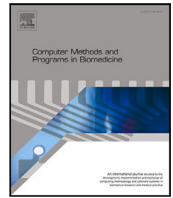




Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: <https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine>

Advancing chest X-ray diagnostics: A novel CycleGAN-based preprocessing approach for enhanced lung disease classification in ChestX-Ray14

Aya Hage Chehade^{a,*}, Nassib Abdallah^b, Jean-Marie Marion^a, Mathieu Hatt^b, Mohamad Oueidat^c, Pierre Chauvet^a^a LARIS, University of Angers, Angers, France^b LaTIM, INSERM UMR 1101, University of Brest, Brest, France^c Faculty of Technology, Lebanese University, Lebanon

ARTICLE INFO

Keywords:

ChestX-Ray14 dataset
Image preprocessing
Image classification
Deep learning
K-means clustering
CycleGAN
Convolutional Block Attention Module (CBAM)

ABSTRACT

Background and objective: Chest radiography is a medical imaging technique widely used to diagnose thoracic diseases. However, X-ray images may contain artifacts such as irrelevant objects, medical devices, wires and electrodes that can introduce unnecessary noise, making difficult the distinction of relevant anatomical structures, and hindering accurate diagnoses. We aim in this study to address the issue of these artifacts in order to improve lung diseases classification results.**Methods:** In this paper we present a novel preprocessing approach which begins by detecting images that contain artifacts and then we reduce the artifacts' noise effect by generating sharper images using a CycleGAN model. The DenseNet-121 model, used for the classification, incorporates channel and spatial attention mechanisms to specifically focus on relevant parts of the image. Additional information contained in the dataset, namely clinical characteristics, were also integrated into the model.**Results:** We evaluated the performance of the classification model before and after applying our proposed artifact preprocessing approach. These results clearly demonstrate that our preprocessing approach significantly improves the model's AUC by 5.91% for pneumonia and 6.44% for consolidation classification, outperforming previous studies for the 14 diseases in the ChestX-Ray14 dataset.**Conclusion:** This research highlights the importance of considering the presence of artifacts when diagnosing lung diseases from radiographic images. By eliminating unwanted noise, our approach enables models to focus on relevant diagnostic features, thereby improving their performance. The results demonstrated that our approach is promising, highlighting its potential for broader applications in lung disease classification.

1. Introduction

Lung diseases are categorized as the third leading cause of death worldwide [1], resulting in a total of five million deaths per year [2]. Pneumonia is one of the most common severe respiratory lung disease, and is considered the major cause of child fatalities around the world [3]. According to the World Health Organization, approximately 740,180 children under the age of five years died of pneumonia in 2019 [4]. Consequently, early diagnosis is crucial to reduce the life threatening aspect of lung diseases and to improve the chance of recovery and survival for people already affected, by allowing rapid and effective treatment.

Clinical exams for thorax diseases include chest X-ray (CXR), computed tomography (CT) scan, and magnetic resonance imaging (MRI).

As a simple, quick, accessible, non-invasive and cost-effective examination method, chest radiography, or chest X-ray, is the most commonly radiological imaging technique used to screen and diagnose various lung diseases [5–8], including pneumonia [9], as well as to evaluate severity and detect possible complications [10].

However, manual observation of CXR images is a time-consuming process and their interpretation can be challenging for radiologists due to the complex nature of chest X-rays [11]. Indeed, distinguishing different types of diseases from chest X-ray images is a difficult task for an expert due to the presence of hundreds of potential lung pathologies, some of which may have similar visual characteristics. This can lead to difficulties in interpretation, missed detections and therefore life-threatening diagnostic errors.

* Corresponding author.

E-mail address: aya.hagechehade@etud.univ-angers.fr (A. Hage Chehade).<https://doi.org/10.1016/j.cmpb.2024.108518>

Received 21 April 2024; Received in revised form 28 October 2024; Accepted 15 November 2024

Available online 25 November 2024

0169-2607/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

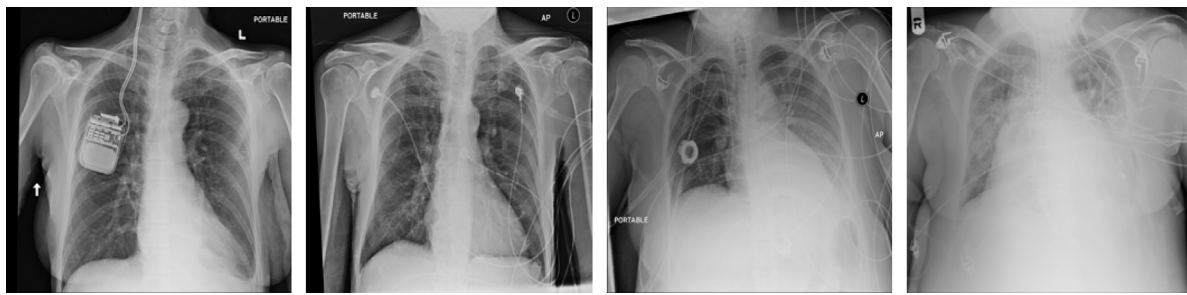


Fig. 1. Examples of chest X-ray images with artifacts and bad quality captures.

Recently, artificial intelligence (AI) technologies have been known for their ability to examine data faster, improve the efficiency of CXR analysis for the detection and classification of lung diseases and make decisions. As a result, there is a critical demand for an artificial intelligence system that can effectively detect chest-related illnesses [11–13]. Therefore, with the advancements of technology, Computer Aided Diagnosis (CAD) systems were proposed to analyze chest radiographs in less time and effort in order to reduce the workload of radiologists and to improve clinical diagnosis [14]. Machine learning (ML) and deep learning (DL) can play a significant role in accurate clinical diagnosis [15–17]. More recently, DL has been widely investigated due to its general applicability to problems involving automated feature extraction and classification. Deep learning-based algorithms have achieved satisfactory performance in a number of computer vision tasks [18,19], including image classification [14,20], disease detection [21–24], medical diagnosis [25,26] and healthcare analysis [27]. Specifically, DL algorithms have been widely used to detect and diagnose lung diseases through chest X-ray images [28].

However, none of the previous studies take into account the presence of artifacts in CXR images. Indeed, as shown in Fig. 1, X-ray images may contain artifacts such as irrelevant objects, medical devices, wires as well as electrodes, and may be low-quality. These objects can introduce unnecessary noise and compromise the quality of the X-ray image, making difficult the distinction of relevant anatomical structures. Indeed, noise makes the image more difficult to interpret, as it can mask the actual lung structures and the important anatomical areas, as well as it introduces irrelevant features within the X-ray image. When an artifact is not correctly identified as such, it can lead to diagnostic errors. For instance, a wire crossing the image can be mistaken for a pulmonary abnormality. This can potentially bias classification results and influence the performance of machine learning models, leading to incorrect medical decisions, such as unnecessary treatments or omission of necessary treatments. Ultimately, image quality and consideration of artifacts are crucial aspects in the interpretation of chest X-ray images, as they can influence the reliability of medical diagnosis. Therefore, it is essential to develop preprocessing strategies to minimize the effect of noise caused by the presence of artifacts in X-ray images. This will improve the classification performance of deep learning models, helping to establish more accurate and reliable diagnoses, and thus improving the quality of healthcare for patients.

It should be noted that this issue of lacking labels for images containing electronic components can arise in any database, emphasizing the general importance of clustering (unsupervised labeling) to the effective processing of image data, particularly in the medical field. This contributes to improved data quality, leading to more accurate and significant results in the process of classifying medical images, thus reinforcing the reliability and efficiency of diagnostic analyses in the healthcare field.

Our work's novelty lies in a unique preprocessing approach for chest X-ray images that distinguishes sharp images from those containing artifacts, subsequently reducing artifact noise by generating sharper images using a CycleGAN model.

Using the NIH ChestX-ray14 dataset, we evaluate the impact of this scientific contribution through the development of a classification model based on DenseNet and attention mechanisms. Notably, the proposed preprocessing approach may also prove valuable for detection and prediction tasks. In clinical practice, radiologists rely not only on image analysis but also on additional information to establish a diagnosis. To fully leverage all dataset information (i.e., age, gender, and view position), we propose a new architecture that integrates this supplementary data with the learned image representation. Experimental results demonstrate that reducing artifact noise in chest X-ray images, coupled with an attention mechanism and the incorporation of clinical variables, enables our approach to achieve high performance, surpassing state-of-the-art results in pneumonia classification. Additionally, we validated our pipeline on a broader set of lung diseases (14 diseases from the ChestX-ray14 dataset), where our approach also exhibits promising performance, outperforming results reported in the literature.

Our contributions are summarized as follows:

- We proposed a novel approach for preprocessing chest X-ray images that deal with artifacts, which begins by detecting images that contain electronic artifacts. Then, we reduce the impact of this electronic noise by generating sharper images using a CycleGAN model.
- We assessed the impact of Convolutional Block Attention Module (CBAM) in maximizing the use of the image's relevant aspects and selectively enhance features strongly related to the lesion area.
- We incorporated clinical characteristics such as patient age, gender and view position.
- We presented experiments on the ChestX-ray14 dataset. The experiments results demonstrated that our approach achieves high performance compared to state-of-the-art approaches.

2. Related work

With the advent of Deep Learning (DL) and the release of the large-scale CXR dataset [29], many researchers have focused on developing deep learning-based techniques for lung disease detection and classification [28].

