge dataset (Stage I) (Radiological Society of North America, 2018) as the source dataset and a Child X-ray dataset (Kermany et al., 2018) as the target domain. The RSNA dataset is constructed by taking 26,684 samples from the ChestX-ray14 dataset (Wang et al., 2017). It contains 6011 pneumonia cases and 20,672 other cases (which contain normal and other diseases). The Child X-ray dataset (Kermany et al., 2018) (anterior-posterior) includes 5863 paediatric patients' X-ray images from Guangzhou Women and Children's Medical Centre.

Multi-class classification: A RSNA+COVID dataset is used as the source domain, and a dataset (named TTSH) we collected from the Tan Tock Seng Hospital is the target domain. Since COVID-19 is a new disease, it is not included in the existing large-scale dataset. The number of publicly available X-rays OF COVID-19 cases is relatively small. We collect 1858 COVID-19 cases by following the instructions of Minaee et al. (2020), Haghanifar et al. (2022) and combine these COVID-19 cases with the RSNA dataset to construct a new dataset named RSNA+COVID that contains three classes (i.e., Pneumonia, COVID, and Others). We collected the samples of this dataset from the hospital, which contains 5192 CXR images for training, validation and testing. Each CXR image was individually and manually reviewed by a radiologist with 14 years of experience interpreting CXRs. The CXRs in the TTSH dataset are labelled into three classes (Pneumonia, COVID, and Others). The CXRs are labelled as COVID if they contain clinical pneumonia signals and their Polymerase Chain Reaction (PCR) test results are positive.

We collect 500 CXR images as the external test to compare our model's performance with that of junior radiologists. The external test set consists of 500 cases performed in the same hospital at different times. It contains 72 COVID cases, 49 pneumonia cases, and 379 other cases. The characteristics of the external test set are summarised in Table 1.

4.2. Experimental settings

In this work, we use $\mu = 3$ weakly-augmented transformations: rotation with the angle between 0 to 10 degrees, horizontal flip, and vertical flip to construct the feature cloud. We adopt a combination of the image rotation (with the angle between 10 to 20 degrees) and the random auto-contrast operation (Cubuk et al., 2019) as the strongly-augmentation transformation to generate samples to expand the feature cloud, and we set $\nu = 3$ and $\epsilon = 0.7$. The optimal value for μ and ν depends on the data scale of the target domain and the available computational resource. Typically, we will set a relatively large value for μ and ν to increase the training data if the target dataset has very limited samples. However, it is notable that the consistency of the weakly-augmented data may be hard to meet if the value of μ is too large and it requires a much larger memory for model training. CDACM employs the popular convolutional layers of the ResNet50 (He et al., 2016), DenseNet121 (Huang et al., 2017) and EfficientNet-B4 (Tan and Le, 2019) as the backbones. The feature extractor's weight parameters of our model are initialised with the corresponding backbone pre-trained on ImageNet (Deng et al., 2009). We train our model using Adam (Kingma and Ba, 2014) with standard parameters (0.9, 0.999). The batch size of the optimiser is 8. We use a learning rate of $1e-4$ for binary classification and a learning rate of $1e-3$ for multi-class classification. The optimal values for

Table 1
Statistical results of the datasets, where n_{train}, n_{val}, n_{test} and n_{ext} denote the numbers of images in the training, validation, test and external test sets, respectively. For the binary classification, we evaluate the performance of transferring knowledge from RSNA to Child X-ray, and for multi-class classification RSNA+COVID to TTSH.

# imgs	Binary classification				Multi-class classification					
	RSNA		Child X-ray		RSNA+COVID			TTSH		
	Pneu	Others	Pneu	Others	COVID	Pneu	Others	COVID	Pneu	Others
n_{train}	4,209	14,471	3,106	1,079	1,301	4,209	14,471	477	448	2,698
n_{val}	601	2,067	777	270	186	601	2,067	58	57	386
n_{test}	1,202	4,134	390	234	371	1,202	4,134	116	114	772
n_{ext}	–	–	–	–	–	–	–	72	49	379
n_{total}	6,011	20,672	4,273	1,583	1,858	6,011	20,672	723	668	4,138

Table 2
Performance comparison of CDACM with the peer methods on Child X-ray in terms of the AUC score (mean ± standard deviation, %).

