J. Vis. Commun. Image R. 85 (2022) 103521

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Full Length Article

Improving lung region segmentation accuracy in chest X-ray images using a two-model deep learning ensemble approach☆

Md Fashiar Rahman [a], Yan Zhuang [b], Tzu-Liang (Bill) Tseng [a,*], Michael Pokojovy [c], Peter McCaffrey [d], Eric Walser [d], Scott Moen [d], Alex Vo [d]

[a] Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at El Paso, TX 79968, USA
[b] Electrical and Computer Engineering, The University of Texas at El Paso, TX 79968, USA
[c] Department of Mathematical Science and Computational Science Program, The University of Texas at El Paso, TX 79968, USA
[d] University of Texas Medical Branch at Galveston, TX 77550, USA

ARTICLE INFO

Keywords:
Lung segmentation
Chest X-ray
Deep learning
UNet
CNNs
Ensemble

ABSTRACT

We propose a deep learning framework to improve segmentation accuracy of the lung region in Chest X-Ray (CXR) images. The proposed methodology implements a "divide and conquer" strategy where the original CXRs are subdivided into smaller image patches, segmented them individually, and then reassembled to achieve the complete segmentation. This approach ensembles two models, the first of which is a traditional Convolutional Neural Network (CNN) used to classify the image patches and subsequently merge them to obtain a pre-segmentation. The second model is a modified U-Net architecture to segment the patches and subsequently combines them to obtain another pre-segmented image. These two pre-segmented images are combined using a binary disjunction operation to get the initial segmentation, which is later post-processed to obtain the final segmentation. The post-processing steps consist of traditional image processing techniques such as erosion, dilation, connected component labeling, and region-filling algorithms. The robustness of the proposed methodology is demonstrated using two public (MC, JPCL) and one proprietary (The University of Texas Medical Branch - UTMB) datasets of CXR images. The proposed framework outperformed many state-of-the-arts competitions presented in the literature.

1. Introduction

Medical imaging is one of the most commonly used techniques in health-care for assisting with diagnosis and treatment of patients [1]. Medical imaging is unique in the diagnostic armamentarium because it is generally non-invasive, natively digital, and highly versatile in enabling detection, monitoring, and even prediction of disease. There are a variety of means that are used in creating images of the human body such as ultrasound, magnetic resonance, computer tomography, X-rays, etc. [2]. Advances in medical imaging have greatly improved the accuracy of screening for these diseases at earlier stages. However, 3D medical images such as computer tomography (CT) and magnetic resonance imaging (MRI) remain relatively expensive and their availability is limited in most parts of the world.

The use of X-rays, on the other hand, is considered the most popular and affordable medical imaging technique worldwide. X-ray imaging is crucial in a variety of medical settings from community clinics to tertiary referral center. Chest X-Rays are widely used to detect infections, such as pneumonia, as well as other relevant pathologies such as lung nodules or pulmonary edema. Due to the effectiveness of chest radiographs, millions of CXRs are generated annually as an initial diagnostic tool, which accounts for about one-third of all medical imaging procedures [3]. However, the use of CXRs for diagnosis is non-trivial and requires a high level of skill, experience, and concentration from the radiologist [4]. It can be a considerable or altogether infeasible diagnostic workload to examine millions of CXRs, especially considering that there is a shortage of radiologists worldwide [5]. According to the Association of American Medical Colleges, it is estimated that the United States could experience such a shortfall of radiologists by 2033 [6]. This necessitates the demand for automated medical diagnostics tools to aid medical professionals.

Given this fact, researchers continue to explore the use of different techniques and algorithms to develop automated and computer-aided

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: btseng@utep.edu (T.-L.(B. Tseng).

https://doi.org/10.1016/j.jvcir.2022.103521
Received 10 July 2021; Received in revised form 21 March 2022; Accepted 16 April 2022
Available online 21 April 2022
1047-3203/© 2022 Published by Elsevier Inc.

methods to assist radiologists in reading chest radiographs [7,8]. Such an attempt was pioneered by Lodwick et. al. [9] through the development of a computer-aided detection (CAD) system. Later, many efforts have been made to improve the accuracy of the CAD system and commercialize it for clinical applications, including CAD4 TB, Riverain, and Delft imaging systems [10,11]. However, due to the complex nature of CXRs, automatic and accurate detection of diseases remains unresolved in most of the existing CAD systems.

CAD systems are mainly divided into four steps: image processing, region of interest (ROI) extraction, ROI features detection, and diseases classification according to the features. It is worth mentioning that classical approaches have involved manual feature extraction within this workflow. Recent development of machine learning algorithms, accumulation of voluminous medical images, and computational power open up new opportunities for building modern CAD systems [12]. For example, in reading chest radiographs, the task of extracting ROIs is now largely replaced by lung segmentation and, once a relevant lung region is segmented, shape irregularities, size, and other abnormalities of the lungs can be analyzed to identify clinical conditions such as pleural effusion, emphysema, and pneumothorax [13]. Hence, lung segmentation is one of the crucial steps in CAD-based disease detection using CXRs and ongoing success in effective segmentation will increase the scale and utility of CAD.

The task of lung segmentation is challenging due to the presence of opacities or consolidation in CXRs, which are typically produced by overlaps between the lung region and intense abnormalities caused by pulmonary diseases, fluid, or bacterial infection. For healthy patients, CAD systems are typically able to provide reliable segmentation results as the contrast between the lung region and the outer area tends to be strong. However, in patients affected by pulmonary disease, there is often poor contrast between the affected lung area and the outer region of the film, making segmentation less accurate in a situation where effective segmentation would be especially important. For this reason, the task of lung segmentation becomes significantly more complex and challenging.

To address the above issues, we propose an automatic method to improve lung segmentation accuracy by utilizing an ensemble of two deep learning models. The proposed work consists of three major steps: (1) CXR preprocessing and patch extraction; (2) generation of two pre-segmentations by deploying patch classification and patch segmentation layers; and (3) merging the two pre-segmentations using the binary disjunction operation and post-processing them to obtain the final segmentation. To classify the patches, we use a deep convolution neural network (CNN), whereas an adapted U-Net architecture is used for patch segmentation. The novelty of this proposed approach lies in the patching technique enabling the framework to be trained with a limited image dataset and the ensemble approach of two deep learning models. Thus, the major contribution of this research includes developing a robust ensemble deep learning algorithm to improve the accuracy of lung segmentation in CXR images. In some cases, where the lung region has less or no bacterial infection, the outcome of this ensemble approach is analogous to other existing segmentation algorithms. However, the major advantage of this proposed approach comes for CXR images where the lung region is vague due to intensive bacterial infection such as pneumonia or COVID-19 positive patients. In such cases, many of the existing algorithms fail to segment some pixels inside or along the boundary of the lung region. Here, the two deep learning branches in the proposed ensemble approach serve as the complement fulfilling the downside of each other to improve the overall accuracy of the lung segmentation. The performance of the proposed methodology outperforms many alternative approaches described in the literature.

The rest of the paper is organized as follows. In Section 2, we present a brief literature review and establish context. In Section 3, we provide a detailed description of the dataset and proposed research framework. In Section 4, we document the experimental setup and the training procedure. In Section 5, we present the results, discuss the findings, and make some critical observations. This section also compares the performance of our proposed method with other related works. Finally, in Section 6 we conclude with some future perspectives.

2. Related works

Lung segmentation is a preliminary and fundamental task in computer-aided diagnostics systems. In light of this fact, the scientific community has shown great interest in this field and contributed to improving the lung segmentation in CXR. In this section, we reviewed some of the most relevant works in this research area. Based on current literature, lung segmentation techniques can be divided into two broad approaches i.e., 1) the classical approach using conventional image processing techniques and 2) convolutional neural networks (CNNs) based approaches.

2.1. Classical approaches

Saad et. al. [14] proposed a method for segmenting lung regions in CXR images using the canny edge filter and image morphology. Before deploying the edge filter, the method utilized the Euler number to improve the accuracy of lung edge detection. An implementation of this method produced segmentation results with an overall Dice score of 0.809 on the JSRT dataset. Another edge-based method was proposed by Xu et. al. in [15], where the global edge and region force (ERF) field-based active shape model (ASM) called ERF-ASM is used to segment the lung field in CXR. This method applied the principal component analysis (PCA) to learn the shape of the lung fields in advance, which is then applied to regularize the ERF-based segmentation and obtained the overall accuracy and sensitivity of 0.955 and 0.912, respectively on the JSRT dataset. Ahmad et. al. [16] presented another method based on an oriented Gaussian derivatives filter with seven orientations, combined with Fuzzy C-Means (FCM) clustering and thresholding to refine the lung region. The performance of this method greatly depends on the initial selection of the number of orientations and thresholds. This method reached a Jaccard index value of 0.870 and an accuracy of 0958 on the JSRT dataset. Due to the effectiveness of accurate thresholding and Gaussian derivatives, Kiran et. al.[17] presented a six-step lung segmentation methodology based on Sauvola thresholding and Gaussian derivative (ST-GD). This method achieved an accuracy of 0.9457 on the JSRT dataset and 0.9075 on the MC dataset. The performance of this method is sensitive to the initial selection of the training images as the shape of detected lung boundaries is refined using the SIFT-flow and graph cuts approaches. Consequently, this method has limitations in generalizing lung boundary detection for a wide variety of CXRs. The nonrigid registration-driven method was proposed by Candemir et. al. in [18], where the authors utilize appropriate retrieval-based patient-specific adaptive lung models to detect lung boundaries. This method reported a Dice value of 0.967 and 0.96 on the JSRT and MC respectively although the performance of this method is very sensitive to initial image selection. Another work that represented the Variation-based Active Contour (TVAC) lung segmentation technique was proposed by Reamaroon et. al. [19]. This method incorporated variation denoising, recursive binarization method, active contour model, and histogram equalization. This algorithm reported the mean value of 0.9501 for DC on the JSRT dataset.

