Neurocomputing 637 (2025) 130089

Contents lists available at ScienceDirect

Neurocomputing

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

Artificial intelligence-powered tuberculosis detection with complementary domain attention model☆

Zeyu Ding [a,c], Razali Yaakob [a,*], Azreen Azman [a], Siti Nurulain Mohd Rum [a], Norfadhlina Zakaria [b], Azree Shahril Ahmad Nazri [a]

[a] Universiti Putra Malaysia, Faculty of Computer Science and Information Technology, Selangor, 43400, Malaysia
[b] Universiti Putra Malaysia, Faculty of Medicine and Health Sciences, Selangor, 43400, Malaysia
[c] Changzhi University, Changzhi, 046000, Shanxi, China

ARTICLE INFO

Communicated by Y. Liu

Keywords:
Tuberculosis
TB
Deep learning
Domain adaptation

ABSTRACT

Artificial intelligence-based X-ray image detection can significantly aid early tuberculosis (TB) detection. However, the varying distribution of X-ray image data across different hospitals has resulted in a decline in the model's performance when transitioning to a new dataset. Domain adaptation techniques can effectively mitigate the impact of this issue. However, current domain adaptation methods align the entire image features between the source and target domains without explicitly focusing on regions containing transferable classification information across domains. Forced alignment of features across the entire image may lead to negative transfer. This paper proposes a complementary domain attention model (CDAM) for TB detection where the feature map is partitioned into domain-shared (DSH) and domain-specific (DSP) features. DSP features are complementary to DSH features. DSH features are responsible for mitigating the impact of domain gaps on classification. Consequently, they focus on areas containing classification information that can be transferred across domains. In contrast, the role of the DSP feature is to maximize the domain gap, concentrating its attention on areas rich in domain information. Given that the DSH and DSP features are complementary, when the DSP feature occupies domain-informative areas, it simultaneously encourages the DSH feature to focus more accurately on areas containing transferable classification information across domains, thereby enhancing classification performance. The objective of CDAM is to fully consider the importance of different regions within the feature map and mitigate negative transfer. The proposed method underwent domain adaptation experiments on the Shenzhen, Montgomery, and TBX11K datasets, achieving average accuracy, sensitivity, and specificity scores of 73.3%, 74.0%, and 72.7%, respectively. This result surpasses existing domain adaptation methods for TB data, providing evidence for the effectiveness and robustness of the proposed approach.

1. Introduction

Tuberculosis (TB) is a prevalent and potentially fatal infectious disease. According to the World Health Organization (WHO) 2023 report, the number of deaths due to tuberculosis is nearly twice that of deaths caused by AIDS [1]. The early detection and treatment of TB are of paramount importance. With the advancement of computer vision technology, artificial intelligence-assisted detection has proven highly beneficial for the early identification of TB, gaining favor among radiologists. However, due to the distinctive distribution of metadata within medical imaging data (such as the model and settings of the scanning device, the patient's geographical location, age, and gender), some algorithms exhibit significant performance degradation when generalized to new datasets (target datasets), even if they perform well on the source dataset. To address this issue, researchers have introduced the concept of domain adaptation, which involves mitigating the distributional differences between source and target domain data. This process facilitates the transfer of a model trained on the source domain to the target domain, thereby enhancing the model's performance in the target domain.

Domain adaptation has played a crucial role in various domains such as cancer diagnosis [2,3], pneumonia detection [4,5], medical

☆ This work was supported by the Ministry of Higher Education under the Fundamentals Research Grant Scheme (FRGS/1/2020/ICT02/UPM/02/5).
* Corresponding author.
E-mail addresses: gs63736@student.upm.edu.my (Z. Ding), razaliy@upm.edu.my (R. Yaakob), azreenazman@upm.edu.my (A. Azman), snurulain@upm.edu.my (S.N.M. Rum), n_fadhlina@upm.edu.my (N. Zakaria), azree@upm.edu.my (A.S.A. Nazri).

https://doi.org/10.1016/j.neucom.2025.130089
Received 6 June 2024; Received in revised form 3 March 2025; Accepted 19 March 2025
Available online 28 March 2025

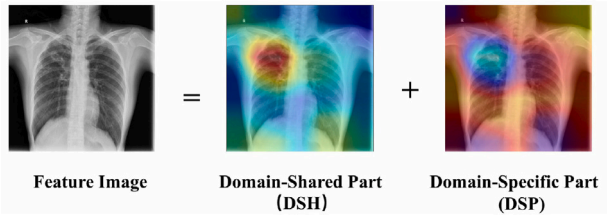

Fig. 1. The feature image can be divided into domain-shared (DSH) and domain-specific (DSP) parts, where warmer colors indicate higher weights. The DSH and DSP features are complementary.

image segmentation [6] and other fields. However, these studies align overall features without considering the varying importance of features in different regions. Aligning features forcefully may lead to negative transfer. Fig. 1 shows the feature image with domain-shared (DSH) and domain-specific (DSP) parts. In medical images, DSH features mainly focus on regions containing classification information, which can be transferred across domains. On the other hand, DSP features focus on less transferable regions across domains. If all features are aligned without considering the distinction between DSH and DSP features, misalignment may occur, leading to negative transfer.

To address this issue, inspired by research [7], this study introduces a complementary domain attention model (CDAM) for AI-based tuberculosis screening. The CDAM model divides the feature maps into DSH and DSP features, and these two features are complementary. DSH features are responsible for TB classification and alleviating the impact of domain divergence to ensure effective classification. As a result, these features primarily concentrate on regions containing classification information that is transferable across domains. Meanwhile, DSP features are used to maximize the domain discrepancy, leading them to focus on regions with richer domain information. As DSH features and DSP features complement each other, with DSP focusing on regions with richer domain information, it enables DSH features to target the regions transferable across domains more accurately, thereby enhancing the domain adaptation effect. Furthermore, this study proposes a novel complementary feature generation method based on a self-attention mechanism. Moreover, the CDAM model outperforms other domain adaptation techniques on the public TB dataset.

2. Related works

Unsupervised domain adaptation (DA) is a specific area of transfer learning that aims to transfer knowledge learned from a source domain to an unlabeled target domain. There are numerous methods for assessing the data distribution between the source domain and the target domain, such as MMD [8], CORAL [9], H-divergence [10], cosine similarity [11], and higher-order moments of matrices [12]. However, these methods are all manually designed features. With the development of generative adversarial networks [13], the concept of adversarial learning has successfully integrated into domain adaptation, such as DANN [14], ADDA [15] and these methods achieve better results compared to manually designed features. The most similar study to this paper is the DIFL [16], which successfully applied adversarial strategies to develop an unsupervised domain adaptation model. It effectively enhanced the generalization ability of TB screening and achieved favorable TB screening results on TB datasets from three different countries. UESM [17] achieved satisfactory results in anatomical structure segmentation by employing self-training and adversarial training schemes. It utilized uncertainty information to enhance the segmentation performance in highly uncertain regions. Some researchers address domain adaptation issues using cycle generative adversarial networks (CycleGAN) [18] methods. CX-DaGAN [19] employs CycleGAN to transfer the style of target domain images to that of source domain images, thereby reducing domain differences. However, there is a risk that the lesion information may be altered during the conversion process. SPA-UDA [20] added feature and label consistency loss to Cycle-GAN to preserve lesion information as much as possible during image translation. This approach achieved good results in domain adaptation tasks for lung disease detection. In addition, there are other approaches that tackle domain adaptation from multiple perspectives. Study [21] designed a feature disentanglement model that separates features into content and style, thus minimizing the impact of domain changes. Lu et al. [22] proposed a low-rank correlation learning (LRCL) method that can prevent noise from interfering with domain adaptation. Research [23] uses a graph-based approach to identify meaningful foreground regions in images from different domains, enabling a better understanding of medical images across domains. The approach proposed in [24] addresses the issue of the class center shift in the target domain during domain adaptation, thereby improving performance. Study [25] introduces a guided discrimination and correlation subspace learning (GDCSL) approach that learns domain-invariant, discriminative, and correlated features to reduce domain shift and enhance cross-domain classification performance. The work in [6] transfers multi-label classification tasks from the source domain to binary classification tasks in the target domain, thereby transferring knowledge from a large source dataset to a smaller target dataset, enhancing pneumonia detection performance. However, the methods mentioned above in domain adaptation focus on the entire image without addressing task-specific regions (such as the TB information region). These approaches may forcefully align mismatched features, resulting in negative transfer. In order to better utilize the importance of different regions in the feature map and reduce the occurrence of negative transfer, attention mechanisms in domain adaptation are proposed.

