SN Computer Science (2023) 4:388
https://doi.org/10.1007/s42979-023-01818-w

ORIGINAL RESEARCH

Deep Learning Methods for Chest Disease Detection Using Radiography Images

Adnane Ait Nasser[1] · **Moulay A. Akhloufi**[1]

Received: 21 December 2022 / Accepted: 4 April 2023 / Published online: 11 May 2023

Abstract

X-ray images are the most widely used medical imaging modality. They are affordable, non-dangerous, accessible, and can be used to identify different diseases. Multiple computer-aided detection (CAD) systems using deep learning (DL) algorithms were recently proposed to support radiologists in identifying different diseases on medical images. In this paper, we propose a novel two-step approach for chest disease classification. The first is a multi-class classification step based on classifying X-ray images by infected organs into three classes (normal, lung disease, and heart disease). The second step of our approach is a binary classification of seven specific lungs and heart diseases. We use a consolidated dataset of 26,316 chest X-ray (CXR) images. Two deep learning methods are proposed in this paper. The first is called DC-ChestNet. It is based on ensembling deep convolutional neural network (DCNN) models. The second is named VT-ChestNet. It is based on a modified transformer model. VT-ChestNet achieved the best performance overcoming DC-ChestNet and state-of-the-art models (DenseNet121, DenseNet201, EfficientNetB5, and Xception). VT-ChestNet obtained an area under curve (AUC) of 95.13% for the first step. For the second step, it obtained an average AUC of 99.26% for heart diseases and an average AUC of 99.57% for lung diseases.

Keywords X-rays · Computer-aided detection · Deep learning · Deep convolutional neural network · Ensemble learning · Vision transformers

Introduction

X-ray radiographies are an affordable and non-invasive method of examining different organs of the body [1]. Recognized as a valuable diagnosis tool for many disorders and abnormalities, X-rays can also be used to monitor diseases during treatment [2]. Around 3.6 billion X-ray images are taken every year worldwide. This number includes over 150 million chest X-ray radiographies (CXR) performed in the United States only. The World Health Organization (WHO) stated that CXRs are the most commonly performed clinical imaging technique worldwide [3]. CXRs are grayscale images generally produced by projecting X-rays onto the human body positioned against a metallic plate. Samples of CXR images are shown in Fig. 1.

Although CXRs play a crucial role in the diagnosis of thoracic disease, visual inspection by radiologists remains complex and error-prone. Previous studies have shown that the risk of misdiagnosis increases with the amount of time it takes for a radiologist to interpret CXR images. In addition, even experienced radiologists were at greater risk of misdiagnosis because of hidden lesions and symptoms in soft tissue and bones [4].

The WHO reports that many chest diseases can be life-threatening and lead to the death of millions of people if not accurately and timely treated [5]. Some chest diseases are of high mortality rates such as tuberculosis that kills around 1.4 million people annually, pneumonia that kills 9 million children under the age of 5 years being the world's leading killer disease, and COVID-19 which caused the death of over 6 million people all over the world as of November 2022.

Moulay A. Akhloufi have contributed equally to this work.

This article is part of the topical collection "Recent Trends on AI for HealthCare" guest edited by Lydia Bouzar- Benlabiod.

✉ Adnane Ait Nasser
eaa3027@umoncton.ca

Moulay A. Akhloufi
moulay.akhloufi@umoncton.ca

[1] Perception, Robotics, and Intelligent Machines (PRIME), Université de Moncton, Moncton, NB E1C 3E9, Canada