2.2. CNN-based approaches

In recent years, CNN based approaches have become very successful in segmenting the lung regions in CXRs. Kholiavchenko et. al. [20] proposed and investigated a contour-aware segmentation framework based on the UNet and LinkNet architecture with a pre-trained ResNeXt (a modified architecture of ResNet) encoder for feature extraction. This framework is contour-aware, meaning that it requires contour information to explicitly indicate the pixels in the lung contour. The overall

segmentation performance produced a Jaccard index of 0.971 for the JSRT dataset. Yan et. al. presented a weekly supervised deep learning framework with squeeze-and-excitation (SE) blocks, multi-map transfer, and max–min pooling for classifying and localizing multi-class diseases lesion in noisy chest X-ray films. Their proposed framework reported comparable area under the curve (AUC) scores for many of the diseases, such as atelectasis, infiltration, Pneumonia, Fibrosis, and other classes [21].

The application of the encoder-decoder architecture is another successful approach in segmenting the lung region. Gaal et. al. [22] presented a novel deep learning approach for lung segmentation. This method used a fully convolutional neural network in conjunction with an adversarial critic model and obtained an average DC value of 0.975 on the JSRT dataset. Saidy and Lee [23] presented a lung segmentation method using encoder-decoder convolutional neural networks and achieved an average DC and sensitivity value of 0.9595 and 0.956, respectively, on the JSRT dataset. Novikov et. al. [24] proposed neural network architecture for automated multiclass segmentation of anatomical organs in CXRs. In this work, the authors mainly investigated different settings and hyper-parameters of various neural networks and obtained the best performing network, which reached the mean Jaccard overlap score of 0.95 on the JSRT dataset. Rashid et. al. [25] used an encoder-decoder based Fully Convolutional Network (FCN) to segment the lung region. The method deployed some post-processing manipulations such as flood-fill algorithm, unwanted object removal, and morphological operation followed by the output of the trained FCN model. The final segmentation obtained an accuracy of 0.971, 0.977, and 0.942 on the JSRT, MC, and a private dataset, respectively. All the proposed models are self-capable to learn the image features to distinguish between the lung and no-lung region.

Besides its more obvious role in data augmentation, generative adversarial networks (GANs) are also being utilized in classification and segmentation applications. Thus, a research framework has recently been developed by Munawar et. al. in [26]. Given an input CXR, a GAN network is trained to generate a mask and, later, a discriminator distinguishes between ground truth and the generated mask. The authors trained four different discriminators, D1, D2, D3, and D4, among which D2 obtained the highest DC value of 0.9780 on the MC dataset. In another work, Chen et. al. [27] proposed semantic-aware GANs for unsupervised domain adaptation called SeUDA. The distinct feature of this method is that it detaches the segmentation deep neural network (DNN) from the domain adaptation process and does not require any labels from the test set. The SeUDA framework conducts image-to-image transformation to generate a source-like image which is directly forwarded to the DNN. This framework achieved a Dice value of 0.9559 and 0.9342 for the right and left lung respectively on JSRT dataset.

Ensemble methodologies, where two or more techniques are combined together, are another successful approach to improve the segmentation accuracy. For example, Gomez et. al. [28] proposed four different convolutional models by introducing structural changes to the original U-Net and InvertedNet model. This model integrates several techniques barely used in segmentation algorithms, such as instance normalization and atrous convolution. In another work, Souza et. al. [29] combined two deep learning networks, where the first network is capable of classifying the CXR patches to generate initial segmentations, later the second network is deployed to reconstruct infected regions. The novelty of this approach greatly depends on the ability of the mask reconstruction phase. This method reported promising performance metrics on MC dataset. However, the reconstruction step could result in increasing the false positive rate due the incorrect mask reconstruction, especially in CXRs images with the presence of extreme abnormalities.

As a matter of fact, several methods exist in the literature that investigate the task of lung segmentation in CXR. Building upon those pursuits, in this paper, we proposed an automatic and hybrid method, where an ensemble of two parallel deep leaning networks is deployed to leverage the lung segmentation of CXR. The novelty of our approach is as follows. First, we crop the CXR patches and demonstrate how small patches can be utilized to increase the training samples and segment lung regions from whole CXRs. Moreover, the benefits of integrating two deep learning networks are analyzed thoroughly with different case studies. Second, ensembling the two networks add robustness which is demonstrated for different cases and levels of bacterial infection. Finally, we validated the obtained result against JSRT, MC, and a dataset derived from a large health system (UTMB) and compared our segmentation performance to 12 other approaches in the literature.

3. Materials and methodology

This section provides all relevant details and a step-by-step explanation of our lung segmentation methodology. The proposed framework consists of five major steps: image pre-processing and patch extraction, CNN-based classifier, an adapted U-Net-based patch segmentation, generation of an initial segmentation, and post-processing to obtain the final segmentation as shown in Fig. 1. The details of each of these steps are explained in the following subsections.

3.1. Description of dataset

The datasets used in this paper have been collected from the Tuberculosis Control Program of the Department of Health and Human Services of Montgomery County (MC), Maryland, USA (https://lhncbc.nlm.nih.gov/publication/pub9931) and Japanese Society of Radiological Technology which is commonly known as the JSRT database (https://db.jsrt.or.jp/eng.php). The MC database and the JSRT database are publicly available and commonly used in various areas of digital imaging including image processing, image enhancement, computer-aided diagnosis, and lung region segmentation. The MC dataset contains 138 posterior-anterior X-rays, of which 80 X-rays are normal and 58 are abnormal with a manifestation of tuberculosis. The X-rays are available in two different sizes either 4020×4892 or 4892×4020 pixels. The JSRT dataset contains 247 images, among which 154 X-rays have lung nodules (100 malignant cases, 54 benign cases) and 93 X-rays are normal without any lung nodules (non-nodules). The image size is 2048×2048. Both the MC and JSRT dataset come with corresponding ground truth masks which make them convenient for supervised learning. We also used another proprietary dataset, which is provided by the University of Texas Medical Branch (UTMB). The UTMB dataset consists of three classes of X-ray images: normal, pneumonia, and COVID-19 cases. In this report, we used 97 images from the UTMB dataset scattered across the three cases. As the UTMB dataset does not provide any ground truth masking, we generated the ground truth manually for all of the 97 images. Later, our physician collaborators from UTMB verified the manual ground truth masks and made necessary adjustments. A brief overview of the dataset is reported in Table 1.

3.2. Image pre-processing and patch extraction

As mentioned in the previous section, we combined X-rays from three different sources which come in different sizes and formats. First, we downscaled all images to 512×512 to reduce the computational cost of model training. While downscaling the images, the MC X-rays were first placed in the center of the 4892×4892 background images and later downscaled to 512×512. Notice that, since the MC dataset has X-ray images in two different sizes, direct resizing could distort image proportion. For JSRT and UTMB datasets, we directly resized the images since the original images have identical width and height. After resizing the images, the pixel values of the X-ray images are normalized, i.e., the pixel intensity is mapped into the range between 0 and 1, where 0 is black and 1 is white. The values between 0 and 1 represent the shades of gray. This is particularly important when combining datasets from separate institutions as the different datasets may have different ranges of pixel values. Moreover, image normalization or enhancement has a

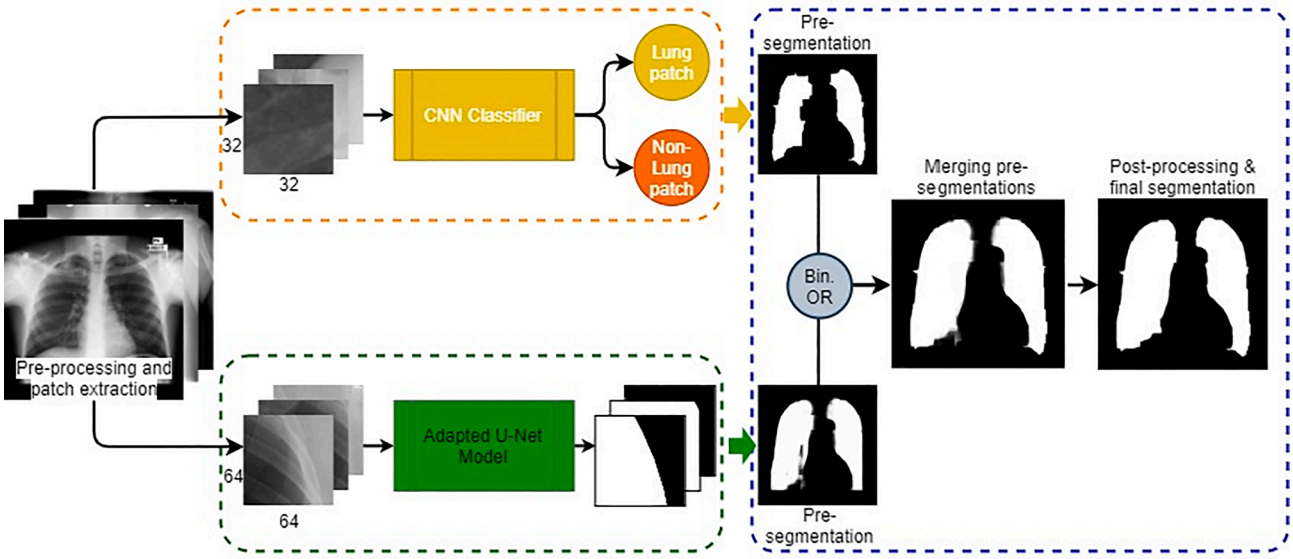


Fig. 1. The proposed ensemble framework for lung segmentation.