Attention mechanisms have made significant progress in computer vision, giving birth to many classical network models such as SENET [26], CBAM [27], Self-attention [28] and DANet [29]. Self-attention was first introduced in research [28] and quickly demonstrated excellent performance in computer vision and natural language processing, becoming one of the most widely used attention mechanisms. Compared to traditional convolutional neural networks, self-attention has a larger field of view by calculating the relevance between different regions, enabling better selection of crucial parts from a global perspective. Due to the outstanding performance of self-attention mechanisms, it has been successfully applied to domain adaptation. Study [30] developed a self-attention-based cross-modal segmentation method that effectively propagates attention information across domains. This method has shown strong performance in heart, abdominal organ, and brain tumor segmentation tasks. Research [31] incorporates self-attention mechanisms into domain adaptation tasks and introduces deformable large kernel attention (D-LKA Attention). This approach enhances contextual understanding of images while simultaneously mitigating computational overhead and reducing overall complexity. Some studies use attention mechanisms to select features that are more easily transferred between domains. For instance, the approach in [32] suggests that not every region in an image has the same transferability, and it leverages the entropy properties to select locally transferable features, achieving good generalization effects. Study [33] also proposed a novel approach to construct transferable attention maps by allocating corresponding weights based on the transfer capabilities of different local regions. This methodology was successfully applied to the task of pneumonia detection in X-ray images. The method in [34] focuses on selecting more transferable channel features by adjusting the weights of feature channels and applying them to unsupervised cross-domain fault diagnosis, achieving impressive results. Some studies have effectively utilized features that are not easily transferable across domains, making the domains more discriminative from each other. For example, in the research conducted by [7], focusing on the pedestrian re-identification challenge, the feature maps are divided into domain-shared and domain-specific parts. The domain-shared part focuses on

the pedestrian's features and is applied to classification, while the domain-specific part emphasizes background regions to avoid negative transfer due to domain divergence. This study drew inspiration from this approach and applied it to the diagnosis of TB.

3. Methods

Let $D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ represent the labeled source dataset and $D_t = \{x_i^t\}_{i=1}^{n_t}$ denote the unlabeled target dataset. x_i^s and x_i^t are the source and target dataset sample, and y_i^s is the label in the source dataset, which takes the value of 1 or 0, representing containing TB and normal, respectively. n_s and n_t represent the number of source and target dataset samples. In order to achieve better generalization performance on the target dataset, this study improved upon the foundation of DANN [14] to develop the complementary domain attention model (CDAM).

Fig. 2 shows the proposed CDAM model, which consists of five parts: feature extraction module, self-attention module, complementary module, domain-shared module, and domain-specific module. Firstly, the feature extraction module utilizes a convolutional neural network to generate a preliminary feature map of size $H * W$, where H and W denote the height and width, respectively. Subsequently, this feature map is flattened, containing $N = H * W$ local features. Next, in the self-attention module, the feature map undergoes a linear transformation to obtain three vectors: Q, K, and V. The product of Q and the transpose of K yields the DSH attention map, which reflects the correlations among the N local features. Then, multiplying the DSH attention map with V produces the DSH feature map. Meanwhile, in the complementary module, the DSP attention map complements the DSH attention map, with corresponding regions exhibiting inverse value trends. Multiplying the DSP attention map with V a generates the DSP feature map. Furthermore, the domain-shared module contains a TB classifier and a domain discriminator. Minimizing classification loss ensures accurate classification while maximizing domain discriminator loss mitigates the impact of domain differences. Consequently, DSH features can focus on regions containing transferable classification information across domains. Additionally, the domain-specific module shares the domain discriminator with the DSH module. Minimizing domain discriminator loss enhances domain differentiation, thus enabling the DSP attention map to concentrate on regions rich in domain information. Through complementary mechanisms, DSH features reduce attention on domain-rich regions, further diminishing the influence of domain differences and thus improving classification performance. The subsequent section will provide a detailed explanation of these modules.

3.1. Feature extraction module

The feature extraction module is mainly responsible for extracting the feature map. According to study [35], DenseNet121 [36] demonstrates excellent performance in extracting TB features from the lung region and achieving outstanding classification performance. This study adopts DenseNet121 as the feature extraction module, extracting the output $F \in R^{W*H*C}$, as the feature map. W, H, and C are the feature map's width, height, and number of channels. The layer before the classification layer of DenseNet121 is chosen as the feature map. The dimensions of the feature map are $7 \times 7 \times 1024$, representing width, height, and channel numbers, respectively.

3.2. Self attention module

After obtaining the feature map F, the role of the self-attention module is to generate a DSH feature F^{DSH} that can focus more on regions containing classification information and can be transferred across domains. In order to enrich this feature with abundant contextual information for improved classification accuracy, this paper draws inspiration from the study [7]. The shape of the feature map F is transformed from $W * H * C$ to $N * C$, where $N = W * H$. W and H represent width and height respectively, and N represents N local features in the feature map F. Then, F undergoes linear transformations (multiplied by three different trainable weight matrices) to obtain three vectors $Q, K, V \in R^{N*C}$. Q, K, V are obtained through linear transformations of F, thus containing the information of the feature map F. Next, we calculate the product of Q and the transpose of K and apply this product to the softmax to obtain the AM^{DSH}. It represents the correlation matrix of N local features, where $AM^{\mathrm{DSH}} \in R^{N*N}$, the equation of AM_{ij}^{DSH} is given by Eq. (1) [28].

$$AM_{ij}^{\mathrm{DSH}} = \frac{\exp\left(Q_i \cdot K_j^T\right)}{\Sigma_{j=1}^{N} \exp\left(Q_i \cdot K_j^T\right)}, \tag{1}$$

AM_{ij}^{DSH} represents the magnitude of the correlation between the ith local feature and the jth local feature. Therefore, the ith row of $AM_i^{\mathrm{DSH}} \in R^{1*N}$ describes the correlation between the ith local feature and all N local features in that row. The final step is to multiply AM^{DSH} by V to obtain F^{DSH}, where $AM^{\mathrm{DSH}} \in R^{N*N}$, as shown in Eq. (2).

$$F^{\mathrm{DSH}} = AM^{\mathrm{DSH}} * V. \tag{2}$$

Fig. 3 describes the process for generating F^{DSH}. The width and height of the feature map F are both 7, resulting in a total of 49 local features. Subsequently, these local features are flattened and undergo linear transformation to obtain the Q, K, and V vectors. The matrix obtained by multiplying Q with the transpose of K, denoted as AM^{DSH}, can reflect the similarity relationships between local features. The first row of AM^{DSH} is denoted as AM_1^{DSH}, representing the similarity coefficients between the first local feature F_1 and other local features $F_1, F_2, F_3, \ldots, F_{49}$. These coefficients are learnable. Then, after multiplying AM_1^{DSH} with V, the first new local feature F_1^{DSH} can be obtained, which is derived from the weighted sum of 49 local features. This allows F_1^{DSH} to integrate information from the global context, providing a broader perspective. The remaining new local features F_2^{DSH}, F_3^{DSH}, $\ldots$, F_{49}^{DSH} can be obtained through similar operations.

3.3. Complementary module

The role of the complementary module is to generate a feature F^{DSP} complementary to F^{DSH}. F^{DSP} focuses on regions rich in domain information. AM^{DSP} also has a complementary relationship with AM^{DSH}. In areas where AM^{DSH} has high weights, AM^{DSP} weights should be low, and in areas where AM^{DSH} has low weights, AM^{DSP} weights should be high. Moreover, the sum of the weight for each row in AM^{DSP} should be 1. The value of AM_{ij}^{DSH} is between 0 and 1. By leveraging the properties of the logarithmic function, larger values of AM_{ij}^{DSH} can be transformed into smaller values of $-\log(AM_{ij}^{\mathrm{DSH}})$, and smaller values of AM_{ij}^{DSH} can be transformed into larger values of $-\log(AM_{ij}^{\mathrm{DSH}})$. Subsequently, use softmax to ensure that the sum of the weight for each row of AM^{DSP} equals 1. Therefore, AM_{ij}^{DSP} can be obtained from Eq. (3).

$$AM_{ij}^{\mathrm{DSP}} = \frac{exp(-\log(AM_{ij}^{\mathrm{DSH}}))}{\sum_{j=1}^{N} exp(-\log(AM_{ij}^{\mathrm{DSH}}))}, \tag{3}$$

where i and j represent the indices of rows and columns in the matrices AM^{DSH} and AM^{DSP}. F^{DSP} is obtained through Eq. (4).

$$F^{\mathrm{DSP}} = AM^{\mathrm{DSP}} * V. \tag{4}$$

3.4. Domain-shared module

The DSH module aims to minimize the difference between features from the source and target domains while ensuring classification performance, so DSH features can focus on regions that contain classification information and can be transferred across domains. As shown in Fig. 2,