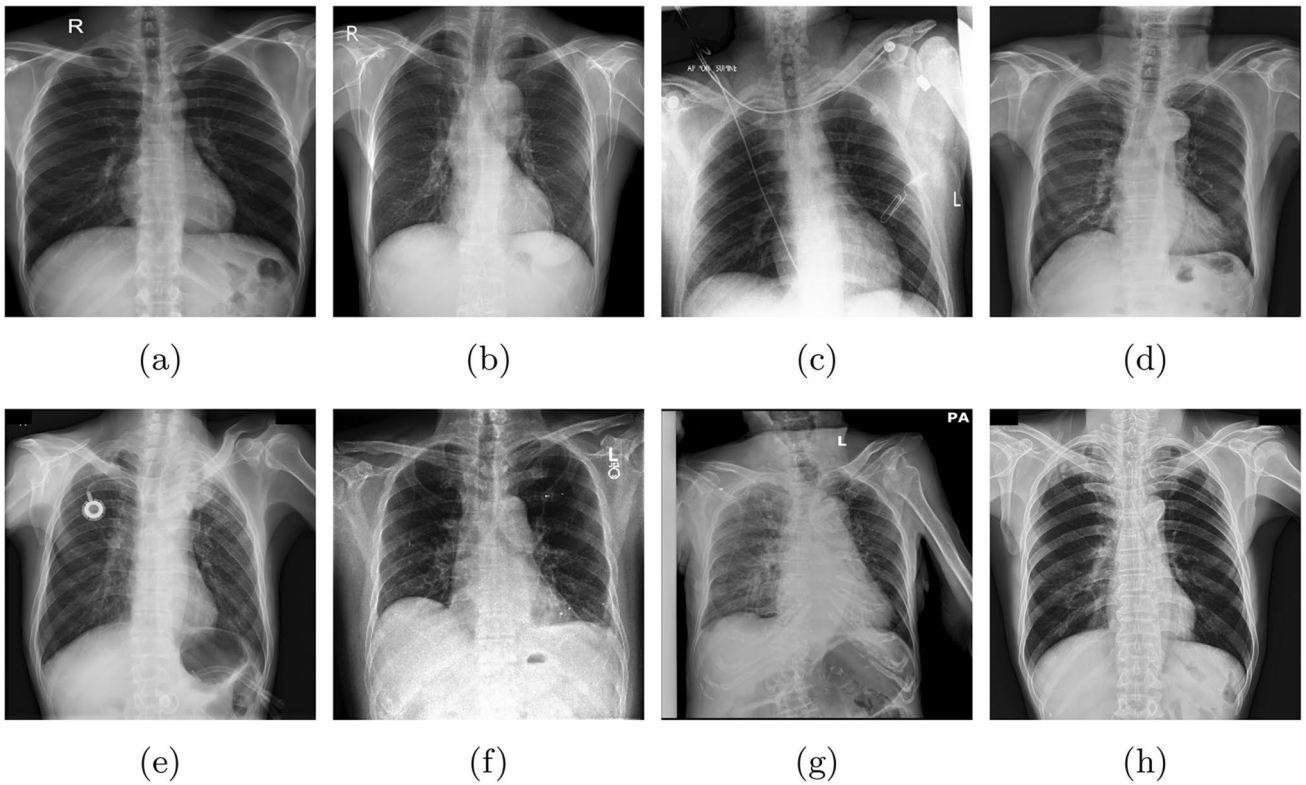

Fig. 1 Samples of CXR images from our consolidated dataset: **a** Normal; **b** Cardiomegaly; **c** Enlarged cardiomediastinum; **d** Aortic enlargement; **e** Atelectasis; **f** Pleural effusion; **g** Pneumothorax; **h** Pulmonary fibrosis

The shortage of radiologists in many countries of the world remains a considerable problem, especially because of the high number of people in need of radiological examinations, faster than it is necessary to train new radiologists. The long waiting lists for diagnosis at hospitals, the high rate of misdiagnosis of CXR images and the spread of life-threatening diseases are the main drivers that highlight the necessity of developing efficient computer aided detection (CAD) systems for early detection of chest diseases.

It is not always evident for radiologists to detect chest diseases on a CXR due to the lack of clarity of this imaging modality and the fact that many diseases are similar. This leads to high error rates for experts who use visual methods to diagnose diseases. A lot of research has been developed to address this challenge using image-based AI systems.

In this paper, we present a dataset we consolidated by combining images from two open-access datasets. We propose a novel two-step approach for the classification of CXR images by implementing two deep learning (DL) methods which are: a DCNN based method (DC-ChestNet) and a vision transformer based method (VT-ChestNet). In the first step of our approach, we classify CXR images from our dataset based on the infected organ into three classes (normal, lung disease, and heart disease). In the second step, we perform a binary classification of the specific diseases of each organ.

The main contributions of our work are as follows:

- A new dataset is consolidated by merging CXR images from two open-access datasets (CheXpert and VinDr-CXR).
- A novel two-step approach is introduced to perform chest disease classification on CXR images.
- Two DL methods (DC-ChestNet and VT-ChestNet) are proposed. DC-ChestNet method employs an ensemble learning of three different DCNN models (EfficientNetB5, DenseNet201, and Xception). VT-ChestNet is based on a modified Swin transformer (M-Swin).

The paper is structured as follows: Section "Related Works" reviews previous works. Section "Materials and Methods" describes the materials and methods used in this work including the consolidated dataset and the proposed methods. In section "Results and Discussion", we discuss the obtained results and highlight the strengths of the two proposed methods. Finally, we summarize this paper in section "Conclusion".

Related Works

DL provides powerful solutions for automating medical imaging diagnosis. It has the potential to reduce the burden of work for specialists in the practice of radiology [6]. The availability of data and high computational resources has allowed the use of neural networks and deep learning to advance the performance of disease detection in medical images. Various studies have investigated the use of DL techniques in order to detect diseases using CXR images from different datasets such as, NIH [7], ChestX-ray14 [8], VinDr-CXR [9], CheXpert [10], PLCO [11], and MIMIC-CXR [12]. The obtained results show that CAD systems based on DL techniques can achieve high performances in detecting these harmful diseases.

To address the problem of overlapping chest diseases and to overcome the challenges facing radiologists, Rajpurkar et al. [13], proposed a modified DenseNet model called CheXNet which has 121 convolutional layers to detect 14 chest abnormalities. ChestX-ray14 dataset was used in this experiment to train and evaluate the model. CheXNet showed impressive results surpassing radiologists level obtaining an average area under curve (AUC) of 84.11% and an F1-score of 43.50% on a test set of 420 images. Sze-To et Wang [14] introduced a DL model named tCheXNet with 122 deep layers to identify pneumothorax using CXR images from CheXpert dataset. tCheXNet employs CheXNet model proposed by Rajpurkar et al. [13] with transfer learning. This model achieved an AUC of 70.80% for the classification of pneumothorax. Khoiriya et al. [15], introduced a custom DCNN model to detect pneumonia. A dataset of 5,856 CXR images collected from Kaggle was used in this experiment. The proposed model used with data-augmentation techniques (rotation, resize, and flip) achieved high results with an accuracy (ACC) of 83.83%. Wang and Xia [16], proposed a two-branches pipeline named ChestNet which incorporates the attention mechanism into its architecture. The first branch in ChestNet is for feature extraction and classification of 14 chest diseases using a ResNet152 as a backbone. The second branch implements an attention mechanism to correlate class labels with disease location. ChestX-ray14 dataset was used to train and evaluate ChestNet which showed high performance, achieving an average AUC of 78.10%.