Table 1
Description of the datasets.

Dataset	Number of images	Case description
MC-Dataset	138	TB positive and Normal
JSRT-Dataset	247	Lung Nodule and Non-Lung Nodule
UTMB-Dataset	97	Normal, Pneumonia, and COVID-19
Total	**482**	

positive effect on the outcome of the deep learning model [30–33].

The three different datasets used in this paper have a total of 535 X-ray images which is not sufficient to adequately train a deep neural network. Hence, we choose to extract patches from the X-ray images to increase the number of training samples. The patches are small images, which are cropped from the original CXR images. We used the patch sizes of 32×32 and 64×64 for the CNN and U-Net models, respectively. The patches for the CNN layer are mutually exclusive, i.e., there is no overlap between two successive patches. Whereas the patches for the U-Net layer are cropped with the stride of 32 pixels, i.e., any two successive patches have an overlap of 32 pixels. The reason behind choosing these two different setups is to extract the same number of patches from each of the images while preserving the necessary details and properties of the lung region. Notice that patches of small size increase the number of training samples but decrease the levels of details. The patch size of 32×32 is suitable for classification purposes. However, for the U-Net model, the patch size should be large enough to be compatible with the down-sampling and up-sampling block in the U-Net architecture. Thus, we chose a patch size of 64×64 for the U-Net layer. Similarly, while cropping the patches from the X-ray images, we also cropped the patches from the corresponding masks. This is done to preserve the ground truth for each of the patches in the supervised training process.

3.3. CNN model

In this subsection, we describe the architecture and settings of the CNN model. The CNN model is employed to classify small image patches into the lung and no-lung categories and subsequently used to paint the segmentation mask. As mentioned earlier, the CNN model takes image patches of size 32×32 pixels as input. While cropping the patches from the original X-ray, we labeled the patches into two classes: lung and non-lung. To this end, the patches are sequentially cropped from the same position of both original X-ray and mask images. Then the cropped patches are compared to determine the proportion of lung and non-lung pixels. A patch is labeled as lung if it contains 20% or more pixels of the lung. Otherwise, it is labeled as a non-lung patch. The threshold for being a lung patch of 20% was determined empirically after several experiments. After cropping and labeling the patches, a CNN model is used to perform the classification of CXR patches. A typical CNN model consists of several convolution, pooling, and activation layers, each having several hyperparameters [34,35]. However, researchers usually experiment with the different network parameters and setups to bring innovation based on the applications and objectives [36–38]. The CNN model we used consisted of four convolution layers with max-pooling after the first, second, and fourth convolution layers as depicted in Fig. 2. It is then followed by two fully connected layers, having 4096 neurons each, and an output layer with the softmax activation for binary classification. The kernel size of 3×3 and 2×2 was set for all convolution and pooling layers, respectively. The number of features for the first, second, third, and fourth convolution layers were increased by 64, 128, 256, and 256, respectively.

The CNN model is trained to classify the patches for any test samples, which are later painted on a background image to obtain a segmentation of whole X-ray images. As mentioned in the previous subsection, the train patches were extracted in a mutually exclusive way, i.e., there was no overlap between two successive patches. However, to obtain a finer lung segmentation mask for test samples, the patches need to be extracted with overlap. In other words, the patches need to be extracted with a close stride such as 1, 2, or 4. The stride value is directly related to the smoothness of segmented lung contour and the computational cost. For example, to generate a segmentation mask of a whole X-ray, the CNN model needs to predict 57,600 patches for 512×512 images with a patch size of 32×32 and a stride of 2. In this work, we used the stride of 4 to achieve a trade-off between smoothness and the run time. A 512×512 background image is painted to obtain the segmentation mask. If an X-ray patch is classified as lung, then the central four pixels of the corresponding mask are painted white on the background images, i.e., we set the pixel values to 1 at the center of the corresponding mask. Thus, we obtain the segmentation of the whole X-ray from the CNN model. It is worth mentioning that the stride of 4 slightly deteriorates the smoothness but it cuts the computation time in half compared to the stride of 2. The smoothness of the contour can be restored by the U-Net layer as described in the later subsection.

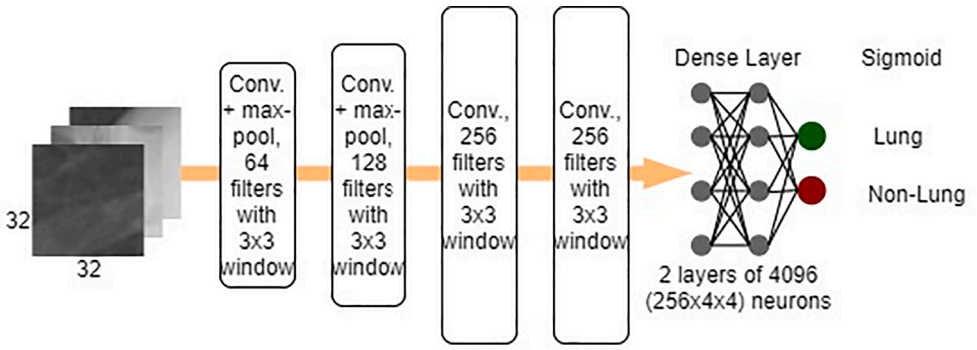


Fig. 2. The architecture of the CNN model.

3.4. U-Net model

This subsection describes the U-Net architecture used for segmenting small patches. The inputs of this model are small cropped patches with the output being their corresponding masks. The U-Net architecture is widely used in segmenting medical images [39–42]. The original U-Net architecture takes a whole image as input and outputs the predicted mask. Usually, input images vary in sizes such as 128×128, 256×256, or 512×512 pixels. These images are contracted and up-sampled through four to five down-sampling and upsampling convolution layers, respectively. However, in this work, we used small cropped patches, which are insufficient for contraction and expansion according to the original U-Net architecture. Hence, the original U-Net architecture is truncated to maintain necessary adjustment of the image size while down-sampling the patches. The modified network is comprised of two down-sampling blocks, one bridging block, and two upsampling blocks. This allows the input patches to contract at size 8×8 and later expand to the original size through the up-sampling blocks. The modified U-Net architecture is shown in Fig. 3.

The network is trained to segment individual patches, which are subsequently merged to generate another pre-segmentation. For any test image (new image), we crop the patches and feed them into the trained model to generate their corresponding masks. At this stage, the patches are cropped with the stride of 64, i.e., there is no overlap between two consecutive patches. Later, the segmented masks are plotted side-by-side on a 512×512 background image, i.e., the segmented masks are placed back at the same position of the respective cropped patches in the original X-ray image. Thus, a pre-segmented mask is generated for the whole image.

3.5. Final segmentation

The final segmentation is obtained by merging the two pre-segmented masks obtained from the CNN and U-Net models. Notice that the pre-segmented masks are binary images having the pixel value of 0 for the background and 1 for the lung region. Hence, we utilize the binary disjunction (logical OR) operation on the two pre-segmented masks to achieve the final segmentation as shown in Fig. 4.

The motivation for merging the pre-segmented masks is to improve segmentation accuracy. Generally, the CNN model or the U-Net presented in Sections 3.3 and 3.4 can segment the lung region with moderate accuracy and capture a major proportion of the lungs. However, the overall segmentation result deteriorates for the X-ray images which are affected by intense consolidations or abnormalities. In this case, merging the two segmentations can help to recover a considerable portion of the missing lung regions. In general, the CNN model shows better performance at capturing the inner region of the lung, whereas the U-Net model appears superior at capturing the lung shape and contour as highlighted in Fig. 4. Thus, these two models complement each other and improve the overall accuracy of segmentation.