Indeed, inspired by the original DenseNet model [30], Rajpurkar et al. [31] developed CheXNet, a 121-layer deep transfer learning model for pneumonia classification from the ChestX-ray14 dataset. As well as classifying pneumonia, they also demonstrated that their model can determine the most symptomatic areas of pneumonia in a chest X-ray. They compared their performance with that of radiologists and demonstrated that the model can outperform the average performance of radiologists in diagnosing pneumonia, achieving an area under the curve (AUC) of 76.8%. Yao et al. [17] proposed an LSTM-based approach which they applied to a variant of DenseNet for extracting features of CXRs, and obtained an AUC of 68.4% for pneumonia detection. Souid et al. [32] proposed the classification of lung pathologies in frontal chest X-rays using the MobileNet V2 model [33], supplemented

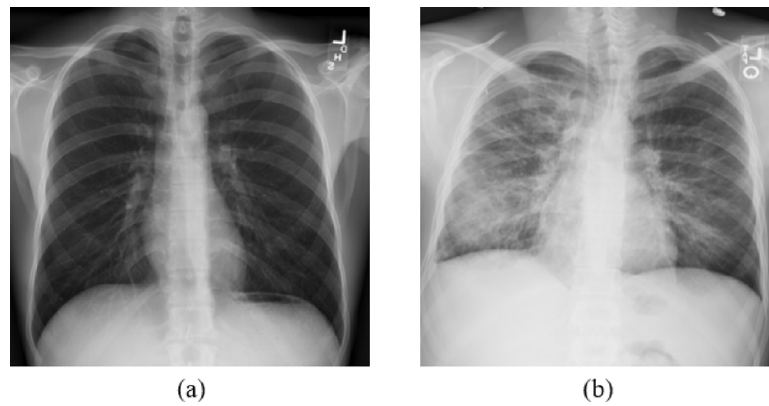


Fig. 2. Sample images of: (a) Normal CXR and (b) Pneumonia collected from the ChestX-ray14 dataset. The pathology in (b) can be easily distinguished from the presence of abnormal opacities in the pulmonary lobe.

by CNN layers. Their approach obtained an AUC of 73.3% for the classification of pneumonia.

Mann et al. [34] used three convolutional neural networks, namely DenseNet-121, ResNet-50, and EfficientNetB1. DenseNet-121 was used as a feature extractor, and a more intricate fully connected layer was trained on it. Only the fully connected layers were modified during the training process, and the rest of the DenseNet-121 architecture remained unchanged. The DenseNet-121 outperformed the other two models and achieved an AUC of 74.3% for the classification of pneumonia.

Some studies have explored the use of attention, demonstrating that forcing the classification model to focus on specific regions of the image yields better performance compared to the use of the entire image. Xu et al. proposed a classification scheme for chest X-ray images based on a multi-scale attention network [35]. The scheme iteratively fuses multi-scale information to focus on regions with high disease probability, to extract more meaningful information from the data. The method obtained an AUC of 77.7% for pneumonia classification. Ma et al. [36] proposed a cross-attention network scheme for thoracic disease classification, using an attention loss that forced the model to focus more accurately on pathogenic areas. The model achieved an AUC of 72.2% for the classification of pneumonia. Zhao et al. [37] used the DenseNet-121 model and incorporated the Convolutional Block Attention Module (CBAM) which includes both a channel attention module and a spatial attention module, achieving 76.92% for pneumonia classification.

In a unique study, Baltruschat et al. demonstrated that using a ResNet-38 architecture and incorporating non-image data gave the best results [38].

Although previous studies yielded good results, none have considered the presence of artifacts in CXR images. For this purpose, the main objective of our study is to develop a preprocessing approach to reduce the impact of artifacts noise on chest X-ray images.

3. Methods

3.1. Dataset

The ChestX-ray14 dataset [29] is extracted from the clinical PACS databases in the hospitals affiliated to National Institutes of Health Clinical Center and includes 112,120 frontal view chest X-ray images from 30,805 unique patients. Each CXR is labeled with binary labels for 14 different diseases. The CXR is labeled as “No finding” if none of these diseases were detected. All CXRs are in PNG format and have a size of 1024×1024 . Beyond the radiographic images themselves, the dataset contains additional patient data, including the gender and age of the patient as well as the view position of the X-ray (i.e. whether the X-ray images were acquired in the posterior–anterior (PA) or anterior–posterior (AP) position).

In this database, 322 images are labeled as pneumonia and 60,361 images are no finding. To address class imbalance, we down-sampled only the no finding images to obtain 354 samples. We chose to maintain a slightly higher number of no finding images compared to disease images to better represent the inherent distribution within the dataset, where no finding images are generally more prevalent. Therefore, we conducted our experiments on a total of 676 images and using their corresponding clinical variables. Sample of radiographic images of pneumonia and no finding that are collected from the ChestX-ray14 dataset are illustrated in Fig. 2.

Images with electronic components are not labeled in the dataset. The use of these images can introduce electronic noise during classification. For this reason, it is crucial to use clustering beforehand and efficiently differentiate between images with and without those artifacts. By identifying images containing electronic components, it becomes possible to reduce the impact of electronic noise, thereby improving classification performance.

3.2. Overview of the methodology

To perform the classification of pneumonia and normal images by addressing the issue of the presence of artifacts, a total of 676 X-ray images were used. Our methodology, as shown in Fig. 3, consists of four major steps: preprocessing, clustering (unsupervised labeling), CycleGAN and classification.

In the preprocessing phase, we used the Histogram Equalization technique to enhance the contrast and improve the visibility of details in an image. After that, we extracted features related to contours and texture from the images and then we performed the clustering using the K-means algorithm. This resulted in two clusters: the first one containing images with artifacts, and the second one containing the sharp images without artifacts. To reduce the noise effect caused by the presence of artifacts, a CycleGAN model was performed on the images from the first cluster containing images with artifacts to generate sharper images without artifacts. After that, the generated images are combined with the second cluster containing images without artifacts to perform the classification. Images were resized to 224×224 pixels, and normalized in the interval $[0; 1]$. To exploit all the information in the dataset, clinical and technical characteristics such as patient age, gender and view position were added.

20% of the data was used as test data, 10% as validation data and 70% was devoted to training the data (randomly chosen). We stratified the data to ensure that both classes are proportionally represented across the training, validation, and test sets. Thereafter, the classification was performed using the DenseNet-121 model with the Convolutional Block Attention Module (CBAM). This type of attention mechanism enables the model to maximize the use of relevant aspects

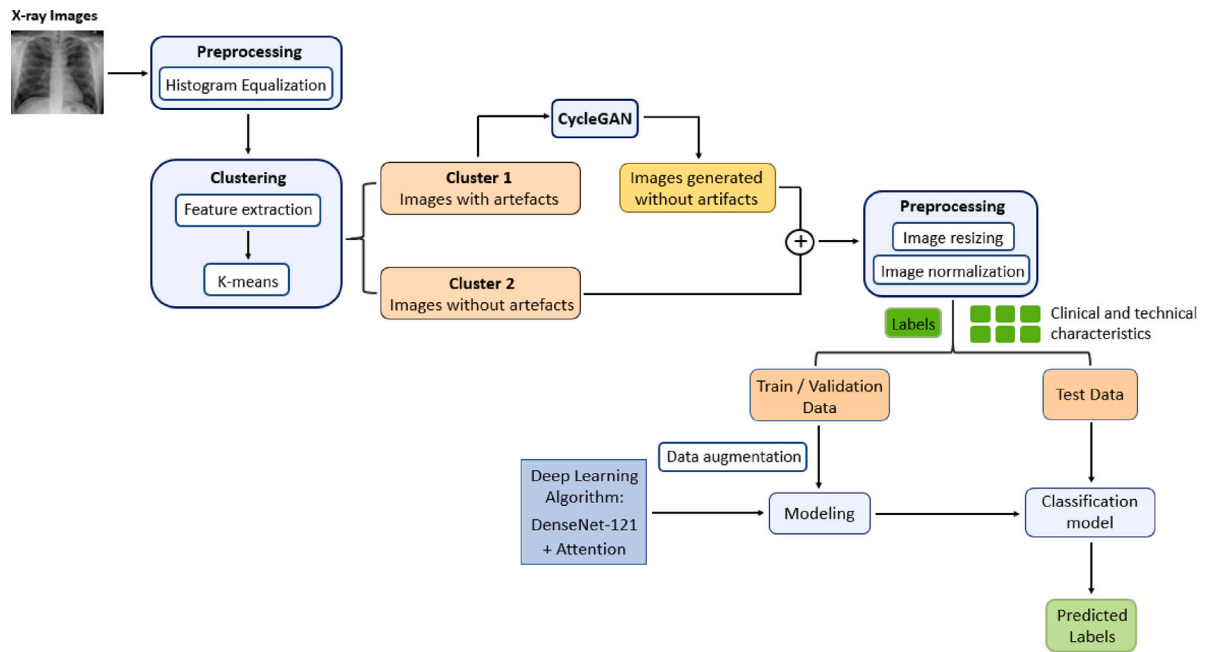


Fig. 3. Overview of the methodology that we propose for X-ray images classification.