Pham et al. [17] introduced a multi-label classification approach using a DL architecture for the detection of 14 thoracic abnormalities. The proposed model achieved high performance on CheXpert dataset. A mean AUC of 94.00% was obtained for the detection of five lung diseases (atelectasis, cardiomegaly, edema, consolidation, and effusion). Gundel et al. [18] presented a DenseNet121 model with a location-aware mechanism to classify 12 chest abnormalities using CXR images from two publicly available datasets (PLCO and ChestX-Ray14). The proposed DenseNet121 model achieved an average AUC of 87.40% outperforming four other models evaluated on the same collection of data including, ResNet50, GoogleNet, VGG16, and AlexNet. Kim et al. [19] proposed an approach based on EfficientNet-V2M used with transfer learning for classification of CXR images from ChestX-ray14 dataset. The model classified images into three classes (normal, pneumonia and pneumothorax). EfficientNet-V2M achieved an overall ACC of 82.15%, a specificity (SPE) of 91.65%, and a sensitivity (SEN) of 81.40%. This model was tested on a dataset with four classes (normal, tuberculosis, pneumothorax, pneumonia). It obtained a mean ACC of 82.20%. Blais and Akhloufi [20] employed different DCNN models to classify CXR images into six classes. The best performing model was Xception with Adam optimizer. It achieved an average AUC of 95.84% surpassing other DCNN models. This model showed high performance when evaluated for the classification of 14 abnormalities from CheXpert dataset. It achieved an overall AUC of 94.90%.

Materials and Methods

In this section, we present the structure of our consolidated CXR dataset as well as the two DL methods proposed to implement the two-step classification approach.

Dataset

In this experiment, we used a consolidated dataset that includes 26,316 CXR images acquired from two open access datasets. The first is CheXpert which is a large collection of CXR images with a total of 224,316 radiology images diagnosed with 14 chest abnormalities. All images in CheXpert are labeled using radiology reports with natural language processing (NLP) algorithms. This dataset was collected by Stanford University Hospital where each image was diagnosed with one or more anomalies. The second is the VinDr-CXR dataset, which comprises a total of 18,000 CXR images, including the location of findings and classification of different thoracic diseases. VinDr-CXR was collected at H108 Hospital and Hanoi Medical University Hospital. Each image in this dataset is diagnosed by one or more experienced radiologists. Therefore, one image can be diagnosed with two different diseases by one or more different radiologists.

In our consolidated dataset, each image corresponds to a single disease. For CXR images acquired from CheXpert, we selected only those diagnosed with a single disease. For

CXR images collected from VinDr-CXR dataset, we retain only those images for which three or more radiologists agree on the presence of the same disease, and we excluded the remaining ones (e.g. if three or more radiologists agree that the diagnosed image has cardiomegaly, that same image will be added to our dataset; otherwise, the image will be excluded). The images in our dataset are grouped in three classes (10,606 normal cases, 8,584 with lung disease, and 7,162 heart disease) for the classification of diseases by infected organs and eight classes (normal, pleural effusion, pulmonary fibrosis, atelectasis, aortic enlargement, enlarged cardiomediastinum, and cardiomegaly) for binary classification of specific diseases. Figure 2 gives an overview on the distribution of CXR images in our consolidated dataset.

Proposed Methods

In our previous research, we found that the detection of CXRs using end-to-end DL pipelines was challenging and showed many weaknesses due to the similarities of symptoms between diseases, which typically manifest as opacities around the infected organ. To this end, we proposed a two-step process for the classification of CXR images. In the first step, we intend to perform a multi-class classification of CXRs on the basis of the infected organ (normal, lung or heart). In the second step, we aim to binary classify CXR images of specific diseases (normal or abnormal), the seven diseases in our dataset are: cardiomegaly, enlarged cardiomediastinum, and aortic enlargement (heart diseases); atelectasis, pneumothorax, pulmonary fibrosis and pleural effusion (lung diseases). Two different methods are proposed in this paper, in order to perform the two-step classification of CXR images. The first method is named DC-ChestNET which is based on DCNN models and the second method is called VT-ChestNet which implements a modified vision transformer model.

DC-ChestNet Method

As a first method, we proposed an architecture called DC-ChestNet for chest disease detection using convolutional neural network (CNN) models and CXR images. We used several DCNN models including ResNet50, DenseNet121, DenseNet169, DenseNet201, EfficientNetB5, EfficientNetB6, EfficientNetB7, VGG16, VGG19, and Xception. After fine-tuning, training, testing and comparing the performance of these models, we selected the best performing ones (DenseNet201, Xception, and EfficientNetB5) for an ensemble learning (EL) pipeline in order to perform an accurate classification. The selected models served as backbones to extract deep features from each of our classes in the two steps of our approach. We used our consolidated CXR dataset that has images from two publicly available sources (CheXpert and VinDr-CXR datasets) to train in parallel the three models in our pipeline.