At this stage, we have obtained a precise contour along with segmentation. However, it can still contain noise and inaccuracies, which are mainly due to the prediction error of an individual X-ray patch from either the CNN model or the U-Net model. For example, a background patch or a portion thereof could be erroneously segmented as lung region, as highlighted in Fig. 4 in orange. Alternatively, a lung region could be segmented as a background patch, as marked in Fig. 4 in red and yellow. These errors are sometimes inevitable as some of the background and lung region patches are difficult to distinguish when a

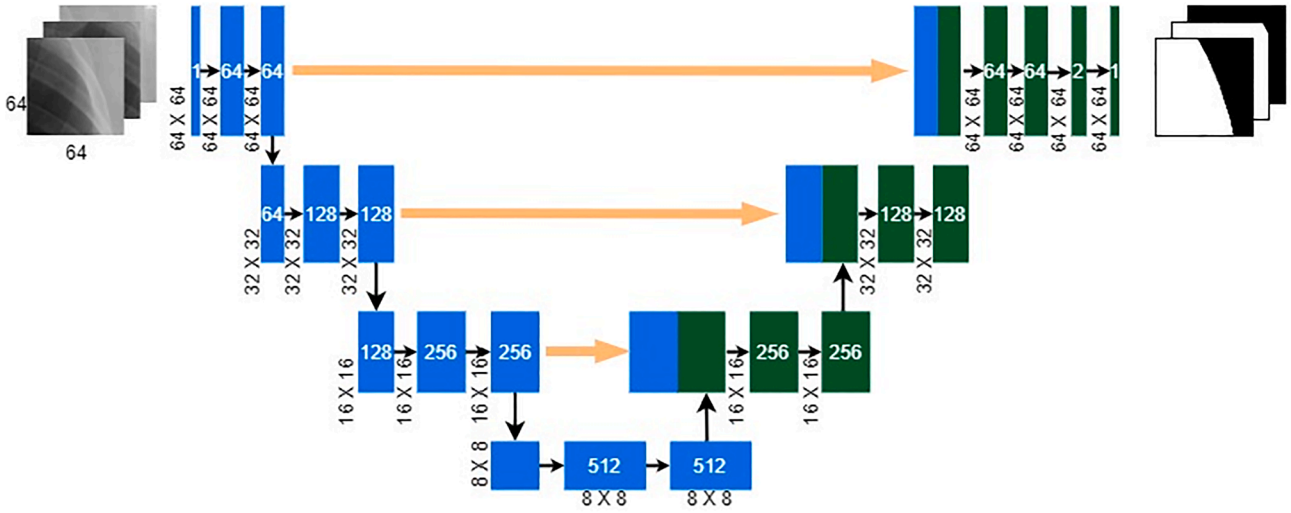


Fig. 3. The modified U-Net architecture used for patch segmentation.

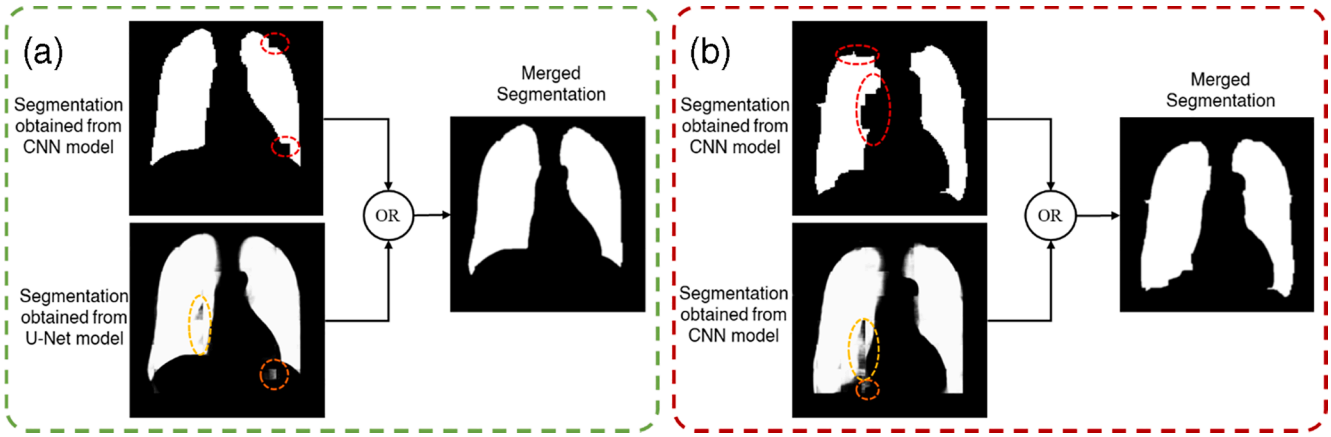


Fig. 4. The process of merging a case (a) with the ideal condition (b) with abnormalities.

whole X-ray image is divided into small patches. This problem is addressed by utilizing several image morphological operations such as erosion, dilation, closing, and connected component labeling algorithm in the post-processing step. The reader is referred to ref. [43–45] for a more detailed description of morphological operations. The erosion operation is used to erode tiny noises and separate noises weakly connected with the lung contour. However, deploying the erosion operation shrinks the lung region, which is recovered by the dilation operation. The connected component algorithm is used to identify the size of the remaining scattered noises after deploying the erosion operation. Subsequently, noises of size less than 3,000 pixels are. The necessity of this step is quite obvious as a good segmentation should only contain the two large lung parts. The region filling algorithm is utilized to fill in any internal gap laying inside the lung region. The structuring elements used in the above morphological operation are set as a square of size 11×11, an ellipse of size 9×9, and an ellipse of size 23×23 for the erosion, dilation, and closing operation, respectively. We choose the kernel size based on several experiments with the segmented masks.

4. Experimental setup and training

We performed experiments to classify and segment X-ray patches, which are subsequently merged to obtain segmentation for whole X-ray images. To this end, we first extract patches from the whole images as described in Section 3.2. The patches are extracted in two different settings. For the CNN model, the patches are extracted only from the original X-ray images and labeled as lung and non-lung in alignment with supervised classification task. Whereas, for the U-Net model, the patches are extracted both from each original X-ray and its corresponding mask (ground truth) images. This is done simultaneously to maintain the right mask for corresponding X-ray patches. In this work, we used a total of 482 X-ray images, which are split as follows: 80% for training, 10% for validation, and 10% for test purposes. Thus, we have only 385, 49, and 48 images for training, validation, and testing, respectively. After the patch extraction, we have sufficient number of samples to train the CNN and U-Net models. The number of patches in each of the settings is reported in Table 2.

Table 2
Dataset split and number of training samples.

Settings	Category	Train (80%)	Validation (10%)	Test (10%)	Total
	Whole images	385	49	48	482
CNN-Layer	Lung patch	32,019	4,129	4,115	40,263
	Non-Lung patch	66,541	8,415	8,173	83,129
U-Net Layer	X-ray patch	98,560	12,544	12,288	123,392
	Mask patch	98,560	12,544	12,288	123,392

The CNN and the U-Net models were trained separately using the Tensorflow and Keras API. Each network was trained for 55 epochs. For the CNN model, we use the categorical loss function and the Adam optimization algorithm with a learning rate of 0.0001. At the same time, the binary cross-entropy loss function and Adam optimizer were used to train the U-Net network. To avoid any possible overfitting issues, the data were augmented using the Keras data generation function. We apply 5% shearing and zooming along with horizontal flipping to the input batch for both networks. The model was trained on a single NVIDIA GTX 1080 Ti Graphics Processing Unit (GPU) with 3.6 GHz Intel(R) Core (TM) i7-7700 CPU and16GB RAM. It took 5 hr 35 min and 11 hr 23 min to complete the entire training for the CNN model and the U-Net model, respectively. After training over 50 epochs, we achieved the training accuracy of 97.8% and validation accuracy of 96.7% for CNN model. For U-Net model, the training accuracy was 97.3% for training data and 96.2% for validation data. Some of the outputs from the CNN model and the U-Net model are shown in Fig. 5.

Clearly, both the trained models can classify and segment the patches with superior accuracy. However, we observe some errors in prediction in certain cases, both for classification and segmentation. Sometimes the trained CNN model erroneously classifies a lung patch as non-lung or vice versa, as circled in Fig. 5(a). Notice that this miss-classified patch is very similar to non-lung patches. In such cases, the trained model sometimes cannot distinguish between lung and non-lung features and makes wrong prediction. Similarly, a patch or some portion thereof can be incorrectly segmented. For example, as shown in Fig. 5(b), some pixels of the patch are segmented as lung pixels, albeit they belong to the background. The classified and segmented patches are merged separately to obtain two pre-segmentation masks for the whole X-ray image. We refer to Section 3.5 for the details on the merging strategy.

5. Result and discussion

This section presents and discusses the results obtained with the proposed lung segmentation method. First, we demonstrate the outcome of pre-segmentation and final segmentation. Later, we evaluate and compare different segmentation outcomes to show superiority of the proposed method. In addition, at the end of this section, the robustness of the proposed methodology is thoroughly discussed using selected case studies from the three different datasets.