Method	EfficientNet-B4	DenseNet121	ResNet50
DANN (Ganin et al., 2016)	85.42 ± 3.24	87.08 ± 1.70	86.24 ± 2.21
CyCADA (Hoffman et al., 2018)	79.76 ± 3.40	79.09 ± 1.89	77.22 ± 7.00
BSWD (Rozantsev et al., 2018)	84.33 ± 1.68	86.79 ± 2.57	86.37 ± 3.13
CDAN (Long et al., 2018)	84.33 ± 1.68	86.46 ± 2.43	86.96 ± 3.31
MCD (Saito et al., 2018)	83.41 ± 4.28	85.43 ± 4.91	86.51 ± 3.19
MDD (Zhang et al., 2019)	88.18 ± 1.72	85.11 ± 2.45	86.77 ± 2.62
FixBi (Na et al., 2021)	84.07 ± 0.86	87.28 ± 1.65	85.88 ± 1.24
MEDM (Wu et al., 2021)	84.65 ± 4.83	81.46 ± 1.63	82.51 ± 0.92
CDACM (ours)	90.57 ± 3.19*	88.33 ± 1.05*	88.08 ± 1.54*

*The symbol indicates that the value of the proposed method is significantly different from all other methods at a 5% level by Wilcoxon's rank sum test.

Table 3
Performance comparison of CDACM with the peer methods with 10% of training and validation data for model training on the Child X-ray dataset in terms of the AUC score (mean ± standard deviation, %).

Method	EfficientNet-B4	DenseNet121	ResNet50
DANN (Ganin et al., 2016)	75.87 ± 4.87	73.77 ± 3.27	73.96 ± 6.49
CyCADA (Hoffman et al., 2018)	67.71 ± 7.87	59.30 ± 6.27	68.61 ± 6.94
BSWD (Rozantsev et al., 2018)	74.31 ± 4.59	54.84 ± 17.23	69.99 ± 17.32
CDAN (Long et al., 2018)	76.25 ± 2.74	74.64 ± 3.20	75.08 ± 8.28
MCD (Saito et al., 2018)	81.69 ± 4.24	81.21 ± 7.53	84.72 ± 4.43
MDD (Zhang et al., 2019)	84.80 ± 2.95	84.23 ± 2.40	84.01 ± 7.30
FixBi (Na et al., 2021)	75.92 ± 3.44	73.65 ± 7.65	81.14 ± 1.78
MEDM (Wu et al., 2021)	79.86 ± 2.37	78.70 ± 0.56	70.76 ± 11.98
CDACM (ours)	86.89 ± 2.70*	85.04 ± 2.61*	87.35 ± 1.58*

*The symbol indicates that the value of the proposed method is significantly different from all other methods at a 5% level by Wilcoxon's rank sum test.

λ_1, λ_2 in Phase I and λ_3 in Phase II are determined using grid search in the set of $\Phi = \{0.001, 0.01, 0.1, 1, 10, 100\}$. We use the model that achieves the highest AUC score on the validation set for the testing evaluation. For the peer methods, we also use the grid search to select the optimal value for the hyper-parameters in the set of $\Omega = \{0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 1, 5, 10, 100\}$.

4.3. Binary classification for child pneumonia diagnosis

We compare CDACM with six different domain adaptation methods, namely Deep Adaptation Neural Network (DANN) (Ganin et al., 2016), Cycle-Consistent Adversarial Domain AdaptationCy (CyCADA) (Hoffman et al., 2018), Beyond Sharing Weights (BSW) (Rozantsev et al., 2018), Conditional Domain Adversarial Network (CDAN) (Long et al., 2018), Maximum Classifier Discrepancy (MCD) (Saito et al., 2018), Margin Disparity Discrepancy (MDD) (Zhang et al., 2019), FixBi (Na et al., 2021), and Entropy Minimisation versus Diversity Maximisation (MEDM) (Wu et al., 2021). In this experiment, the RSNA and Child X-ray datasets are used as the source and target datasets, respectively. We evaluate the performance under three different backbones, i.e., ResNet50 (He et al., 2016), DenseNet121 (Huang et al., 2017) and EfficientNet-B4 (Tan and Le, 2019).