Through using an EL of DCNN models from different families, the power of each model can be leveraged while extracting features which can also improve the overall prediction task of the models. This technique is often more efficient, robust, and has computational advantage over a single model. Using an EL approach also reduces the generalization error of the prediction. In addition, EL techniques correct inter-model errors and avoid overfitting, especially in the case of scarce data. The three DCNN models in our pipeline are described in the following.

DenseNet201 is a DCNN model with 201 deep layers and 20 million parameters. It was introduced by Huang et al. [21] in order to overcome the decreased accuracy resulting from the vanishing gradient in deep neural networks. This model connects all layers (each layer to every other layer) in a feed-forward manner using dense connections through dense blocks. DenseNet models showed high performance on ImageNet dataset [22].

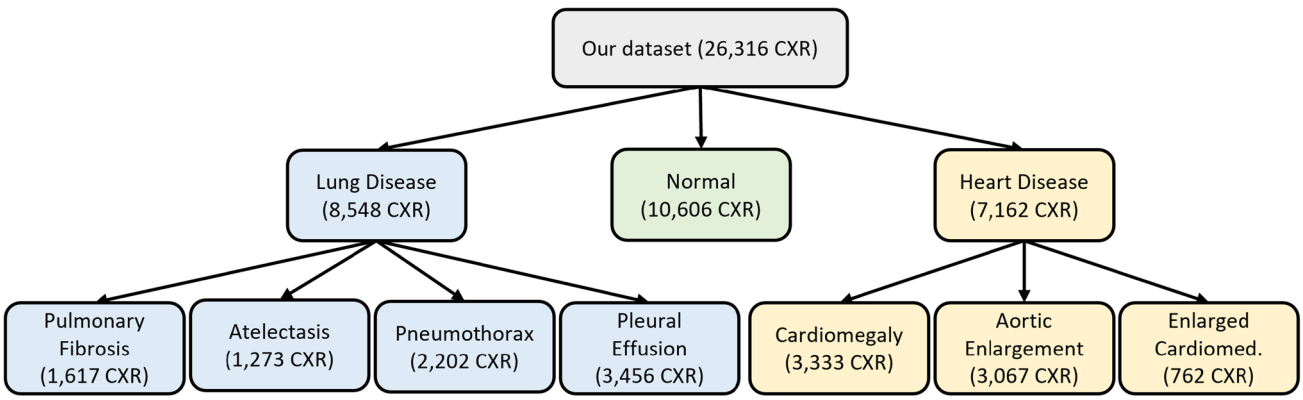

Fig. 2 Distribution of X-ray images in consolidated dataset

Xception also called "extreme inception", is a DCNN model with 71 deep layers and a total of 22.8 million parameters [23]. It is an extensive version of the Inception architecture with the implementation of depth-wise separable convolutions instead of the standard convolutions in the original model. The convolutional layers in Xception architecture have few parameters than regular layers in Inception which makes the Xception model less likely to overfit. Xception showed high performance on the ImageNet dataset, outperforming multiple models including Inception and ResNet models.

EfficientNetB5 is part of a family of eight DCNN models called EfficientNet, introduced by Google AI [24]. The eight models of EfficientNet range from B0 to B7 where the largest is B7. EfficientNets showed higher accuracy and better efficiency in comparison to existing CNNs. The EfficientNet architectures are based on a scaling approach that uses a compound coefficient to consistently scale the three dimensions (resolution, depth, and width). This results in higher performance and greater accuracy of the models.

In addition to the three DCNN models in DC-ChestNet method, two layers were added to this architecture which are an average pooling layer to calculate the mean values for the feature map patches and dense layer with a softmax activation function for the first step (multi-class classification based on infected organs). For the second step of our approach, a dense layer with a sigmoid activation function was used to classify CXR images of specific diseases into healthy or unhealthy. Moreover, 20% of the hidden layer neurons were randomly ignored when training our model using a dropout function. This reduces the dependency between neurons and avoids overfitting. Figure 3 depicts the architecture of the DC-ChestNet architecture.

VT-ChestNet Method

Models based on convolutional neural networks have long dominated the field of computer vision. After the release of the AlexNet model [25] and its impressive results on the ImageNet challenge, several DCNN models were released with architectures varying in terms of complexity, size, depth, number of trainable and non-trainable parameters. The concept of convolution in these models was a key success factor as it helps to recognize the existing patterns and extract the local features of a given image. Nevertheless, DCNN models remain limited in terms of the global context modeling, the high computational cost, and the spatial invariance to the input data. This makes the models in many instances incapable of detecting new features at non-trained locations and lacking in ability to understand long-range dependencies in images.

To overcome these drawbacks, the first vision transformer (ViT) was proposed by Dosovitskiy et al. [26]. ViT employs a self-attention mechanism that connects positions of a single patch to calculate a representation of that same patch. This mechanism enables long-range dependencies to be encoded and facilitates the learning of high expressive representations. Vision transformers are usually based on splitting images into several patches, and embedding positions as input to the transformer encoder.