5.1. Whole image segmentation

As mentioned earlier, for any new image (e.g., a test image), the final segmentation is obtained in three steps. Initially, the patches are crop-

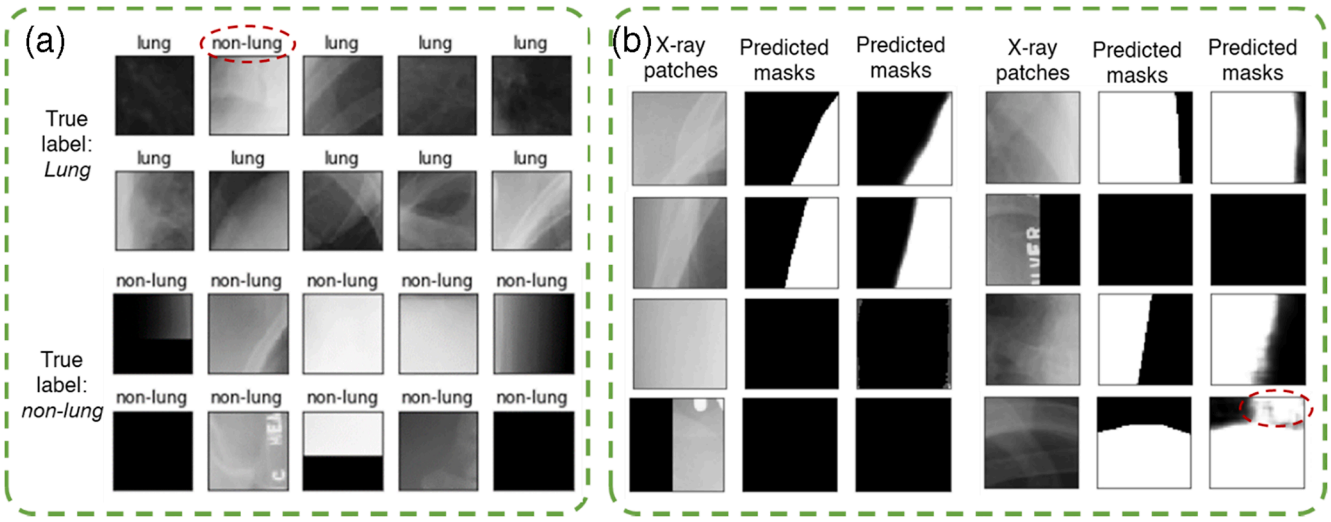


Fig. 5. Prediction results from the trained models (a) patch classification using the CNN model (b) patch segmentation using the U-Net model.

ped and passed through the trained CNN classifier model and U-Net segmentation model. Subsequently, individual classified and segmented patches are plotted on a 512×512 background image to obtain two separated pre-segmented masks. Finally, pre-segmented masks are combined using binary OR operation, which is subsequently post-processed to obtain the final segmentation. Some segmentation results from each of the steps are demonstrated in Fig. 6.

From Fig. 6 it is obvious that the integration of two separate trained models can generate excellent segmentation masks. In fact, any individual trained model i.e., either CNN model or U-Net model, can provide reasonable segmentation with a moderate error in "good" scenarios, where the lung region is visually apparent. However, under poor conditions, where the shape of lung region is vague due to consolidation, the segmentation result deteriorates. This mainly happens due to the presence of unavoidable prediction errors, such as missing regions along the lung contour, some little gaps in the lung region, scattered white pixels, or even some extraneous non-lung pixels with the lung boundary as shown in Fig. 7. The first two errors are mitigated following the merging of two pre-segmented masks as described in Section 3.5. The other errors are resolved by deploying image processing techniques in the post-processing step.



Fig. 6. Segmentation results at each step of the proposed method (a) original CXR images (b) pre-segmentations obtained from the CNN model (c) pre-segmentations obtained from the U-Net model (d) Combined segmentations before post-processing (e) final segmentations (f) ground truth masks.

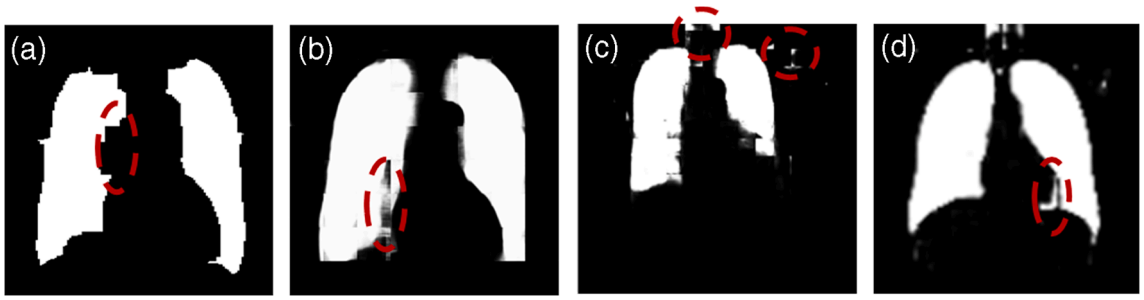


Fig. 7. Examples of prediction error (a) missing boundary (b) small gap in the lung region (c) scattered white pixels (d) some extraneous non-lung pixels.

5.2. Performance evaluation

We used Sensitivity, Specificity, Accuracy, Dice Coefficient (DC), and Intersection-Over-Union (IoU) to quantitatively evaluate segmentation results at each step of the proposed methodology. These metrics are widely used to evaluate the performance of semantic segmentation in medical imaging [46,47]. The sensitivity measures the proportion of lung pixels that are correctly identified, whereas the specificity measures the proportion of correctly identified non-lung pixels. The IoU metric, also commonly known as Jaccard index, is a method to quantify the overlap percentage between the target mask and the predicted mask. Similarly, the DC is two times of the number of overlapping pixels divided by the total number of pixels in both the target mask and the predicted mask. Using the notation True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), the evaluation metrics are calculated as follows:

$$Sen = \frac{TP}{TP + FN}, \tag{1}$$

$$Spec = \frac{TN}{TN + FP}, \tag{2}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{4}$$

$$DC = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{5}$$

where, TP indicates the number of correctly identified lung pixels, FP indicates the number of incorrectly identified lung pixels, TN represents the number of correctly identified non-lung pixels, and FN represents the number of incorrectly identified non-lung pixels. The values of all of the metrics range from 0 to 1 (0–100%) with 0 representing no overlap and 1 indicating perfect overlap between the target mask and predicted mask. We calculated sensitivity, specificity, accuracy, IoU, and DC values for all of the text X-rays at each stage of the proposed method. Later, we averaged them to obtain the overall performance, as reported in Table 3.

Table 3 clearly demonstrates the effectiveness of the proposed methodology. Integration of the CNN and U-Net models significantly improves the final segmentation result. One can observe that the average sensitivity of the merged segmentation attains the value of 96.12%, which is approximately 14% and 3% improvement compared to the individual CNN and U-Net models, respectively. At the same time, we also observe outstanding amelioration of the average DC and IoU values after merging the two pre-segmentations. The post-processing step further enhances the segmentation result by removing some unexpected noises and errors. Notice that after post-processing, the final segmentation result is improved by approximately 1% in terms of sensitivity, DC, and IoU values, whereas the accuracy is increased by nearly 2% on average. The proposed segmentation method can improve segmentation results in most cases. However, in some "ideal" cases, initial pre-segmentation (either generated by CNN or U-Net model) alone may provide extremely precise contours, as illustrated in Fig. 10. In such cases, the performance of the proposed methodology remains stable compared to individual pre-segmentations.

Table 3
Segmentation evaluation of each individual model on the test data.

Layer-wise Segmentation	Avg. Sen.	Avg. Spec.	Avg. Acc.	Avg. DC	Avg. IoU
CNN Layer (Pre-Segmentation 1)	0.8215	0.9811	0.9391	0.8877	0.8235
U-Net Layer (Pre-Segmentation 2)	0.9368	0.9698	0.9559	0.9301	0.8986
Merged Seg. (Before Post-Processing)	0.9612	0.9688	0.9609	0.9427	0.9158
Final Segmentation	0.9667	0.9683	0.9688	0.9532	0.9315

5.3. Case studies

To make the contribution of the proposed method clear, this subsection presents a set of case studies and analyzes some most common application scenarios of our lung segmentation method. Based on the experiments in this work, we observed the following three common situations as (1) the output of the U-Net model is sufficient for segmenting the lung region, however, incorporating the CNN outcome slightly improves the overall segmentation quality (most common cases); (2) the output of the CNN model helps fill in missing pixels in the segmentation result generated by U-Net model and improve the segmentation result (frequently observed cases); and (3) either the CNN or U-Net model can precisely segment the lung region, but there is no downside to incorporating both outcomes (ideal cases).

To illustrate the aforementioned three situations, we present three case studies from three different datasets used in this work. Case Study 1 uses an X-ray image from the MC dataset (MCUCXR_0108_1). This CXR presents the exam of a patient with "Tuberculosis positive". In this case, the U-Net model alone is capable of determining a moderate lung contour, whereas the CNN model does not make any significant contribution. However, after merging the two pre-segmentations and deploying the post-processing step, the overall sensitivity is slightly improved ($\approx$ 1%) compared to the individual segmentation with the U-Net model and becomes 97.78%. Other evaluation metrics remain almost unchanged, for example, the segmentation of U-Net model obtained a sensitivity of 96.67%, an accuracy of 98.77%, DC of 97.32%, and an IoU of 94.79%, while the final segmentation produced 97.78% of sensitivity, 98.57% of accuracy, 96.92% of DC, and 94.03% of IoU. The segmented lung contours at each stage are depicted in Fig. 8, where red, yellow, blue, and green indicate the contours of Pre-segmentation 1, Pre-segmentation 2, final segmentation, and ground truth.