In the first experiment, we randomly split the original training set of the Child X-ray dataset (Kermany et al., 2018) into a training set and a validation set with the ratio of 80% : 20% for the model training. Table 2 reports the statistical results of the method with five runs, from which we can find that:

- Even though the target dataset is unlabelled, most of the peer methods and CDACM can achieve an AUC score higher than 85%. These results show the promising performance of unsupervised domain adaptation for pneumonia diagnosis.
- The adversarial learning-based methods (i.e., DANN, CDAN, and CDACM) are more effective in transferring knowledge than directly minimising the Margin Disparity Discrepancy of the two domains in Saito et al. (2018). It indicates the potential effectiveness of adversarial learning for domain adaptation. However, CyCADA is inferior to other methods by a large margin. The potential reason is that the images between the two domains are challenging to be aligned with the cycle-consistency constraint.
- CDACM outperforms all peer methods. Specifically, it improves the AUC score of the second-best performing algorithms by 2.39%, 1.05%, and 1.12% for the backbones of ResNet50, DenseNet121, and EfficientNet-B4, respectively. CDACM can achieve a mean AUC score of 90.57% with a standard deviation of 3.19%. These results indicate that our proposed domain adaptation method is effective for binary automated pneumonia diagnosis. Wilcoxon's rank sum test results at the 5% level show that our method can significantly outperform all other methods.

To evaluate the algorithm's performance under the scenario where the target domain has a limited amount of data, in the second experiment, we randomly sample 10% of the data in the training set and the validation set of the first experiment for model training. Table 3 reports the statistical results of the method with five runs, from which we can have:

- All the methods suffer performance degradation, especially for most adversarial learning-based methods. These results show that a large amount of data is essential for the high performance of deep learning methods for pneumonia diagnosis.
- The adversarial learning-based methods of DANN and CDAN have a high value of standard deviation. It indicates that the performance of adversarial learning-based methods can be unstable when the training data is limited.
- The selection of the backbone has a high impact on the performance of the trained model. Different methods may have different optimal backbones for the pneumonia diagnosis. For instance, DANN and FixBi achieve their highest AUC scores when using DenseNet121 as the backbone. MDD and CDACM perform the best when they use the backbone of EfficientNet-B4 and ResNet50, respectively.
- CDACM can still achieve a mean AUC score of 86.89%, 85.04%, and 87.35% when using the backbone of EfficientNet-B4, DenseNet121, and ResNet50, respectively. These results indicate that our proposed domain adaptation method is effective in handling the

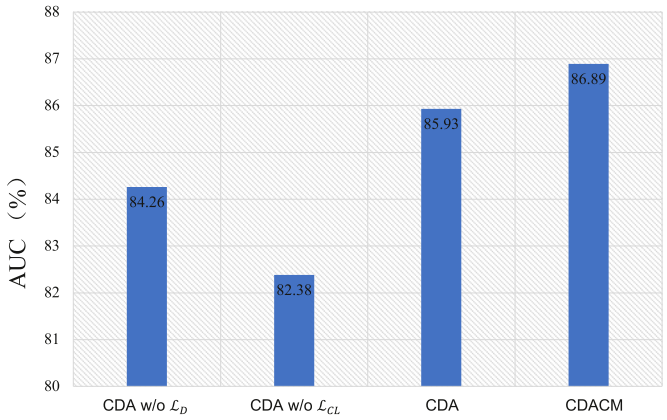


Fig. 4. The comparison among our proposed method and its three variants.

data deficiency problem in the target domain for automated pneumonia diagnosis.

4.4. Impact of different components

To investigate the effectiveness of Phase II, we test a variant of CDACM (denoted as CDA), which only contains CDACM's Phase I. Furthermore, to evaluate the impact of the two terms $\mathcal{L}_D(\theta_f, \theta_d)$ and $\mathcal{L}_{CL}(\theta_f)$ in Eq. (1), we test two variants of CDA: CDA without $\mathcal{L}_D(\theta_f, \theta_d)$, and CDA without $\mathcal{L}_{CL}(\theta_f)$. In this experiment, we use the EfficientNet-B4 as the backbone and the training data in the second experiment to train the three variants of CDACM and itself. The mean AUC score of 5 runs is shown in Fig. 4. Comparing the AUC scores of the full CDACM and CDA, we find that Phase II can improve the AUC score by 0.96%. By comparing CDA with its two variants, we can find that CDA achieves the highest AUC score, which indicates that both $\mathcal{L}_D(\theta_f, \theta_d)$ and $\mathcal{L}_{CL}(\theta_f)$ in Eq. (1) contribute to the final performance. In summary, these results demonstrate that all our designed components are important to deliver the final high classification score in automated pneumonia diagnosis.