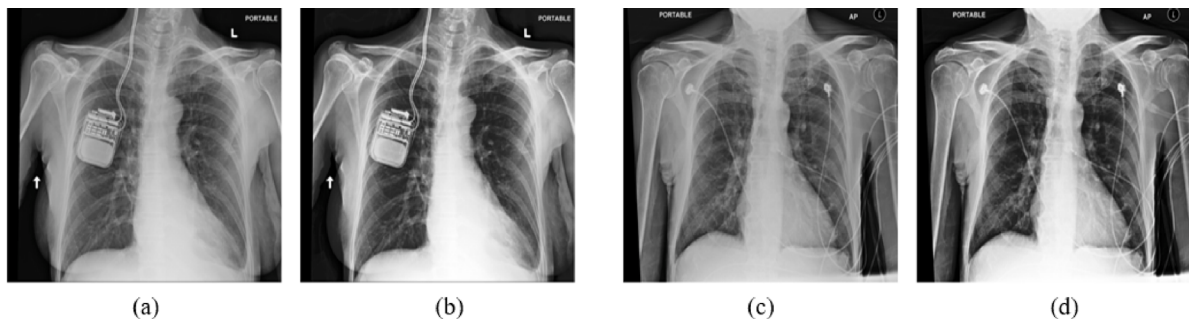


Fig. 4. A sample of chest X-ray image: (a) and (c) original image, (b) and (d) sharpened image using the Histogram Equalization technique.

of the image by selectively reinforcing features in the input data that are strongly related to the lesion zone [39].

The test set images were used to evaluate model performance, which was quantified by the Area Under the Curve (AUC). The AUC was chosen as the evaluation metric to compare our results with previous studies using the same dataset. In the literature, AUC remains a predominant metric for evaluating the performance of lung disease classification models, which justifies its choice in the context of our study. In addition, to further verify and validate the results, we computed other evaluation metrics such as Precision, Recall and F1-score for the different classes using the approach that gave the best results.

3.3. Image preprocessing

Image preprocessing is essential to improve image quality, enabling better detection of patterns and relevant features in the image by the learning algorithm. We enhanced the contrast of each image using the Histogram Equalization method.

The histogram equalization technique aims to distribute the gray levels uniformly within an image. Thus, each gray level has an equal chance of occurring. Histogram equalization changes the brightness and contrast of the dark and low contrast images to enhance image

quality [40–42]. This allows to enhance details, sharpen contours and improve the visual quality of the image. Hence, histogram equalization is an intensity transformation process. It gives important information about the grayscale and contrast of the image. Fig. 4 illustrates the result of enhancing a X-ray image using the Histogram Equalization method.

3.4. Clustering

After preprocessing, given that images containing electronic components are not labeled in the dataset and could introduce noise during classification, we proposed an image preprocessing module to separate the images into two clusters: one with artifacts and one without. This was achieved by first extracting relevant texture and contour features from the images, followed by clustering them into two groups using the K-means clustering method.

3.4.1. Features extraction

Features extraction involve identifying specific characteristics or patterns in the images that can be used for analysis. Feature extraction, especially related to the contours and texture of X-ray images, is crucial to interpret and understand the topology of the information

contained in the images. By identifying the subtle details of contours and texture, it enables K-means to accurately distinguish sharp images from those altered by artifacts. For this purpose, four characteristics were computed in this paper (resulting in a vector of 580,648 features for each image), including:

- **Variance of Gradients feature:** This feature calculates the image’s horizontal and vertical partial derivatives using the Sobel filter. These partial derivatives help quantify the intensity variations of pixels in the image and provide information on how brightness changes in different directions of the image. Then, it combines these derivatives to calculate the magnitude of the gradient. More precisely, it evaluates the variance of gradient magnitude values across the entire image. Regions of an image with a complex or rough texture may have a higher gradient variance than regions with a more homogeneous texture. Consequently, the variance of gradients feature can be sensitive to local texture variations. So this characteristic provides an indication of the overall variation in gradients within the whole image, which can be valuable for assessing the general texture of the image.
- **Lung sharpness feature:** This feature evaluates the sharpness of lung contours in a X-ray image. To do this, it uses the Sobel filter to compute the horizontal and vertical partial derivatives of the image, highlighting the variations in pixel intensity related to contours. It then combines these partial derivatives to calculate the gradient magnitude, which represents the sharpness of the image contours, measuring how sharp and precise the brightness changes are. The higher the gradient magnitude, the sharper and more well-defined the contours. To ensure that the sharpness measurement focuses specifically on the lung region of interest, it creates a mask that isolates this area. This eliminates unnecessary information outside the lungs, thus improving measurement efficiency. Once the lung mask is created, it is applied to the previously calculated contours. This means that only contours within the lung region are taken into account when assessing sharpness. A high value for this feature indicates that lung contours are well-defined and sharp. Reduced sharpness may be the result of artifacts or unwanted objects that affect image quality.
- **Black–white ratio feature:** This feature measures the distribution of black and white pixels in an image. By calculating the percentage of black pixels (corresponding to low intensity) and white pixels (corresponding to high intensity), this characteristic quantifies the density of dark structures compared to light areas within the image. In the context of CXR images, it can help differentiate lung areas (which are generally darker) from non-lung areas (which are generally lighter), thus contributing in the identification of lungs and electronic objects in the image. A high black ratio may indicate high tissue density, while a high white ratio may be associated with artifacts, overexposure or the presence of unwanted objects that modify the brightness distribution.
- **Histogram of Oriented Gradients (HOG) feature:** HOG features are a powerful representation of gradient patterns in an image. They are based on a process of dividing the image into cells and calculating the gradients in each cell, which allows capturing the local orientations of contours present in the image [43]. Then it computes the histogram of oriented gradients, normalizes the obtained results by a block-wise pattern, concatenates the 2×2 grid cells, and returns the HOG descriptor at each grid location [44]. Specifically, HOG features analyze how changes in brightness occur in different directions within the image. They are particularly effective for object detection and pattern recognition [45–47]. They can identify specific gradient patterns that are characteristic of certain structures or objects, making them a valuable tool for the segmentation and identification of elements of interest within X-ray images.

Table 1
Number of images of pneumonia and no finding classes in each cluster obtained using k-means clustering.

	Pneumonia	No finding	Total
Cluster 1: Images with artifacts	152	133	285
Cluster 2: Images without artifacts	170	221	391

3.4.2. K-means

After extracting key features from the X-ray images, these were normalized and used as input data for the unsupervised machine learning algorithm known as K-means clustering. The objective of K-means clustering is to group data by maximizing the similarity of features within groups and maximizing the differences between existing groups [48, 49].

In our context, this unsupervised learning method has the power to reveal hidden structures within our dataset, thus enabling the effective separation of sharp images from those with artifacts. In accordance with the objectives of this study, K-means allowed to efficiently group images based on their extracted features, creating two clusters: one cluster contained images that were perceived as sharp, while the other cluster contained images with artifacts and unwanted objects. This automated strategy proved to be a valuable asset in our research aiming to improve classification performance, and consequently the precision of medical diagnoses based on radiographic images. In our study, k-means was applied to both no finding and pneumonia images, resulting in two clusters for each class. Table 1 presents the number of images of pneumonia and no finding classes in each cluster that are obtained using k-means clustering algorithm.

To evaluate our unsupervised labeling approach, we labeled the images of pneumonia and no finding manually as with and without artifacts. These annotations were used as the reference ground truth for our evaluation. Then we computed the confusion matrix for each class that are shown in Fig. 5, and calculated the performance metrics for each class. The clustering of pneumonia obtained an accuracy of 83.54% and an F1-score of 84.37%. Also, the clustering of no finding achieved an accuracy of 82.49% and an F1-score of 85.91%. From the confusion matrix (a) and (b), we notice that only 53 and 62 samples out of 322 and 354 images were misclassified, respectively. Overall, we can say that the clustering (unsupervised labeling) is very accurate in identifying clear images from images containing artifacts, and this can be done in less time and effort than manual labeling by a physician and can be applied to a larger number of images.

It is crucial to emphasize that the obtained clusters cannot be directly combined due to potential bias that could affect later classification results. To address the issue of image quality variation, particularly between images with and without artifacts, we implemented a data harmonization approach. In our study, we utilized the CycleGAN model, known for its capability to transfer information from a reference domain to the target domain. This approach facilitates the extraction of common features between the two image clusters, enabling their integration while reducing noise from electronic artifacts. Specifically, CycleGAN allows us to adjust the characteristics of images in the first cluster (those with artifacts) to more closely resemble those in the second cluster (images without artifacts), thereby effectively minimizing noise caused by artifacts. This data harmonization technique significantly reduces the impact of artifacts, enhancing the overall quality of the images. By mitigating variations that could potentially impact classification performance, this approach aims to yield more reliable and accurate classification outcomes.

3.5. Cycle-Consistent Adversarial Network (CycleGAN)

Proposed by Zhu et al. [50], Cycle-Consistent Adversarial Network, often abbreviated as “CycleGAN”, is a specific variant of Generative

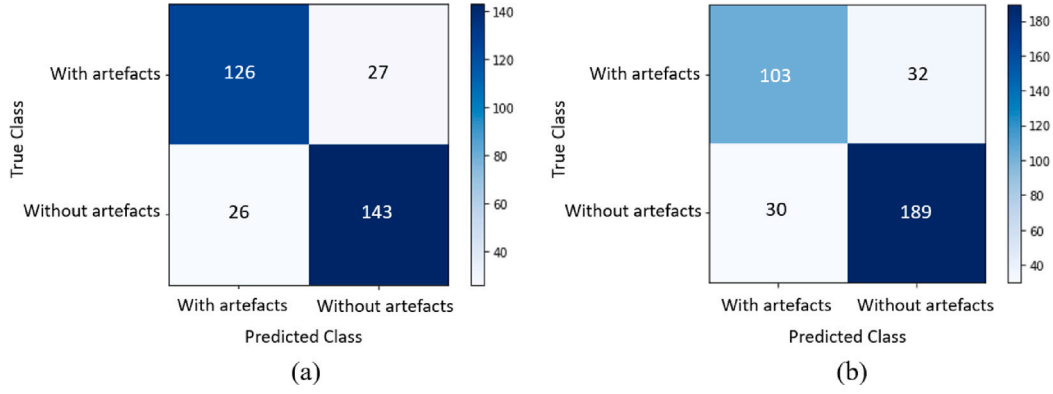


Fig. 5. The confusion matrix of clustering for: (a) pneumonia class and (b) no finding class.