In Case Study 2, we consider a sample from the JSRT dataset (JPCLN120). This example represents a case of a patient having a lung nodule in CXR. In this case, it is observed that the segmentation obtained

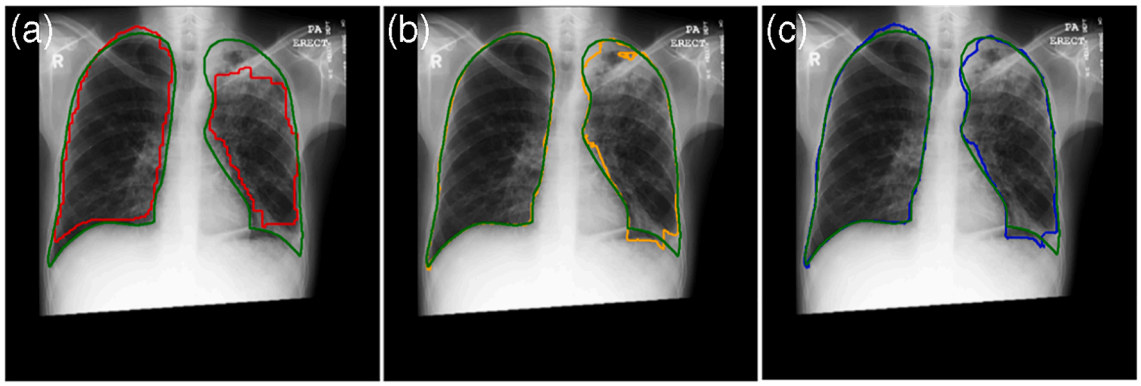


Fig. 8. Case study 1 – MCUCXR_0108_1 (a) pre-segmentation 1 in red vs ground truth in green, (b) pre-segmentation 2 in yellow vs ground truth in green, and (c) final segmentation in blue vs ground truth in green.

from U-Net model misses some inner pixels and lower tail in the left lung, as depicted in Fig. 9(b). On the other hand, the CNN output can capture these missing regions but misses some pixels along the lung contours. Thus, a combination of these two segmentations yields a significant improvement in the final segmentation. We observe more than 10% and 3% improvement of sensitivity in the final segmentation compared to pre-segmentation 1 and pre-segmentation 2, respectively. Similarly, the accuracy, DC, and IoU in the final segmentation amount to 97.72%, 97%, and 94.17%, compared to 86.65%, 94.5%, 92.34%, and 85.77% for CNN and 93.21%, 97.19%, 96.21%, and 92.7% for U-Net, respectively. This is an obvious "success case", where both the CNN and U-Net model contribute to improving the segmentation result. Such contribution becomes apparent when the internal lung region is heavily affected by abnormalities. The segmented result at each stage is shown in Fig. 9.

Case Study 3 is an example of an ideal case, where the lung region is visually apparent and no abnormalities are present in the CXR. This sample is taken from the proprietary UTMB dataset (UTMB_0047_0). In this case, the segmentation obtained from either the CNN model or the U-Net model is sufficient to precisely extract the lung contour. However, integration of the two pre-segmentations does not adversely affect contour extraction. The segmentation from the U-Net model obtained a sensitivity of 99.66%, an accuracy of 98.04%, a DC of 96.63%, and an IoU of 93.49%, while the final segmentation produced 99.82% of sensitivity, 97.17% of accuracy, 96.21% of DC, and 93.11% of IoU. In this particular case, the final segmentation result is slightly inferior compared to the U-Net result but an improvement is observed over the CNN outcome, though it is not necessarily true for other cases. In general, this is still considered a successful case and maintained a great degree of consistency extracting the lung boundary from the CXR. The stepwise segmented result is shown in Fig. 10.

5.4. Robustness analysis

A good segmentation method should perform well at segmenting the lung region from a wide variety of cases. This section reviews segmentation performance for a variety of test cases to illustrate the robustness of the proposed methodology. The trained model is deployed on three different datasets, where the CXRs have various diseases abnormalities caused by tuberculosis, lung nodule, pneumonia, COVID-19, etc. The segmentation results are shown in Fig. 11. For all of the cases, the segmentation results are quantitively evaluated using the aforementioned metrics of sensitivity, specificity, accuracy, Dice coefficient, and IoU. The CXRs in the JSRT dataset originates from a medical exam to identify the presence of lung nodules and non-nodules. These CXRs are not affected by the bacterial infection and hence, the lung boundary is comparatively obvious in the images. As expected, our proposed method can successfully segment the lung region in the UTMB CXRs both for lung nodule and non-nodule cases. From the first row in Fig. 11, we observe high values for all evaluation metrics, especially the sensitivity and DC value of more than 97% for all cases. The second row in Fig. 11 represents the segmentation result for the MC dataset. The CXRs in the MC dataset document the cases of tuberculosis (TB), positive and negative, which is a bacterial disease. The last digit of the file name indicates the disease label, i.e., 1 for "TB positive" and 0 for "TB negative". Due to the bacterial tuberculosis infection, the lung boundary in the CXRs sometimes remains vague. However, our proposed method can effectively extract the lung boundary in different situations. In a few cases, where the lung is seriously affected by the bacterial infection, the method fails to segment parts of the lung region. For example, the

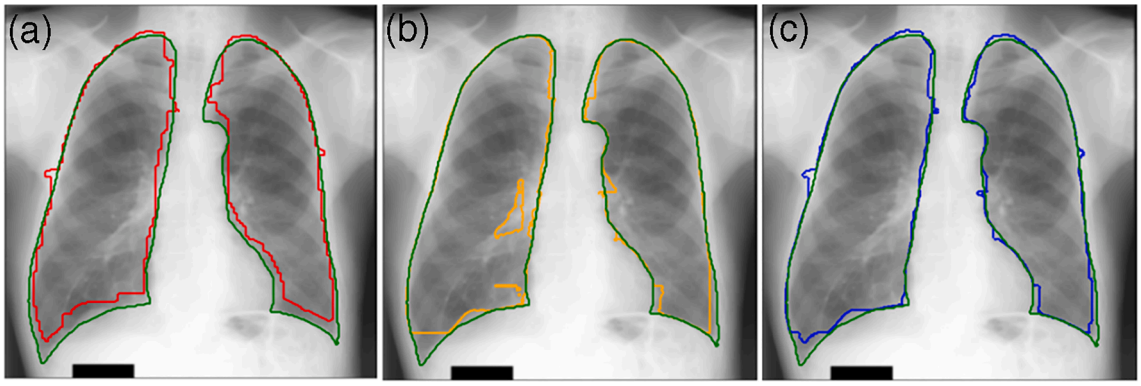


Fig. 9. Case study 2 – JPCLN120 (a) pre-segmentation 1 in red vs ground truth in green, (b) pre-segmentation 2 in yellow vs ground truth in green, and (c) final segmentation in blue vs ground truth in green.

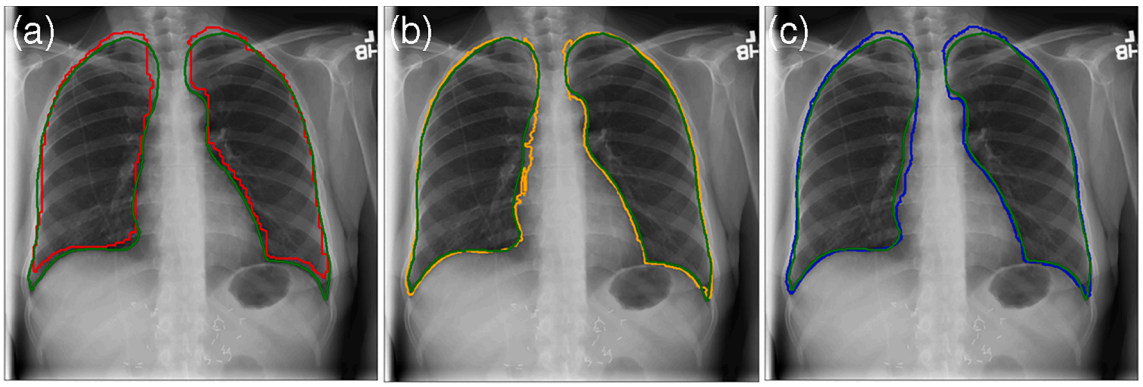


Fig. 10. Case study 3 – UTMB_0047_0 (a) pre-segmentation 1 in red vs ground truth in green, (b) pre-segmentation 2 in yellow vs ground truth in green, and (c) final segmentation in blue vs ground truth in green.