4.5. Visualisation of representations from different domains

To visually understand the effectiveness of our proposed method on feature representation learning, we map the data representations of images that come from different domains in the high-level feature space into a 2D plane by applying the t-SNE method (Maaten and Hinton, 2008). In this experiment, we randomly sample 400 samples for each class in the test sets of the target domain and the source domain, respectively. Fig. 5 plots the results of the representations of samples from the source domain and the target domain. Fig. 5(a) displays the visualisation results of the data representations obtained by using a variant of CDACM without domain adaptation; Fig. 5(b) shows the visualisation results of the data representations obtained by using CDACM.

From Fig. 5(a), we see that the data representations of the two domains are almost located in two different groups with a considerable distance. It indicates that the two domains have a large gap from the feature representation perspective. However, the representations of the samples from different categories are mixed. Therefore, a precise classification boundary does not exist that can distinguish different categories accurately for both the source domain and the target domain.

Differently, in Fig. 5(b), the positive pneumonia samples from the source domain (denoted with green squares) are well mixed with the positive pneumonia samples from the target domain (denoted with blue stars). Similarly, the non-pneumonia samples from the two domains have overlapped. It means that the domain gap is well mitigated.

Table 4
Performance comparison of CDACM with the peer methods on the TTSH dataset. AUC_C is the AUC score (%) for COVID vs. Pneumonia+Others, AUC_P is the AUC score for Pneumonia vs. COVID+Others, AUC_O is the AUC score for Others vs. COVID+Pneumonia, and AUC_W denotes the $weighted\ AUC$ score for all classes.

Method	AUC_C	AUC_P	AUC_O	AUC_W
DANN (Ganin et al., 2016)	78.23	74.50	87.22	84.73
CyCADA (Hoffman et al., 2018)	68.38	76.55	69.52	52.32
BSW (Rozantsev et al., 2018)	65.07	58.47	76.99	73.50
CDAN (Long et al., 2018)	72.85	**89.72**	86.21	85.06
MCD (Saito et al., 2018)	68.94	81.86	80.18	79.07
MDD (Zhang et al., 2019)	76.57	85.47	85.46	84.43
FixBi (Na et al., 2021)	69.69	50.55	71.99	73.55
MEDM (Wu et al., 2021)	70.59	85.89	80.30	83.48
CDACM (ours)	**82.65**	86.27	**89.41**	**88.27**

Furthermore, pneumonia and non-pneumonia samples from the target domain are much easier to classify than in Fig. 5(a). Lastly, we can see that the accurate classifiers for the source and target domains may differ. These results show that CDACM can accurately mitigate the gap between the two domains and learn discriminative features to distinguish positive and non-positive pneumonia cases.

4.6. Multi-class classification

To evaluate the effectiveness of CDACM on multi-class classification, we test it for the COVID-19 and pneumonia diagnosis tasks. We use the RSNA+COVID and TTSH datasets as the source and target datasets, respectively. In addition, we adopt EfficientNet-B4 as the backbone.

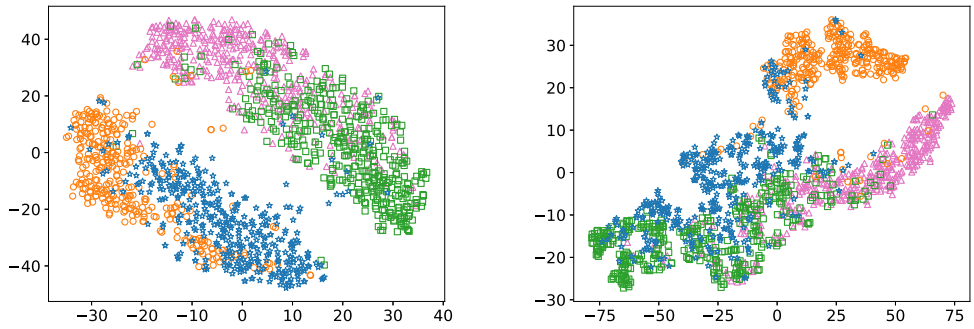
Table 4 reports the results of CDACM and its peer methods, from which we have the following observations:

- The methods of DANN, CDAN, MDD, and CDACM have a weighted AUC score higher than 84%. It demonstrates the power of unsupervised domain adaptation for pneumonia and COVID-19 screening.
- CDACM achieves a higher score than other methods in terms of AUC. For instance, CDACM outperforms CDAN, DANN, and MDD with a margin of 3.21%, 3.54%, and 3.84% in terms of the weighted AUC score for all classes.
- Even CDAN has higher accuracy (3.45%) than CDACM on AUC for pneumonia vs COVID+others, it has a much lower accuracy (10.2%) than CDACM for COVID vs pneumonia+others. It means that our method has a much higher capability to detect COVID from chest X-rays.

4.7. Comparison with junior radiologists

To investigate the effectiveness of CDACM by comparing it with three junior radiologists (JRs). Three JRs (with 6-month, 1-year, and 2-year experience in CXRs interpretation, respectively) perform on CXRs interpretation (in three classes: COVID-19 pneumonia, other pneumonia, and non-pneumonia) of the external test set. All three JRs were blinded to any clinical information and had recent experiences with COVID-19 CXR interpretation. The results are reported in Table 5, from which we can see:

- CDACM outperforms all of the three JRs for pneumonia and COVID diagnosis in terms of $Sensitivity$, $Specificity$, and AUC scores. Specifically, CDACM outperforms the three JRs with 6-month, 1-year, and 2-year experience in CXRs interpretation with a margin of 12.68%, 24.06%, and 7.56% respectively, in terms of the AUC score for pneumonia diagnosis.
- Three JRs have obtained higher AUC scores than our method for the category of others. It indicates that if the JRs can check the predicted results from our method will significantly reduce the false alarm.



(a) Without domain adaptation (b) CDACM (full version)

Fig. 5. Visualisation of data representations using the t-SNE method. The pink triangle and green square denote the non-pneumonia samples and the positive pneumonia samples in the source domain, respectively. The orange circle and the blue star represent the non-pneumonia cases and the positive pneumonia cases in the target domain. (a) The feature representations are obtained without domain adaptation. (b) The feature representations obtained by CDACM.

Table 5
Performance comparison of CDACM with three JRs on the external test set of the TTSH dataset. The *AUC* score (%) is the value for the corresponding class vs all other classes.

		Sensitivity	*Specificity*	*AUC*
JR1 ~6 months	COVID	38.89	91.59	65.24
	Pneu	67.35	73.17	72.06
	Others	71.50	99.01	81.21
JR2 ~1 year	COVID	50.00	91.59	70.79
	Pneu	55.10	82.26	68.68
	Others	81.53	90.08	85.81
JR3 > 2 years	COVID	62.50	91.12	76.81
	Pneu	61.22	89.14	75.18
	Others	86.81	92.56	89.68
CDACM (ours)	COVID	64.49	80.56	77.24
	Pneu	72.06	75.51	82.74
	Others	63.64	85.22	77.82

5. Conclusion

In this paper, by considering domain adaptation and data deficiency problems of the target domain simultaneously, we proposed a novel method called CDACM. It optimises the domain adaptation loss and contrastive loss in a jointly learning manner in Phase I. Specifically, CDACM conquers the data efficiency problem of the target domain by constructing a feature cloud for each target sample using contrastive learning. Furthermore, the gap between the source and target domains is mitigated by the knowledge transfer between the large-scale source samples and the generated feature clouds. To further improve the model's generalisation ability, we designed the adaptive feature cloud expansion module to push the classifier's boundary into a low-density area in Phase II. Specifically, we generate new samples to expand the feature cloud if there is a consistency of high prediction confidence on existing samples in the feature cloud. Also, the predictions of newly generated samples are enforced to match the pseudo-label of the feature cloud, which is derived from a consistency of high confidence scores on one of the classes. The experimental results show that CDACM can achieve a mean AUC score of 90.57% for child pneumonia diagnosis and a weighted AUC score of 88.27% for COVID-19 and pneumonia diagnosis. These results indicate that our proposed CDACM can transfer the knowledge effectively to improve automated pneumonia and COVID diagnosis and screening.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the Agency for Science, Technology and Research (A*STAR) through its AME Programmatic Funding Scheme Under Project A20H4b0141. Joey Tianyi Zhou is supported by A*STAR SERC Central Research Fund (Use-inspired Basic Research) and the Singapore Government's Research, Innovation and Enterprise 2020 Plan (Advanced Manufacturing and Engineering domain) under Grant A18A1b0045.