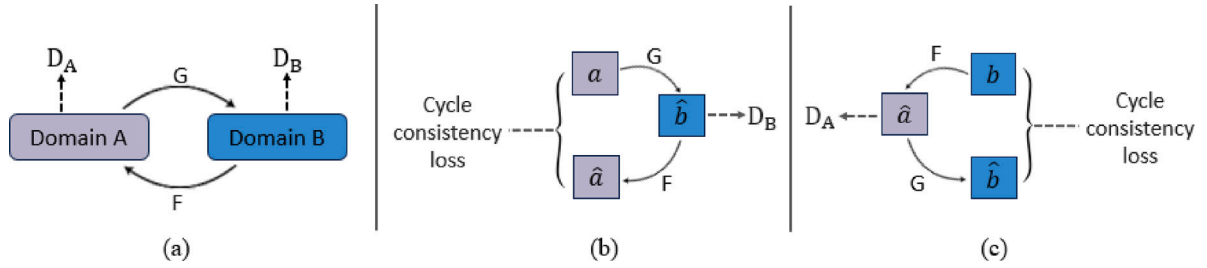


Fig. 6. (a) CycleGAN model contains two mapping functions $G : A \rightarrow B$ and $F : B \rightarrow A$, and associated adversarial discriminators D_B and D_A . D_B encourages G to translate A into outputs indistinguishable from domain B , and vice versa for F . (b) Forward cycle-consistency loss: for each image from the domain A : $a \rightarrow G(a) = \hat{b} \rightarrow F(G(a)) = \hat{a} \approx a$. (c) Backward cycle-consistency loss: $b \rightarrow F(b) = \hat{a} \rightarrow G(F(b)) = \hat{b} \approx b$.

Adversarial Networks (GANs) designed for unpaired image-to-image translation tasks.

As shown in Fig. 6, the CycleGAN is characterized for using two generators and two discriminators. The generator is tasked with producing synthetic images, whereas the discriminator evaluates whether these generated images possess a realistic appearance. Consequently, the generator is trained to create synthetic samples so realistic that they are deemed by the discriminator to be real samples [51]. Moreover, CycleGAN integrates the concept of “cycle consistency” into the GAN framework to ensure forward and backward cycle consistencies, and enable the translation of images from one domain (A) to another domain (B) without the need for paired training data. It enforces that an image that propagates through both generators will be reconstructed back to the original input image [52]. So CycleGAN has two types of losses: the cycle consistency loss and the adversarial loss. The adversarial loss is utilized across both mapping directions, leveraging the least-square loss to determine the difference between the expected and the predicted outputs [51]. CycleGAN also defines the identity loss. For example, assuming that the mapping function $G : A \rightarrow B$ is designed to translate images from domain A to domain B , it is plausible to assume that the function will also return a synthetic image from domain B when it receives a real image from domain B ($a \rightarrow F(a) \approx a$, and $b \rightarrow G(b) \approx b$). Consequently, it is also reasonable to assume that the two images should be the same.

Concerning the architectures, both generators are configured with an encoder-decoder architecture, while the discriminators utilize an encoder. Both generators use ResNet with nine residual blocks (denoted as ResNet-9). The discriminator has a 70×70 PatchGAN architecture which aims to classify whether 70×70 overlapping image patches are real or fake.

Fig. 7 presents the CycleGAN workflow to translate images with artifacts to images without artifacts. The architecture is composed of two generators and two discriminators. We have a X-ray image with artifacts taken from the set A containing a large number of images with artifacts. The generator G_{AB} converts this image into an image without

artifacts. The discriminator B takes this generated image and decides whether the generated image is fake or not based on observation of the large number of images from the set B containing a large number of images without artifacts. Thus, it produces an output of 0 or 1, depending on whether this particular image is close to an image without artifacts or not. Then, this generated image, intended to be classified as class B , is provided as input for the generator G_{BA} which reconverts this image without artifacts into an image with artifacts.

So G_{BA} performs the reverse task to generate a synthetic image intended to be classified as class A . Therefore, the final generated image, designated as cycle A , must closely resemble the original real input image from domain A . This reconstruction is evaluated with the cycle consistency loss which shows that if we have an image of A , convert it into an image of B and then convert it back into an image of A , how close these two images are to each other. If they are very close to each other, the quality of the reconstruction is very good.

This CycleGAN model is applied to the 285 images containing artifacts and unwanted objects in order to generate sharper images and reduce the effect of artifact noise in the images. Fig. 8 illustrates examples of image translation by the CycleGAN model, where the generated images show a significant reduction in the impact of artifacts and electronic elements which become almost invisible. Furthermore, the reconstructed images also demonstrate the model’s ability to restore the images to their original state, highlighting the efficiency of the CycleGAN model.

Given that we lack information about the images without artifacts, we evaluated the effectiveness of our preprocessing approach by assessing its impact on classification performance, with classification metrics serving as the primary evaluation criteria. For the rest of the study, the generated images are combined with the second cluster containing images without artifacts to perform the classification.

3.6. Classification

To perform the classification of normal and pneumonia images, we used the DenseNet-121 model. We resized the images to 224×224

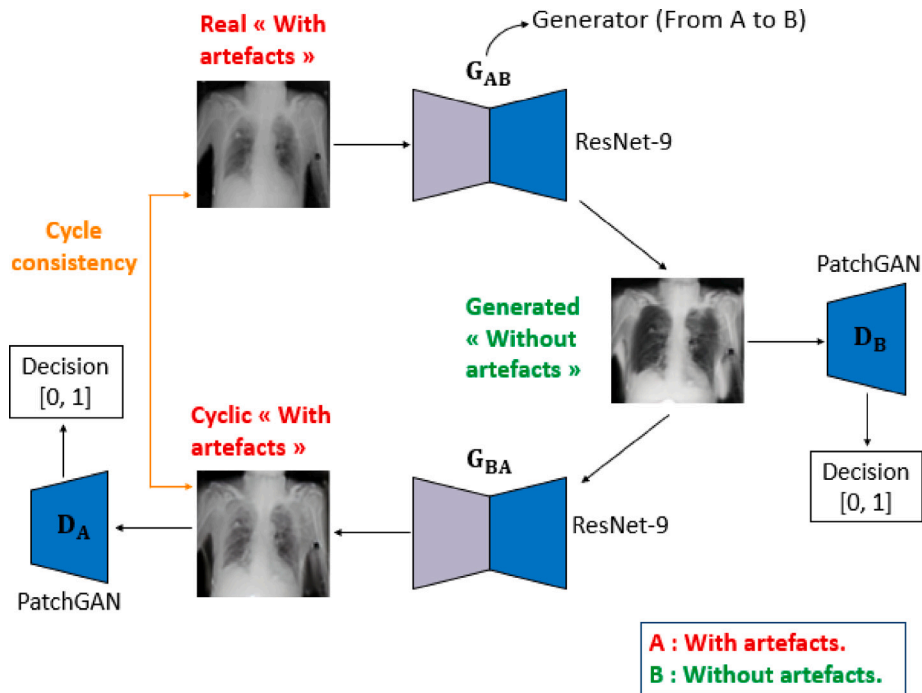


Fig. 7. The CycleGAN architecture that was adapted for the experiment of this work for the scenario images with artifacts vs images without artifacts.

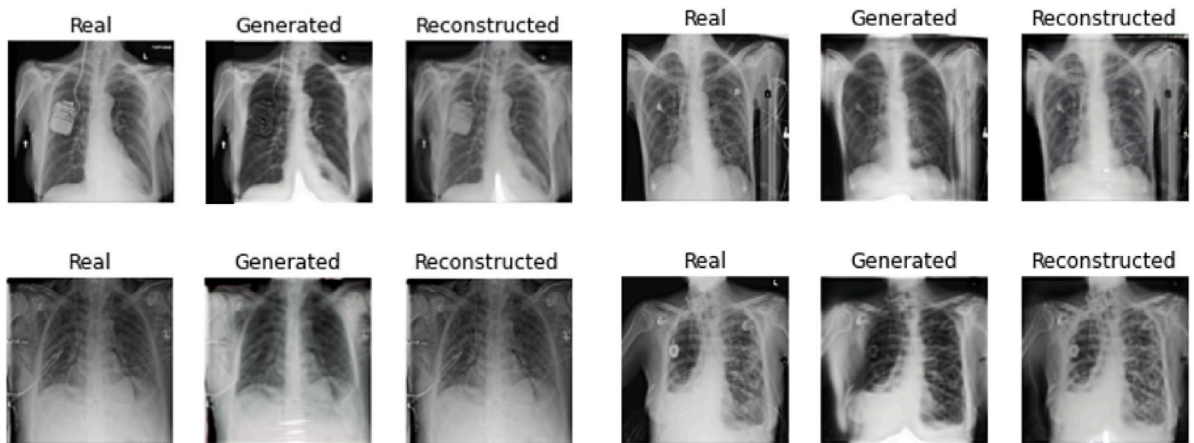


Fig. 8. Examples of image translation by the CycleGAN model.

pixels, then normalized them to be fed into the pre-trained DenseNet-121 model. During training, we initialized the DenseNet-121 with the weights pre-trained on ImageNet. Then we froze all the weights from the lower convolutional layers, and we added two dense layers with 1024 and 512 hidden neurons respectively, followed by a dropout layer with a rate of 0.5. Finally, we replaced the final fully connected layer with a fully connected layer of a 2-dimensional output and a sigmoid activation function that classifies the output image as pneumonia or normal.