Recently, vision transformer models started to play a prominent role in many applications in the computer vision field such as image classification [27], object detection [28],

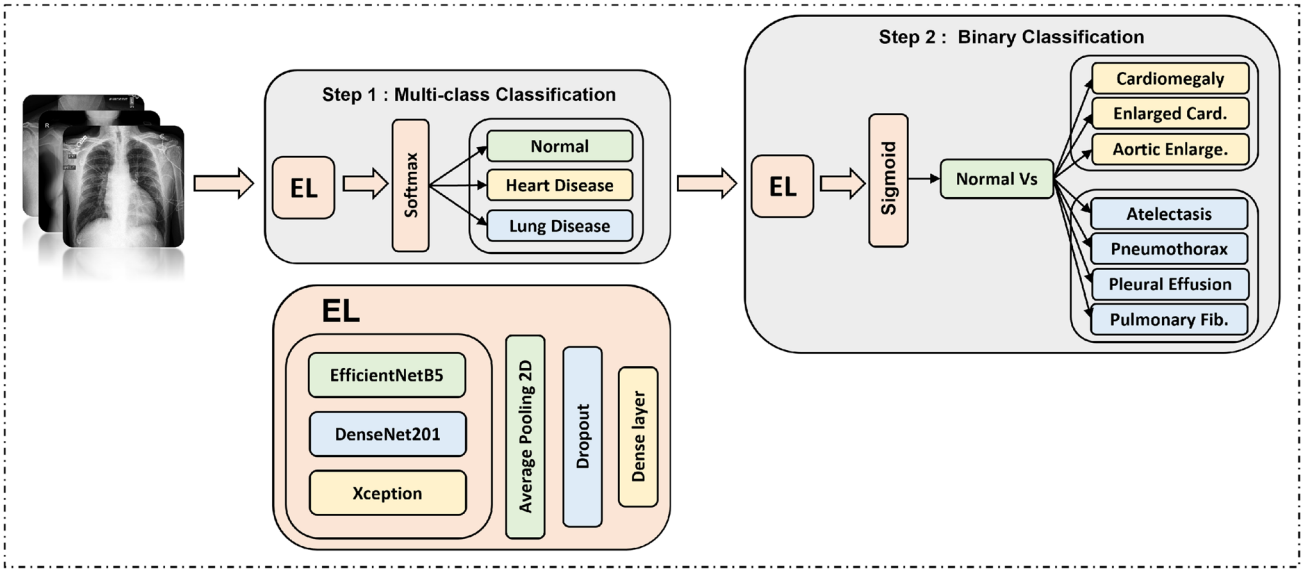

Fig. 3 The proposed DC-ChestNet architecture

semantic segmentation [29]. They showed impressive results compared to DCNN approaches.

To explore the potential of vision transformer, we proposed a novel method, named VT-ChestNet, in order to perform the two-step classification approach using our consolidated CXR dataset. VT-ChestNet method is based on a modified Swin transformer (M-Swin) architecture, as illustrated in Fig. 4. M-Swin utilizes a shifted window partitioning technique to shift between two partitioning configurations in its successive blocks. It generates hierarchical feature maps with linear computational complexity for the input CXR images by merging the image patches in its deep layers, thereby addressing the high computational complexity problem when using high resolution input images.

M-Swin has four stages (stage 1, stage 2, stage 3, and stage 4) each of them implements two blocks that are Swin Transformer Block and Patch Merging block. Swin Transformer block includes two sub-blocks. Each sub-block comprises an attention mechanism, a normalization layers, and a MLP layer. The first sub-block employs a W-MSA (Window standard multi-head self-attention) mechanism. The second adopts a SW-MSA (Shifted Window standard multi-head self-attention) technique. The Patch Merging block combines the group of neighboring patches and concatenates their characteristics [30]. A dropout strategy was added to the M-Swin hidden layers. Then, a fully connected layer (FCL) was added to adapt the number of output classes by applying a softmax activation function for the first step (multi-classification) and a sigmoid activation function for the second step (binary classification).

Results and Discussion

Two DL based methods were proposed. The first called DC-ChestNet which employs three DCNN models using an EL technique. The second named VT-ChestNet is a transformer based method that implements an M-Swin architecture. For the implementation of both solutions, various hyper-parameters were utilized, including a batch size of 32, and Rectified Adam as an optimizer. Both methods were developed using a TensorFlow framework on a machine with a Tesla P100 PCIE GPU. The consolidated dataset was divided into three subsets by dedicating 80% of the CXR images to training, 10% to validation and 10% to testing. Various data-augmentation techniques (randomly rotating between −15 and 15°, shearing between 70% and 100%, flipping, and translating 20 pixels of CXR images in four directions) were performed to diversify the content of the dataset and to prevent the overfitting. This generated a collection of 84,208 CXR images to be used as input of our models.

