scientific reports

OPEN

Addressing cross-population domain shift in chest X-ray classification through supervised adversarial domain adaptation

Aminu Musa[1,4,5]✉, Rajesh Prasad[1,3,5] & Monica Hernandez[2,5]

Medical image analysis, empowered by artificial intelligence (AI), plays a crucial role in modern healthcare diagnostics. However, the effectiveness of machine learning models hinges on their ability to generalize to diverse patient populations, presenting domain shift challenges. This study explores the domain shift problem in chest X-ray classification, focusing on cross-population variations, especially in underrepresented groups. We analyze the impact of domain shifts across three population datasets acting as sources using a Nigerian chest X-ray dataset acting as the target. Model performance is evaluated to assess disparities between source and target populations, revealing large discrepancies when the models trained on a source were applied to the target domain. To address with the evident domain shift among the populations, we propose a supervised adversarial domain adaptation (ADA) technique. The feature extractor is first trained on the source domain using a supervised loss function in ADA. The feature extractor is then frozen, and an adversarial domain discriminator is introduced to distinguish between the source and target domains. Adversarial training fine-tunes the feature extractor, making features from both domains indistinguishable, thereby creating domain-invariant features. The technique was evaluated on the Nigerian dataset, showing significant improvements in chest X-ray classification performance. The proposed model achieved a 90.08% accuracy and a 96% AUC score, outperforming existing approaches such as multi-task learning (MTL) and continual learning (CL). This research highlights the importance of developing domain-aware models in AI-driven healthcare, offering a solution to cross-population domain shift challenges in medical imaging.

The discipline of medical image analysis has experienced a profound transformation through the integration of artificial intelligence (AI) technologies[1]. AI models have demonstrated promising capabilities in analyzing medical images, thereby assisting medical professionals in achieving accurate diagnoses and devising effective treatment plans[2]. Notably, recent advancements in computer vision have facilitated the utilization of deep learning in medical image analysis, fostering the development of accurate and efficient models for disease diagnosis[3]. However, the application of AI techniques to medical imaging analysis in racially diverse populations and datasets presents critical challenges, particularly regarding generalization, biases, and safety concerns, which are amplified when racial and ethnic representation is insufficient[4].

The study of domain shift in medical imaging is essential for building AI models that are robust across diverse populations and healthcare environments. Domain shifts occur due to variations in patient demographics, imaging protocols, equipment, and environmental factors, which can lead to significant performance degradation when models trained on one population are applied to another. This issue is particularly pressing for underrepresented regions such as Nigeria, where limited access to diverse and representative datasets can impede the development of fair and accurate AI-driven healthcare solutions. Collecting data from such regions is crucial, as it enables the training of models that reflect a broader range of patient characteristics, thereby reducing the risk of biased predictions and enhancing the generalizability of AI models across racially and ethnically diverse populations.

[1]Present address: Deparment of Computer Science, African University of Science and Technology, Abuja 900107, Nigeria. [2]Deparment of Computer Science, University of Zaragoza, Zaragoza 50018, Spain. [3]Department of Computer Science and Engineering, Ajay Kumar Garg Engineering College, Ghaziabad 201015, India. [4]Department of Computer Science, Federal University Dutse, Dutse, Nigeria. [5]Aminu Musa, Rajesh Prasad and Monica Hernandez contributed equally to this work. ✉email: musa.aminu@fud.edu.ng

Addressing these disparities not only advances global health equity but also ensures that AI models are more reliable in various clinical settings, ultimately supporting better patient outcomes worldwide.

Chest X-ray analysis plays a crucial role in the diagnosis and management of a wide range of respiratory and cardiovascular conditions[5]. Nowadays, AI models are increasingly used with this imaging modality for a wide range of tasks such as image segmentation, image registration, analysis of respiration motion, detection of anatomical features, disease diagnosis, and prognosis[6]. The application of AI models in health is being widely explored, and their performance is commendable for these tasks[7].

Traditionally, the analysis and interpretation of medical images have been conducted by expert radiologists and physicians[3]. However, in developing countries, a shortage of skilled radiologists, combined with the immense pressure on physicians due to overwhelming workloads, increases the likelihood of errors in human interpretation. This not only extends the time needed for initial diagnosis but also leads to delays in follow-up care and a higher risk of misdiagnosis. AI models are known to be accurate in classification and object detection tasks[8]. Therefore, integrating AI models in chest X-ray analysis and interpretation may greatly improve the process and offer accurate and timely analysis to support decision-making[9–11]. However, differences in patient demographics, imaging protocols, and equipment across various populations can lead to substantial domain shifts, which may significantly impair the performance of AI models trained on one population when applied to another, particularly if the populations differ in racial or ethnic composition[4].

Many research works focus on investigating domain shifts as a result of different equipment, different hospitals, or cross modality[7,12,13]. These models often exhibit bias when tested on samples from underrepresented groups within the dataset. A major contributing factor to this bias is the lack of diversity in the training data, where minority populations are significantly underrepresented [4]. Addressing this issue is further complicated by the limited availability of publicly accessible datasets from these underrepresented groups, making it even more challenging to develop fair and robust models.

In this paper, we study the critical issue of domain shift in chest X-ray image classification, particularly concerning racially diverse populations. We recognize that conventional machine learning and deep learning models, even when pre-trained on extensive datasets, may encounter challenges in adapting to new populations due to inherent domain differences. Our focus extends to quantifying population-based domain shifts within Nigerian chest X-ray datasets, where unique demographic and clinical characteristics can introduce domain-specific complexities in accurately classifying X-ray images. By thoroughly examining and mitigating domain shift, we aim to enhance the generalizability and clinical applicability of AI-driven chest X-ray image classification across diverse populations

To achieve this goal, we propose and evaluate a supervised adversarial domain adaptation (ADA) approach aimed at mitigating the adverse effects of domain shift. ADA prioritizes the adjustment of invariant features within the target domain to narrow the disparity between the pre-trained source domain and the Nigerian target domain. Our analysis entails a comprehensive examination of domain shifts in chest X-ray image classification, utilizing a meticulously curated dataset representative of the Nigerian population. We delve into the intricacies of feature-level domain adaptation techniques, elucidating their application in enhancing the performance of AI models across disparate domains. Through extensive experimentation, we demonstrate the efficacy of our proposed approach in addressing domain shift and enhancing the accuracy of chest X-ray image classification. Our contributions advance the realm of AI-driven healthcare solutions by fostering adaptability to racially diverse populations.

The primary objective of this study is twofold: first, to analyze the impact of domain shift on chest X-ray classification accuracy across different populations, and second, to propose and evaluate the ADA approach in mitigating the adverse effects of domain shift. We foresee that traditional machine learning and deep learning models, even when pre-trained on large datasets, often struggle to adapt to new populations due to domain discrepancies.

This paper is an extended version of our work previously presented at the Miccai conference[14], with significant enhancements, including theoretical details of our proposed method, and a thorough comparison with the state-of-the-art methods that may compete with our approaches, such as multi-task learning or continual learning[15,16].