We trained the networks with mini-batches of size 16. The ADAM optimizer was used as an optimization algorithm during training with a learning rate of 10^{-4} . To regularize the network, early stopping strategy was performed with a patience of 5 epochs, which is a method used to detect convergence of training and thus prevent the networks from overfitting. To compile the model, 'binary_crossentropy' was used as a loss function.

We applied data augmentation to increase the size of the dataset, prepare the model to increase the efficiency of test images in a different

orientation and improve the model's generalization ability. Augmentation includes a rotation range of 10, a horizontal flipping, a width shift range of 0.1, a height shift range of 0.1, a zoom range equal to 0.1 and setting the fill-mode to nearest.

Our model's performance was assessed by calculating the Area Under the Curve (AUC) which indicates the model's ability to accurately distinguish between classes.

3.6.1. Clinical and technical characteristics

The ChestX-ray14 dataset contains data such as the gender and the age of the patient and the view position of the X-ray (i.e. whether the X-ray images were acquired in the posterior–anterior (PA) or anterior–posterior (AP) position). In clinical practice, radiologists use further information beyond the image to conclude the presence or not of pathologies and to establish a diagnosis.

To exploit all the information in the dataset (i.e. age, gender and view position), we proposed integrating this additional information to the learned image representation. We concatenate the image feature vector which is the output of the last layer of our DenseNet-121 model

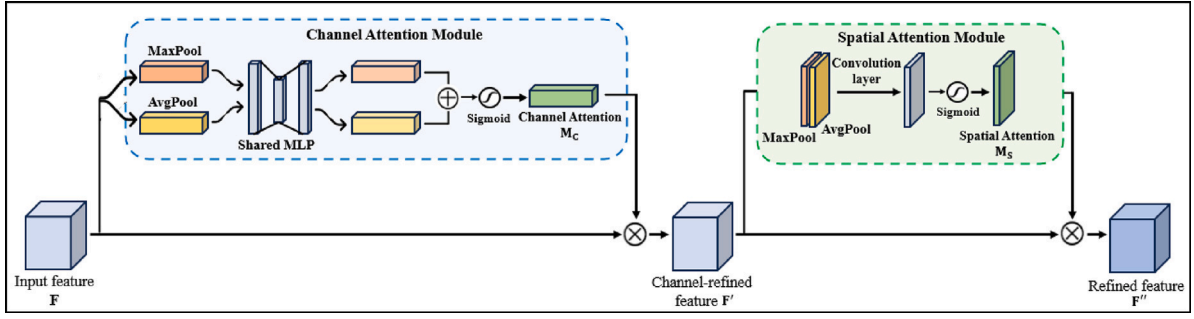


Fig. 9. Convolutional Block Attention Module (CBAM).

with the new clinical feature vector (of dimension 3×1). Specifically, gender and view position are encoded in binary form as $\{0, 1\}$ and the age is scaled linearly within the interval $[0, 1]$, to avoid a bias towards features with a wide range of values.

3.6.2. Convolutional block attention module (CBAM)

In recent years, the attention mechanism has attracted considerable interest and proved to be useful in pixel-wise computer vision tasks [53, 54]. In addition, it has shown its superiority in a variety of medical image analysis tasks [55].

In general, diseases are characterized by lesion areas, which contain key information for disease diagnosis. An attention-based CAD system focuses on lesion features of specific regions within CXR images. Instead of treating all parts of the image equally, the attention mechanism enables the model to assign higher weights or attention to areas that contain relevant semantic features associated with lesions. This selective attention allows the model to filter out irrelevant information, thereby improving its efficiency in disease diagnosis.

Convolutional block attention module (CBAM) is a type of attention mechanism that can be used in a convolutional neural network to improve its performance [56]. CBAM enables the attention module to be applied to both the channel dimension and the spatial dimension. As illustrated in Fig. 9, the basic structure of a CBAM is based on two modules: a channel attention module and a spatial attention module.

A given feature map $F \in \mathbb{R}^{C \times H \times W}$ (where C represents the number of channels, H represents the height, and W the width, of the feature map) is processed by the channel attention module and the spatial attention module as follows:

$$F' = M_C(F) \otimes F, F'' = M_S(F') \otimes F' \quad (1)$$

where \otimes denotes the element-wise multiplication, F' represents the result of the feature map multiplying by the channel attention map, and F'' represents the result of the spatial attention map multiplying by F' or the final refined output.

- **Channel attention module:** As each channel of the feature map is considered as a specific feature detector, channel attention focuses on what is significant, and allows to extract the channels that contained useful information. The channel attention module squeezed the feature map in the spatial dimension by using both global average pooling layer and global max pooling layer, as shown in the blue part of Fig. 9. The global average pooling layer provided the overall information, while the global maximum pooling layer captured information about feature differences. Then, the squeezed feature maps of the two pooling layers F_{avg}^C and F_{max}^C were sent to the shared network which was composed of multi-layer perceptron (MLP) with one hidden layer. The MLP was set to a certain compression ratio to reduce parameter overhead. After applying the shared network to each descriptor, the output feature vectors are merged

using element-wise summation. Then, the sigmoid function computed the channel attention map $M_C(F) \in \mathbb{R}^{C \times 1 \times 1}$. Therefore, the channel attention is computed as follows:

$$\begin{aligned} M_C(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \end{aligned} \quad (2)$$

where σ denotes the sigmoid function, W_0 and W_1 belong to the weights of the MLP perceptron.

Thus, the channel attention module focuses on feature information within each channel of the input data. It does so by learning to assign varying degrees of importance to different channels, considering inter-channel relationships.

- **Spatial attention module:** In contrast to channel attention, the spatial attention focuses on where is an informative part of the feature map. To compute spatial attention, as shown in the green part of Fig. 9, average pooling and max pooling operations are applied along the channel axis generating two 2D feature maps: F_{avg}^S and F_{max}^S . Then, these two 2D feature maps are concatenated to generate an efficient feature descriptor for convolution. Applying pooling operations along the channel axis proves to be effective in highlighting informative regions [56]. After the convolution operation, the sigmoid function computed the 2D spatial attention map $M_S(F) \in \mathbb{R}^{H \times W}$, which indicates where to emphasize or suppress. Therefore, the spatial attention is computed as follows:

$$\begin{aligned} M_S(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7 \times 7}([F_{avg}^S; F_{max}^S])) \end{aligned} \quad (3)$$

where σ denotes the sigmoid function and $f^{7 \times 7}$ represents the convolution operation with the filter size of 7×7 .

Thus, the spatial attention module learns to highlight important spatial locations while downplaying less significant ones. By doing so, it enables the network to selectively attend to specific spatial regions within each feature map, emphasizing the location information that is most pertinent to the task.

Therefore, the two modules help the CBAM to adaptively adjust its attention both across feature channels and spatially within feature maps, ultimately enhancing the network's capability to capture meaningful patterns and features in the input data.

In our experiment, we applied the CBAM to the last convolution layer of the DenseNet-121 model, allowing the attention module to be applied to both the channel and spatial dimensions.

4. Results

This section reveals the achieved results from the techniques investigated in this paper. To assess their performance, the models were evaluated using test data. We performed ten different random splits of the dataset into training, validation and testing sets to ensure the stability of the results, and presented the results in terms of mean and standard deviation. The ablation study presented in Table 2 demonstrates the effectiveness of our design choice.

Table 2

AUC of the binary classification of pneumonia and no finding images using the different approaches.

		Binary classification (Pneumonia/No Finding)
Without clustering	DenseNet-121	77.96% \pm 1.5%
With clustering, on the cluster containing images without artifacts (unsupervised labeling)	DenseNet-121	79.58% \pm 1.6%
With clustering, on the cluster containing images without artifacts (manual labeling)	DenseNet-121	81.47% \pm 1.8%
With clustering (unsupervised labeling) and CycleGAN	DenseNet-121	81.94% \pm 1.4%
	DenseNet-121 + clinical variables	82.66% \pm 1.5%
	DenseNet-121 + CBAM + clinical variables	83.87% \pm 1.4%

Table 3

Precision, Recall, F1 score and overall accuracy of the DenseNet-121 model with the CBAM and the clinical variables, applied to the images after clustering and CycleGAN, for the different classes (pneumonia and no finding).

	Precision	Recall	F1-score	Accuracy
No finding	83% \pm 1%	78% \pm 1.3%	80% \pm 1.1%	80.48% \pm 1.5%
Pneumonia	76% \pm 1.9%	81% \pm 1.6%	78% \pm 1.7%	

As shown in Table 2, applying the DenseNet-121 model to the cluster containing sharp images only (i.e. with clustering (unsupervised labeling)) improves the results for pneumonia classification by 1.62% compared with the base DenseNet-121 model (i.e. without clustering). The DenseNet-121 model applied to the cluster without artifacts obtained by manual labeling achieves an AUC of 81.47%. In addition, applying CycleGAN on images containing artifacts such as electronic materials to generate images without artifacts, then combining them with the cluster of sharp images, leads to a 3.98% improvement in the classification results for pneumonia, in comparison with the base DenseNet-121 model (i.e. without clustering). Table 2 also presents the results for the binary classification of pneumonia with the integration of clinical characteristics, as well as the addition of the CBAM which is composed of the channel and spatial attention modules. The greatest results are obtained using clustering (unsupervised labeling), CycleGAN then applying the DenseNet-121 model with CBAM, and combining the clinical variables with model features. Table 3 shows the precision, recall and F1-score of this best approach for the different two classes on the test data. Table 4 provides the results of the binary classification of another lung disease, namely consolidation. Also, the highest results are obtained using clustering (unsupervised labeling), CycleGAN, CBAM and the clinical variables. Table 5 shows the precision, recall and F1-score of this best approach for the different two classes on the test data. From Table 6, we can see that our approach outperforms the literature for the classification of lung diseases using the same ChestX-ray14 dataset.

