SN Computer Science (2023) 4:414
https://doi.org/10.1007/s42979-023-01848-4

ORIGINAL RESEARCH

Vision Transformers for Lung Segmentation on CXR Images

Rafik Ghali[1] · **Moulay A. Akhloufi**[1]

Received: 20 December 2022 / Accepted: 17 April 2023 / Published online: 24 May 2023

Abstract
Accurate segmentation of the lungs in CXR images is the basis for an automated CXR image analysis system. It helps radiologists in detecting lung areas, subtle signs of disease and improving the diagnosis process for patients. However, precise semantic segmentation of lungs is considered a challenging case due to the presence of the edge rib cage, wide variation of lung shape, and lungs affected by diseases. In this paper, we address the problem of lung segmentation in healthy and unhealthy CXR images. Five models were developed and used in detecting and segmenting lung regions. Two loss functions and three benchmark datasets were employed to evaluate these models. Experimental results showed that the proposed models were able to extract salient global and local features from the input CXR images. The best performing model achieved an F1 score of 97.47%, outperforming recent published models. They proved their ability to separate lung regions from the rib cage and clavicle edges and segment varying lung shape depending on age and gender, as well as challenging cases of lungs affected by anomalies such as tuberculosis and the presence of nodules.

Keywords Lung segmentation · Deep learning · Chest X-rays · Vision transformers · Medical image analysis

Introduction

Chest X-ray (CXR) image is the most popular diagnosis technique among existing medical imaging methods due to its low cost, ease of acquisition, and wide availability [1]. Each CXR image can identify multiple anomalies simultaneously. It requires manual validation by radiologists. However, the analysis of a large number of CXRs is a heavy task on medical personnel, especially radiologists. It takes time thus slowing the process of patient diagnosis.

Accurate segmentation of the lungs in CXR images is the basis of an automated CXR image analysis system, as the lungs are the region of interest for many pulmonary and thoracic diseases, such as emphysema, tuberculosis, cardiomegaly, pneumothorax, and lung cancer. Lung contours can also display useful and important information for diagnosis of life-threatening illnesses [2].

Lung segmentation has shown impressive results using deep learning models to detect the area of the lungs from other organs. Nevertheless, the presence of the rib cage and clavicles as well as wide variations of lung shape due to age and gender still present challenges.

To address this challenge, we use vision Transformers, which employ an attention mechanism in extracting important features and removing noisy and irrelevant features. Five methods, namely TransM, ARSeg [3], TransUNet [4], Medical Transformer (MedT) [5], and UNeXt [6], were developed to segment the lungs from CXR images. Two loss functions that are Dice loss [7] and Combo loss (Cross Binary Entropy Dice) [7, 8] were used to reduce detection errors using three public datasets, JSRT (Japanese Society of Radiological Technology) [9], Shenzhen [10, 11], and MC (Montgomery) [10, 11].

The main contributions of this paper are:

1. We develop and use novel semantic segmentation methods in detecting and segmenting lung areas in CXR images to address the limitations of state-of-the-art work.

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

✉ Rafik Ghali
rafik.ghali@umoncton.ca

Moulay A. Akhloufi
moulay.akhloufi@umoncton.ca

[1] Perception, Robotics, and Intelligent Machines (PRIME), Department of Computer Science, Université de Moncton, Moncton, NB E1A 3E9, Canada

2. We evaluate and compare the performance of these methods using two loss functions (Dice loss and Combo loss).
3. Experiments on three public datasets (MC, JSRT, and Shenzhen) demonstrate the potential of these methods, giving a promising performance compared to recent published work.
4. We demonstrate the potential of these methods to distinguish the lung areas from other tissues or organs and to surpass challenging cases such as the presence of the clavicles and rib cage, lung affected with tuberculosis and nodules (benign and malignant), and the large variation in lung shape due to age and gender.

The paper is structured as follows: the next section reviews previous work on semantic segmentation of lungs in CXR images. The subsequent section introduces the proposed methods, the dataset used, and the metrics employed in training and testing. In the penultimate section, we present and discuss the experimental results. The final section summarizes this work.

Related Works

With the development of AI-based computer vision tasks, deep learning models have become the most widely used technology for lung segmentation. Many studies were proposed in the literature as shown in Table 1. For instance, Hwang et al. [12] introduced a deep CNN (Convolutional Neural Network) adopting atrous convolutions to detect and segment lung areas. An accurate Dice coefficient of 98% was achieved using the JSRT dataset and a network-wise learning strategy. Islam et al. [13] developed a lung segmentation approach based on the encoder–decoder U-Net [14]. A Dice score of 98.6% was obtained surpassing previous work on two datasets, Shenzhen and MC. Liu et al. [15] proposed an improved U-Net to accurately identify and segment lung zones. This model used a pretrained EfficientNet-B4 [16] as an encoder. The leaky activation function [17] and the residual block [18] were employed in the decoder. They obtained Jaccard scores of 95.5%, 95.8%, and 97.4% on MC, JSRT, and NIH datasets, respectively. In particular, this model proved its ability to segment lung areas in difficult cases such as abnormal cases (pleural effusion, lung deformation, etc.) and blurred CXR images. Dai et al. [19] presented a novel segmentation network, SCAN (Structure Correcting Adversarial Network), which contains two models: FCN (fully convolutional network) and a critic network. First, the FCN generates a predicted mask. Then, the critic network model guides the FCN to obtain a segmentation resembling the ground truth. This method achieved an IoU (Intersection-Over-Union) of 95.1%, better than existing models on MC and JSRT datasets. Chen et al. [20] exploited the U-Net as a lung segmenter to detect lung regions and generate indices for different types of lung diseases such as cardiomegaly, emphysema, lung nodules, etc. This segmenter was used as the first step in the multi-label CXR image classification task. The performance of U-Net was evaluated using three available CXR datasets that are MC, JSRT, and NIH Chest X-ray dataset [21]. It showed its efficiency to improve the result of the classification task on CXR images. Mittal et al.

Table 1 Existing models for lung segmentation

References	Methodology	Dataset	Results (%)
[12]	Deep CNN	JSRT: 247 images	Dice = 98.0
[13]	U-Net	Shenzhen and MC: 753 images	Dice = 98.6
[15]	Improved U-Net	MC: 138 images JSRT: 247 images NIH: 2785 images	Jaccard = 95.5 Jaccard = 95.8 Jaccard = 97.4
[20]	U-Net	MC, JSRT, and NIH: 485 images	Dice = 97.3
[19]	SCAN	MC and JSRT: 385 images	IoU = 95.1
[22]	LF-SegNet	MC and JSRT: 385 images	Accuracy = 98.73
[23]	U-Net with VGG-16	JSRT: 247 images	Jaccard = 96.1
[28]	InvertedNet	JSRT: 247 images	Jaccard = 95.0
[30]	XNet+	JSRT: 247 images	Dice = 97.8
[31]	Encoder–decoder with VAE	Shenzhen and MC: 704 images	Accuracy = 88.15
[33]	XLSor	MC, JSRT, and NIH: 485 images	Accuracy = 97.6
[35]	Attention UW-Net	NIH: 200 images	F1-score = 95.7
[40]	U-Net and transfer learning	General COVID and CHUAC: 6502 images	Accuracy = 97.61
[42]	Deep LF-Net	MC: 138 images JSRT: 247 images Shenzhen: 662 imagesIndian: 668 images	Dice = 94.19 Dice = 96.85 Dice = 90.54 Dice = 97.20
[36]	DCNN	MC: 138 images JSRT: 247 images Shenzhen: 658 images	Jaccard = 96.6 Jaccard = 96.8 Jaccard = 96.7