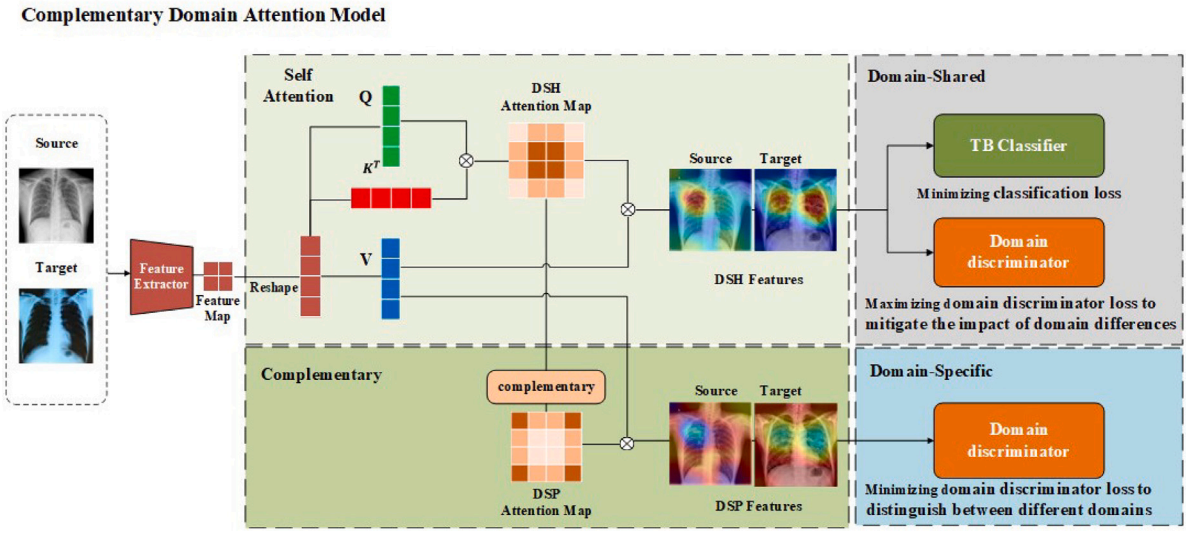

Fig. 2. The overall architecture of the complementary domain attention model (CDAM).

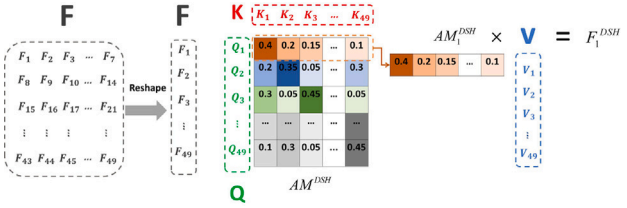

Fig. 3. An example of using the self-attention mechanism to generate a domain-shared features.

the DSH module takes F^{DSH} as input, and consists of a classifier G_y and a domain discriminator G_d, with parameters θ_y and θ_d, respectively. G_y is responsible for classifying images as TB or normal, while G_d identifies which domain the image belongs to. Both G_y and G_d are binary classifiers, hence their network structures are identical. The inputs for G_y and G_d are first processed by a linear layer to reduce the dimensionality from 1024 to 256, followed by ReLU activation and dropout regularization. Subsequently, another linear layer is employed to further reduce the dimensionality to 2, thereby fulfilling the binary classification task. The loss functions for the classifier and domain discriminator are denoted as L_y and L_d, both of which are cross-entropy loss functions. To correctly classify images, it is necessary to minimize the TB classifier G_y's loss. In order to make the domain discriminator G_d incapable of discerning whether the data comes from the source domain or the target domain, it is necessary to maximize the discriminator's loss. Hence, the training process should be executed following the outlined steps, which need to be repeated for a certain number of epochs.

1. **Train the domain discriminator G_d first, and minimize the domain discriminator loss to enable it to distinguish which domain the data comes from.:**
   Input F^{DSH} and the domain labels d_i into G_d, minimizing the loss of G_d, while fixing θ_f and updating θ_d.
2. **Train the TB classifier G_y while maximizing the loss of domain discriminator G_d to make it cannot distinguish which domain the data comes from:**
   Input F^{DSH} and category labels y_i into G_y,minimizing the TB classification loss; Input F^{DSH} and domain labels d_i into G_d, maximizing the loss of domain discriminator G_d. Fix θ_d and update θ_f, θ_y based on Eq. (5) [14].

$$\mathrm{E}\left(\theta_f,\theta_d^{\mathrm{DSH}},\theta_y\right)=\frac{1}{n_s}\sum_{F_i^{\mathrm{DSH}}\in D_s}L_y\left(G_y\left(F_i^{\mathrm{DSH}}\right),y_i\right)-\frac{\lambda_1}{n}\sum_{F_i^{\mathrm{DSH}}\in D_s\cup D_t}L_d\left(G_d\left(F_i^{\mathrm{DSH}}\right),d_i\right),\tag{5}$$

$n=n_s+n_t$, where n_s and n_t represent the number of source and target dataset samples, and n represents the sum of the two quantities. y_i and d_i are the class label and domain label corresponding to x_i. λ_1 is a hyperparameter used to adjust and balance the components. The integration of the feature extraction module and the self-attention module can be considered as a novel feature extractor, with its parameters denoted as θ_f.

3.5. Domain-specific module

The DSP module, with F^{DSP} as input, aims to discriminate between data from the source and target domains effectively, so DSP features can focus on regions rich in domain information. With the aid of complementary mechanisms, DSH features are able to reduce their focus on domain-rich regions, thereby enhancing classification performance. The DSP module and the DSH module share the same domain discriminator G_d with common parameters θ_d. G_d takes F^{DSP} and the domain labels d_i as inputs, minimizing the loss of G_d, and simultaneously updating θ_f. The overall loss function requires the joint consideration of the classification loss, domain discriminator loss (in DSH module), and domain discriminator loss (in DSP module), as shown in Eq. (6).

$$\mathrm{E}\left(\theta_f,\theta_d^{\mathrm{DSH}},\theta_y\right)=\frac{1}{n_s}\sum_{F_i^{\mathrm{DSH}}\in D_s}L_y\left(G_y\left(F_i^{\mathrm{DSH}}\right),y_i\right)-\frac{\lambda_1}{n}\sum_{F_i^{\mathrm{DSH}}\in D_s\cup D_t}L_d\left(G_d\left(F_i^{\mathrm{DSH}}\right),d_i\right)+\frac{\lambda_2}{n}\sum_{F_i^{\mathrm{DSP}}\in D_s\cup D_t}L_d\left(G_d\left(F_i^{\mathrm{DSP}}\right),d_i\right),\tag{6}$$

λ_2 is a hyperparameter used to adjust and balance the components.

4. Experiment

There are three datasets for tuberculosis binary classification: Shenzhen [37], Montgomery [37], and TBX11K datasets [38]. The Shenzhen dataset was collected and labeled at Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen, China, comprising 662 chest

Table 1
The distribution of common diseases in the CheXpert and Chest X-ray14 datasets.

Disease	CheXpert	ChestX-ray14
Atelectasis	33,456	11,559
Consolidation	14,716	4,667
Pneumothorax	19,456	5,302
Edema	52,291	2303
Pneumonia	6047	1431
Cardiomegaly	27,068	2776
Total	116,719	24,199

X-rays (CXR), with 336 containing TB and the remaining 326 being normal CXRs. The Montgomery dataset, collected and annotated by the U.S. Department of Health and Human Services in Montgomery County, consists of 138 X-ray images, with 80 normal cases and the remaining 58 TB cases. The TBX11K dataset consists of 11,200 CXRs, divided into training, validation, and test sets. The test set does not include annotations, so only the training and test sets are used in this study. A total of 630 CXRs with active TB and 630 healthy CXRs were selected for this research. This paper split the three datasets into training and test sets in an 80% to 20% ratio.

To validate the performance of CDAM in handling multi-classification tasks, this study selected two large-scale chest X-ray datasets, namely CheXpert[1] [39] and Chest X-ray14 [2][40]. Since the disease categories in these two datasets are not entirely consistent, only six common diseases were selected, which are atelectasis, consolidation, pneumothorax, edema, pneumonia, and cardiomegaly. The distribution of each disease in this dataset is shown in Table 1. Since each X-ray may contain multiple diseases, the sum of samples for each disease may exceed the total number of images in the dataset. The performance is evaluated using five-fold cross-validation.

The experiments in this paper were conducted on a computer equipped with an Intel(R) Core(TM) i7-7700K CPU, 16 GB RAM, and an NVIDIA GeForce RTX 3060 12 GB GPU. The operating system used is Ubuntu 20.04.3 LTS, with PyTorch version 1.12.1 and Python version 3.9. Both G_y and G_d are composed of fully connected networks. The training was stopped at the 25th epoch as there was no significant decrease in the loss value. The learning rate for θ_f is set to 5e−6, while the learning rates for θ_y and θ_d are set to 2e−5, with a halving of the learning rates every five epochs. After extensive experimentation, the optimal values for λ_1 and λ_2 were set to 0.5 and 0.2, respectively.