The remainder of this paper is organized as follows: Section "Literature review" provides a review of the relevant literature, highlighting the significance of domain adaptation in medical image analysis. In Section "Datasets and methods", we detail the methodology adopted, including the protocols for the Nigerian dataset acquisition and preprocessing. Section "Experimental results and discussions" presents the experimental results and the discussion, followed by an exploration of the implications of our findings. We conclude our study in Section "Conclusion and future work", by summarizing our contribution.

Literature review

In recent years, the application of artificial intelligence (AI) in medical image analysis has paved the way for transformative advancements in healthcare diagnostics[17]. Recent advancements in deep learning-based medical image analysis have significantly improved diagnostic performance and model interpretability. Works such as DeepXplainer[18] and Explainable AI-driven IoMT Fusion[19] highlight the role of interpretable AI models in enhancing clinical trust and decision-making. Additionally, transformer-based architectures like TransResUNet[20] have demonstrated success in medical image segmentation, suggesting potential applications for feature extraction in domain adaptation. Furthermore, attention-based mechanisms such as FGA-Net[21] offer promising insights into feature selection and domain-invariant learning, which could be beneficial in domain adaptation settings for robust disease prediction. These recent advancements align with our study's objectives, emphasizing the need for AI models that are both domain-adaptive and interpretable in medical imaging. Chest X-ray image classification holds particular significance for its role in diagnosing a wide range of respiratory and cardiovascular conditions[22]. Several X-ray diagnosis systems based on machine learning have been proposed, achieving promising performance[5,23–26]. However, the successful application of AI models in this domain often hinges on their adaptability to account for the distinct characteristics of racially diverse patient populations[27].

The term cross-population domain shift describes variations in data distributions across different populations[14]. This is a widely recognized challenge in medical image analysis, especially with the advent of deep-learning[28]. Extensive studies underscore the impact of domain shifts on the performance of AI models. For example, Rajpurkar et al. highlighted that models trained on data from one population may not generalize effectively to others, leading to wide disparities in performance[29]. Given the unique challenges inherent in chest X-ray image classification across diverse datasets[30], it becomes evident that domain shift originating from variations across different patient populations could introduce patient-specific discrepancies. Addressing this pressing issue is crucial for ensuring the generalizability and effectiveness of AI-based diagnostic systems in clinical practice[31,32].

Several AI-driven medical diagnosis have explored various optimization techniques to improve model performance and generalizability, mitigating bias. For instance, the study "An optimized ensemble model based on meta-heuristic algorithms for effective detection and classification of breast tumors"[33] demonstrates how ensemble learning combined with meta-heuristic optimization can enhance classification accuracy in medical imaging. Similarly, the work of Saber et. al.[34] highlights the effectiveness of knowledge distillation and advanced optimization techniques in improving model efficiency and convergence. These studies align with our research on domain adaptation by emphasizing the importance of robust feature learning and optimization strategies for better generalization in medical image classification.

Domain adaptation (DA) techniques have emerged as a promising avenue to tackle the issue of domain shift[16]. Feature-level adaptation techniques have gained significant attention, with methods such as Maximum Mean Discrepancy (MMD) and adversarial training, demonstrating their ability to align feature distributions[35].

Ganin et al.,[36] proposed domain-adversarial training and unsupervised domain adaptation methods[37], which have demonstrated success in mitigating domain discrepancies. Additionally, He et al.,[38] introduced a classification-aware semi-supervised domain adaptation technique, which shows promise for leveraging limited labeled data effectively. Techniques such as data augmentation through latent space interpolation[39] have also emerged as valuable tools for enhancing model generalization. While the concept of domain adaptation has been well-explored in computer vision, a real value can be seen in the context of medical imaging applications in healthcare domains[40].

Although DA techniques adapt to variations in data from different populations are highly pertinent, they are relatively limited. Some preliminary studies have explored DA techniques on different medical imaging datasets. For instance, Feng et al., applied domain adaptation methods to different chest X-ray datasets, highlighting the potential of these techniques to adapt models to new populations and domains[41]. Seyyed-Kalantari et al., investigated the issue of bias in machine learning-assisted diagnosis systems in the underserved patient population[42]. Their work highlights the presence of cross-population variability in medical imaging analysis. Some studies provide a comprehensive survey and valuable insights into the landscape of domain adaptation techniques in medical image analysis, offering a roadmap for future research directions in this rapidly evolving field[43,44].

Domain shift is one of the major issues that undermine the generalization potential of deep learning-based diagnosis systems[45]. DA can be categorized into two broad categories:

(1) Unsupervised DA, which tries to align data distributions of the features in the feature space with a labeled source data, mapped to an unlabelled target dataset[46,47]. Unsupervised deep DA gained significant interest in medical image analysis, because of its advantage of not needing any labeled target data. This can be done in two ways, either through feature alignment, using Domain Adversarial Neural Networks (DANN) or image alignment, which mostly uses Generative Adversarial Networks (GANs)[48].

(2) Supervised methods, which align the distribution gap between the labeled source and target domains[49,50]. Supervised DA involves transferring model knowledge acquired in the source domain to the target domain through fine-tuning[51]. Numerous studies concentrate on employing shallow domain adaptation models for medical image analysis utilizing deep features extracted by convolutional neural networks (CNN), and then using an adaptation technique such as Transfer Component Analysis (TCA)[52], Correlation Alignment (CORAL)[53], or Balanced Distribution Adaptation (BDA)[54].

Semi-supervised learning was also proposed to augment the limited labels on target data using GANs. A GAN-based DA framework for classifying chest X-ray images is proposed by Madani et al.[55]. Unlike traditional GANs, this model accepts as inputs generated images, unlabelled target data, and labelled source data to align the data in a semi-supervised approach.

Other works emphasize the importance of measuring domain shift for deep learning models in medical images[56]. In the work of Stacke et al., the authors proved the existence of domain shift in medical image data as a result of different data acquisition pipelines, different medical facilities, or over time. To further confirm their hypothesis, the authors experimented on domain shift in tumor classification of hematoxylin and eosin stained images, using different datasets and models[57]. Results from their experiments demonstrate how the proposed model outperforms existing methods for measuring domain shift and uncertainty by having a strong association with performance decline when testing a model across a wide range of different types of domain changes[57].

Although supervised domain adaptation leverages transfer learning and fine-tuning with labelled target domain data, its performance may degrade when there are substantial discrepancies between the source and target domain distributions[58]. In contrast, our proposed ADA method explicitly tackles these discrepancies by aligning feature distributions through adversarial training, offering a more robust adaptation to domain shifts.