[22] developed a novel fully convolutional encoder–decoder, called LF-SegNet, to detect and segment lungs from CXR images. The experimental results showed a high accuracy with 98.73% on the MC and JSRT datasets. The U-Net model was also adopted by Frid-Adar et al. [23] with three methods (FCN [24], Dilated Residual Networks [25], and Fully Convolutional DenseNet [26]). This model achieved a Jaccard score of 96.1% using a pre-trained VGG-16 [27] as the backbone and the JSRT dataset. Novikov et al. [28] evaluated three fully convolutional methods (InvertedNet: fully-convolutional network with fewer parameters, All-Convolutional: simplifying fully convolutional network by learning pooling, and All-Dropout: fully convolutional network with restrictive regularization) in detecting and segmenting multi-organs that are lung, heart, and clavicles. Based on the Jaccard score result, InvertedNet was selected as the best performing model with a Jaccard score of 95% using 247 CXR images of the JSRT dataset, surpassing the result of the human observer [29] and U-Net as baseline methods. Gómez et al. [30] also addressed the problem of segmenting multi-organs that are lungs, hearts, and clavicles on CXR images. Four deep convolutional networks, XNet, XNet+, RX-Net as a simplification of X-Net, and RX-Net+, were employed by modifying the baseline of U-Net, and InvertedNet methods. XNet+ reached the best Dice score of 97.8% for lung segmentation. Selvan et al. [31] developed a novel lung segmentation method trained on normal CXR images to segment lungs on high opacity CXR images. This method is an encoder–decoder with VAE (variational autoencoders) [32], which was applied to perform data imputation. An accuracy of 88.15% was achieved using 704 images from the MC and Shenzhen datasets. It demonstrated the possibility of extending this method to extreme anomaly cases. Tang et al. [33] proposed a novel architecture, XLSor (X-ray Lung Segmentor), which includes Criss-Cross Attention [34] module and data augmentation technique via abnormal CXR pairs construction. Experimental results with an accuracy of 97.6% validate the robustness of XLSor for the lung segmentation task. Pal et al. [35] proposed a new method, called attention UW-Net, to improve the segmentation performance of small lesion areas. Attention UW-Net is an encoder–decoder architecture consisting of many densely connected convolutional layers, which connect the encoder and its corresponding decoder using skip connections. Using a NIH dataset (200 manually annotated CXR images), the results proved that the UW-Net model reached an efficient performance (an F1-score of 95.7%) in segmenting lungs as well as small lesions. Maity et al. [36] presented a DCNN (Deep Convolution Neural Network) as a semantic lung segmentation method from posteroanterior or anteroposterior view CXR images. DCNN was developed based on UNet++ [37] with EfficientNet-B4 as backbone and residual blocks as decoder to solve the degradation issues and to increase the performance with fewer parameters and layers. It achieved high Jaccard scores with 96.8%, 96.6%, and 96.7% using JSRT, MC, and Shenzhen datasets, respectively, employing CLAHE (Contrast Limited Adaptive Histogram Equalization) [38] and Top-Bottom-Hat transform [39] as a preprocessing techniques and data augmentation techniques (rotation, width and height shift, shearing, zoom, and flip). Vidal et al. [40] also proposed a system to segment lung regions, especially lungs affected by COVID-19 on CXR images collected from portable X-ray devices. This system used the U-Net model as the lung segmenter and two transfer learning stages. Using two COVID-19 datasets (the general COVID dataset [41] and the CHUAC dataset acquired from portable devices [40]), it obtained a high accuracy of 97.61% for COVID-19 patients. It demonstrated its robustness in segmenting lung areas from portable X-ray devices overcoming the low quality of CXR images and sample scarcity. Singh et al. [42] also developed a novel semantic lung semantic method, Deep LF-Net, based on DeepLabv3+ [43] architecture. A novel dataset was presented consisting of 688 CXR images [42]. This dataset contains healthy and unhealthy CXR images of patients contaminated with tuberculosis, pleural effusion, interstitial lung disease, lung cancer, and chronic obstructive pulmonary disease. Deep LF-Net obtained Dice values of 96.85%, 94.19%, 90.54%, and 97.20% with JSRT, MC, Shenzhen, and their proprietary datasets, respectively. They used MobileNetV2 [44] as a backbone.

Materials and Methods

In this section, we first present our proposed semantic segmentation models for lung segmentation. Then, we introduce the public datasets employed in this work. Finally, we describe the metrics used in this paper.

Proposed Methods

We used five models, namely ARSeg, TransM, Medical Transformer (MedT), TransUNet, and UNeXt to determine the precise lung segmentation masks.

ARSeg

ARSeg [3] is an encoder–decoder architecture with four skip connections, as shown in Fig. 1. Two attention gates are used to remove the non-pertinent and noisy characteristics transmitted by skip connections. The encoder adopted RegNetY-32GF model [45] as a backbone to extract and generate the characteristics of lungs. It comprises five convolutional blocks (Stem, stage1, stage2, stage3, and stage4) consisting of 3×3 and 1×1 convolutional layers, pooling layers, ReLU

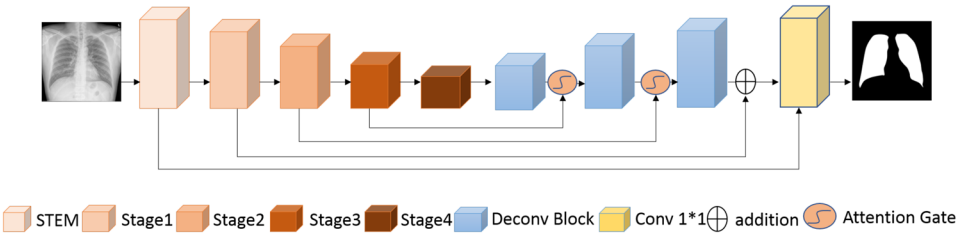

Fig. 1 The proposed ARSeg architecture

activation, batch normalization, and Squeeze-and-Excitation block. The decoder contains three Deconv blocks, which consist of a transposed convolution layer, a batch normalization layer, and a ReLU activation, then, a 1×1 convolutional layer, which determines the mask of lungs as output.

Medical Transformer (MedT)

Medical Transformer (MedT) [5] is an encoder–decoder model. First, two convolutional blocks, which contains convolutional layers, ReLU activation, and batch normalization layers are used to extract feature from input CXR images and their patches. Then, two branches (local branch and global branch) are fed by the generated feature maps. The local branch contains five encoders and five decoders. The global branch consists of two encoders and two decoders. Each encoder comprises 1×1 convolutional layers, normalization layers, and two gated multi-head attentions. Each decoder also includes a convolutional layer, a ReLU activation, and a batch normalization layer. Finally, a 1×1 convolutional layer generates the output lung mask. Figure 2 presents the architecture of the proposed MedT.

TransM

TransM is a modified MedT architecture with a dropout strategy, as illustrated in Fig. 3. It was developed to reduce the number of MedT parameters and help with the MedT memory problem. TransM employs gated position-sensitive axial attention and a Local-Global training methodology, which uses local and global branches to improve the performance of lung segmentation. The global branch contains one encoder and one decoder. The local branch also consists of five encoders and five decoders.

TransUNet

TransUNet [4] is an encoder–decoder based on U-Net architecture, as depicted in Fig. 4. It employs a pretrained ResNet-50-ViT as a backbone and three skip connections between the encoder and decoder. The encoder is a hybrid CNN-Transformer. It includes Multihead Self-Attention (MSA), Multi-Layer Perceptron (MLP) blocks, and normalization layers. The decoder contains five 3×3 convolutional layers followed by ReLU activation and four upsampling operators.