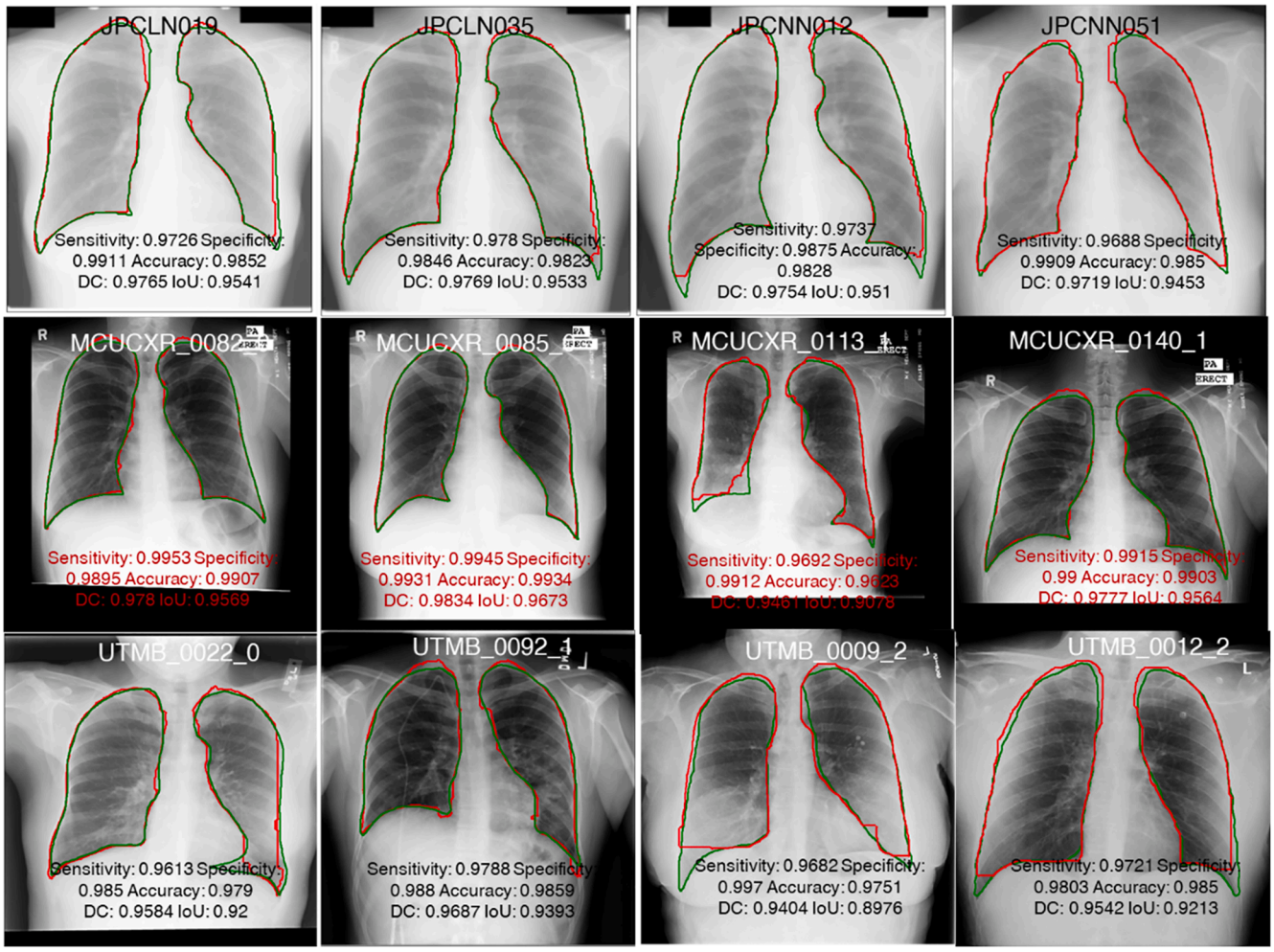


Fig. 11. Performance analysis on three different datasets: first row - JSRT CXRs with Lung-nodule and non-nodule; second row – MC CXRs with TB positive and TB negative cases; and third row – UTMB CXRs with normal, pneumonia, and COVID-19 positive cases.

performance on the MCUCXR_0113_1 CXR (a TB positive case with a serious presence of abnormalizes) is comparatively low, with a sensitivity of 96.92%, an accuracy of 96.12%, a DC of 94.61% of DC, and an IoU of 90.78%. On the other hand, in spite TB positivity, we achieved higher values for all evaluation metrics on MCUCXR_0140_**1** CXR. Finally, the third row in Fig. 11, demonstrates segmentation of some UTMB images. This dataset has three types of CXRs, which are labeled using the last digit of the file name (0 for normal, 1 for pneumonia, and 2 for COVID-19 positive). We observed that the proposed method also performs well on UTMB CXRs, especially for normal cases. For pneumonia and COVID-19 positive cases, the performance may slightly deteriorate sometimes. For example, UTMB_0009_**2** CXR resulted in a sensitivity of 96.82%, an accuracy of 97.51%, a DC of 94.04%, and an IoU of 89.76%. Notice that this is a COVID-19 positive case, where the lower part of the lung region is heavily affected and obscure. In contrast, we achieved high segmentation accuracy for UTMB_0092_**1** and

UTMB_0012_**2** CXR, though they are pneumonia and COVID-19 positive cases, respectively. In general, the proposed method shows superior performance on different datasets and various cases and types of diseases. In some difficult cases, the overall performance drops, but the method can still segment a major part of the lung region. Sometimes, segmentation can miss the eccentric regions of the lung contour such as the costodiaphragmatic recess in the lung region, which is a limitation of this proposed method. However, the overall performance of the segmentation maintains strong consistency with expert judgment.

5.5. Comparison with related works

In this subsection, we provide a comparative analysis of our proposed method and other related works mentioned in Section 2. A comparison of different evaluation metrics is reported in Table 4. Despite a good accuracy of segmentation results, the "classical" approaches mentioned in Section 2.1 are semi-automatic and require manual specification of features for their successful implementation. Some of the methods are sensitive to the selection of appropriate tuning parameters and thresholds. Moreover, these approaches employ sophisticated image processing techniques to eliminate clavicle, artifacts, and refine the lung contour. In general, these approaches are limited in their capacity to generalize the task of lung segmentation and the performance is very sensitive variation in images. In contrast, our proposed method is completely automatic and does not require manual feature specification. In addition, the sensitivity, accuracy, DC, and Jaccard index are approximately 97%, 97%, 96%, and 94%, respectively, on both JSRT and MC datasets, which outperforms the competitions.

In general, CNN based approaches have a better capability to generalize the task of lung segmentation. Despite the advantages of deep learning approaches, there are some disadvantages and limitations as well. The most common limitations are that this method requires a voluminous amount of images. Hence, in this work, we used a patching technique and investigated its performance in lung segmentation. In comparison to the encoder-decoder architecture, an advantage of our model is that we used a patch classification model in parallel to an adapted U-Net model to capture possible anomalies through a concurrent effort of patch classification and segmentation. Thus, our method is more robust for the task of CXR segmentation of patients with bacterial infections such as tuberculosis, pneumonia, or COVID-19. In some cases, generative adversarial networks report good segmentation accuracy, however, the GANs need special arrangement of inputs and levels. Though GANs are very powerful tools, sometimes they exhibit excessive sensitivity to small datasets. In contrast, our method is simple, straightforward, and easy to implement. Moreover, we achieved a DC value of 0.9609 on MC dataset. Compared to the hybrid approach presented by Souza et. al. in [29], our approach minimizes the risk of inflating the false positive rate by replacing the reconstruction step with an adapted U-Net model. While Souza et. al. [29] reported average sensitivity, accuracy, DC, and Jaccard values of 0.9754, 0.9697, 0.9356, and 0.8807, respectively, our method achieved those of 0.9711, 0.9724, 0.9609, and 0.9429, respectively, on the MC dataset.

Table 4 summarizes the performance of different methodologies and algorithms. Obviously, over the years, segmentation methods have become very robust exceeding 90% in all evaluation metrics. However, it is worth stressing again that our method is fully automatic and utilizes two convolutional neural networks, capable to capture missing pixels caused by abnormalities resulting from bacterial infection such as tuberculosis, pneumonia, and COVID-19. In addition, our patch-based approach is suitable to utilize small datasets in training the CNNs. We also report the average execution time and compare it with other methods. The execution times are recorded based on the 512×512 CXR images and the type of processing unit used to deploy the algorithms i.e., the central processing unit (CPU) or graphical processing units (GPU). In general, the algorithms deployed using a CPU take a longer execution time than GPU. The advancement of GPU allows us to use more

Table 4
Comparison of segmentation performance of the proposed method and related works.