5. Results and analysis

For the binary classification experiment on TB, this paper reproduces several mainstream domain adaptation methods on the Shenzhen, Montgomery, and TBX11K datasets, including MMD [8], DANN [14], DIFL [16], CX-DaGAN [19], SPA-UDA [20], and AGPDA [33]. The results of CDAM and these methods are shown in Table 2. Here, 'S' stands for the Shenzhen dataset, 'M' denotes the Montgomery dataset, and 'T' signifies the TBX11K dataset. The symbol "→" denotes the transfer of knowledge acquired from a source domain to an unlabeled target domain. For example, S→M denotes transferring knowledge learned from the Shenzhen dataset to the unlabeled Montgomery dataset. The metrics used to evaluate domain adaptation effectiveness include accuracy (Acc), sensitivity (Sen), and specificity (Spe), with all numbers in the table representing percentages. CI stands for a 95% Confidence Interval, which is a range used in statistics to estimate a population parameter. It means that if repeated samples were taken, 95% of the confidence intervals would contain the true value of the population parameter. It reflects the reliability of our estimate. To better demonstrate the performance improvement of each domain adaptation method, the "No DA" and "Up" groups in Table 2 represent the lower and upper bounds of domain adaptation, respectively. "No DA" represents training on the source dataset and testing on the target dataset. And "Up" represents training on and testing on the target dataset. The performance of "No DA" group is the lowest among all methods, indicating a significant performance drop due to domain divergence. In contrast, the results of the "Up" group are the highest across all methods, as its training and testing datasets are consistent, representing the upper limit of model performance.

From the results of the domain adaptation experiment, it can be observed that: (1) All domain adaptation methods have shown improvements in accuracy, sensitivity, and specificity compared to the 'No DA' group, indicating that these methods can effectively utilize information from the source domain data, making it easier for the model to adapt to target domain data and enhancing data transferability. (2) Adversarial-based domain adaptation methods (DANN and DIFL) demonstrate more significant performance improvements than Cycle-GAN-based methods (CX-DaGAN and SPA-UDA) and MMD. AGPDA, which incorporates an attention mechanism into the original MMD, also shows substantial performance improvement over the original MMD approach. The CDAM method achieves the best performance among all methods, indicating that it is the most suitable for TB X-ray image detection tasks. (3) In TB screening, high accuracy means the model can accurately identify individuals with the disease and correctly exclude those without it, demonstrating the effectiveness and reliability of the screening. As shown in Table 2, CDAM surpasses all other methods in terms of accuracy, with the average accuracy increasing from 59.9% to 73.3% compared to methods without domain adaptation. High sensitivity means the model can better identify samples with the disease, reducing the chance of missed diagnoses. In disease screening and diagnosis, high sensitivity is crucial as it ensures that most actual cases are identified. CDAM achieves the best sensitivity among all methods, with the average sensitivity increasing from 59.9% to 74.0% compared to methods without domain adaptation, indicating that the CDAM method can effectively learn from the source domain and identify disease features in the target domain. High specificity means the model can better distinguish normal samples, reducing the chance of misdiagnosis. CDAM performs well in specificity, although it did not achieve the best results in some datasets such as S→T, T→S, M→T, but the differences are insignificant. Although some methods have higher specificity than CDAM in these datasets, this is at the expense of sensitivity. Overall, the CDAM method's performance in TB's domain adaptation task is impressive, especially in maintaining high accuracy, high sensitivity, and relatively high specificity. CDAM is a powerful candidate for addressing domain shift issues in medical image analysis.

5.1. Ablation study

To investigate the effectiveness of each module in the proposed CDAM model, this paper will conduct an ablation study focusing on the components of self-attention (SA), DSH and DSP. The results are shown in Table 3. MOD1 represents the model with only the DSH module, MOD2 includes both the DSH and SA modules, and MOD3 contains the DSH, SA, and DSP modules. When the model only contains the DSH module, the feature map bypasses the SA module and directly computes the classification loss and domain discriminator's loss (in the DSH module). Thus, its function is equivalent to methods like DANN. When DSH incorporates the SA module, the average accuracy, sensitivity, and specificity increased by 1.7%, 1.4%, and 1.6%, respectively, indicating that the SA module can better extract global features and semantic information of lesions. When DSH incorporates the SA and DSP modules, the average accuracy, sensitivity, and specificity improves by 2.5%,4.0% and 1.4%, demonstrating a significant performance enhancement and affirming the effectiveness of the DSP modules.

[1] https://www.kaggle.com/datasets/ashery/chexpert.
[2] https://www.kaggle.com/datasets/nih-chest-xrays/data.

Table 2
Experimental results of different domain adaptation methods on different source-target domain pairs.

Method		S→M	S→T	T→S	T→M	M→S	M→T	Avg
No DA	Acc[CI]	65.8[57.9,73.7]	71.2[68.7,73.7]	53.0[49.2,56.8]	46.2[37.9,54.5]	60.2[56.5,63.9]	62.7[60.0,65.4]	59.9
	Sen[CI]	60.3[47.7,72.9]	67.1[63.4,70.8]	59.5[54.3,64.7]	47.3[34.5,60.1]	63.9[58.8,69.0]	61.5[57.7,65.3]	59.9
	Spe[CI]	69.8[59.7,79.9]	75.3[71.9,78.7]	46.3[40.9,51.7]	45.4[34.5,56.3]	56.4[51.0,61.8]	63.9[60.1,67.7]	59.5
UP	Acc[CI]	83.1[76.8,89.4]	99.1[98.6,99.6]	88.7[86.3,91.1]	83.1[76.8,89.4]	88.7[86.3,91.1]	99.1[98.6,99.6]	90.3
	Sen[CI]	81.3[71.3,91.3]	99.2[98.5,99.9]	83.9[80.0,87.8]	81.3[71.3,91.3]	83.9[80.0,87.8]	99.2[98.5,99.9]	88.1
	Spe[CI]	84.4[76.4,92.4]	99.1[98.4,99.8]	93.6[90.9,96.3]	84.4[76.4,92.4]	93.6[90.9,96.3]	99.1[98.4,99.8]	92.4
MMD	Acc[CI]	66.1[58.2,74.0]	73.7[71.3,76.1]	62.8[59.1,66.5]	57.4[49.1,65.7]	66.8[63.2,70.4]	65.7[63.1,68.3]	65.4
	Sen[CI]	64.8[52.5,77.1]	70.9[67.4,74.4]	70.5[65.6,75.4]	58.2[45.5,70.9]	67.5[62.5,72.5]	64.9[61.2,68.6]	66.1
	Spe[CI]	67.0[56.7,77.3]	76.5[73.2,79.8]	54.9[49.5,60.3]	56.8[45.9,67.7]	66.1[61.0,71.2]	66.5[62.8,70.2]	64.6
DANN	Acc[CI]	69.7[62.0,77.4]	77.5[75.2,79.8]	67.1[63.5,70.7]	60.2[52.0,68.4]	71.8[68.4,75.2]	68.5[65.9,71.1]	69.1
	Sen[CI]	68.2[56.2,80.2]	74.7[71.3,78.1]	68.2[63.2,73.2]	61.3[48.8,73.8]	71.2[66.4,76.0]	67.9[64.3,71.5]	68.6
	Spe[CI]	70.8[60.8,80.8]	80.3[77.2,83.4]	66.0[60.9,71.1]	59.4[48.6,70.2]	72.4[67.5,77.3]	69.1[65.5,72.7]	69.7
DIFL	Acc[CI]	71.0[63.4,78.6]	78.2[75.9,80.5]	67.3[63.7,70.9]	59.2[51.0,67.4]	80.0[77.0,83.0]	67.4[64.8,70.0]	70.5
	Sen[CI]	70.3[58.5,82.1]	76.7[73.4,80.0]	64.9[59.8,70.0]	69.3[57.4,81.2]	77.8[73.4,82.2]	64.3[60.6,68.0]	70.6
	Spe[CI]	70.8[60.8,80.8]	79.7[76.6,82.8]	69.8[64.8,74.8]	51.9[41.0,62.8]	**82.3 [78.2,86.4]**	70.5[66.9,74.1]	70.8
CXDaGAN	Acc[CI]	67.2[59.4,75.0]	76.9[74.6,79.2]	66.8[63.2,70.4]	58.7[50.5,66.9]	70.4[66.9,73.9]	67.5[64.9,70.1]	67.9
	Sen[CI]	65.6[53.4,77.8]	75.4[72.0,78.8]	68.4[63.4,73.4]	60.5[47.9,73.1]	69.1[64.2,74.0]	69.2[65.6,72.8]	68.0
	Spe[CI]	68.4[58.2,78.6]	78.4[75.2,81.6]	65.2[60.0,70.3]	57.4[46.6,68.2]	71.7[66.9,76.6]	65.8[62.1,69.5]	67.8
SPA-UDA	Acc[CI]	68.3[60.5,76.1]	77.6[75.3,79.9]	67.0[63.4,70.6]	58.7[50.5,66.9]	72.3[68.9,75.7]	67.6[65.0,70.2]	68.6
	Sen[CI]	66.7[54.6,78.8]	78.4[75.2,81.6]	65.5[60.4,70.6]	60.2[47.6,72.8]	70.3[65.4,75.2]	68.9[65.3,72.5]	68.3
	Spe[CI]	69.5[59.4,79.6]	76.8[73.5,80.1]	**68.5 [63.5,73.6]**	57.6[46.8,68.4]	74.4[69.6,79.1]	66.3[62.6,70.0]	68.8
AGPDA	Acc[CI]	71.1[63.5,78.7]	78.9[76.6,81.2]	68.1[64.5,71.7]	60.5[52.3,68.7]	73.5[70.1,76.9]	68.9[66.3,71.5]	70.2
	Sen[CI]	70.0[58.2,81.8]	77.3[74.0,80.6]	70.5[65.6,75.4]	58.5[45.8,71.2]	70.3[65.4,75.2]	66.4[62.7,70.1]	68.8
	Spe[CI]	71.9[62.0,81.7]	**80.5 [77.4,83.6]**	65.6[60.5,70.8]	62.0[51.3,72.6]	76.8[72.2,81.4]	**71.4 [67.9,74.9]**	71.4
CDAM	Acc[CI]	**71.4 [63.9,78.9]**	**79.5 [77.3,81.7]**	**69.3 [65.8,72.8]**	**67.9 [60.1,75.7]**	**80.2 [77.2,83.2]**	**71.5 [69.0,74.0]**	**73.3**
	Sen[CI]	**70.4 [58.7,82.1]**	**80.2 [77.1,83.3]**	**70.2 [65.3,75.1]**	**70.3 [58.5,82.1]**	**78.3 [73.9,82.7]**	**74.7 [71.3,78.1]**	**74.0**
	Spe[CI]	**72.1 [62.3,81.9]**	78.8[75.6,82.0]	68.4[63.4,73.4]	**66.2 [55.8,76.6]**	82.2[78.0,86.4]	68.3[64.7,71.9]	**72.7**