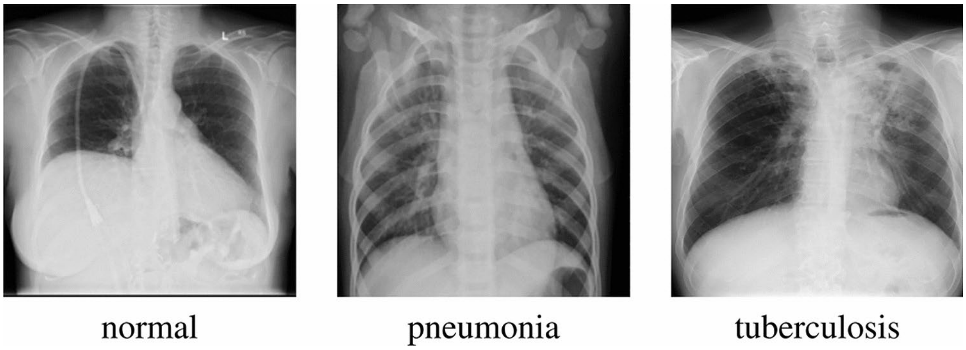


Fig. 1. Representative images from the Nigerian dataset. Left, image from the healthy normal group. Center, image from the pneumonia group. Right, image from the tuberculosis group.

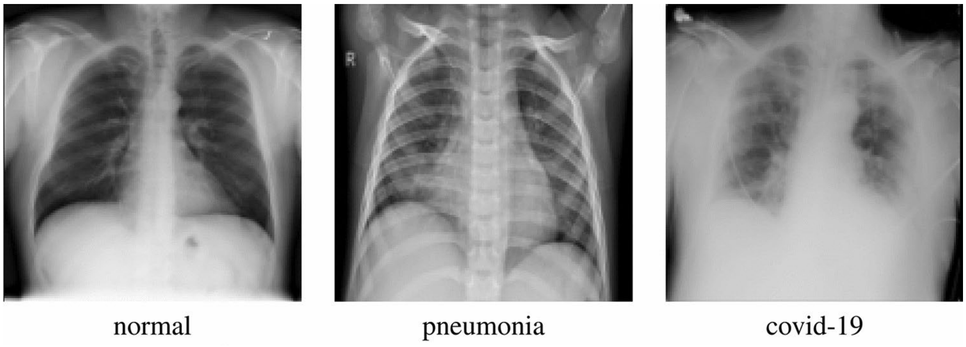


Fig. 2. Sample images from Doha dataset.

Datasets and methods

This section introduces the Nigerian chest X-ray dataset used as the target domain and describes the methodology to investigate the existence of domain shifts in deep learning-based chest X-ray image classification. Finally, the section presents the proposed ADA technique employed to mitigate the impact of domain shift.

Nigerian X-ray dataset

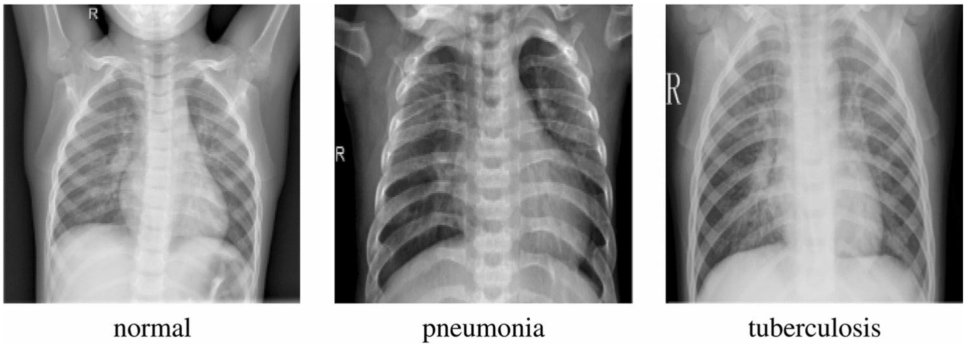
The target domain dataset used in this research is a locally collected dataset of chest X-rays from the Radiology Centre of Aminu Kano Teaching Hospital (AKTH), Nigeria. The dataset consists of 6345 X-ray images. The images were annotated by three different physicians into three different categories: pneumonia, tuberculosis, and normal X-rays. The distribution of the images per category is 2340 X-ray images diagnosed with tuberculosis by physicians, 1445 diagnosed with pneumonia, and 2560 termed as normal X-rays. The image dimension was 299 × 299 pixels with 72 pixels/inch DPI. The dataset is released for public use under Apache license 2 on Kaggle.[1] Figure 1 shows representative sample images from this dataset.

Baseline X-ray datasets

Three different chest X-ray datasets are selected as source datasets for our study. The first dataset is the chest X-ray classification dataset, collected at Guangzhou Medical Center, China[17]. The dataset is made of 5863 images. The images were annotated by expert physicians into pneumonia and normal X-rays. sample X-rays images can be seen from Fig. 2. The second dataset is VinDr-CXR, which consists of 18,000 images annotated by 17 different physicians into 6 different classes of diagnosis including pneumonia and tuberculosis[59]. The dataset was collected in Hanoi Medical University Hospital, Vietnam. Sample images from the dataset are given in Fig. 3. The third dataset is the COVID-19 radiography database, organized by a team of researchers from Qatar University, Doha[60]. The dataset has 1345, pneumonia X-rays, 1341 normal X-rays, and 1200 COVID-19 X-rays. sample images are presented in Fig. 4. In the following, we will refer to the datasets as China, Vietnam, and Doha. We show sample images from each source dataset alongside the annotated classes in Figs. 2, 3, and 4, respectively.

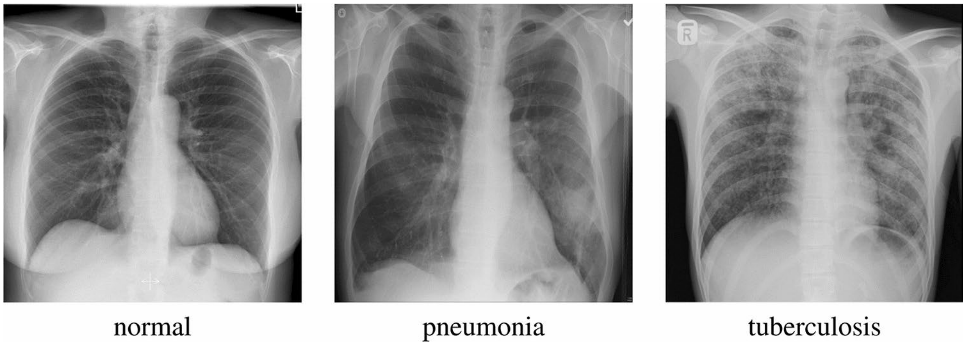
The datasets were selected to examine cross-population domain shifts in chest X-ray classification. We chose source datasets from China, Vietnam, and Doha (Qatar) due to their geographic, ethnic, and imaging diversity. These datasets represent populations from Asia and the Middle East, offering a broad distribution of patient demographics. The Nigerian dataset was selected as the target domain because African populations are

[1] Dataset Link: https://www.kaggle.com/datasets/aminumusa/nigeria-chest-x-ray-dataset.



normal pneumonia tuberculosis

Fig. 3. Sample images from China dataset.



normal pneumonia tuberculosis

Fig. 4. Sample images from Vietnam dataset.

significantly underrepresented in publicly available medical AI datasets. This choice allows us to investigate how models trained on non-African populations generalize to an African dataset, addressing critical issues of fairness and model adaptability in AI-driven healthcare. Additionally, the datasets vary in imaging protocols, equipment, and disease distributions, providing a realistic scenario for studying domain shifts in medical imaging.