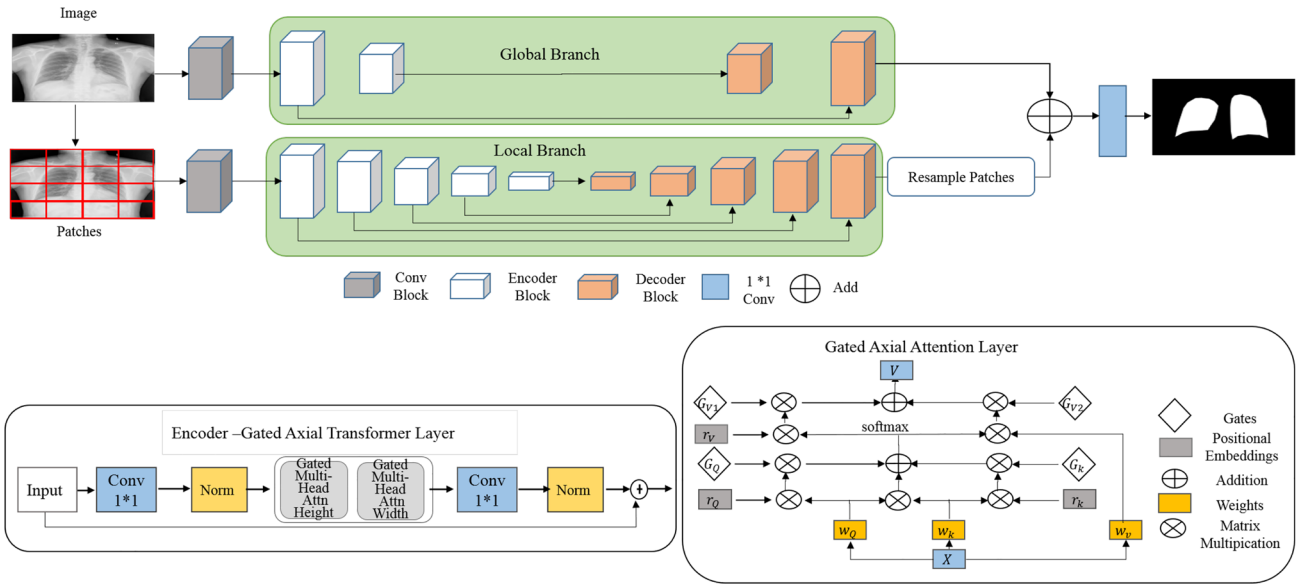

Fig. 2 The proposed MedT architecture

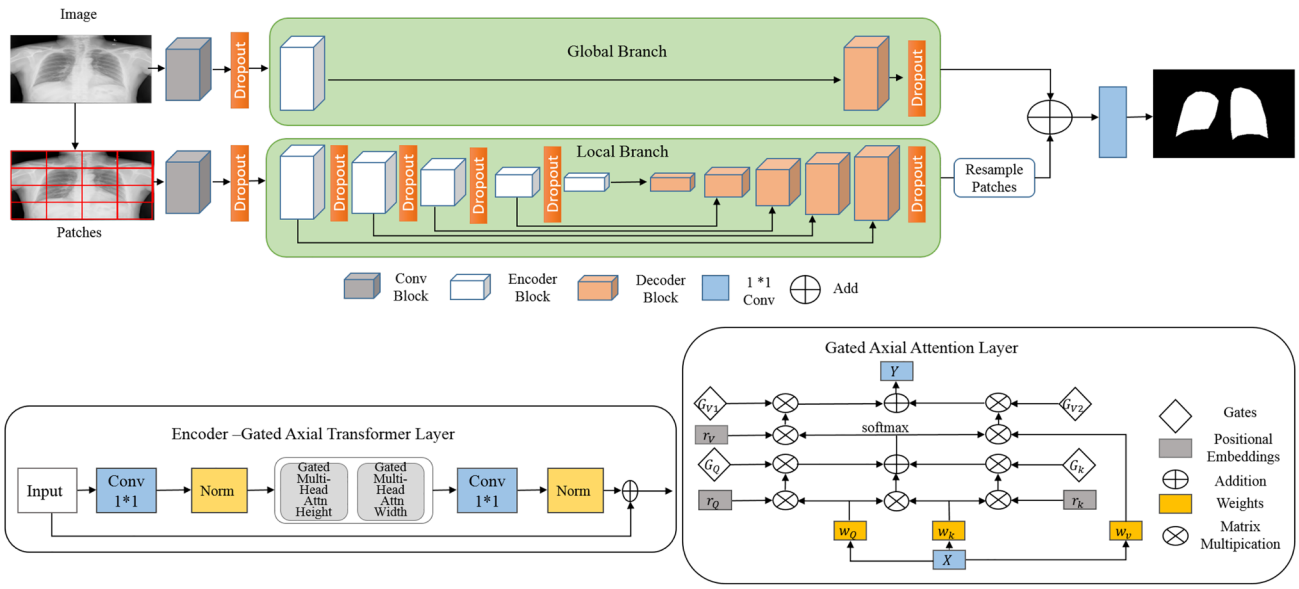

Fig. 3 The proposed TransM architecture

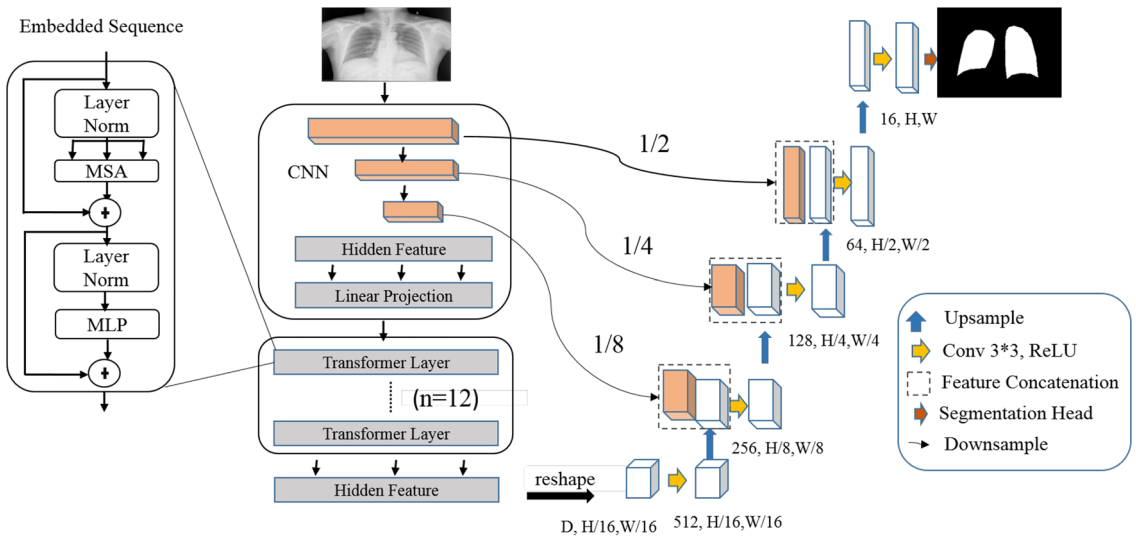

Fig. 4 The proposed TransUNet architecture

UNeXt

UNeXt [6] is an encoder–decoder architecture with four skip connections, as shown in Fig. 5. It employs MLP (multilayer perceptron) and convolutional networks to reduce the number of parameters and the computational complexity as well as improve the performance of medical image segmentation [6]. First, the encoder generates a feature map using three convolutional blocks followed by two tokenized MLP blocks. Then, the extracted features were passed through the decoder, which comprises two tokenized MLP blocks, upsampling operator, and three convolutional blocks. Each convolutional block includes a 3×3 convolutional layer, ReLU activation, max-pooling layer, and batch normalization layer. Tokenized MLP block consists of Shifted MLP blocks, a depth-wise convolutional layer, GELU activation, and normalization layers.