Table 3
Ablation study on DSH, SA and DSP modules. MOD1 (DSH), MOD2 (DSH+SA), MOD3 (DSH+SA+DSP).

Method		S→M	S→T	T→S	T→M	M→S	M→T	Avg
No DA	Acc[CI]	65.8[57.9,73.7]	71.2[68.7,73.7]	53.0[49.2,56.8]	46.2[37.9,54.5]	60.2[56.5,63.9]	62.7[60.0,65.4]	59.9
	Sen[CI]	60.3[47.7,72.9]	67.1[63.4,70.8]	59.5[54.3,64.7]	47.3[34.5,60.1]	63.9[58.8,69.0]	61.5[57.7,65.3]	59.9
	Spe[CI]	69.8[59.7,79.9]	75.3[71.9,78.7]	46.3[40.9,51.7]	45.4[34.5,56.3]	56.4[51.0,61.8]	63.9[60.1,67.7]	59.5
MOD1	Acc[CI]	69.7[62.0,77.4]	77.5[75.2,79.8]	67.1[63.5,70.7]	60.2[52.0,68.4]	71.8[68.4,75.2]	68.5[65.9,71.1]	69.1
	Sen[CI]	68.2[56.2,80.2]	74.7[71.3,78.1]	68.2[63.2,73.2]	61.3[48.8,73.8]	71.2[66.4,76.0]	67.9[64.3,71.5]	68.6
	Spe[CI]	70.8[60.8,80.8]	**80.3 [77.2,83.4]**	66.0[60.9,71.1]	59.4[48.6,70.2]	72.4[67.5,77.3]	69.1[65.5,72.7]	69.7
MOD2	Acc[CI]	70.3[62.7,77.9]	78.2[75.9,80.5]	67.9[64.3,71.5]	63.2[55.2,71.2]	75.6[72.3,78.9]	69.4[66.9,71.9]	70.8
	Sen[CI]	69.6[57.8,81.4]	77.8[74.6,81.0]	66.8[61.8,71.8]	62.3[49.8,74.8]	74.9[70.3,79.5]	68.5[64.9,72.1]	70.0
	Spe[CI]	70.9[60.9,80.9]	78.6[75.4,81.8]	**69.0 [64.0,74.0]**	62.7[52.1,73.3]	76.4[71.8,81.0]	**70.2 [66.6,73.8]**	71.3
MOD3	Acc[CI]	**71.4 [63.9,78.9]**	**79.5 [77.3,81.7]**	**69.3 [65.8,72.8]**	**67.9 [60.1,75.7]**	**80.2 [77.2,83.2]**	**71.5 [69.0,74.0]**	**73.3**
	Sen[CI]	**70.4 [58.7,82.1]**	**80.2 [77.1,83.3]**	**70.2 [65.3,75.1]**	**70.3 [58.5,82.1]**	**78.3 [73.9,82.7]**	**74.7 [71.3,78.1]**	**74.0**
	Spe[CI]	**72.1 [62.3,81.9]**	78.8[75.6,82.0]	68.4[63.4,73.4]	**66.2 [55.8,76.6]**	**82.2 [78.0,86.4]**	68.3[64.7,71.9]	**72.7**

This paper highlights the role of DSP by comparing the attention maps and heat maps of DSH+SA and DSH+SA+DSP more intuitively. Fig. 4 shows the attention maps and heat maps of two TB images and two normal images in the S→T task using the DSH+SA and DSH+SA+DSP methods. In the top title of each column of images, AM_{DSH} represents DSH attention map, HM_{DSH} represents DSH heat map, AM_{DSP} represents DSP attention map, and HM_{DSP} represents DSP heat map. The weights in the attention map are obtained from the row vectors of AM^{DSH} and AM^{DSP}. They represent the importance of various local features learned by the model and are reshaped into a 7×7 shape. In heat maps, regions with a deeper red color indicate a higher probability of containing richer classification information, while in attention maps, brighter regions signify richer classification information. The blue boxes in Fig. 4(b), (d), (j) and (n) denote the true locations of TB lesions.

DSH+SA part: From Fig. 4(b), there is a part of the red area inside the blue box and a part of the red area at the bottom of the image. The red area at the bottom incorrectly predicted the location of the TB lesion. In Fig. 4(d), the red area only appears inside the left blue box, and it does not appear in the right blue box, which also has errors. Fig. 4(f) and (h) are heat maps for normal CXRs, where the red areas indicate that the features in these areas contribute more to the classification result as normal.

DSH+SA+DSP part: This part adds the DSP attention map and DSP heat map. Fig. 4(j) shows that the red area falls perfectly within the blue box area. The red area in Fig. 4(n) is also within the two blue box areas. Compared to Fig. 4(b) and (d), these two figures have greatly improved, with the red areas being able to indicate the location of the TB lesion more accurately. The red areas in Fig. 4(l) and (p) are outside the lung area, where there is a lot of domain information. When the DSP features try to occupy the areas rich in domain information, with the help of the complementary mechanism, they can enhance Fig. 4(k) and (o) to prioritize areas containing TB and can be transferred across different domains. In Fig. 4(r) and (v), the red areas are outside the lung area. They also contain categorization information and contribute greatly to the recognition results being normal, and are easy to transfer across domains. Fig. 4(t) and (x) contain rich domain information, and the complementary mechanism also prompts Fig. 4(r) and (v) to focus more precisely on the regions transferable across domains.