References

Chadha, A., Andreopoulos, Y., 2019. Improved techniques for adversarial discriminative domain adaptation. IEEE Trans. Image Process. 29, 2622–2637.

Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.-A., 2019. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33 no.01. pp. 865–872.

Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A., 2020. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Trans. Med. Imaging 39 (7), 2494–2505.

Cicek, S., Soatto, S., 2019. Unsupervised domain adaptation via regularized conditional alignment. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1416–1425.

Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V., 2019. Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 113–123.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255.

Dou, Q., Ouyang, C., Chen, C., Chen, H., Heng, P.-A., 2018. Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. In: Proceedings of the International Joint Conference on Artificial Intelligence. pp. 691–697.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. J. Mach. Learn. Res. 17 (1), 2030–2096.

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J., 2019. CE-Net: Context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging 38 (10), 2281–2292.

Gupta, A., Ayhan, M., Maida, A., 2013. Natural image bases to represent neuroimaging data. In: Proceedings of the International Conference on Machine Learning. PMLR, pp. 987–994.

Haghanifar, A., Majdabadi, M.M., Choi, Y., Deivalakshmi, S., Ko, S., 2022. COVID-CXNet: Detecting COVID-19 in frontal chest X-ray images using deep learning. Multimedia Tools Appl. 1–31.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A., Darrell, T., 2018. CyCADA: Cycle consistent adversarial domain adaptation. In: Proceedings of the International Conference on Machine Learning.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708.

Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al., 2017. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Proceedings of the International Conference on Information Processing in Medical Imaging. Springer, pp. 597–609.

Kang, G., Jiang, L., Yang, Y., Hauptmann, A.G., 2019. Contrastive adaptation network for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4893–4902.

Ke, A., Ellsworth, W., Banerjee, O., Ng, A.Y., Rajpurkar, P., 2021. CheXtransfer: Performance and parameter efficiency of ImageNet models for chest X-Ray interpretation. In: Proceedings of the Conference on Health, Inference, and Learning. pp. 116–124.

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

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. CoRR abs/1412.6980URL https://arxiv.org/abs/1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Proc. Adv. Neural Inf. Process. Syst. 25, 1097–1105.

Lakhani, P., Sundaram, B., 2017. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 284 (2), 574–582.

Liu, P., Kong, B., Li, Z., Zhang, S., Fang, R., 2019. CFEA: Collaborative feature ensembling adaptation for domain adaptation in unsupervised optic disc and cup segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 521–529.

Long, M., Cao, Y., Wang, J., Jordan, M.I., 2015. Learning transferable features with deep adaptation networks. In: Proceedings of the International Conference on Machine Learning. pp. 97–105.

Long, M., Cao, Z., Wang, J., Jordan, M.I., 2018. Conditional adversarial domain adaptation. Proceedings of the Advances in Neural Information Processing Systems 31.

Long, M., Zhu, H., Wang, J., Jordan, M.I., 2017. Deep transfer learning with joint adaptation networks. In: Proceedings of the International Conference on Machine Learning. pp. 2208–2217.

Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. 29 (2), 102–127.

Luo, Y., Liu, P., Guan, T., Yu, J., Yang, Y., 2020a. Adversarial style mining for one-shot unsupervised domain adaptation. Proc. Adv. Neural Inf. Process. Syst. 33, 20612–20623.