5. Discussion

In this study, we aimed to classify pneumonia by taking into account the presence of artifacts in the CXR images. The use of the images containing artifacts can introduce noise during classification. Since these images are not labeled in the dataset, we proposed an image

Table 4

AUC of the binary classification of consolidation and no finding images using the different approaches.

		Binary classification (consolidation/no finding)
Without clustering	DenseNet-121	78.52% \pm 1.4%
With clustering, on the cluster containing images without artifacts (unsupervised labeling)	DenseNet-121	80.26% \pm 1.3%
With clustering (unsupervised labeling) and CycleGAN	DenseNet-121	82.89% \pm 1.6%
	DenseNet-121 + clinical variables	83.65% \pm 1.4%
	DenseNet-121 + CBAM + clinical variables	84.96% \pm 1.5%

Table 5

Precision, Recall, F1 score and overall accuracy of the DenseNet-121 model with the CBAM and the clinical variables, applied to the images after clustering and CycleGAN, for the different classes (consolidation and no finding).

	Precision	Recall	F1-score	Accuracy
No finding	83% \pm 1.5%	82% \pm 1.4%	82% \pm 1.5%	81.67% \pm 1.7%
Consolidation	79% \pm 1.8%	81% \pm 2.1%	80% \pm 1.9%	

preprocessing module that will distinguished sharp images from those containing artifacts.

For this purpose, various contour and texture features were extracted to be used as input data for the k-means clustering algorithm. As a result, as shown in Table 1, two clusters are created for each class: one cluster containing sharp images without artifacts, and the other cluster containing images affected by artifacts. We assessed the impact of the artifacts noise by performing the classification using the DenseNet-121 model in two experiments: in the first experiment we applied the model to all the images without clustering (all images with and without artifacts), and in the second experiment we applied the model after clustering (unsupervised labeling) only to the cluster containing the sharp images without artifacts and eliminated the cluster containing the images with artifacts. The AUC for each case was calculated. To ensure the stability of the results, we performed ten different random splits of the dataset into training, validation and testing sets and presented the results in terms of mean and standard deviation. As indicated in Table 2, by applying the model to all images with and without artifacts (without clustering), we obtained an AUC of 77.96% for the classification of pneumonia and no finding images. However, with clustering (unsupervised labeling), we reached an AUC of 79.58% for pneumonia classification solely on the cluster containing images without artifacts. This underscores that eliminating images containing artifacts significantly enhances classification performance, highlighting the benefits of clustering in improving classification by 1.62%. Additionally, the lower performance in the non-clustered case illustrates the negative impact of artifacts on classification, underscoring the importance of addressing them in medical imaging tasks. Moreover, the DenseNet-121 model applied to the cluster without artifacts obtained by manual labeling achieves an AUC of 81.47%. The difference between the AUC values of the model with unsupervised labeling and manual labeling is not very significant. This indicates that our clustering (unsupervised labeling) approach is very accurate in identifying clear images from images containing artifacts. It can be used to reduce the workload of medical professionals, separating images with and without artifacts more quickly and with less effort compared to manual labeling by

Table 6
Comparison of AUC results between our best approach and other existing methods in the literature using the ChestX-ray14 dataset.

Disease	Wang et al. (2017) [29]	Yao et al. (2017) [17]	Baltruschat et al. (2019) [38]	Xu et al. (2021) [35]	Chen et al. (2022) [7]	Yang et al. (2022) [57]	Our approach
Atelectasis	77.3%	77.2%	76.7%	83.1%	75.9%	78.5%	86.74% \pm 1.1%
Cardiomegaly	85.4%	90.4%	88.3%	91%	89.83%	89.2%	93.58% \pm 1.2%
Consolidation	77%	78.8%	74.9%	81.7%	75.12%	74.7%	84.96% \pm 1.5%
Edema	86.1%	88.2%	84.6%	90.2%	86.36%	83.7%	93.15% \pm 1.4%
Effusion	86.1%	85.9%	82.2%	88.6%	82.66%	83.6%	90.36% \pm 1.3%
Emphysema	73.6%	82.9%	89.5%	93.4%	88.98%	92.5%	95.67% \pm 1.5%
Fibrosis	73.9%	76.7%	81.6%	85.6%	83.59%	83.8%	88.73% \pm 1.3%
Hernia	74.6%	91.4%	93.7%	94.7%	94.04%	90.5%	95.82% \pm 1.4%
Infiltration	63.6%	69.5%	69.4%	71.5%	70.99%	71%	81.27% \pm 1.5%
Mass	76.1%	79.2%	82%	85.5%	83.2%	82.6%	87.92% \pm 1.1%
Nodule	66.4%	71.7%	74.7%	79.8%	75.33%	75.5%	84.16% \pm 1.2%
Pleural_Thickening	74.9%	76.5%	76.3%	79.1%	78.94%	78.5%	86.51% \pm 1.3%
Pneumonia	66.4%	71.13%	71.4%	77.7%	73.56%	73.5%	83.87% \pm 1.4%
Pneumothorax	79.9%	84.1%	84%	88.6%	87.6%	84.7%	91.38% \pm 1.3%

a practitioner. Additionally, it can be applied to a larger number of images.

To further use the images containing artifacts in the classification, we applied CycleGAN to the images containing artifacts obtained by the unsupervised labeling method in order to reduce the noise effect caused by the presence of artifacts in the images. So, CycleGAN was performed on the images from the cluster containing images with artifacts to generate images without artifacts. After that, the generated images are combined with the second cluster containing images without artifacts to perform the classification. As shown in Table 2, using clustering (unsupervised labeling) and CycleGAN we achieved an AUC of 81.94% for pneumonia classification, leading to a 3.98% improvement in classification results, compared with the basic DenseNet-121 model (i.e. without clustering). This improvement underscores the importance of preprocessing and reducing artifacts in chest X-ray images, as these steps enhance image quality and, consequently, classification performance.

Furthermore, to fully exploit all the information contained in the dataset (i.e. age, gender and view position), we concatenated the image feature vector which is the output of the last layer of our DenseNet-121 model with the clinical feature vectors. The results show that these non-image features improved the classification results. But their incorporation did not significantly increase model performance, perhaps because the image features already encode information on clinical characteristics, i.e. non-image data. As presented in Table 2, we obtained an AUC of 81.94% using the DenseNet-121 model after applying clustering (unsupervised labeling) and CycleGAN. But after integrating the clinical variables, we attained an AUC of 82.66% for pneumonia classification.

In another experiment, after applying clustering (unsupervised labeling) and CycleGAN, we introduced an attention mechanism, namely the CBAM, into our DenseNet model, and integrated the clinical variables. This approach achieved the highest AUC of 83.87% for the binary classification of pneumonia and normal images, as indicated in Table 2.

To further validate our results, we computed additional evaluation metrics for the configuration that delivered the best results, utilizing the DenseNet-121 model with CBAM and clinical variables after applying clustering and CycleGAN. As shown in Table 3, our proposed approach achieved strong performance in the binary classification of pneumonia and no finding images, with an accuracy of 80.48%. These results underscore the robustness of our approach, as the combination of attention mechanisms, clinical variables, and artifact reduction which plays a crucial role in enhancing image quality, significantly enhances classification performance, leading to more reliable and accurate diagnostic outcomes.

Therefore, it is crucial to consider the presence of artifacts when diagnosing lung diseases on the basis of radiographic images. The artifacts, whether they result from technical flaws in the X-ray equipment, unwanted interference or other factors, can significantly affect the quality of medical images. In a field where precision is crucial, these

artifacts can distort the interpretation and detection of lung anomalies, leading to inaccurate diagnoses and inappropriate therapeutic decisions. By performing clustering (unsupervised labeling) and reducing the effect of artifacts noise by using CycleGAN, our approach improves the classification performance of the deep learning model. By eliminating the unwanted noise, it allows the model to focus on relevant diagnostic features, thereby increasing its performance. Furthermore, the addition of the CBAM enables the model to automatically adjust its attention to focus on the most relevant features of the images, at both channel and spatial levels. This selective attention allows the model to filter out irrelevant information, enabling it to learn to automatically focus on disease-specific areas in the input image. It contributes to more accurate classification and noise reduction, improving the model's performance and efficiency in disease diagnosis. This means more accurate detection of lung diseases which can have a significant impact on clinical outcomes. In the medical context, this enables radiologists to optimize the detection of abnormalities, contribute to more effective treatment decisions and, eventually, improve the quality of diagnoses.

Thereafter, we generalized our approach for the classification of consolidation and no finding (2,751 images). We have taken all the consolidation images from the ChestX-ray14 dataset, of which 1,310 are marked as consolidation. Then we have down-sampled the no finding images to have 1,441 images that will be used for the classification. After that, we performed the clustering (unsupervised labeling) using k-means on the proposed features extracted from the images. We obtained two clusters for each class: the first cluster contains sharp images, while the second one contains images with artifacts. So we have a total of 1,328 consolidation and no finding images that are clear without artifacts, and 1,423 images containing artifacts. As shown in Table 4, applying the DenseNet-121 model to all images with and without artifacts (without clustering) resulted in an AUC of 78.52% for the classification of consolidation and no finding images. In contrast, by utilizing clustering (unsupervised labeling) exclusively on the cluster containing images without artifacts, we achieved an AUC of 80.26% for consolidation classification. This demonstrates that eliminating images with artifacts significantly enhances classification performance by 1.74%.