5.2. Determination of optimal values for hyperparameters λ_1 and λ_2

This study conducted experiments to determine the optimal values of the hyperparameters λ_1 and λ_2. From Table 3, it can be observed that with the introduction of the DSH module, there were improvements of 9.9%, 8.7%, and 10.2% in average accuracy, average sensitivity, and average specificity, respectively. Similarly, introducing the DSP module led to performance enhancements of 2.5%, 4%, and 1.4%, respectively. Therefore, the DSH module was predominant in the domain adaptation

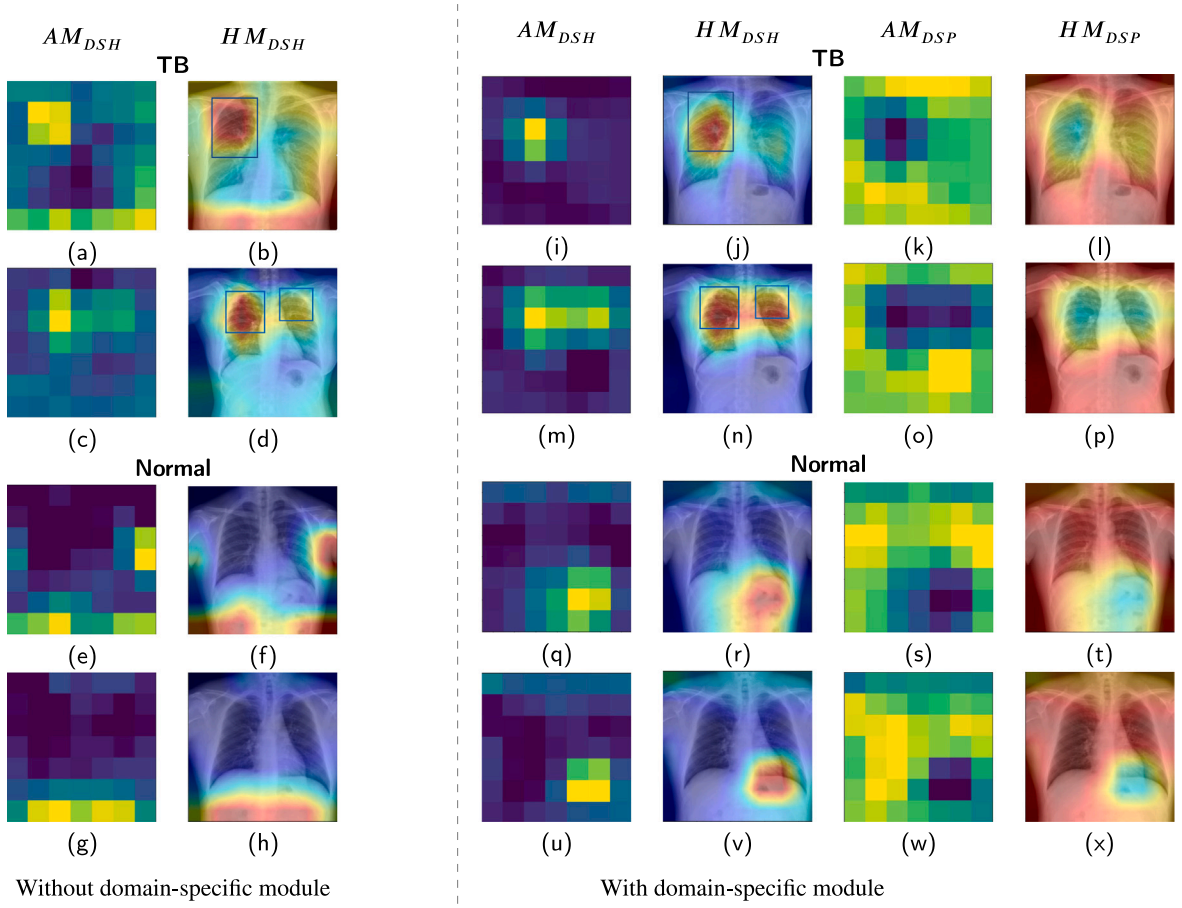

Fig. 4. The visualization comparison before and after adding the domain-specific module.

task, while the DSP module served as an auxiliary function. Consequently, this research initially focused on evaluating the impact of λ_1's value on domain adaptation performance. As illustrated in Fig. 5(a), the value of λ_1 was varied from 0 to 1, with increments of 0.1. Its average accuracy, average sensitivity, and average specificity reached their highest values of 70.8%, 70%, and 71.3%, respectively, when λ_1 was 0.5. Subsequently, with λ_1 fixed at 0.5, λ_2's value was varied from 0 to 0.5, with increments of 0.1. As depicted in Fig. 5(b), the peak values for average accuracy (73.3%), average sensitivity (74.0%), and average specificity (72.7%) were observed at $\lambda_2 = 0.2$. Furthermore, λ_2 values beyond 0.5 exhibited a declining trend. Consequently, the optimal hyperparameter values were $\lambda_1 = 0.5$ and $\lambda_2 = 0.2$.

5.3. Domain adaptation visualization

To more intuitively demonstrate the performance of adaptation methods in various domains, this paper visualizes the feature distributions of source and target data in domain adaptation tasks using t-SNE in Fig. 6. Red points indicating normal samples from the source dataset (source_Normal), and green points indicating TB samples from the source dataset (source_TB). Orange points represent normal samples from the target domain (target_Normal), and blue points represent TB samples from the target domain (target_TB). In Fig. 6(a) and (d), it can be observed that there is a noticeable discrepancy between the distributions of the blue points and green points, as well as a considerable difference between the distributions of the red points and yellow points. This indicates a lack of good alignment between the source and target domains when employing the MMD and CX-DaGAN method, which in turn contributes to its poorer performance, as shown in Table 2. Apart from the MMD and CX-DaGAN method, it is evident that the distributions of sample points between the source and target domains are roughly consistent for the remaining methods. Among them, the proposed CDAM exhibits the best performance, showing the most robust consistency in the distribution between the source and target domains, with clear boundaries. These results demonstrate the effectiveness of the method proposed in this paper.

5.4. Multi-class classification tasks

To evaluate the performance of different models in multi-class classification tasks, this study conducted domain adaptation experiments on the CheXpert and Chest X-ray14 datasets. The experiments applied several methods that have shown strong performance in binary classification, including DIFL, AGPDA, and CDAM. The results, presented in Table 4, indicate that all domain adaptation methods outperformed the 'No DA' group, demonstrating their effectiveness in transferring knowledge from the source domain to the target domain. Among these methods, CDAM achieved the highest performance, with an average accuracy of 70.3%, an average precision of 71.8%, and an average specificity of 69.6% on the CheXpert and Chest X-ray14 datasets. This result confirms its ability to maintain strong performance in multi-class classification tasks, similar to its success in TB binary classification. These findings highlight the potential of CDAM for broader applications in multi-disease classification.

6. Discussion

The experimental results show that the CDAM method performs the best, followed by the adversarial-based domain adaptation methods (DANN and DIFL) and AGPDA in the second tier. The Cycle-GAN-based

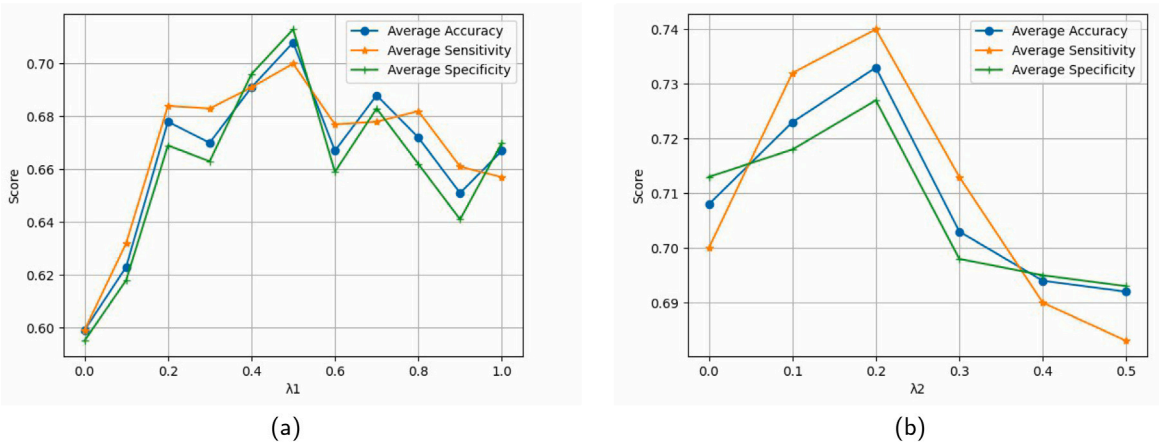

Fig. 5. The variation of domain adaptation performance with respect to λ_1 and λ_2. (a) The impact of λ_1 on domain adaptation performance. (b) Variation of domain adaptation performance with λ_2 when λ_1 is fixed at 0.5.