Dataset

Three datasets, namely Shenzhen [10, 11], MC (Montgomery) [10, 11], and JSRT (Japanese Society of Radiological

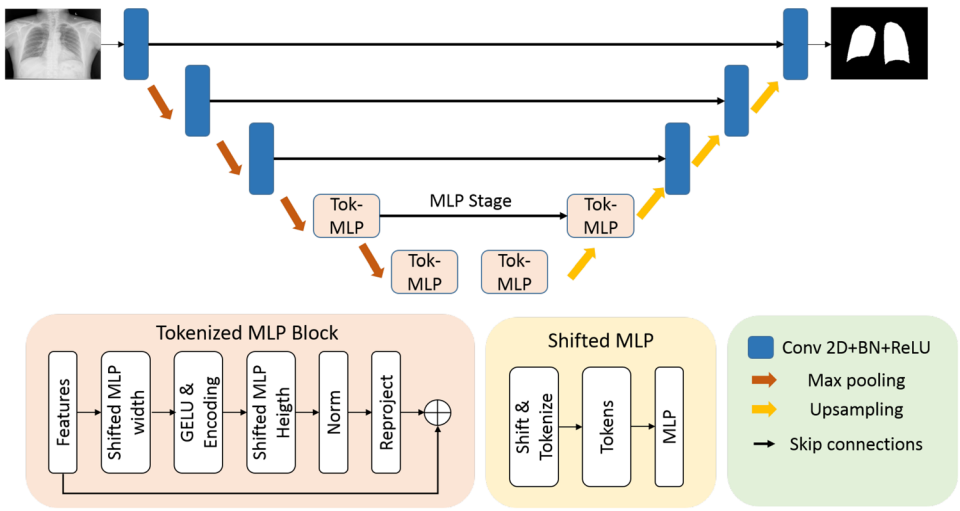

Fig. 5 The proposed UNeXt architecture

Technology) [9], were used in learning and evaluating the proposed models.

1. Shenzhen dataset is a public dataset collected by Shenzhen No.3 People's Hospital and Guangdong Medical College in China in 2012. It includes 662 CXR images with a resolution of 3000×3000 and their binary masks, annotated manually by the radiologist. It consists of 336 images with tuberculosis disease and 326 normal images. Figure 6 presents examples of Shenzhen dataset.
2. The MC dataset was developed by the Montgomery County Department of Health and Human Services, USA. It contains 138 CXR images with a resolution of 4892×4020 or 4020×4892 and their corresponding masks. It was annotated by the supervision of a radiologist. It consists of 58 images with tuberculosis disease and 80 normal images. Figure 7 shows an example of MC dataset.
3. JSRT dataset was created by the Japanese Society of Radiological Technology and the Japanese Radiological Society. It comprises 247 CXR images with a resolution of 2048x2048, including 93 normal images and 154 images with lung nodules (54 images with a benign nodule and 100 images with a malignant nodule). It was labeled by two human observers and radiologist. Figure 8 depicts some examples of the JSRT dataset.