A learning rate that starts at 0.001 and multiplies by 0.9 while the model is stagnating in learning was used. In addition, an early stopping technique was employed to

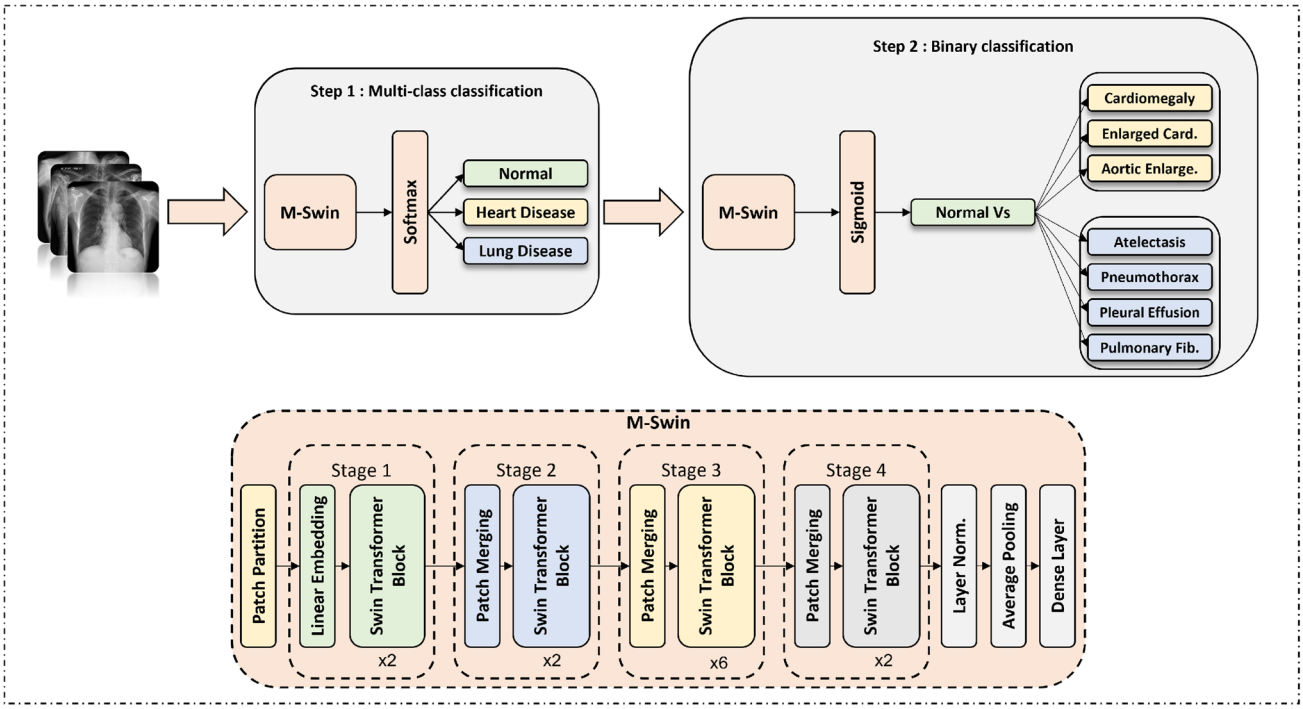

Fig. 4 The proposed VT-ChestNet architecture

Table 1 Obtained results (in %) of the proposed methods and four other models trained and tested separately on our dataset

Methods	F1-score	Specificity	Sensitivity	Accuracy	AUC
EfficientNet-B5	78.87	92.22	72.96	78.91	93.21
DenseNet-201	76.54	93.75	72.83	75.05	92.84
Xception	72.52	97.03	59.55	73.63	90.19
DenseNet-121	71.86	97.40	58.97	73.10	89.92
DC-ChestNet	**81.00**	95.65	74.35	81.10	94.89
VT-ChestNet	79.34	**98.60**	**84.21**	**82.36**	**95.13**

Bold highlight the best results or the average results obtained by each architecture

terminate training after 15 epochs of unimprovement during the validation process of the model. The three DCNN models (EfficientNetB5, DenseNet201, and Xception) in DC-ChestNet method are pretrained on ImageNet dataset. The M-Swin transformer in VT-ChestNet is pretrained on ImageNet22k dataset.

Performance of Proposed Methods on Multi-class Classification

In this first step, we evaluate the performance of the two proposed methods (DC-ChestNet and VT-ChestNet) with four different state-of-the-art DCNN models, three of which are implemented in DC-ChestNet as backbones to extract features. These models (EfficientNet-B5, Xception, DenseNet-121, and DenseNet-201) were selected for their high performance in similar classification challenges. They were trained, validated, and tested separately on the same CXR dataset.

As presented in Table 1, the two proposed methods achieved higher results compared to single DCNN models. VT-ChestNet method showed the highest performance with an SPE of 98.60%, an SEN of 84.21%, an ACC of 82.36%, and an AUC rate of 95.13% outperforming all other experiments including DC-ChestNet method which achieved a slightly better F1-score with 81.00%. DC-ChestNet which implements an EL of DCNN models outperformed the single DCNN models for this multi-class classification of CXR images based on infected organs. DC-ChestNet obtained an AUC of 94.89%, an ACC of 81.10%, an SPE of 95.65%, an SEN of 74.35%, and an F1-score of 81.00%.

Both methods DC-ChestNet and VT-ChestNet showed strengths over single DCNN models. DC-ChestNet which employs three different models as backbones in an EL architecture allowed extracting a wide variety of characteristics for the three classes at this step (normal, heart disease, and lung diseases). VT-ChestNet method with M-Swin architecture showed better results thanks to the mechanism of self-attention with shifted windowing that lowered latency and performed at a linear complexity.