Models training strategy

We employ a pre-trained DenseNet201 CNN model to obtain the baseline models[61]. The architecture is known as one of the best in classifying X-ray images. All images are resized to 224 × 224 pixels to yield input images compatible with DenseNet.

The baseline models were trained in their respective datasets, and their performance was measured using standard metrics: accuracy, AUC, and precision. The models were initialized with ImageNet weights, and the initial layers were frozen to ensure that only the weights of the unfrozen layers were updated during training. This approach allows the models to preserve generic features while learning domain-specific representations through the unfrozen layers. We replaced the final layer with a dense layer using the softmax activation function for multiclass classification.

All the models utilized a similar hyper-parameter setup: employing categorical cross-entropy as the loss function, restricting the computation of the loss (for training and validation) to the labels from the target domain. Due to its effectiveness in multi-class classification and its stability in optimizing deep learning models, cross-entropy loss is particularly suitable for our task as it ensures proper probability calibration through the softmax activation function, which is critical for accurate chest X-ray classification. Additionally, in adversarial domain adaptation, the classifier and domain discriminator require a robust loss function that facilitates smooth gradient flow and stable convergence. While alternatives such as Kullback-Leibler (KL) divergence and mean squared error (MSE) were considered, KL divergence is more suited for comparing probability distributions rather than direct classification, and MSE tends to suffer from gradient saturation in categorical tasks. Empirical evaluations confirmed that cross-entropy loss led to faster convergence and improved classification accuracy in both source and target domains, making it the optimal choice for our framework. Stochastic gradient descent was used as the optimizer, with an initial learning rate of 0.001, and a mini-batch size of 32. After each epoch, we reduced the learning rate by a factor of 10 if the validation loss did not improve. The batch images underwent data augmentation using common techniques such as scaling, rotation around the image center, translation relative to the image extent, and zooming. Finally, all the models were evaluated on the target domain data. Five-fold cross-validation was used, utilizing 10% of the samples from the source data as a validation set. In our fivefold cross-validation setup, the source data is split into five equal parts. In each iteration, four parts are used for training, and the remaining part is used for validation. This process is repeated five times, ensuring that each part serves as the test set once.

The fivefold cross-validation is performed only on the three source datasets to ensure robust model training and validation. The Nigerian dataset is not included in the data splitting for fivefold cross-validation. It is used separately as the target test dataset to evaluate the performance of the models.

To ensure the ADA model does not overfit during training, we employed multiple regularization strategies. Dropout layers were introduced in the classifier and feature extractor to prevent co-adaptation of neurons. Additionally, L2 regularization (weight decay) was applied to stabilize weight updates. Early stopping was used based on validation loss to halt training when no further improvement was observed. Furthermore, extensive data augmentation techniques, including random rotations, scaling, translations, and flipping, were incorporated to increase dataset diversity. Importantly, our adversarial domain adaptation strategy inherently mitigates overfitting by enforcing domain-invariant feature learning through the domain discriminator, ensuring the model does not memorize dataset-specific artifacts but instead generalizes effectively across populations

Proposed adversarial domain adaptation

Figure 5 illustrates the general workflow of our proposed ADA technique, as a two-stage process. In the first stage, the feature extractor is trained on the source dataset using a cross-entropy loss function to predict the labels in the source domain. This step is performed independently without any adversarial training. The 20% of the target dataset is used to align feature distributions. The remaining 80% of the target dataset is reserved exclusively for testing. The goal of aligning feature distributions using adversarial training is to make the feature distributions of the source and target domains similar. This alignment helps the model generalize better to the target domain by ensuring that the learned features are domain-invariant. In the second stage, the domain adaptation begins. The feature extractor is frozen, and an adversarial domain discriminator is introduced to distinguish between the source and target domains. This design choice ensures that the model retains general feature representations relevant to medical imaging while preventing catastrophic forgetting. Freezing the feature extractor also enhances adversarial training by providing a stable feature space, allowing the domain discriminator to focus on aligning the target domain features effectively. However, we acknowledge that this approach may limit adaptability to significantly different target domains. The labeled target data is incorporated in the process for supervised adaptation. By separating the data used for feature alignment and testing, we ensured that our evaluation metrics accurately reflect the generalization capability of the models to new, unseen data from the Nigerian population.

The main idea of leveraging adversarial learning is to create domain-invariant feature representations. This is achieved in three-steps:

1. Feature Extractor and Label Predictor: This step works similarly to a conventional deep learning architecture. The feature extractor transforms the input data (images) into a feature space, while the label predictor performs the primary task of image classification.
2. Domain Discriminator: A neural network performs the task of determining the domain (source or target) of the feature representations extracted by the feature extractor.
3. Classification head: the extracted features are fed into the task classifier, which is trained to perform the main task (classification) using labeled source domain data. The goal for the task classifier is to learn task-relevant features that also generalize well to the target domain.

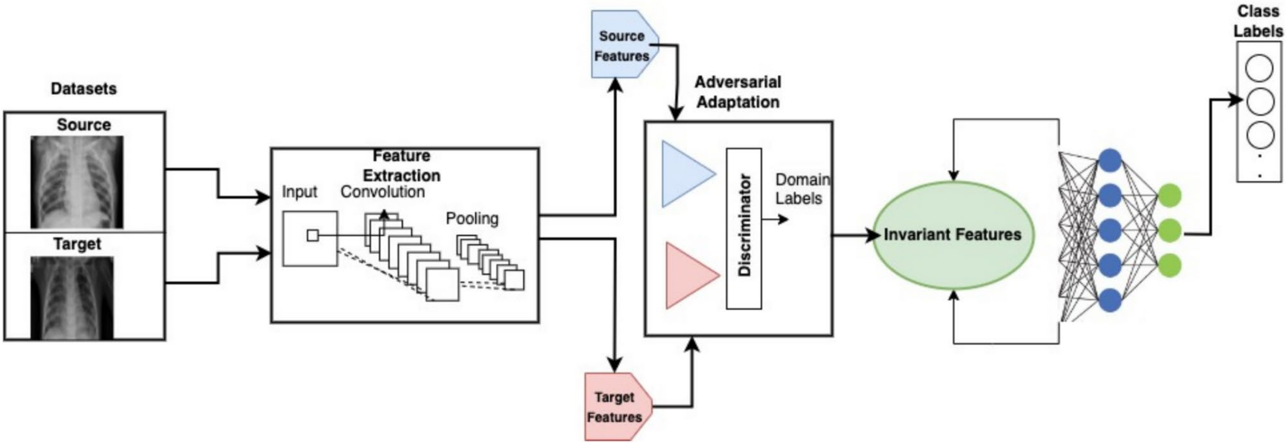


Fig. 5. Architecture of our proposed Adversarial Domain Adaptation (ADA). The input data comes from two distinct datasets (labeled source and target) and is fed into the feature extraction block. Then, a discriminator model performs adversarial training and generates invariant features before being passed to a classifier. The Domain Adversarial (DA) training encourages the feature extractor to learn domain-invariant representations, enhancing the model's ability to generalize across domains.