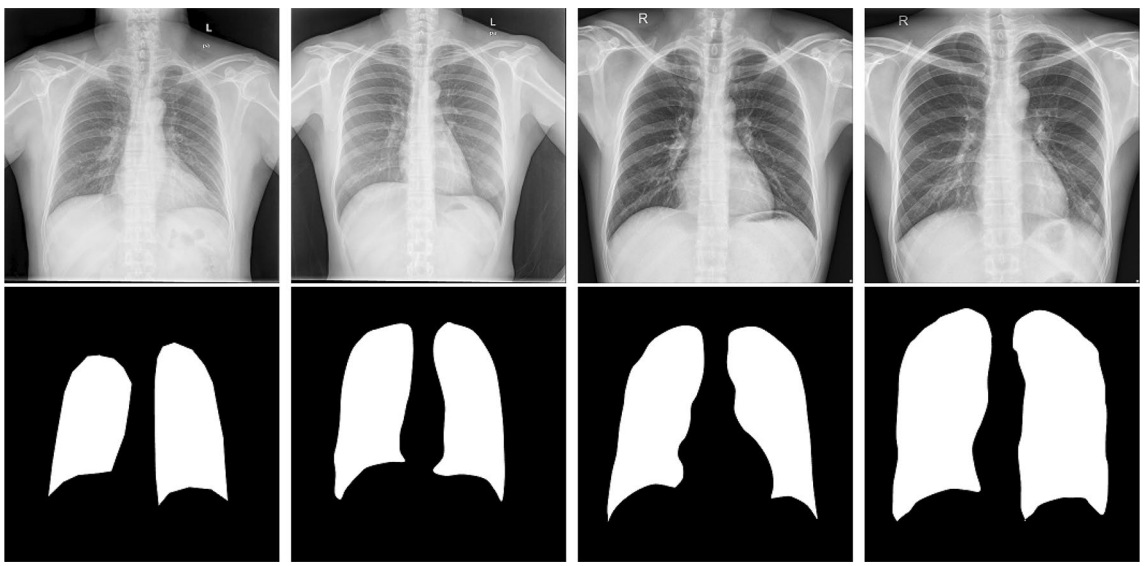

Fig. 6 Shenzhen dataset example. Top: CXR images; bottom: their binary masks

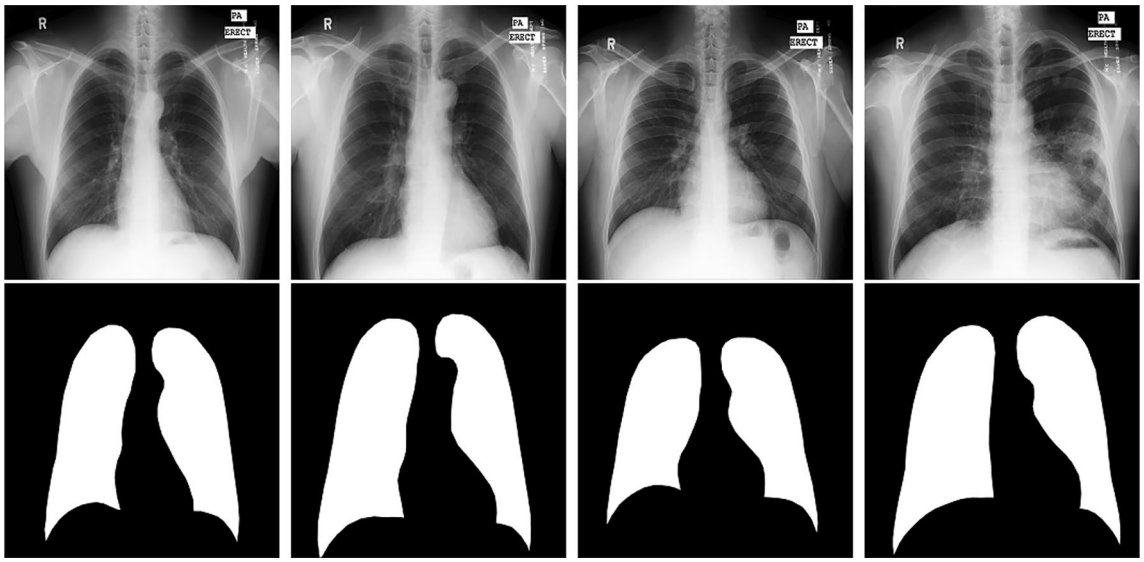

Fig. 7 Montgomery dataset example. Top: CXR images; bottom: their binary masks

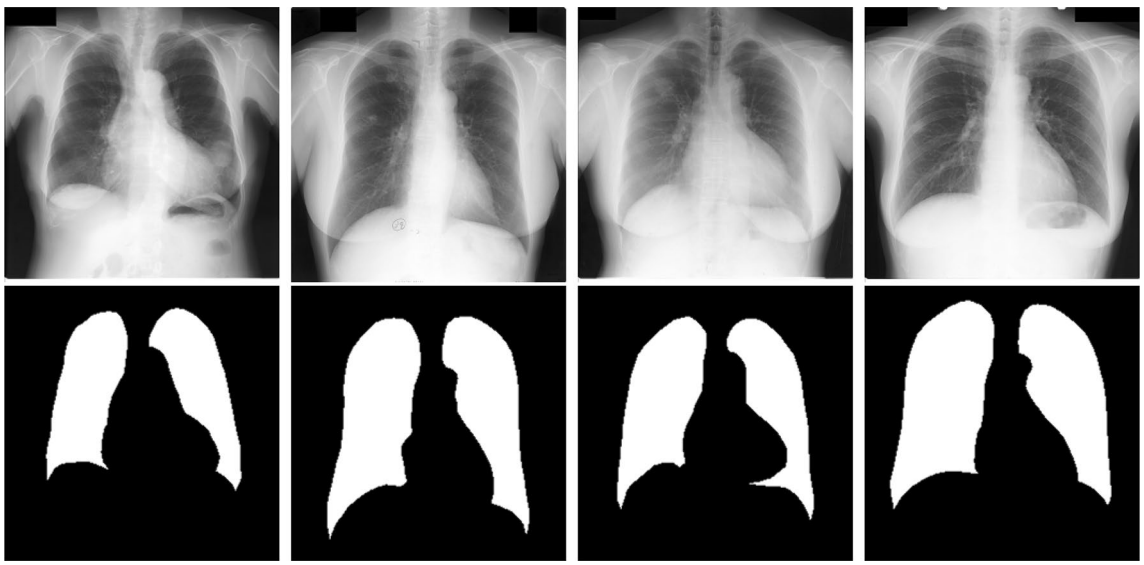

Fig. 8 JSRT dataset example. Top: CXR images; bottom: their binary masks

Evaluation Metrics

We employed two metrics (accuracy and F1-score) to evaluate our proposed segmentation methods with published works, using true-positive rate (TP), false-positive rate (FP), true-negative rate (TN), and false-negative rate (FN).

- Accuracy is the proportion of correct pixels predictions over the total number of predictions obtained by the proposed approach, as given by Eq. (1).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

- F1 score combines precision and recall metrics to measure the performance of the proposed approach, as shown by Eq. (2).

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}, \tag{2}$$

where $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$.

Results and Discussion

We developed the proposed semantic segmentation models using PyTorch [46] on a machine with GPU NVIDIA A100-SXM. We also mixed and randomly split our learning datasets (Shenzhen, MC, and JSRT) into three subsets (training: 635 CXR images, validation: 105 CXR images, and test: 211 CXR images).

The proposed models (ARSeg, MedT, TransM, and UNeXt) were evaluated from scratch (no pre-training). The proposed TransUNet model was also trained using a pre-trained ResNet50-ViT-16 on a ImageNet dataset [47] as a backbone.

For our experiments, we used as input CXR image with a resolution of 224×224, and two data augmentation techniques, namely rotation of +20 degree and horizontal flip, to diversify and augment our learning data. In addition, we employed two loss functions, Dice loss [7] and Combo loss, to reduce segmentation errors. Dice loss (DC) determines the similarity between the predicted CXR images and the input CXR mask (see Eq. 3). The Combo loss (BCD) combines the Dice loss and the Cross Binary Entropy loss [8], as shown in Eq. 4.

$$DC = 1 - \frac{2 \mid W \bigcap Y \mid}{\mid W \mid + \mid Y \mid}, \tag{3}$$

where W is the input CXR mask, Y is the generated mask, and $\bigcap$ is the intersection of the input and generated masks.

$$BCD = DC - \sum_{c=1}^{N} m_{b,c} \log (p_{b,c}), \tag{4}$$

where N is the classes' number (2 classes, Lung and Non-Lung, in our case), m is the binary indicator, and p is the predicted probability.

We first evaluate the proposed techniques with two loss functions and compare their performance to the baseline model, U-Net. Then, we show their performance in three challenging cases: healthy CXR images of varying lungs, CXR images of lungs affected by tuberculosis, and CXR images with lung nodules.

Quantitative Results

Table 1 reports the performance (accuracy and F1-score) of the proposed ARSeg, TransM, TransUNet, MedT, and UNeXt, using different loss functions (DC and BCD). We can see that learning using DC loss gives better results for ARSeg, TransUNet, and UNeXt. In addition, TransM and MedT show high performance with the BCD loss. We can also see that all the proposed models achieve better accuracy and F1 score than the baseline model U-Net. They also demonstrate their potential in detecting and segmenting lung regions overcoming challenging cases such as wide variation in lung shape, presence of rib cage and clavicles, and lungs affected with tuberculosis and nodules (benign and malignant).

Vision Transformers (TransM, MedT, and TransUNet) achieve higher performances compared to UNeXt, ARSeg, and U-Net thanks to their potential in extracting finer features from input CXR images and in determining long-range interactions within generated features. The proposed TransM reaches the best F1 score of 97.58% and an accuracy of 98.22% thanks to its ability in extracting rich feature maps using the global branch and local branch. MedT also obtains an excellent F1-score of 97.40% compared to TransUNet, UNeXt, ARSeg, and U-Net models. However, it requires a larger computational capacity and time when learning. Using a hybrid CNN- Transformer to extract global and local features, TransUNet obtained an F1-score of 96.80% and an accuracy of 98.36% outperforming UNeXt, ARSeg, and U-Net. However, it needs a pretrained backbone on a large dataset. UNeXt and ARSeg obtained F1 score values of 96.26% and 95.36%, respectively. They performed better than U-Net.