To further use all the images, CycleGAN was performed on the 1,423 images with artifacts to generate images without artifacts. After that, the generated images are combined with the second cluster containing images without artifacts to perform the classification. The results in Table 4 demonstrated that the classification results with the use of clustering (unsupervised labeling) and CycleGAN are better than without clustering, as confirmed by the AUC: the DenseNet-121 model applied to all images (without clustering) resulted in an AUC of 78.52%. But after employing clustering (unsupervised labeling) and with CycleGAN, the model achieved an AUC of 82.89%. This indicates that CycleGAN leads to a 4.37% improvement in classification results compared with the basic model without clustering. These results underscore the importance of our preprocessing approach, as it effectively reduces the impact

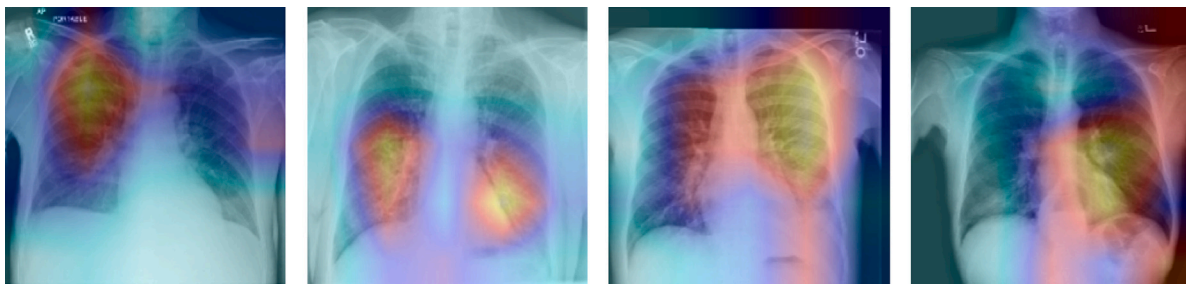


Fig. 10. Visualization examples of the disease localization on the test images of pneumonia and consolidation. Higher response is indicated with yellow, and lower response with blue.

of artifacts, leading to improved image quality and more accurate classification results.

Subsequently, we incorporated the CBAM along with the clinical variables. This configuration, starting with clustering (unsupervised labeling) and CycleGAN, followed by the addition of CBAM and clinical variables, enabled us to achieve the highest AUC, as detailed in Table 4: we obtained an AUC of 84.96% for the classification of consolidation and no finding images. Additionally, our proposed approach demonstrated strong performance in the binary classification of consolidation and no finding images, achieving an accuracy of 81.67%, as indicated in Table 5. These results underscore the effectiveness of our approach in enhancing image quality, thereby significantly improving classification performance, ultimately leading to more accurate diagnoses and better clinical outcomes.

Moreover, we tested our approach on additional lung diseases and we compared it with other studies in the literature using the ChestX-ray14 dataset. As shown in Table 6, our approach outperforms previous studies for the classification of all lung diseases. For instance, Chen et al. [7], who employed CBAM with DenseNet-121, reported an AUC of 73.56% for pneumonia, and 75.12% for consolidation. Baltruschat et al. [38], in their study integrating clinical variables with the ResNet-38 model, achieved AUC of 71.4% for pneumonia and 74.9% for consolidation. Our proposed approach, which focuses on reducing the impact of image artifacts, demonstrates a significant improvement in classification performance, exceeding these previous studies (6.17% and 3.26% for pneumonia and consolidation classification, respectively). These findings underscore the importance of our preprocessing approach in improving image quality, ultimately leading to more accurate and reliable classification outcomes.

Visualizing attention enables researchers and physicians to better understand how machine learning models make decisions. By examining which parts of the image the model prioritizes and focuses its attention on when classifying lung diseases, we can gain insight into its decision-making logic. By overlapping the output of the attention layer with the original image, we can observe which regions of the image are of greatest importance to the model when performing classification.

Fig. 10 shows exemplary localization results of our proposed diagnosis model for pneumonia and consolidation diseases. This visualization and localization enables the explainability of chest X-ray images and the interpretability of the model. It allows interpret the inner workings of the model and understand the specific parts of the image on which it focuses to perform classification. Indeed, attention visualization contributes to increasing confidence in model predictions. For instance, medical professionals may feel more confident in using AI models if they can visually check where the model is focusing its attention when making a classification. Thus, this can help build confidence in the model's decisions, validate results and provide additional information for physicians when interpreting diagnostic results.

Moreover, to validate our results, we presented these images to a physician, who confirmed and validated that the areas targeted by the model indeed correspond to the regions affected by the disease.

This medical validation reinforces the credibility and robustness of our approach.

Our study presents some limitations. First, since the images with artifacts are not labeled in the ChestX-ray14 dataset, we utilized an unsupervised classification approach (K-means clustering). While the classification accuracy of K-means is promising, improvements can be made to further enhance the separation between images with and without artifacts. Indeed, the effectiveness of clustering relies on the quality of the extracted features. Although we obtained promising results with our current set of features, exploring other feature extraction methods could potentially enhance the clustering performance and improve the separation of images with and without artifacts. Another limitation concerns the quality of the generated images. Although CycleGAN demonstrates effectiveness in reducing artifacts, the sharpness of the generated images may not always be optimal. Additional refinements to the CycleGAN architecture might be necessary to further enhance the quality of these generated images. Indeed, in our experiments, we used the default hyperparameters from the original CycleGAN paper [50]. While these hyperparameters yielded promising results, further hyperparameter optimization could lead to even better artifact reduction and potentially produce higher resolution images with improved clarity. Furthermore, we acknowledge that while we used the ChestX-ray14 dataset which includes a large number of images collected from multiple acquisition centers (hospitals affiliated to National Institutes of Health Clinical Center), there could be a limitation in generalizability. Indeed, variability may be introduced due to different types of scanners used in image acquisition, which can impact classification performance. Future work could involve addressing the harmonization of data obtained from different acquisition centers or scanners to better account for this variability and enhance the model's adaptability across diverse datasets.

6. Conclusion

In this paper, we presented a novel preprocessing approach to reduce the impact of artifacts, such as wires or electronic objects, in chest X-ray images using the ChestX-ray14 dataset. Since the images containing artifacts are not labeled in the dataset, we employed an unsupervised classification approach (K-means clustering) to differentiate between images with and without artifacts. After separating the two clusters, we developed a CycleGAN-based approach to generate, from the images with artifacts, the images without these artifacts. This contribution demonstrates a significant impact later on the classification process using the DenseNet-121 model, adding the CBAM and incorporating the clinical variables. The experimental results demonstrate our preprocessing approach's efficacy, yielding a 5.91% and 6.44% improvement in AUC for pneumonia and consolidation, respectively, compared to baseline models evaluated on all images, including those with and without artifacts.

Therefore, it is critical to address the issue of artifacts when diagnosing lung diseases from chest X-ray images. It allows the model to focus on relevant diagnostic features and improve lung disease

classification performance. Besides, combining the strengths of deep learning and attention mechanism improves the model's performance in detecting disease-specific regions within the lungs. Indeed, the CBAM enables the model to automatically adjust its attention to focus on the most relevant features of the images, at both channel and spatial levels. Consequently, it enables ensure accurate treatment, minimizes diagnostic errors and guarantee a consistent results interpretation. The validation by a medical expert, confirming that the areas specifically identified by the model significantly coincide with regions affected by the disease, affirms the reliability and effectiveness of our approach.

CRedit authorship contribution statement

Aya Hage Chehade: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Nassib Abdallah:** Writing – review & editing, Visualization, Validation, Supervision, Methodology. **Jean-Marie Marion:** Writing – review & editing, Visualization, Validation, Supervision, Methodology. **Mathieu Hatt:** Writing – review & editing. **Mohamad Oueidat:** Supervision. **Pierre Chauvet:** Supervision.

Ethical approval

This study uses public database cited in the references and therefore this section is not applicable.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used AI-assisted technologies in order to enhance the English language quality. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors have no relevant financial or non-financial interests to disclose. The authors declare that they have no conflict of interest.