Luo, L., Yu, L., Chen, H., Liu, Q., Wang, X., Xu, J., Heng, P.-A., 2020b. Deep mining external imperfect data for chest X-ray disease screening. IEEE Trans. Med. Imaging 39 (11), 3583–3594.

Maaten, L.v.d., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (Nov), 2579–2605.

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., Soufi, G.J., 2020. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. Med. Image Anal. 65, 101794.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. CoRR abs/1411.1784URL https://arxiv.org/abs/1411.1784.

Na, J., Jung, H., Chang, H.J., Hwang, W., 2021. FixBi: Bridging domain spaces for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1094–1103.

Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., Rueckert, D., 2019. Data efficient unsupervised domain adaptation for cross-modality image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 669–677.

Patel, V.M., Gopalan, R., Li, R., Chellappa, R., 2015. Visual domain adaptation: A survey of recent advances. IEEE Signal Process. Mag. 32 (3), 53–69.

Radiological Society of North America, 2018. RSNA pneumonia detection challenge. URL https://www.kaggle.com/c/rsna-pneumonia-detection-challenge.

Redko, I., Morvant, E., Habrard, A., Sebban, M., Bennani, Y., 2019. Advances in Domain Adaptation Theory. Elsevier.

Rozantsev, A., Salzmann, M., Fua, P., 2018. Beyond sharing weights for deep domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell. 41 (4), 801–814.

Saenko, K., Kulis, B., Fritz, M., Darrell, T., 2010. Adapting visual category models to new domains. In: Proceedings of the European Conference on Computer Vision. pp. 213–226.

Saito, K., Ushiku, Y., Harada, T., Saenko, K., 2019. Strong-weak distribution alignment for adaptive object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6956–6965.

Saito, K., Watanabe, K., Ushiku, Y., Harada, T., 2018. Maximum classifier discrepancy for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3723–3732.

Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D., 2019. Attention gated networks: Learning to leverage salient regions in medical images. Med. Image Anal. 53, 197–207.

Schroff, F., Kalenichenko, D., Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 815–823.

Shin, H.-C., Lu, L., Kim, L., Seff, A., Yao, J., Summers, R.M., 2016a. Interleaved text/image deep mining on a large-scale radiology database for automated image interpretation. J. Mach. Learn. Res. 17 (107), 1–31.

Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016b. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35 (5), 1285–1298.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.-L., 2020. FixMatch: Simplifying semi-supervised learning with consistency and confidence. Proc. Adv. Neural Inf. Process. Syst. 33.

Sun, B., Feng, J., Saenko, K., 2016. Return of frustratingly easy domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30 no. 1.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X., 2020. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Med. Image Anal. 63, 101693.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the International Conference on Machine Learning. PMLR, pp. 6105–6114.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7167–7176.

Varsavsky, T., Orbes-Arteaga, M., Sudre, C.H., Graham, M.S., Nachev, P., Cardoso, M.J., 2020. Test-time unsupervised domain adaptation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 428–436.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M., 2017. ChestX-ray8: Hospital-scale chest X-Ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

World Health Organization, 2019. Pneumonia. https://www.who.int/news-room/fact-sheets/detail/pneumonia.

Wu, X., Zhang, S., Zhou, Q., Yang, Z., Zhao, C., Latecki, L.J., 2021. Entropy minimization versus diversity maximization for domain adaptation. IEEE Trans. Neural Netw. Learn. Syst..

Yi, X., Walia, E., Babyn, P., 2019. Generative adversarial network in medical imaging: A review. Med. Image Anal. 101552.

Zhang, Y., Liu, T., Long, M., Jordan, M., 2019. Bridging theory and algorithm for domain adaptation. In: Proceedings of the International Conference on Machine Learning. pp. 7404–7413.

Zhang, Y., Miao, S., Mansi, T., Liao, R., 2018. Task driven generative modeling for unsupervised domain adaptation: Application to X-ray image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 599–607.

Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., et al., 2020. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Trans. Med. Imaging 39 (7), 2531–2540.

Zhou, J., Jing, B., Wang, Z., Xin, H., Tong, H., 2020. SODA: Detecting COVID-19 in chest X-rays with semi-supervised open set domain adaptation. IEEE/ACM Trans. Comput. Biol. Bioinform. to be published.

Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J., 2019. Confidence regularized self-training. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5982–5991.