Qualitative Results

Similar to the numerical results shown in Table 2, we can see in Fig. 9 that the proposed models (TransUNet, TransM, MedT, UNeXt, and ARSeg) with DC loss correctly distinguish the normal lungs from the background including tissue and other organs. They determine the precise region of normal lungs very close to the input CXR mask annotated by a radiologist. They were able to separate the lung areas from the edge of the rib cage and clavicles and to segment varying lungs according to age and gender. We can also note

Table 2 Comparative analysis of TransUNet, TransM, MedT, UNeXt, and ARSeg using two loss functions

Model	F1 score (%)	Accuracy (%)	Loss Function
TransUNet	96.80	**98.36**	**DC loss**
TransUNet	96.24	98.34	BCD loss
TransM	97.47	98.14	DC loss
TransM	**97.58**	98.22	**BCD loss**
MedT	97.40	98.09	DC loss
MedT	97.35	98.04	BCD loss
UNeXt	96.26	98.10	DC loss
UNeXt	95.86	98.04	BCD loss
ARSeg	95.36	95.69	DC loss
ARSeg	95.32	95.33	BCD loss
U-Net [3]	93.00	93.75	DC loss

Best results are in bold

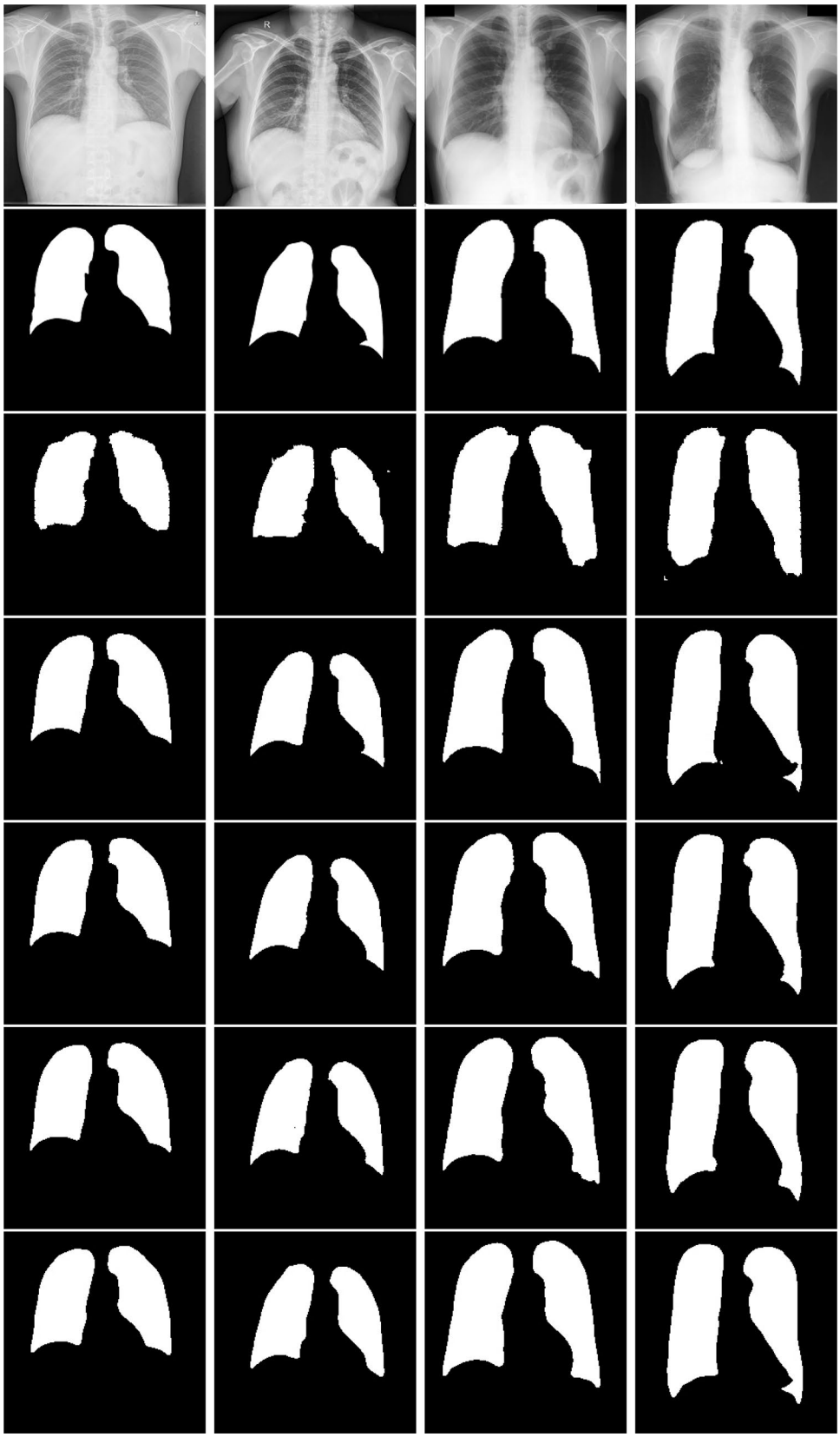

Fig. 9 Results of proposed models using normal CXR images. From top to bottom: CXR images, their corresponding mask, ARSeg results, TransUNet results, TransM results, MedT results, and UNeXt results

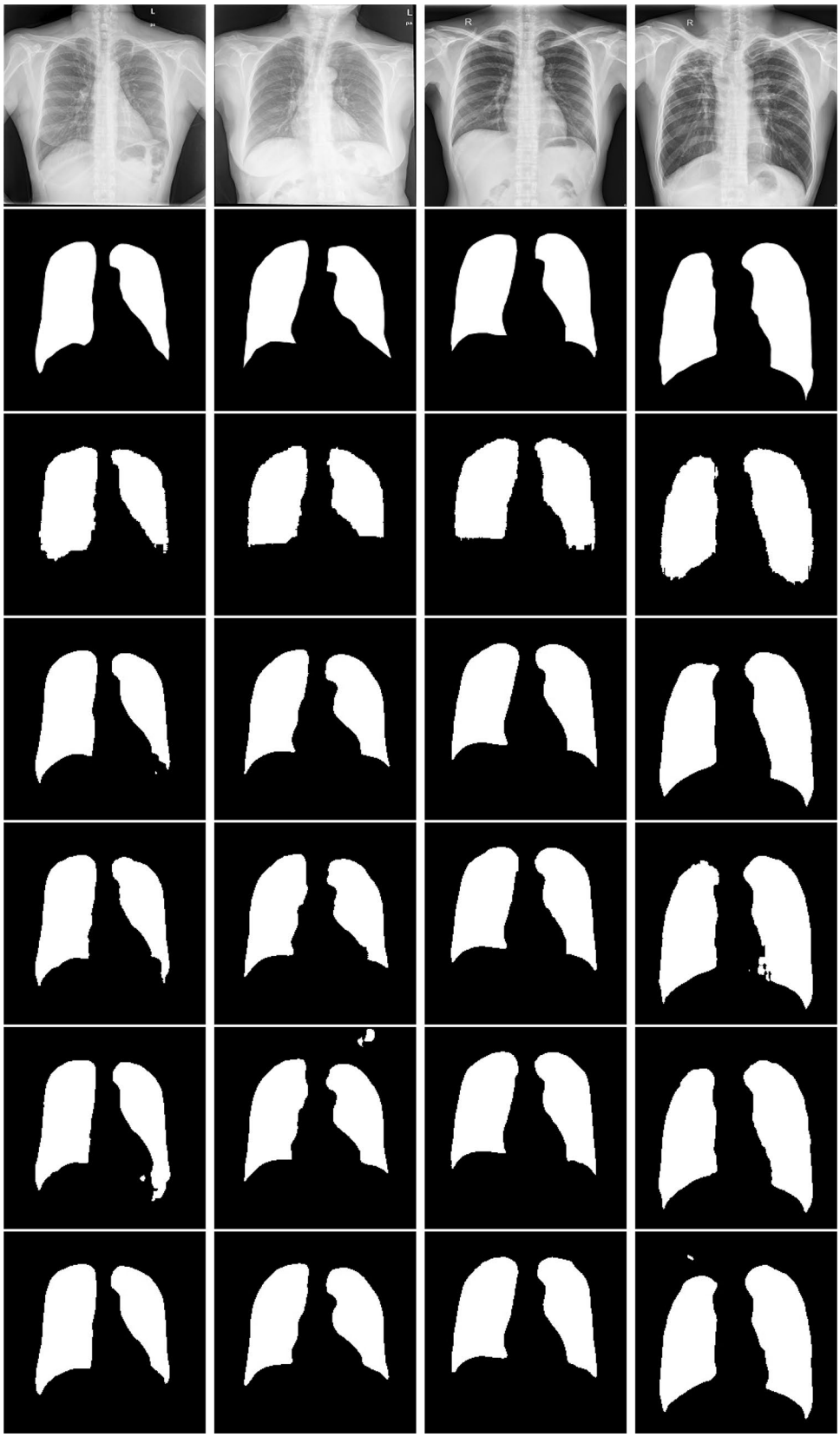

Fig. 10 Results of proposed models using CXR images with tuberculosis. From top to bottom: CXR images, their corresponding mask, ARSeg results, TransUNet results, TransM results, MedT results, and UNeXt results