Table 2 Binary classification results (in %) of heart and lung diseases using DC-ChestNet method

Pathology	F1-score	Specificity	Sensitivity	Accuracy	AUC
Aortic Enlarge.^{h}	94.55	97.08	85.94	94.59	98.92
Cardiomegaly^h	94.64	97.44	86.13	94.69	98.94
Enlarged Cardio.^{h}	99.81	99.92	99.91	99.79	99.91
Average^h	**96.33**	**98.14**	**90.66**	**96.36**	**99.26**
Atelectasis^l	99.41	99.90	95.38	99.41	99.97
Pneumothorax^l	99.45	100.00	96.72	99.45	99.98
Pleural Eff.^{l}	96.82	99.35	88.61	96.87	99.31
Pulmonary Fib.^{l}	94.18	99.05	87.94	94.60	99.01
Average^l	**97.46**	**99.57**	**92.16**	**97.58**	**99.57**

Bold highlight the best results or the average results obtained by each architecture

Note: ^{h} for heart disease, ^{l} for lung disease

Performance of Proposed Methods on Binary Classification

For this step, we performed binary classification of CXR images belonging to eight classes, including normal cases, three heart diseases (cardiomegaly, aortic enlargement, and enlarged cardiomediastinum), and four lung diseases (pulmonary fibrosis, pneumothorax, atelectasis, and pleural effusion) using DC-ChestNet and VT-ChestNet methods. As shown in Table 2, DC-ChestNet achieved the following average scores for heart diseases: an F1-score of 96.33%, an SPE of 98.14%, an SEN of 90.66%, an ACC of 96.36% and AUC of 99.26%. For the lung diseases classification using DC-ChestNet method, the obtained average scores are as follows: an F1-score of 97.46%, an SPE of 99.57%, an SEN of 92.16%, an ACC of 97.58%, and an AUC of 99.57%. The second proposed method (VT-ChestNet) showed better performance compared to DC-ChestNet as shown in Table 3. This method has shown high average scores for both heart and lung diseases outperforming DC-ChestNet. The obtained results are as follows: an F1-score of 97.04%, an SPE of 98.47%, an SEN 95.07%, an ACC of 96.72%, and an AUC of 99.35% for heart diseases, and an F1-score of 98.79%, an SPE of 99.60%, an SEN 95.72%, an ACC of 98.55%, and an AUC of 99.83% for lung diseases.

The use of ensemble learning with three different CNN models in a single pipeline (DC-ChestNet) improved

Table 3 Binary classification results (in %) of heart and lung diseases using VT-ChestNet method

Pathology	F1-score	Specificity	Sensitivity	Accuracy	AUC
Aortic Enlarge.^{h}	95.14	98.23	92.07	94.89	99.12
Cardiomegaly^h	96.23	97.31	93.21	95.40	99.06
Enlarged Cardio.^{h}	99.77	99.89	99.95	99.88	99.89
Average^h	**97.04**	**98.47**	**95.07**	**96.72**	**99.35**
Atelectasis^l	99.79	99.89	96.74	99.81	99.88
Pneumothorax^l	99.42	99.87	96.83	99.14	99.94
Pleural Eff.^{l}	98.56	99.93	94.18	98.09	99.51
Pulmonary Fib.^{l}	97.37	98.70	95.14	97.17	99.97
Average^l	**98.79**	**99.60**	**95.72**	**98.55**	**99.83**

Bold highlight the best results or the average results obtained by each architecture

Note: ^{h} for heart disease, ^{l} for lung disease

accuracy, robustness, generalization, feature representation, and reduced overfitting for both steps of our approach. The pipeline achieved high results for multi-classification of CXR images based on the infected organ at the first step and showed high potential for detecting specific heart and lung diseases at the second step, thanks to error compensation between the three models. Compared to DC-ChestNet, the second method (VT-ChestNet) demonstrated slightly better performance due to the attention mechanism implemented in the modified Swin transformer. Additionally, the use of data augmentation techniques showed high potential in generating a large number of samples with a wide variety of characteristics, which helps to train the models effectively resulting in a significant improvement in the overall performance of the two proposed methods.

Conclusion

In this paper, we proposed a novel approach composed of two steps for chest disease classification using new DL architectures. In the first step (multi-classification), we classified chest X-ray (CXR) images into three classes (normal, lung disease, and heart disease). In the second step (binary classification), we classified CXR images into specific diseases to predict whether it is a normal or abnormal case. A dataset of 26,316 CXR images was consolidated by merging images from two public datasets (VinDr-CXR and CheXpert) to train, validate, and test our methods. For this work, we implemented two DL methods to perform our two-step classification approach. The first is called DC-ChestNet which is based on an ensemble learning (EL) of three deep convolutional neural network (DCNN) models. The second method named VT-ChestNet is based on a modified Swin transformer (M-Swin). Our two methods showed high performance outperforming state-of-the-art models trained and tested individually on our consolidated dataset including, DenseNet121, DenseNet201, EfficientNetB5, and Xception. VT-ChestNet outperformed DC-ChestNet by obtaining the best results on our dataset for the two-steps of our approach. VT-ChestNet achieved an area under curve (AUC) of 95.13% for the first step. For the second step, it obtained an average AUC of 99.26% for heart diseases and an average AUC of 99.57% for lung diseases.

In future work, we intend to investigate the use of different datasets with multi-labeled images. We plan to examine the potential of more transformer-based architectures and implement an explainability algorithm to show the features that our models focused on to classify the images. We aim to explore the ability of the proposed methods to be generalized to other diseases and test their ability to perform multiple classification of chest diseases.