References

- [1] T. Han, V. Nunes, L. Souza, A. Marques, I. Silva, M. Junior, J. Sun, P. Reboucas Filho, Internet of medical things—based on deep learning techniques for segmentation of lung and stroke regions in CT scans, *IEEE Access* 8 (2020) 71117–71135.
- [2] R. Tobias, L. De Jesus, M. Mital, S. Lauguico, M. Guillermo, E. Sybingco, A. Bandala, E. Dadios, CNN-based deep learning model for chest X-ray health classification using tensorflow, in: 2020 RIVF International Conference on Computing and Communication Technologies, RIVF, 2020, pp. 1–6.
- [3] A. Acharya, R. Satapathy, A deep learning based approach towards the automatic diagnosis of pneumonia from chest radio-graphs, *Biomed. Pharmacol. J.* 13 (2020) 449–455.
- [4] World Health Organization, Pneumonia in children, 2022, (Accessed on 20 June 2023).
- [5] S. Akter, F. Shamrat, S. Chakraborty, A. Karim, S. Azam, COVID-19 detection using deep learning algorithm on chest X-ray images, *Biology* 10 (2021) 1174.
- [6] G. Liang, L. Zheng, A transfer learning method with deep residual network for pediatric pneumonia diagnosis, *Comput. Methods Programs Biomed.* 187 (2020) 104964.
- [7] L. Chen, T. Mao, Q. Zhang, Identifying cardiomegaly in chest X-rays using dual attention network, *Appl. Intell.* 52 (2022) 11058–11067.
- [8] Y. Lee, S. Huang, R. Chang, CheXGAT: A disease correlation-aware network for thorax disease diagnosis from chest X-ray images, *Artif. Intell. Med.* 132 (2022) 102382.
- [9] T. Rahman, M. Chowdhury, A. Khandakar, K. Islam, K. Islam, Z. Mahbub, M. Kadir, S. Kashem, Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray, *Appl. Sci.* 10 (2020) 3233.
- [10] T. Franquet, Imaging of community-acquired pneumonia, *J. Thorac. Imaging* 33 (2018) 282–294.
- [11] K. Islam, S. Wijewickrema, A. Collins, S. O'Leary, A deep transfer learning framework for pneumonia detection from chest X-ray images, in: *VISIGRAPP (5: VISAPP)*, 2020, pp. 286–293.
- [12] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett, S. Valaee, Synthesizing chest X-ray pathology for training deep convolutional neural networks, *IEEE Trans. Med. Imaging* 38 (2018) 1197–1206.
- [13] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, J. Barfett, Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2018, pp. 990–994.
- [14] F. Shamrat, S. Azam, A. Karim, K. Ahmed, F. Bui, F. De Boer, High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images, *Comput. Biol. Med.* 155 (2023) 106646.
- [15] S. Bharati, P. Podder, M. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, *Inform. Med. Unlocked* 20 (2020) 100391.
- [16] K. Sripor, C. Tsai, C. Tsai, P. Wang, Analyzing lung disease using highly effective deep learning techniques, *Healthcare* 8 (2020) 107.
- [17] L. Yao, E. Poblentz, D. Dagunts, B. Covington, D. Bernard, K. Lyman, Learning to diagnose from scratch by exploiting dependencies among labels, 2017, *arXiv: 1710.10501*.
- [18] Z. Cui, F. Xue, X. Cai, Y. Cao, G. Wang, J. Chen, Detection of malicious code variants based on deep learning, *IEEE Trans. Ind. Inform.* 14 (2018) 3187–3196.
- [19] Y. Wang, X. Qiao, G. Wang, Architecture evolution of convolutional neural network using monarch butterfly optimization, *J. Ambient Intell. Humaniz. Comput.* (2022) 1–15.
- [20] A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 25 (2012).
- [21] H. Behzadi-Khormouji, H. Rostami, S. Salehi, T. Derakhshande-Rishehri, M. Masoumi, S. Salemi, A. Keshavarz, A. Gholamrezanezhad, M. Assadi, A. Batouli, Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images, *Comput. Methods Programs Biomed.* 185 (2020) 105162.
- [22] V. Fernandes, G. Junior, A. Paiva, A. Silva, M. Gattass, Bayesian convolutional neural network estimation for pediatric pneumonia detection and diagnosis, *Comput. Methods Programs Biomed.* 208 (2021) 106259.
- [23] D. Kuzinkovas, S. Clement, The detection of covid-19 in chest x-rays using ensemble cnn techniques, *Information* 14 (2023) 370.
- [24] V. Ravi, V. Acharya, M. Alazab, A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images, *Cluster Comput.* 26 (2023) 1181–1203.
- [25] S. Xie, Z. Yu, Z. Lv, Multi-disease prediction based on deep learning: A survey, *CMES Comput. Model. Eng. Sci.* 128 (2021).
- [26] A. Tekerek, I. Al-Rawe, A novel approach for prediction of lung disease using chest X-ray images based on DenseNet and MobileNet, *Wirel. Pers. Commun.* (2023) 1–15.
- [27] L. Kondaka, M. Thenmozhi, K. Vijayakumar, R. Kohli, An intensive healthcare monitoring paradigm by using IoT based machine learning strategies, *Multimedia Tools Appl.* 81 (2022) 36891–36905.
- [28] A. Hage Chehade, N. Abdallah, J. Marion, M. Hatt, M. Oueidat, P. Chauvet, A systematic review: Classification of lung diseases from chest X-ray images using deep learning algorithms, *SN Comput. Sci.* 5 (2024) 405.
- [29] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. 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, CVPR*, 2017, pp. 2097–2106.
- [30] G. Huang, Z. Liu, L. Van Der Maaten, K. Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [31] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning, 2017, *arXiv:1711.05225*.
- [32] A. Souid, N. Sakli, H. Sakli, Classification and predictions of lung diseases from chest x-rays using mobilenet v2, *Appl. Sci.* 11 (2021) 2751.
- [33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
- [34] M. Mann, R. Badoni, H. Soni, M. Al-Shehri, A. Kaushik, D. Wei, Utilization of deep convolutional neural networks for accurate chest X-ray diagnosis and disease detection, *Interdiscip. Sci.: Comput. Life Sci.* (2023) 1–19.
- [35] J. Xu, H. Li, X. Li, MS-ANet: Deep learning for automated multi-label thoracic disease detection and classification, *PeerJ Comput. Sci.* 7 (2021) 541.

- [36] C. Ma, H. Wang, S. Hoi, Multi-label thoracic disease image classification with cross-attention networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 730–738.
- [37] J. Zhao, M. Li, W. Shi, Y. Miao, Z. Jiang, B. Ji, A deep learning method for classification of chest X-ray images, *J. Phys. Conf. Ser.* 1848 (2021) 012030.
- [38] I. Baltruschat, H. Nickisch, M. Grass, T. Knopp, A. Saalbach, Comparison of deep learning approaches for multi-label chest X-ray classification, *Sci. Rep.* 9 (2019) 1–10.
- [39] V. Huy, C. Lin, An improved densenet deep neural network model for tuberculosis detection using chest X-ray images, *IEEE Access* (2023).
- [40] S. Kansal, S. Purwar, R. Tripathi, Image contrast enhancement using unsharp masking and histogram equalization, *Multimedia Tools Appl.* 77 (2018) 26919–26938.
- [41] M. Veluchamy, B. Subramani, Image contrast and color enhancement using adaptive gamma correction and histogram equalization, *Optik* 183 (2019) 329–337.
- [42] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. Kashem, M. Islam, S. Al Maadeed, S. Zughaier, M. Khan, Others, Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images, *Comput. Biol. Med.* 132 (2021) 104319.
- [43] M. Jawahar, J. Prassanna, V. Ravi, L. Anbarasi, S. Jasmine, R. Manikandan, R. Sekaran, S. Kannan, Computer-aided diagnosis of COVID-19 from chest X-ray images using histogram-oriented gradient features and random forest classifier, *Multimedia Tools Appl.* 81 (2022) 40451–40468.
- [44] T. Ho, J. Gwak, Multiple feature integration for classification of thoracic disease in chest radiography, *Appl. Sci.* 9 (2019) 4130.
- [45] B. Li, K. Cheng, Z. Yu, Others, Histogram of oriented gradient based gist feature for building recognition, *Comput. Intell. Neurosci.* (2016).
- [46] A. Salau, S. Jain, Feature extraction: a survey of the types, techniques, applications, in: 2019 International Conference on Signal Processing and Communication, ICSC, 2019, pp. 158–164.
- [47] A. Ayalew, A. Salau, B. Abeje, B. Enyew, Detection and classification of COVID-19 disease from X-ray images using convolutional neural networks and histogram of oriented gradients, *Biomed. Signal Process. Control* 74 (2022) 103530.
- [48] M. Darwis, L. Hasibuan, M. Firmansyah, N. Ahady, R. Tiaharyadini, Implementation of K-means clustering algorithm in mapping the groups of graduated or dropped-out students in the management department of the national university, *JISA (Jurnal Informatika Dan Sains)* 4 (2021) 1–9.
- [49] F. Putri, N. Wibowo, H. Mustofa, Clustering of tuberculosis and normal lungs based on image segmentation results of chan-veye and canny with K-means, *Indonesian J. Artif. Intell. Data Min.* 6 (2023) 18–28.
- [50] J. Zhu, T. Park, P. Isola, 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.
- [51] D. Moris, J. Moura Ramos, J. Bujan, M. Hortas, Data augmentation approaches using cycle-consistent adversarial networks for improving COVID-19 screening in portable chest X-ray images, *Expert Syst. Appl.* 185 (2021) 115681.
- [52] A. Bar-El, D. Cohen, N. Cahan, H. Greenspan, Improved cyclegan with application to COVID-19 classification, *Med. Imaging 2021: Image Process.* 11596 (2021) 296–305.
- [53] S. Singh, M. Kumar, A. Kumar, B. Verma, S. Shitharth, Pneumonia detection with QCSA network on chest X-ray, *Sci. Rep.* 13 (2023) 9025.
- [54] J. Li, K. Jin, D. Zhou, N. Kubota, Z. Ju, Attention mechanism-based CNN for facial expression recognition, *Neurocomputing* 411 (2020) 340–350.
- [55] B. Chen, J. Li, G. Lu, D. Zhang, Lesion location attention guided network for multi-label thoracic disease classification in chest X-rays, *IEEE J. Biomed. Health Inf.* 24 (2019) 2016–2027.
- [56] S. Woo, J. Park, J. Lee, I. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.
- [57] M. Yang, H. Tanaka, T. Ishida, Performance improvement in multi-label thoracic abnormality classification of chest X-rays with noisy labels, *Int. J. Comput. Assist. Radiol. Surg.* (2022) 1–9.