Domain adaptation is achieved through adversarial learning. Specifically, the feature distribution between source and target domains is effectively aligned through adversarial adaptation in the shared feature space. The feature extractor learns to produce domain-invariant features, making the model adaptable to new domains different from the source population datasets, without requiring extensive retraining of the model on the target datasets.

While Adversarial Discriminative Domain Adaptation (ADDA) is popularly used in unsupervised settings where the datasets are unlabelled[56], we modify the architecture and the loss function to propose a supervised ADA variant. In our proposed approach, we leverage adversarial training to align the feature distributions of the source and target domains. The domain discriminator's loss function is updated based on its ability to distinguish between features from the source and target domains. This adversarial objective ensures that the feature extractor learns domain-invariant features. For the classification head, the objective is to minimize the combined classification loss from both domains. The concept can be formalized as follows:

We denote with (X_s, Y_s) the labelled source dataset and with (X_t, Y_t) the labelled target dataset. In our proposed supervised ADA, we utilize a series of loss functions to ensure robust domain adaptation. The classification loss for the source domain L^s_{cls} is defined as

$$L^s_{\text{cls}} = \mathbb{E}_{(x_s, y_s) \sim (X_s, Y_s)} \left[\ell(C(F_s(x_s)), y_s) \right], \tag{1}$$

where the expectation $\mathbb{E}$ is computed over the samples (x_s, y_s) from the source domain distribution, ℓ represents the cross-entropy loss, *C* represents the shared classifier that will predict the labels from the extracted features, F_s is a feature extractor for the source. The target domain loss function L^t_{cls} is defined analogously.

The domain discrimination loss L_{adv} is defined as

$$L_{\text{adv}} = -\mathbb{E}_{x_s \sim X_s} \left[\log D(F_s(x_s)) \right] - \mathbb{E}_{x_t \sim X_t} \left[\log(1 - D(F_t(x_t))) \right] \tag{2}$$

where *D* is a domain discriminator intended to distinguish between source and target features. The goal of this loss is to align the feature distributions, making the features extracted from F_s and F_t indistinguishable.

The fooling loss L_{fool} is defined as

$$L_{\text{fool}} = -\mathbb{E}_{x_t \sim X_t} \left[\log D(F_t(x_t)) \right] \tag{3}$$

with the intention to maximize the confusion of the discriminator and the adversarial system.

The labeled target data is used to train the feature extractors F_s and F_t and the classifier C from the minimization of the total loss

$$L_{\text{total}} = L^s_{\text{cls}} + \lambda L^t_{\text{cls}} + \beta(L_{\text{adv}} + L_{\text{fool}}). \tag{4}$$

The parameters λ and β balance the contribution of the target and adversarial losses to the total loss.

This comprehensive loss formulation ensures effective supervised domain adaptation by aligning feature distributions while leveraging labeled data from both source and target domains. We compute the classification loss using labeled samples from both the source and target datasets. The loss function penalizes the model for incorrect classifications across all labels present in both domains. This dual contribution of domain discriminator loss from the target domain and classification loss from both domains ensures robust adaptation and improved generalization across diverse populations.

```
Input: Source data $(X_s, Y_s)$, target data $(X_t, Y_t)$, feature extractors $F_s$, $F_t$, classifier $C$,
domain discriminator $D$, hyperparameters $\lambda$, $\beta$.
Output: Trained feature extractors $F_s$, $F_t$ and classifier $C$.
Initialize feature extractors $F_s$, $F_t$, classifier $C$, and domain discriminator $D$.
while not converged
  Step 1: Train source classifier
    $L^s_{\text{cls}} \leftarrow \mathbb{E}_{(x_s,y_s)\sim(X_s,Y_s)}[\ell(C(F_s(x_s)),y_s)]$
    Update $F_s$ and $C$ to minimize $L^s_{\text{cls}}$
  Step 2: Train target classifier
    $L^t_{\text{cls}} \leftarrow \mathbb{E}_{(x_t,y_t)\sim(X_t,Y_t)}[\ell(C(F_t(x_t)),y_t)]$
    Update $F_t$ and $C$ to minimize $L^t_{\text{cls}}$
  Step 3: Train domain discriminator
    $L_{\text{adv}} \leftarrow -\mathbb{E}_{x_s\sim X_s}[\log D(F_s(x_s))] - \mathbb{E}_{x_t\sim X_t}[\log(1-D(F_t(x_t)))]$
    Update $D$ to minimize $L_{\text{adv}}$
  Step 4: Train target feature extractor to fool discriminator
    $L_{\text{fool}} \leftarrow -\mathbb{E}_{x_t\sim X_t}[\log D(F_t(x_t))]$
    Update $F_t$ to minimize $L_{\text{fool}}$
  Step 5: Combine losses
    $L_{\text{total}} \leftarrow L^s_{\text{cls}} + \lambda L^t_{\text{cls}} + \beta(L_{\text{adv}} + L_{\text{fool}})$
    Update $F_s$, $F_t$, $C$, and $D$ to minimize $L_{\text{total}}$
Return Trained $F_s$, $F_t$, and $C$.
end
```

Algorithm 1. Supervised Domain Adaptation using ADA.

The proposed method is described in Algorithm 1, which provides a step-by-step procedure for training the feature extractors, the classifier, and the domain discriminator using supervised domain adaptation within the ADA framework.

The algorithm begins with initializing feature extractors for both the source and target domains, a shared classifier, and a domain discriminator. Then, the method operates in an iterative manner, updating the source classifier to minimize the classification loss on the labeled source data (L_s^{cls}), then, it updates the target classifier using labeled target data to minimize the corresponding target classification loss (L_t^{cls}). Subsequently, the domain discriminator is updated to distinguish between source and target domain features (L_{adv}), and finally, the target feature extractor is updated to fool the discriminator by minimizing L_{fool}. The overall loss is a weighted combination of classification and adversarial losses (L_{total}), which ensures alignment of feature distributions across domains while maintaining classification accuracy. The loop continues until meeting typical convergence criteria, at which point the trained feature extractors and classifier are returned.

Experimental results and discussions

To evaluate the effectiveness of our domain adaptation approach, we used the Nigerian chest X-ray dataset for both feature alignment and testing. Specifically, the 80% of the Nigerian dataset was employed as test data, while the 20% was used during the feature alignment process. This subset was used to train the domain discriminator and align the feature distributions between the source and target domains.

We assess the performance of the DenseNet models trained in the baseline datasets and tested in our Nigerian dataset. Tables 1 and 2 gather the performance results of the three models tested on both the source and target domains, respectively. Our analysis reveals a significant decline in performance, with accuracies ranging from 79.70 to 96.70% in the source domains, contrasting sharply with the 62.45–71.70% accuracy observed in the

Dataset	Accuracy (%)	Precision (%)	F1-score (%)	AUC
China	79.58 ± 0.72	74.0 ± 3.74	78.20 ± 2.74	0.88 ± 0.14
Vietnam	96.32 ± 0.12	93.0 ± 0.44	96.12 ± 0.74	0.97 ± 0.04
Doha	94.89 ± 1.74	92.22 ± 7.74	95.07 ± 1.16	0.95 ± 0.03

Table 1. Performance of DenseNet models trained and tested on China, Vietnam, and Doha datasets, respectively.