Funding This work has been supported in part by the New Brunswick Health Research Foundation (NBHRF) and the New Brunswick Innovation Foundation (NBIF), New Brunswick Priority Occupation Student Support Fund (NBPOSS) POF2021-006.

Availability of Data and Materials This work uses a consolidated dataset of X-ray images from two open-access datasets, VinDr-CXR and CheXpert, see reference [9, 10] for data availability. More details about the data are available under Sect. 3.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Ambati A, Dubey SR. Ac-covidnet: Attention guided contrastive cnn for recognition of covid-19 in chest x-ray images. In: International Conference on Computer Vision and Image Processing, pp. 71–82 (2022)
2. Russo P. Handbook of X-ray Imaging: Physics and Technology. Series in Medical Physics and Biomedical Engineering. CRC Press, New York (2017)
3. World Health Organization: To X-ray or not to X-ray. https://www.who.int/news-room/feature-stories/detail/to-x-ray-or-not-to-x-ray- (Accessed: November 2022)
4. Nasser AA, Akhloufi M. Chest diseases classification using cxr and deep ensemble learning. In: 19th International Conference on Content-based Multimedia Indexing, pp. 116–120 (2022)
5. Marciniuk DD, Schraufnage DE, Ferkol T, Fong KM, Joos G, Varela VL, Zar H. The Global Impact of Respiratory Disease. European Respiratory Society, ??? (2017)
6. Tsuneki M. Deep learning models in medical image analysis. J Oral Biosci. 2022;64(3):312–20.

7. Demner-Fushman D, Kohli MD, Rosenman MB, Shooshan SE, Rodriguez L, Antani S, Thoma GR, McDonald CJ. Preparing a collection of radiology examinations for distribution and retrieval. J Am Med Inform Assoc. 2016;23(2):304–10.
8. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017; pp. 2097–2106
9. Nguyen H, Lam K, Le L, Pham H, Tran D, Nguyen D, Le D, Pham C, Tong H, Dinh D, et al. Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. 2021; preprint arXiv:2012.15029
10. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghgoo B, Ball R, Shpanskaya K, Seekins J, Mong D, Halabi S, Sandberg J, Jones R, Larson D, Langlotz C, Patel B, Lungren M, Ng A. Chexpert : A large chest radiograph dataset with uncertainty labels and expert comparison. In: AAAI Conference on Artificial Intelligence, 2019; pp. 590–597
11. Zhu CS, Pinsky PF, Kramer BS, Prorok PC, Purdue MP, Berg CD, Gohagan JK. The prostate, lung, colorectal, and ovarian cancer screening trial and its associated research resource. JNCI. 2013;105(22):1684–93.
12. Johnson AE, Pollard TJ, Berkowitz SJ, Greenbaum NR, Lungren MP, Deng C-Y, Mark RG, Horng S. Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. Sci Data. 2019;6(1):1–8.
13. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding DY, Bagul A, Langlotz CP, Shpanskaya KS, Lungren MP, Ng AY. Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. 2017; arXiv preprint arXiv:1711.05225
14. Sze-To A, Wang Z. tchexnet: Detecting pneumothorax on chest x-ray images using deep transfer learning. In: International Conference on Image Analysis and Recognition, 2019; pp. 325–332
15. Khoiriyah SA, Basofi A, Fariza A. Convolutional neural network for automatic pneumonia detection in chest radiography. In: International Electronics Symposium (IES), 2020; pp. 476–480
16. Wang H, Xia Y. Chestnet: A deep neural network for classification of thoracic diseases on chest radiography. 2018; arXiv preprint arXiv:1807.03058
17. Pham HH, Le TT, Tran DQ, Ngo DT, Nguyen HQ. Interpreting chest X-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels. Neurocomputing. 2021;437:186–94.
18. Guendel S, Grbic S, Georgescu B, Liu S, Maier A, Comaniciu D. Learning to recognize abnormalities in chest X-rays with location-aware dense networks. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2019; pp. 757–765.
19. Kim S, Rim B, Choi S, Lee A, Min S, Hong M. Deep learning in multi-class lung diseases' classification on chest X-ray images. Diagnostics. 2022;12(4):915.
20. Blais M-A, Akhloufi MA. Deep learning and binary relevance classification of multiple diseases using chest X-ray images. In: 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 2021; pp. 2794–2797.
21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017; pp. 4700–4708
22. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009; pp. 248–255
23. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017; pp. 1251–1258
24. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, vol. 97, 2019; pp. 6105–6114
25. Krizhevsky A, Sutskever I. H. geoffrey e.,"alex net,". Adv. Neural Inf. Process. Syst 2012; 25, 1–9
26. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. 2020; arXiv preprint arXiv:2010.11929
27. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H. Training data-efficient image transformers & distillation through attention. In: International Conference on Machine Learning, 2021; pp. 10347–10357
28. Hong W, Lao J, Ren W, Wang J, Chen J, Chu W. Training object detectors from scratch: An empirical study in the era of vision transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; pp. 4662–4671
29. Ghali R, Akhloufi MA, Mseddi WS. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. Sensors. 2022;22(5):1977.
30. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021; pp. 10012–10022.