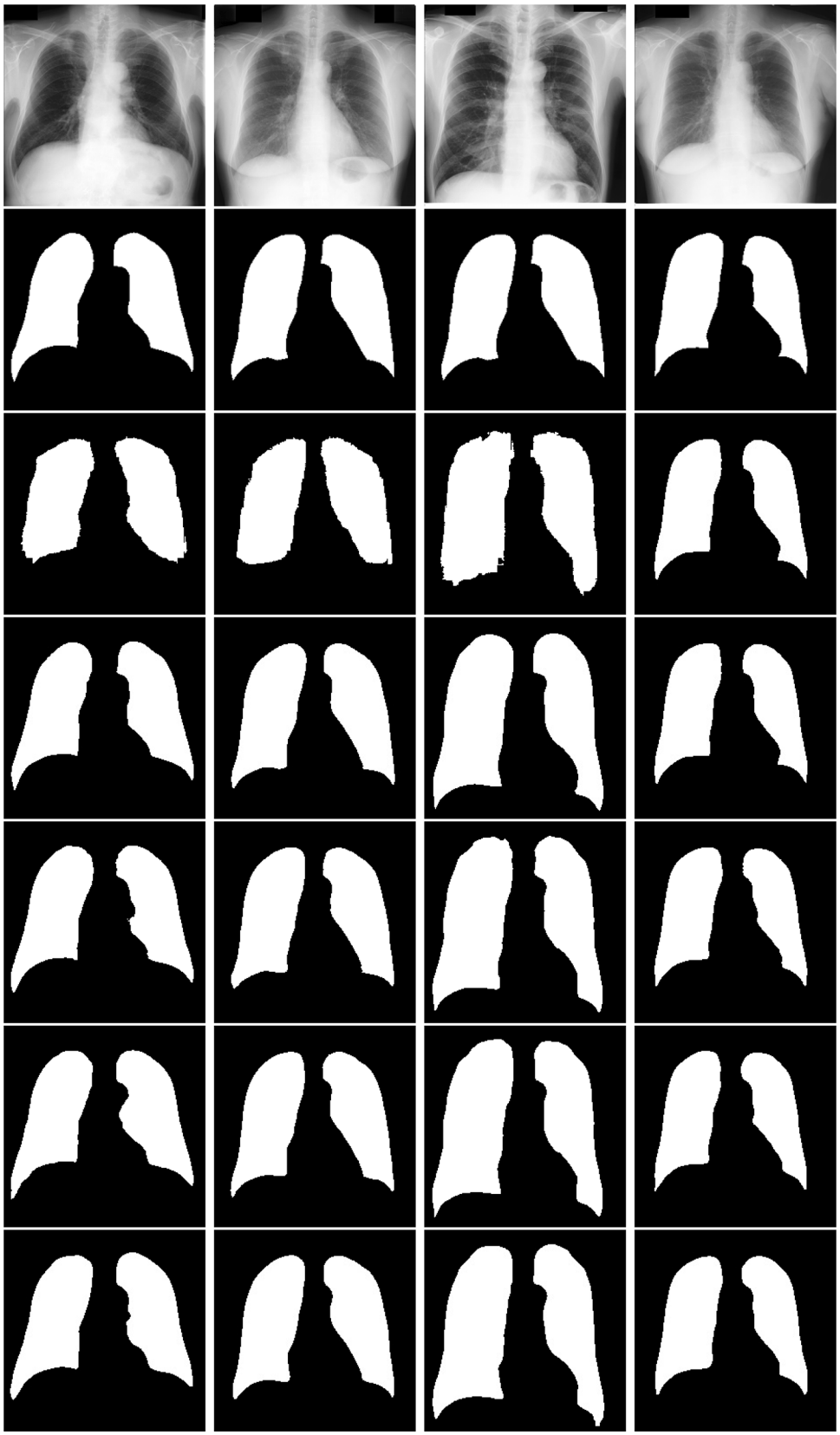

Fig. 11 Results of proposed models using CXR images with benign and malignant nodule. From top to bottom: CXR images, their corresponding mask, ARSeg results, TransUNet results, TransM results, MedT results, and UNeXt results

that TransM, TransUNet, MedT, and UNeXt produce smooth and clear boundaries of the lung zones compared to ARSeg.

Figure 10 shows some semantic segmentation results of TransUNet, MedT, TransM, ARSeg, and UNeXt using CXR images with tuberculosis disease. We can see that all proposed methods demonstrate a good performance in detecting and segmenting lung areas in a challenging case (CXR images of patients affected by tuberculosis). More precisely, TransM and TransUNet produce a lung mask much closer to the input mask of the lung affected by tuberculosis. ARSeg, MedT, and UNeXt also prove their reliability in segmenting lung areas in this challenging case. However, they are still incorrectly segmenting some small areas of the background as lung zones.

Figure 11 illustrates some lung masks generated by TransUNet, MedT, TransM, ARSeg, and UNeXt using unhealthy CXR images including severe lung anomalies (benign and malignant lung nodules) as input. We can note that the masks predicted by the proposed models are very similar to the manually labeled ground truths, thus, confirming their adequate binary segmentation of the lungs in a difficult case of CXR images with lung nodules.

Conclusion

In this paper, we proposed new models for lung segmentation in healthy and unhealthy CXR images. Five models, namely TransM, MedT, TransUNet, UNeXt, and ARSeg, were used in determining and segmenting precise lung zones. We evaluate the performance of the proposed models using two loss functions (Dice loss and Combo loss), two evaluation metrics (F1 score and accuracy), and public data, consisting of 951 CXR images collected from JSRT, MC, and Shenzhen datasets. Based on the F1 score, our proposed TransM achieved the best F1 score with 97.47%, compared to MedT, TransUNet, UNeXt, and ARSeg, which reached F1 score values of 97.40%, 96.80%, 96.26%, and 95.36%, respectively. These models also outperformed the popular U-Net model. They showed their great potential in differentiating lung zones from the rib cage and clavicles, and in segmenting varying lung shapes due to the age and gender of patients and lungs affected by diseases such as tuberculosis and nodules. In future work, we plan to use the developed vision Transformers to detect and segment the heart, lungs, and clavicles as a multi-class segmentation task.

Funding This research was enabled in part by support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2018-06233, the New Brunswick Health Research Foundation (NBHRF), and the New Brunswick Innovation Foundation (NBIF).

Availability of data and materials This work uses three publicly available datasets, JSRT, MC, and Shenzhen, see Refs. [9–11] for data availability. More details about the data are available under Sect. "Dataset".