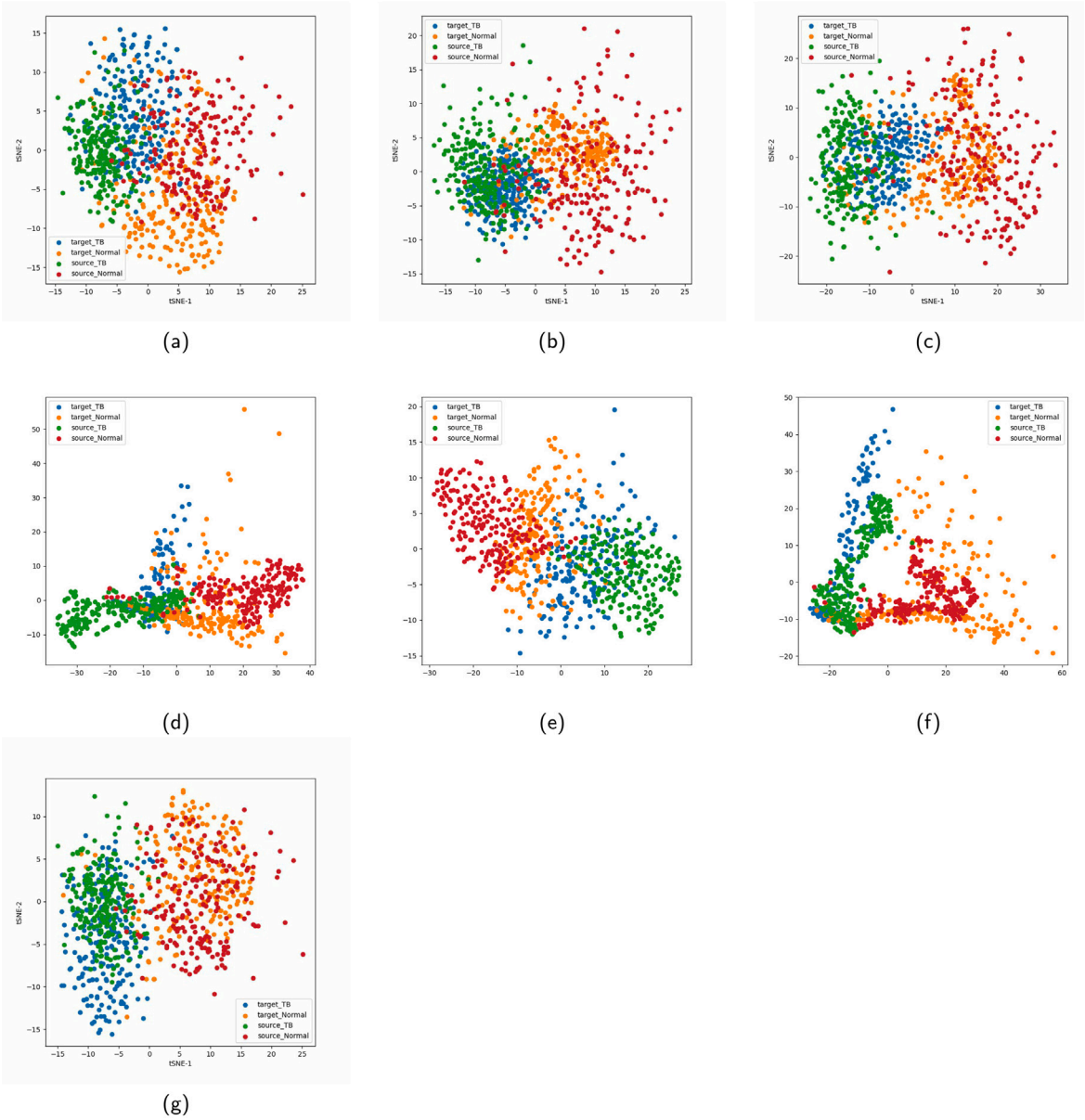

Fig. 6. Visualization images using t-SNE for (a) MMD, (b) DANN, (c) DIFL, (d) CX-DaGAN, (e) SPA-UDA, (f) AGPDA and (g) CDAM methods.

Table 4
Multi-class domain adaptation experimental results. A represents CheXpert, and B represents Chest X-ray14.

Method	Metric	A → B	B → A	Avg
No DA	Acc($\mu \pm \sigma$)	59.8 ± 0.7	60.3 ± 1.1	60.1
	Sen($\mu \pm \sigma$)	59.5 ± 0.9	58.9 ± 1.4	59.2
	Spe($\mu \pm \sigma$)	60.0 ± 1.3	61.2 ± 0.8	60.6
DIFL	Acc($\mu \pm \sigma$)	68.3 ± 0.3	65.4 ± 0.8	66.9
	Sen($\mu \pm \sigma$)	69.4 ± 1.1	64.3 ± 1.1	66.9
	Spe($\mu \pm \sigma$)	66.9 ± 0.7	66.7 ± 0.4	66.8
AGPDA	Acc($\mu \pm \sigma$)	70.7 ± 1.2	66.7 ± 0.9	68.7
	Sen($\mu \pm \sigma$)	71.1 ± 0.7	67.3 ± 0.8	69.2
	Spe($\mu \pm \sigma$)	68.7 ± 1.0	64.9 ± 1.1	66.8
CDAM	Acc($\mu \pm \sigma$)	**72.3** ± 0.4	**68.3** ± 1.1	**70.3**
	Sen($\mu \pm \sigma$)	**73.8** ± 0.6	**69.8** ± 1.0	**71.8**
	Spe($\mu \pm \sigma$)	**71.4** ± 0.9	**67.7** ± 1.2	**69.6**

methods (CX-DaGAN and SPA-UDA) come next, while the MMD method performs the worst. The fundamental limitation of MMD lies in its reliance on handcrafted features, rendering it incapable of adaptively learning the optimal feature representation, thereby limiting its performance potential. In contrast, adversarial training paradigms possess the ability to automatically acquire the optimal feature representation, effectively ensuring consistency between the feature distributions of the source and target domains, circumventing the need for manual feature engineering. Cycle-GAN-based methods aim to reduce domain discrepancy by transforming the style of target domain images to match that of source domain images. However, during this transformation, there may be changes in lesion information. Despite the use of feature and label consistency mechanisms, new errors can still be introduced, limiting the improvement in domain adaptation performance. AGPDA builds upon MMD by incorporating an attention mechanism, which allows the model to focus better on areas more likely to propagate across domains, thereby enhancing performance. After incorporating the complementary domain attention mechanism, the CDAM methodology demonstrates a clear advantage by specifically targeting task-relevant regions. This targeted approach allows CDAM to address the distributional differences between the source and target domains more effectively, thereby minimizing negative transfer. As a result, CDAM achieves better transfer performance than other methods.

While CDAM has demonstrated promising results, its computational resource requirements are relatively high. The computational complexity of the CDAM model is primarily determined by the self-attention module, which involves large-scale matrix multiplications. The complexity is given by $O(N \times C \times (H \times W)^2)$, where N is the batch size, C is the number of channels, and H, W are the dimensions of the feature map. Under the experimental settings ($H = W = 7$, $C = 1024$, batch size = 16), the total computation amounts to approximately 79.2G FLOPs. In terms of memory usage, CDAM consumes 4.78 GB of memory during training. It is recommended to use a GPU with at least 6 GB of memory to ensure stable training. However, the high computational demand and memory requirements pose challenges for regions with limited computational resources. To reduce the demand for computational resources, future work will explore methods such as low-rank approximation to mitigate the computational load of attention matrix calculations. Alternatively, incorporating mechanisms such as local attention or hierarchical structures could be considered to limit the scope of computations, thereby reducing the need for global attention across the entire feature map and decreasing the overall computational complexity. These approaches are expected to enhance the model's computational efficiency, making it more suitable for larger-scale datasets and resource-constrained environments.

Despite these limitations, CDAM presents a powerful approach to mitigating domain shifts in medical image analysis, offering a promising solution for improving the generalization of AI-based tuberculosis screening systems across diverse datasets.

7. Conclusion

The CDAM model partitions the feature map into DSH and DSP features. DSH features focus on regions containing transferable TB classification information across domains, alleviating the impact of domain divergence. Conversely, DSP features target areas rich in domain information, maximizing domain discrepancy. Crucially, their complementary nature prompts DSH to accurately localize transferable regions as DSP occupies domain-rich areas, reducing negative transfer and enhancing cross-domain generalization capability for AI-based TB screening. The proposed method underwent domain adaptation experiments on the Shenzhen, Montgomery, TBX11K, CheXpert and ChestX-ray 14 datasets, demonstrating superior performance compared to existing domain adaptation methods.

Overall, the CDAM model presents a powerful and robust approach to mitigating domain shifts in medical image analysis, offering a promising solution for improving the generalization capabilities of AI-based tuberculosis screening systems across diverse datasets. Future research directions include extending the model to other medical imaging tasks and evaluating its performance on a broader range of datasets to further validate its robustness and generalization capabilities.

CRediT authorship contribution statement

Zeyu Ding: Writing – original draft, Data curation, Conceptualization. **Razali Yaakob:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Azreen Azman:** Supervision, Methodology. **Siti Nurulain Mohd Rum:** Resources, Methodology. **Norfadhlina Zakaria:** Supervision, Resources, Methodology. **Azree Shahril Ahmad Nazri:** Validation, Supervision, Methodology.

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.

Acknowledgment

This work was supported by the Ministry of Higher Education under the Fundamentals Research Grant Scheme (FRGS/1/2020/ICT02/UPM/02/5).