Dataset	Accuracy (%)	Precision (%)	F1-score (%)	AUC
China	62.45 ± 3.31	54.34 ± 3.33	60.45 ± 1.08	0.65 ± 0.02
Vietnam	66.02 ± 2.77	45.67 ± 2.08	65.43 ± 3.22	0.66 ± 0.03
Doha	71.70 ± 2.54	68.41 ± 8.21	69.00 ± 5.05	0.73 ± 0.04

Table 2. Performance of DenseNet model trained on China, Vietnam, and Doha datasets and tested on the Nigerian dataset.

Dataset	Accuracy (%)	Precision (%)	F1-score (%)	AUC
China± Nigeria	81.85 ± 8.59	79.07 ± 3.47	83.72 ± 2.28	0.93 ± 0.03
Vietnam→ Nigeria	89.40 ± 0.44	83.02 ± 9.89	90.01± 0.61	0.92 ± 0.08
Doha→ Nigeria	88.70 ± 7.46	78.44 ± 3.17	70.70 ± 5.46	0.94 ± 0.16
China + Doha +				
Vietnam → Nigeria	90.08 ± 2.25	87.29 ± 0.34	89.75± 0.93	0.96 ± 0.01

Table 3. Performance of ADA evaluated on the Nigerian dataset.

Reference	Method	Accuracy	Sensitivity	Specificity	AUC
Tang et al.[25]	TUNA-NET	93.01	92.09	91.10	0.96
He et al.[26]	WDDM	74.03	73.08	74.04	0.80
Thiam et al.[62]	UDA	86.83	85.71	87.97	0.89
Tzeng et al.[56]	ADDA	88.01	88.01	88.02	0.91
Proposed Method	ADA	90.08	87.29	89.75	0.96

Table 4. Comparison of competing state-of-the-art methods evaluated with accuracy, sensitivity, specificity, and AUC on the Nigerian dataset.

target domain. This notable decrease underscores a cross-population domain shift among the datasets, with the China model demonstrating the most pronounced decline. Consequently, it becomes evident that DenseNet struggles to generalize effectively to our target domain when solely trained on the baseline source domains.

Table 3 presents the results of our proposed model and state of the art metthods on the Nigerian test set. The metrics show considerable performance improvement across all the datasets, with an accuracy now ranging from 81.85 to 89.4% when tested in the target domain. The accuracy of the ADA model evaluated on the Nigerian test set reached 90.08%, close to Vietnam and Doha's performance. These results indicate the effectiveness of ADA on handling the cross-population domain shifts.

To further validate the performance of our proposed approach, we compared the ADA results in our dataset with competing state-of-the-art techniques: Task-oriented UNsupervised Adversarial Network (TUNA-NET[25]), Wasserstein Distance and Discrepancy Metric (WDDM[26]), Unsupervised Domain Adaptation (UDA[62]), and Adversarial Discriminative Domain Adaptation (ADDA[56]). The results are presented in Table 4.

Our proposed ADA method obtains competitive results with an accuracy of 90.08%, sensitivity of 87.29%, specificity of 89.75%, and an AUC of 0.96, outperforming most of the established state of the art methods. Only TUNA-NET showed a superior performance. TUNA-NET uses a CycleGAN to create synthetic images from the source domain while preserving class-specific semantic information of lesions or abnormalities. The synthetic images are used in the process of distribution alignment, taking advantage over other methods which rely on extracting invariant features. However, the method is computationally more expensive and relies on generating high quality synthetic images. While ADDA and UDA obtain also competitive results, ADA's superior specificity and AUC indicate a better capacity to generalize and accurately classify chest X-ray images across domains. The balance between sensitivity and specificity underlines ADA's effectiveness for domain adaptation in medical image analysis within our context.

Finally, we compared the effectiveness of ADA to competing techniques such as Multi-task learning (MTL)[63], Continual Learning (CL)[16] and Unsupervised Domain Adaptation (UDA)[62]. For MTL, we introduced an additional task of domain classification alongside disease classification[63], leveraging a shared DenseNet-201 backbone with task-specific output heads. Sharing the model across tasks may enhance the feature extraction process and improve generalization across datasets. For CL, we adopted a joint training framework[16], where the model was first trained on the Vietnam dataset, followed by Qatar, China, and Nigeria datasets, simulating real-world scenarios of continual data availability and domain shifts. Table 5 shows the performance metrics compared with our approach.

Multi-task learning (MLT) showed the lowest performance gains across our diverse dataset scenarios, likely due to the shared representations failing to enhance generalization across domains. In contrast, Continual Learning (CL) delivered reasonable performance but faced challenges with catastrophic forgetting, especially

Method	Accuracy	Sensitivity	Specificity	AUC
MTL	77.96	79.88	77.96	0.73
CL	83.76	76.69	88.68	0.91
UDA	86.83	85.71	87.97	0.89
ADA	90.08	87.29	89.75	0.96

Table 5. Comparison of the performance of our proposed ADA with multi-task learning (MTL) and continual learning (CL) on the Nigerian dataset.

when transitioning between different datasets. Nonetheless, CL managed to achieve competitive results by retaining knowledge from previously learned tasks. Our proposed ADA approach proved to be the most effective for domain adaptation, successfully aligning feature distributions across domains. Its superior performance on the target Nigerian dataset, compared to both MLT and CL, underscores ADA's effectiveness in mitigating domain shifts and enhancing cross-domain robustness.

Another notable strength of ADA is its scalability with larger and more diverse datasets. Compared to UDA and TUNA-NET, ADA benefits from supervised target domain data, making it more robust in real-world clinical settings where some labeled data is available. Unlike TUNA-NET, which depends on synthetic image generation for domain alignment, ADA operates at the feature level, improving computational efficiency as dataset size increases. While UDA scales well due to its unsupervised nature, the lack of labeled target data can limit performance in highly diverse datasets. Beyond the baselines evaluated in this study, several other domain adaptation techniques exists. Methods such as Maximum Mean Discrepancy (MMD)[26,64] and Correlation Alignment (CORAL)[65] offer non-adversarial approaches to domain adaptation by aligning feature distributions statistically. However, these methods may struggle with complex feature transformations needed for robust adaptation. Future enhancements to ADA could include progressive domain adaptation, where models adapt to new datasets incrementally, and self-supervised learning, reducing reliance on labeled target data while improving feature generalization across populations. These improvements would enhance ADA's adaptability to the rapidly evolving landscape of medical imaging

Conclusion and future work

In this work, we investigated the challenges of domain shifts due to cross-population faced by modern deep-learning models in a relevant medical domain such as X-ray classification. Our research introduces a new chest X-ray data for the Nigerian population and uncovers cross-population domain shifts in deep-learning-based X-ray classification models. We investigate the extent of domain shift between different sources and target populations and propose the use of supervised domain adversarial networks as a domain adaptation strategy. Through empirical evaluation and analysis, we have quantified the extent of the downgrade in performance due to the domain shifts. Our results revealed notable disparities in classification performance across the different adaptation scenarios. The scenarios selected for this study were particularly challenging due to the source and target populations exhibiting substantial demographic and clinical diversity, with evident shifts in the domain. These findings underscore the necessity of domain adaptation techniques for enhancing model generalizability across diverse populations. By thoroughly characterizing domain shift, we illuminate the path toward mitigating its effects and improving the robustness of chest X-ray image classification models.