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Kieu STH, Bade A, Hijazi MHA, Kolivand H. A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions. J Imaging. 2020;6(12):131. https://doi.org/10.3390/jimaging6120131.
2. Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest x-ray analysis: a survey. Med Image Anal. 2021;72: 102125. https://doi.org/10.1016/j.media.2021.102125.
3. Ghali R, Akhloufi MA. Arseg: an attention regseg architecture for cxr lung segmentation. In: IEEE 23rd international conference on information reuse and integration for data science (IRI). 2022. p. 291–6.
4. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y. Transunet: transformers make strong encoders for medical image segmentation. 2021. arXiv preprint arXiv:2102.04306
5. Valanarasu JMJ, Oza P, Hacihaliloglu I, Patel VM. Medical transformer: gated axial-attention for medical image segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2021. 2021. p. 12901.
6. Valanarasu JMJ, Patel VM. Unext: Mlp-based rapid medical image segmentation network. 2022, arXiv preprintarXiv:2203.04967 abs/
7. Sudre C, Li W, Vercauteren T, Ourselin S, Cardoso J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. 2017. p. 240–8.
8. Yi-de M, Qing L, Zhi-bai Q. Automated image segmentation using improved pcnn model based on cross-entropy. In: Proceedings of international symposium on intelligent multimedia, video and speech processing. 2004. p. 743–6.
9. Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K-I, Matsui M, Fujita H, Kodera Y, Doi K. Development of a digital image database for chest radiographs with and without a lung nodule. Amer J Roentgenol. 2000;174(1):71–4. https://doi.org/10.2214/ajr.174.1.1740071.
10. Jaeger S, Candemir S, Antani S, Wáng Y-XJ, Lu P-X, Thoma G. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. Quantit Imaging Med Surg. 2014;4(6):475–7. https://doi.org/10.3978/j.issn.2223-4292.2014.11.20.
11. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang Y-X, Lu P-X, McDonald CJ. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging. 2014;33(2):233–45. https://doi.org/10.1109/TMI.2013.2284099.
12. Hwang S, Park S. Accurate lung segmentation via network-wise training of convolutional networks. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. 2017. p. 92–9.
13. Islam J, Zhang Y. Towards robust lung segmentation in chest radiographs with deep learning. 2018. arXiv preprint arXiv:1811.12638

14. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention-MICCAI. 2015. p. 234–41.
15. Liu W, Luo J, Yang Y, Wang W, Deng J, Yu L. Automatic lung segmentation in chest x-ray images using improved u-net. Sci Reports. 2022;12(1):8649. https://doi.org/10.1038/s41598-022-12743-y.
16. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th international conference on machine learning. 2019. p. 6105–14.
17. Liu Y, Wang X, Wang L, Liu D. A modified leaky relu scheme (mlrs) for topology optimization with multiple materials. Appl Math Comput. 2019;352:188–204. https://doi.org/10.1016/j.amc.2019.01.038.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2016. p. 770–8.
19. Dai W, Dong N, Wang Z, Liang X, Zhang H, Xing EP. Scan: structure correcting adversarial network for organ segmentation in chest x-rays. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. 2018. p. 263–73.
20. Chen B, Zhang Z, Lin J, Chen Y, Lu G. Two-stream collaborative network for multi-label chest x-ray image classification with lung segmentation. Pattern Recogn Lett. 2020;135:221–7. https://doi.org/10.1016/j.patrec.2020.04.016.
21. 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. p. 3462–71.
22. Mittal A, Hooda R, Sofat S. Lf-segnet: a fully convolutional encoder-decoder network for segmenting lung fields from chest radiographs. Wirel Pers Commun. 2018;101(1):511–29. https://doi.org/10.1007/s11277-018-5702-9.
23. Frid-Adar M, Ben-Cohen A, Amer R, Greenspan H. Improving the segmentation of anatomical structures in chest radiographs using u-net with an imagenet pre-trained encoder. In: Image analysis for moving organ, breast, and thoracic images. 2018. p. 159–68.
24. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2015. p. 3431–40.
25. Yu F, Koltun V, Funkhouser T. Dilated residual networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017. p. 472–80.
26. Jegou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) workshops. 2017. p. 11–19.
27. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2018. arXiv preprint arXiv:1409.1556
28. Novikov AA, Lenis D, Major D, Hladůvka J, Wimmer M, Bühler K. Fully convolutional architectures for multiclass segmentation in chest radiographs. IEEE Trans Med Imaging. 2018;37(8):1865–76. https://doi.org/10.1109/TMI.2018.2806086.
29. van Ginneken B, Stegmann MB, Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med Image Anal. 2006;10(1):19–40. https://doi.org/10.1016/j.media.2005.02.002.
30. Gómez O, Mesejo P, Ibáñez O, Valsecchi A, Cordón O. Deep architectures for high-resolution multi-organ chest x-ray image segmentation. Neural Comput Appl. 2020;32(20):15949–63. https://doi.org/10.1007/s00521-019-04532-y.
31. Selvan R, Dam EB, Detlefsen NS, Rischel S, Sheng K, Nielsen M, Pai A. Lung segmentation from chest x-rays using variational data imputation. 2020. arXiv preprint arXiv:2005.10052
32. Hoshen Y. Non-adversarial mapping with vaes. In: Advances in neural information processing systems. 2018. p. 7528–37.
33. Tang Y-B, Tang Y-X, Xiao J, Summers RM. Xlsor: a robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation. In: Proceedings of the 2nd international conference on medical imaging with deep learning. 2019. p. 457–67.
34. Huang Z, Wang X, Huang L, Huang C, Wei Y, Liu W. Ccnet: criss-cross attention for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV). 2019. p. 603–12.
35. Pal D, Reddy PB, Roy S. Attention uw-net: a fully connected model for automatic segmentation and annotation of chest x-ray. Comput Biol Med. 2022;150: 106083. https://doi.org/10.1016/j.compbiomed.2022.106083.
36. Maity A, Nair TR, Mehta S, Prakasam P. Automatic lung parenchyma segmentation using a deep convolutional neural network from chest x-rays. Biomed Signal Process Control. 2022;73: 103398. https://doi.org/10.1016/j.bspc.2021.103398.
37. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: a nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. 2018. p. 3–11.
38. Sonali Sahu S, Singh AK, Ghrera SP, Elhoseny M. An approach for de-noising and contrast enhancement of retinal fundus image using clahe. Opt Laser Technol. 2019;110:87–98. https://doi.org/10.1016/j.optlastec.2018.06.061.
39. Widyantara IMO, Asana IMDP, Wirastuti NMAED, Adnyana IBP. Image enhancement using morphological contrast enhancement for video based image analysis. In: 2016 international conference on data and software engineering (ICoDSE). 2016. p. 1–6
40. Vidal PL, de Moura J, Novo J, Ortega M. Multi-stage transfer learning for lung segmentation using portable x-ray devices for patients with covid-19. Expert Syst Appl. 2021;173: 114677. https://doi.org/10.1016/j.eswa.2021.114677.
41. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. Covid-19 image data collection: prospective predictions are the future. 2020. arXiv preprint arXiv:2006.11988
42. Singh A, Lall B, Panigrahi BK, Agrawal A, Agrawal A, Thangakunam B, Christopher DJ. Deep lf-net: semantic lung segmentation from Indian chest radiographs including severely unhealthy images. Biomed Signal Process Control. 2021;68: 102666. https://doi.org/10.1016/j.bspc.2021.102666.
43. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer Vision–ECCV 2018. 2018. p. 833–51.
44. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2018. p. 4510–20.
45. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollar P. Designing network design spaces. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2020. p. 10428–36.
46. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. Pytorch: an imperative style, high-performance deep learning library. In: Advances in neural information processing systems 32: annual conference on neural information processing systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver. 2019. p. 8024–35.

47. 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. 2009. p. 248–55.