Data availability

Data will be made available on request.

References

[1] W.H. Organization, et al., Global Tuberculosis Report 2023, World Health Organization, 2023.
[2] Y. Zhang, H. Chen, Y. Wei, P. Zhao, J. Cao, X. Fan, X. Lou, H. Liu, J. Hou, X. Han, et al., From whole slide imaging to microscopy: Deep microscopy adaptation network for histopathology cancer image classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 360–368.
[3] Y. Gao, Y. Zhang, Z. Cao, X. Guo, J. Zhang, Decoding brain states from fMRI signals by using unsupervised domain adaptation, IEEE J. Biomed. Heal. Inform. 24 (6) (2019) 1677–1685.
[4] Y. Feng, Z. Wang, X. Xu, Y. Wang, H. Fu, S. Li, L. Zhen, X. Lei, Y. Cui, J.S.Z. Ting, et al., Contrastive domain adaptation with consistency match for automated pneumonia diagnosis, Med. Image Anal. 83 (2023) 102664.
[5] N. Karani, E. Erdil, K. Chaitanya, E. Konukoglu, Test-time adaptable neural networks for robust medical image segmentation, Med. Image Anal. 68 (2021) 101907.
[6] Y. Feng, X. Xu, Y. Wang, X. Lei, S.K. Teo, J.Z.T. Sim, Y. Ting, L. Zhen, J.T. Zhou, Y. Liu, et al., Deep supervised domain adaptation for pneumonia diagnosis from chest x-ray images, IEEE J. Biomed. Heal. Inform. 26 (3) (2021) 1080–1090.

[7] Y. Huang, P. Peng, Y. Jin, Y. Li, J. Xing, Domain adaptive attention learning for unsupervised person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 11069–11076.
[8] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: Maximizing for domain invariance, Comput. Sci. (2014).
[9] B. Sun, K. Saenko, Deep coral: Correlation alignment for deep domain adaptation, in: Computer Vision–ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14, Springer, 2016, pp. 443–450.
[10] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J.W. Vaughan, A theory of learning from different domains, Mach. Learn. 79 (2010) 151–175.
[11] I. Chung, D. Kim, N. Kwak, Maximizing cosine similarity between spatial features for unsupervised domain adaptation in semantic segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 1351–1360.
[12] C. Chen, Z. Fu, Z. Chen, S. Jin, Z. Cheng, X. Jin, X.-S. Hua, Homm: Higher-order moment matching for unsupervised domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 3422–3429.
[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[14] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (59) (2016) 1–35.
[15] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176.
[16] N. Ravin, S. Saha, A. Schweitzer, A. Elahi, F. Dako, D. Mollura, D. Chapman, Mitigating domain shift in AI-based TB screening with unsupervised domain adaptation, IEEE Access 10 (2022) 45997–46013.
[17] C. Bian, C. Yuan, J. Wang, M. Li, X. Yang, S. Yu, K. Ma, J. Yuan, Y. Zheng, Uncertainty-aware domain alignment for anatomical structure segmentation, Med. Image Anal. 64 (2020) 101732.
[18] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
[19] K. Sanchez, C. Hinojosa, H. Arguello, D. Kouamé, O. Meyrignac, A. Basarab, CX-DaGAN: Domain adaptation for pneumonia diagnosis on a small chest X-ray dataset, IEEE Trans. Med. Imaging 41 (11) (2022) 3278–3288.
[20] X. Qin, F. Bui, Z. Han, Semantically preserving adversarial unsupervised domain adaptation network for improving disease recognition from chest x-rays, Comput. Med. Imaging Graph. 107 (2023) 102232.
[21] D. Mahapatra, S. Korevaar, B. Bozorgtabar, R. Tennakoon, Unsupervised domain adaptation using feature disentanglement and GCNs for medical image classification, in: European Conference on Computer Vision, Springer, 2022, pp. 735–748.
[22] Y. Lu, W.K. Wong, C. Yuan, Z. Lai, X. Li, Low-rank correlation learning for unsupervised domain adaptation, IEEE Trans. Multimed. (2023).
[23] C. Chen, J. Wang, J. Pan, C. Bian, Z. Zhang, GraphSKT: graph-guided structured knowledge transfer for domain adaptive lesion detection, IEEE Trans. Med. Imaging 42 (2) (2022) 507–518.
[24] Y. Lu, D. Li, W. Wang, Z. Lai, J. Zhou, X. Li, Discriminative invariant alignment for unsupervised domain adaptation, IEEE Trans. Multimed. 24 (2021) 1871–1882.
[25] Y. Lu, W.K. Wong, B. Zeng, Z. Lai, X. Li, Guided discrimination and correlation subspace learning for domain adaptation, IEEE Trans. Image Process. 32 (2023) 2017–2032.
[26] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[27] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.
[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[29] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.
[30] W. Ji, A.C. Chung, Unsupervised domain adaptation for medical image segmentation using transformer with meta attention, IEEE Trans. Med. Imaging (2023).
[31] R. Azad, L. Niggemeier, M. Hüttemann, A. Kazerouni, E.K. Aghdam, Y. Velichko, U. Bagci, D. Merhof, Beyond self-attention: Deformable large kernel attention for medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 1287–1297.
[32] X. Wang, L. Li, W. Ye, M. Long, J. Wang, Transferable attention for domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 5345–5352.
[33] W. Liu, Z. Ni, Q. Chen, L. Ni, Attention-guided partial domain adaptation for automated pneumonia diagnosis from chest x-ray images, IEEE J. Biomed. Heal. Inform. (2023).
[34] Y. Shi, A. Deng, M. Deng, M. Xu, Y. Liu, X. Ding, J. Li, Transferable adaptive channel attention module for unsupervised cross-domain fault diagnosis, Reliab. Eng. Syst. Saf. 226 (2022) 108684.
[35] D. Zeyu, R. Yaakob, A. Azman, S.N.M. Rum, N. Zakaria, A.S.A. Nazri, A Grad-CAM-based knowledge distillation method for the detection of tuberculosis, in: 2023 International Conference on Information Management, ICIM, IEEE, 2023, pp. 72–77.
[36] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[37] S. Jaeger, S. Candemir, S. Antani, Y.-X.J. Wáng, P.-X. Lu, G. Thoma, Two public chest X-ray datasets for computer-aided screening of pulmonary diseases, Quant. Imaging Med. Surg. 4 (6) (2014) 475.
[38] Y. Liu, Y.-H. Wu, Y. Ban, H. Wang, M.-M. Cheng, Rethinking computer-aided tuberculosis diagnosis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2646–2655.
[39] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, et al., Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 590–597.
[40] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2097–2106.

Zeyu Ding obtained his Bachelor's degree in Electronic Information Engineering from Hefei University of Technology, China, in 2013. In 2016, he earned a Master's degree in Control Engineering from the University of Science and Technology of China. Since 2022, he has been pursuing a Ph.D. at Universiti of Putra Malaysia in the field of Intelligent Systems. His research focuses on artificial intelligence and medical image processing.

Razali Yaakob received the Bachelor Degree in Computer Science in 1996 and Master in Computer Science from in 1999, from Universiti Putra Malaysia, and Ph.D. from University of Nottingham, United Kingdom in 2008. Currently, he is a Associate Professor at the Faculty of Computer Science and IT, Universiti Putra Malaysia. His research areas include artificial neural network, pattern recognition, and evolutionary computation in game playing. He is a member of the Intelligent Computing Group at the faculty. Orcid id: 0000-0002-8228-5753.

Azreen bin Azman received the Diploma degree in software engineering from the Institute of Telecommunication and Information Technology, in 1997, the Bachelor of Information Technology degree in information systems engineering from Multimedia University, Malaysia, in 1999, and the Ph.D. degree in computing science specializing in information retrieval from the University of Glasgow, Scotland, in September 2007. He is currently an Associate Professor at University Putra Malaysia. His current research interests include information retrieval, text mining, natural language processing, and intelligent systems.

Siti Nurulain binti Mohd Rum received her Masters and Doctorate in Computer Science in 2012 and 2017, both from the University of Malaya (UM). Her bachelor's degree is from Universiti Teknologi Malaysia (UTM). She is currently a senior lecturer in the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM). Her research interests include, but are not limited to, the areas of artificial intelligence, database processing, data science analytic, semantic web, and social media analytics.

Nor Fadhlina binti Zakaria obtained her Master of Medicine degree in 2010 from Universiti Kebangsaan Malaysia and her Doctor of Medicine degree in 2014 from Universiti Kebangsaan Malaysia. Her area of specialization is internal medicine and nephrology.

Azree Nazri received the PhD. degrees in Mathematics from the University of Cambridge, UK. He is a senior lecturer in University Putra Malaysia. His research interests include the development of artificial general intelligence.