Our solution leverages the adversarial adaptation framework ADA to address the challenge. We showed the effectiveness of ADA in mitigating domain shift and greatly improving classification performance for diverse patient populations, contributing to the advancement of AI-driven healthcare solutions.

While this study evaluates the ADA model using the Nigerian dataset as the target domain, future work will focus on testing the approach on additional external datasets from different healthcare systems. Expanding the evaluation to datasets from North America, Europe, and other regions would provide further validation of ADA's robustness and generalizability. Additionally, incorporating datasets with varying imaging modalities and disease distributions could help assess the adaptability of ADA in more complex real-world settings. These extensions will enhance the applicability of ADA in global AI-driven healthcare solutions.

Data availability

The dataset is released for public use under Apache license 2 on Kaggle.com. Dataset link: https://www.kaggle.com/datasets/aminumusa/nigeria-chest-x-ray-dataset.

Code availability

The supervised adversarial adaptation model (ADA) developed during this study can be made available from the corresponding author on a reasonable request.

Received: 8 January 2025; Accepted: 20 March 2025
Published online: 03 April 2025

References

1. Bajwa, J., Munir, U., Nori, A. & Williams, B. Artificial intelligence in healthcare: Transforming the practice of medicine. *Future Healthc J* **8**, e188–e194. https://doi.org/10.7861/fhj.2021-0095 (2021).

2. Singh, V. K. *et al.* A computer-aided diagnosis system for breast cancer molecular subtype prediction in mammographic images. *Elsevier eBooks* 153–178, https://doi.org/10.1016/b978-0-12-819740-0.00008-5 (2021).
3. Reardon, S. Rise of robot radiologists. *Nature* **576**, S54–S58. https://doi.org/10.1038/d41586-019-03847-z (2019).
4. Ricci, A., Echeveste, R. & Ferrante, E. Addressing fairness in artificial intelligence for medical imaging. *Nature Commun.* https://doi.org/10.1038/s41467-022-32186-3 (2022).
5. Vidal, C. Chest X-ray, https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/chest-xray (2019).
6. Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **19**, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442 (2017).
7. Luo, Y., Zheng, L., Guan, T., Yu, J. & Yang, Y. Category-level adversaries for semantics consistent domain adaptation. arXiv.org (Taking a closer look at domain shift, 2018).
8. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255 (IEEE, 2009).
9. Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K. G. & Murphy, K. Deep learning for chest X-ray analysis: A survey. *Med. Image Anal.* **72**, 102125. https://doi.org/10.1016/j.media.2021.102125 (2021).
10. Wang, X. et al. 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 (CVPR)* https://doi.org/10.1109/CVPR.2017.369 (2017).
11. Johnson, A. E. W. et al. Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. *Sci. Data* **6**, 317. https://doi.org/10.1038/s41597-019-0322-0 (2019).
12. Hong, J., Yu, S.C.-H. & Chen, W. Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning. *Appl. Soft Comput.* **121**, 108729. https://doi.org/10.1016/j.asoc.2022.108729 (2022).
13. Xing, F., Bennett, T. D. & Ghosh, D. Adversarial domain adaptation and pseudo-labeling for cross-modality microscopy image quantification. *Lect. Notes Comput. Sci.* **11840**, 740–749. https://doi.org/10.1007/978-3-030-32239-7_82 (2019).
14. Musa, A., Ibrahim Adamu, M., Kakudi, H. A., Hernandez, M. & Lawal, Y. Analyzing cross-population domain shift in chest X-ray image classification and mitigating the gap with deep supervised domain adaptation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 585–595 (Springer, 2024).
15. Weninger, L., Liu, Q. & Merhof, D. Multi-task learning for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5*, 327–337 (Springer, 2020).
16. Lenga, M., Schulz, H. & Saalbach, A. Continual learning for domain adaptation in chest X-ray classification (2020).
17. Kermany, D. S. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **172**, 1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010 (2018).
18. Wani, N. A., Kumar, R. & Bedi, J. Deepxplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. *Comput. Methods Programs Biomed.* **243**, 107879 (2024).
19. Wani, N. A., Kumar, R., Bedi, J., Rida, I. *et al.* Explainable ai-driven iomt fusion: Unravelling techniques, opportunities, and challenges with explainable ai in healthcare. *Information Fusion* 102472 (2024).
20. Rasool, N., Bhat, J. I., Wani, N. A., Ahmad, N. & Alshara, M. Transresunet: Revolutionizing glioma brain tumor segmentation through transformer-enhanced residual unet. *IEEE Access* (2024).
21. Miglani, A. Fga-net: Feature-gated attention for glioma brain tumor segmentation in volumetric MRI images. *Art. Intell. Knowl. Process.* 66 (2024).
22. Musa, A., Adam, F. M., Ibrahim, U. & Zandam, A. Y. Learning from small datasets: An efficient deep learning model for covid-19 detection from chest X-ray using dataset distillation technique. In *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)*, 1–6, https://doi.org/10.1109/NIGERCON54645.2022.9803131 (2022).
23. Won Jo, S. & Seok, J. A study on deep learning-based classification for pneumonia detection. In *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 1496–1498, https://doi.org/10.1109/ICTC55196.2022.9952562 (2022).
24. Vinoth, R., Subalakshmi, S. & Thamaraichandra, S. Pneumonia detection from chest X-ray using alexnet image classification technique. In *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1307–1312, https://doi.org/10.1109/ICICCS56967.2023.10142405 (2023).
25. Tang, Y., Tang, Y., Sandfort, V., Xiao, J. & Summers, R. M. Tuna-net: Task-oriented unsupervised adversarial network for disease recognition in cross-domain chest X-rays. arXiv (Cornell University) https://doi.org/10.48550/arxiv.1908.07926 (2019).
26. He, B., Chen, Y., Zhu, D. & Xu, Z. Domain adaptation via Wasserstein distance and discrepancy metric for chest X-ray image classification. *Sci. Rep.* https://doi.org/10.1038/s41598-024-53311-w (2024).
27. Ranjan, V., Harit, G. & Jawahar, C. Domain adaptation by aligning locality preserving subspaces. https://doi.org/10.1109/ICAPR.2015.7050715 (2015).
28. arXiv.org. On the limits of cross-domain generalization in automated X-ray prediction. (MIDL, 2020).
29. Rajpurkar, P. *et al.* Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv (Cornell University) https://doi.org/10.48550/arxiv.1711.05225 (2017).
30. Shelke, A. et al. Chest X-ray classification using deep learning for automated covid-19 screening. *Sn Comput. Sci.* **2**, 300. https://doi.org/10.1007/s42979-021-00695-5 (2021).
31. Chen, C. *Improving the Domain Generalization and Robustness of Neural Networks for Medical Imaging*. Ph.D. thesis, Imperial College London (2021).
32. Drukker, K. et al. Toward fairness in artificial intelligence for medical image analysis: Identification and mitigation of potential biases in the roadmap from data collection to model deployment. *J. Med. Imaging* **10**, 061104–061104 (2023).
33. Hassan, E., Saber, A. & Elbedwehy, S. Knowledge distillation model for acute lymphoblastic leukemia detection: Exploring the impact of nesterov-accelerated adaptive moment estimation optimizer. *Biomed. Signal Process. Control* **94**, 106246. https://doi.org/10.1016/j.bspc.2024.106246 (2024).
34. Saber, A., Elbedwehy, S., Awad, W. A. & Hassan, E. An optimized ensemble model based on meta-heuristic algorithms for effective detection and classification of breast tumors. *Neural Comput. Appl.* 1–14 (2024).
35. Zhao, T. Seismic facies classification using different deep convolutional neural networks. *Seg Tech. Program Expand. Abstr.* https://doi.org/10.1190/segam2018-2997085.1 (2018).
36. Ganin, Y. et al. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **17**, 1–35 (2016).
37. Ganin, Y. & Lempitsky, V. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, 1180–1189 (PMLR, 2015).
38. He, G., Liu, X., Fan, F. & You, J. Classification-aware semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 964–965 (2020).
39. Liu, X. *et al.* Data augmentation via latent space interpolation for image classification. In *2018 24th International Conference on Pattern Recognition (ICPR)*, 728–733 (IEEE, 2018).
40. Wachinger, C. et al. Domain adaptation for Alzheimer's disease diagnostics. *Neuroimage* **139**, 470–479 (2016).
41. Feng, Y. et al. Deep supervised domain adaptation for pneumonia diagnosis from chest S-ray images. *IEEE J. Biomed. Health Inform.* **26**, 1080–1090. https://doi.org/10.1109/jbhi.2021.3100119 (2022).

42. Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y. & Ghassemi, M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat. Med.* **27**, 2176–2182. https://doi.org/10.1038/s41591-021-01595-0 (2021).
43. Guan, H. & Liu, M. Domain adaptation for medical image analysis: A survey. *IEEE Trans. Biomed. Eng.* **69**, 1173–1185 (2021).
44. Yu, M., Guan, H., Fang, Y., Yue, L. & Liu, M. Domain-prior-induced structural mri adaptation for clinical progression prediction of subjective cognitive decline. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 24–33 (Springer, 2022).
45. Pooch, E. H. P., Ballester, P. & Barros, R. C. Can we trust deep learning based diagnosis? the impact of domain shift in chest radiograph classification. In Petersen, J. *et al.* (eds.) *Thoracic Image Analysis*, 74–83 (Springer International Publishing, Cham, 2020).
46. Long, M., Cao, Y., Wang, J., Jordan, M. & Edu, J. Learning transferable features with deep adaptation networks (2015).
47. Thiam, P. et al. Unsupervised domain adaptation for the detection of cardiomegaly in cross-domain chest X-ray images. *Front. Artific. Intell.* **6**, 1056422 (2023).
48. Kamnitsas, K. *et al.* Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25*, 597–609 (Springer, 2017).
49. He, B., Chen, Y., Zhu, D. & Xu, Z. Domain adaptation via wasserstein distance and discrepancy metric for chest X-ray image classification. *Research Square (Research Square)* https://doi.org/10.21203/rs.3.rs-3122415/v1 (2023).
50. He, B., Chen, Y., Zhu, D. & Xu, Z. Domain adaptation via Wasserstein distance and discrepancy metric for chest X-ray image classification. *Sci. Rep.* **14**, 2690 (2024).
51. Ghafoorian, M. *et al.* Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, 516–524 (Springer, 2017).
52. Pan, S. J., Tsang, I. W., Kwok, J. T. & Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **22**, 199–210 (2010).
53. Sun, B., Feng, J. & Saenko, K. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30 (2016).
54. Wang, J., Chen, Y., Hao, S., Feng, W. & Shen, Z. Balanced distribution adaptation for transfer learning. In *2017 IEEE International Conference on Data Mining (ICDM)*, 1129–1134 (IEEE, 2017).
55. Madani, A., Moradi, M., Karargyris, A. & Syeda-Mahmood, T. Semi-supervised learning with generative adversarial networks for chest X-ray classification with ability of data domain adaptation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 1038–1042 (IEEE, 2018).
56. Tzeng, E., Hoffman, J., Saenko, K. & Darrell, T. Adversarial discriminative domain adaptation.
57. Stacke, K., Eilertsen, G., Unger, J. & Lundström, C. Measuring domain shift for deep learning in histopathology. *IEEE J. Biomed. Health Inform.* **25**, 325–336. https://doi.org/10.1109/jbhi.2020.3032060 (2021).
58. Sun, S., Shi, H. & Wu, Y. A survey of multi-source domain adaptation. *Inf. Fusion* **24**, 84–92 (2015).
59. Vindr-cxr: An open dataset and benchmarks for disease classification and abnormality localization on chest radiographs | vindr (2020).
60. Rahman, T. Covid-19 radiography database (2020).
61. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2261–2269, https://doi.org/10.1109/cvpr.2017.243 (2017).
62. Thiam, P. et al. Unsupervised domain adaptation for the detection of cardiomegaly in cross-domain chest X-ray images. *Front. Artific. Intell.* https://doi.org/10.3389/frai.2023.1056422 (2023).
63. Imran, A.-A.-Z. & Terzopoulos, D. Semi-supervised multi-task learning with chest X-ray images. In *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, 151–159 (Springer, 2019).
64. Wang, W. et al. Rethinking maximum mean discrepancy for visual domain adaptation. *IEEE Trans. Neural Netw. Learn. Syst.* **34**, 264–277. https://doi.org/10.1109/TNNLS.2021.3093468 (2023).
65. Ouyang, L. & Key, A. Maximum mean discrepancy for generalization in the presence of distribution and missingness shift. arXiv preprint arXiv:2111.10344 (2021).

Acknowledgements

The authors would like to acknowledge the anonymous reviewers for their willingness to revise the manuscript. We also would like to acknowledge the authors of the backbone deep-learning methods and the owners of the source datasets for the publication of the models and data. This work was partially supported by the national research grant PID2022-138703OB-I00 (Trust-B-EyE project) and Government of Aragon grant PROY_B50_24, and Group Reference T64_23R (COS2MOS research group). The funders had no role in study design, data collection, and analysis, the decision to publish, or the preparation of the manuscript.

Author contributions

All authors contributed to the study's conception and design. Conceptual research design was carried out by Monica Hernandez and Aminu Musa. Data collection, model development, and experimentation were performed by Aminu Musa and Rajesh Prasad. Project supervision by Rajesh Prasad, funding acquisition by Monica Hernandez. The first draft of the manuscript was written by Aminu Musa and Monica Hernandez. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Additional information

Correspondence and requests for materials should be addressed to A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025