Work	Dataset	Infection	Approach	Sen.	Acc.	Dice	Jaccard	Time cost (s)
Saad et. al. [14]	JSRT	No	Classical Semi-Automatic			0.809		
Ahmad et. al. [16]	JSRT	No	Classical Semi-Automatic		0.958		0.870	10–15 (CPU)
Kiran et. al. [17]	JSRT	No	Classical		0.9457			
	MC	Yes	Semi-automatic		0.9075			
Xu et. al. [15]	JSRT	No	Classical Semi-automatic	0.912	0.955			0.38,0.35* (CPU)
Candemir et. al. [18]	JSRT	No	Classical			0.967	0.954	85–90 (CPU)
	MC	Yes	Semi-automatic			0.96	0.941	
Reamaroon et. al. [19]	JSRT	No	Classical			0.9501		
	MC	Yes	Semi-Automatic			0.9569		
Kholiavchenko et. al. [20]	JSRT	No	CNN, Automatic				0.971	
Gaal et. al. [22]	JSRT	No	CNN, automatic			0.975		
Munawar et. al. [26]	MC	Yes	CNN, automatic			0.9780		0.05 (GPU)
Chen et. al. [27]	JSRT	No	CNN, automatic			0.9450		
Novikov et. al. [24]	JSRT	No	CNN, automatic				0.95	4.1, 0.03† (CPU/GPU)
Souza et. al. [29]	MC	Yes	CNN, automatic	0.9754	0.9697	0.9356	0.8807	
Saidy & Lee [23]	JSRT	No	CNN, automatic	0.956		0.9595		
Rashid et. al. [25]	JSRT	No	CNN,		0.971	0.951		
	MC	Yes	automatic		0.977	0.954		
	Private	Yes			0.942	0.880		
Our ensemble method	**JSRT**	**No**	**CNN, automatic**	**0.9791**	**0.9738**	**0.9648**	**0.9501**	**0.043 + 0.009¥**
	MC	**Yes**		**0.9711**	**0.9724**	**0.9609**	**0.9429**	**(GPU)**
	UTMB	**Yes**		**0.9501**	**0.9588**	**0.9341**	**0.9016**	

* The execution time represents for separate segmentation of left Lung and right lung, respectively.
† The execution times are based on 256×256 CXR using CPU and CPU.
¥ The execution time shows including both segmentation + post processing steps using GPU.

sophisticated deep learning algorithms producing the expected result with minimal time cost. As shown in Table 4, our algorithm demonstrates comparable time cost having the average execution time of 0.043 for initial segmentation and 0.009 for post-processing step. This algorithm utilizes two separate simple deep learning architectures in a novel setup and ensemble them using parallel computing, which makes the algorithm time-efficient compared to Munawar et. al. [26] and Novikov et. al. [24], especially for the segmentation step. Notice that the time reported for Novikov et. al. [24] is for 256×256 CXR images. On the downside, the algorithm adds an extra portion of time for the post-processing step to remove the noises and inaccuracies after the initial segmentation, which sometimes may not need in ideal cases where lung region has less or no bacterial infections. CXRs having intensive bacterial infection must deploy the post-processing to get better lung contour in compromise of additional time cost.

6. Conclusions

In this work, we proposed a fully automated deep learning framework to increase the accuracy of lung segmentation in CXRs. The framework is divided into three major stages. First, we extract small patches from the original CXR images. Second, the extracted patches are classified and segmented by deploying an ensemble of a CNN model and an adapted U-Net model. Later, the patches are merged back to obtain two pre-segmentation masks. Subsequently, the pre-segmentation masks are combined together using the binary OR operation, which is later post-processed to generate the final segmented mask. The novelty of this framework is that the two concurrently employed parallel deep learning models complement each other by capturing missing pixels in two pre-segmentation masks. Thus, merging the pre-segmented masks helps to increase the overall segmentation accuracy, especially for the CXRs from patients affected by pulmonary disease. It is worth mentioning that, this framework can efficiently utilize small datasets for training the CNNs. The performance and robustness of our proposed framework are thoroughly investigated with different cases studies and datasets. The framework outperforms many of the state-of-the-art methods presented in the literature. In our experiments, we achieved the values of 96.67%, 96.83%, 96.88%, 95.32%, and 93.15% for average sensitivity, specificity, accuracy, Dice coefficient, and intersection-over-union, respectively.

The proposed method can be further improved and adapted to better capture some other practical aspects. For example, this work only involved the posteroanterior (PA) and anteroposterior (AP) view of chest X-rays. In the future, we plan to extend this work to segment the lung region from lateral view CXRs and different kinds of radiographs. Since segmentation is the first step of reading the chest radiographs in the CAD systems, we also aim to adapt our methods to develop a CAD-based automatic diagnostics system via proper and accurate analysis of abnormalities, lung size, and other phenomena.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the National Science Foundation (ECR-PEER-1935454) and (ERC-ASPIRE-1941524). The authors wish to express sincere gratitude for their financial support. We also acknowledge to the medical doctors from the University of Texas at Medical Branch for their continuous support, information, and providing the proprietary dataset used in this research.

References

[1] W.R. Hendee, et al., Addressing overutilization in medical imaging, Radiology 257 (1) (2010) 240–245.
[2] R.S. Mackay, Medical images and displays. Comparisons of nuclear magnetic resonance, ultrasound, X-rays, and other modalities, United States, 1984.
[3] B. Van Ginneken, B.T.H. Romeny, M.A. Viergever, Computer-aided diagnosis in chest radiography: a survey, IEEE Trans. Med. Imaging 20 (12) (2001) 1228–1241.
[4] L.G. Quekel, A.G. Kessels, R. Goei, J.M. van Engelshoven, Miss rate of lung cancer on the chest radiograph in clinical practice, Chest 115 (3) (1999) 720–724.
[5] G. Torres-Mejía, et al., Radiographers supporting radiologists in the interpretation of screening mammography: a viable strategy to meet the shortage in the number of radiologists, BMC Cancer 15 (1) (2015) 1–12.
[6] W. J. Palmer. Physician Specialty Shortage - Including Radiologist - Continue to Climb. Diagnostic Imaging. Available at: https://www.diagnosticimaging.com/view/the-impact-of-price-transparency-in-radiology (accessed March 1, 2021).
[7] S. Katsuragawa, K. Doi, Computer-aided diagnosis in chest radiography, Comput. Med. Imaging Graph. 31 (4–5) (2007) 212–223.
[8] N. Mahomed, et al., Computer-aided diagnosis for World Health Organization-defined chest radiograph primary-endpoint pneumonia in children, Pediatric Radiol. 50 (4) (2020) 482–491.
[9] G.S. Lodwick, T.E. Keats, J.P. Dorst, The coding of roentgen images for computer analysis as applied to lung cancer, Radiology 81 (2) (1963) 185–200.
[10] A. Zakirov, R. Kuleev, A. Timoshenko, A. Vladimirov, Advanced approaches to computer-aided detection of thoracic diseases on chest X-rays, Appl. Math. Sci. 9 (88) (2015) 4361–4369.
[11] R.S. Lee, F. Gimenez, A. Hoogi, K.K. Miyake, M. Gorovoy, D.L. Rubin, A curated mammography data set for use in computer-aided detection and diagnosis research, Sci. Data 4 (1) (2017) 1–9.
[12] M.K. Santos, J.R. Ferreira Júnior, D.T. Wada, A.P.M. Tenório, M.H.N. Barbosa, P.M. D.A. Marques, Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine, Radiol. Brasileira 52 (6) (2019) 387–396.
[13] A.H. Dallal, C. Agarwal, M.R. Arbabshirani, A. Patel, G. Moore, Automatic estimation of heart boundaries and cardiothoracic ratio from chest x-ray images, in: Medical Imaging 2017: Computer-Aided Diagnosis, 2017, vol. 10134, International Society for Optics and Photonics, p. 101340K.
[14] M.N. Saad, Z. Muda, N.S. Ashaari, H.A. Hamid, Image segmentation for lung region in chest X-ray images using edge detection and morphology, in: 2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014), 2014, IEEE, pp. 46-51.
[15] T. Xu, M. Mandal, R. Long, I. Cheng, A. Basu, An edge-region force guided active shape approach for automatic lung field detection in chest radiographs, Comput. Med. Imaging Graph. 36 (6) (2012) 452–463.
[16] W.S.H.M.W. Ahmad, W.M.D.W. Zaki, M.F.A. Fauzi, Lung segmentation on standard and mobile chest radiographs using oriented Gaussian derivatives filter, Biomed. Eng. Online 14 (1) (2015) 1–26.
[17] M. Kiran, I. Ahmed, N. Khan, A.G. Reddy, Chest X-ray segmentation using Sauvola thresholding and Gaussian derivatives responses, J. Ambient Intell. Hum. Comput. 10 (10) (2019) 4179–4195.
[18] S. Candemir, et al., Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration, IEEE Trans. Med. Imaging 33 (2) (2013) 577–590.
[19] N. Reamaroon, et al., Robust segmentation of lung in chest x-ray: applications in analysis of acute respiratory distress syndrome, BMC Med. Imaging 20 (1) (2020) 1–13.
[20] M. Kholiavchenko, et al., Contour-aware multi-label chest X-ray organ segmentation, Int. J. Comput. Assist. Radiol. Surg. 15 (3) (2020) 425–436.
[21] C. Yan, J. Yao, R. Li, Z. Xu, J. Huang, Weakly supervised deep learning for thoracic disease classification and localization on chest x-rays, in: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 2018, pp. 103-110.
[22] G. Gaál, B. Maga, A. Lukács, Attention u-net based adversarial architectures for chest x-ray lung segmentation, arXiv preprint arXiv:2003.10304, 2020.
[23] L. Saidy, C.-C. Lee, Chest X-ray image segmentation using encoder-decoder convolutional network, in: 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 2018, IEEE, pp. 1-2.
[24] A.A. Novikov, D. Lenis, D. Major, J. Hladůvka, M. Wimmer, K. Bühler, Fully convolutional architectures for multiclass segmentation in chest radiographs, IEEE Trans. Med. Imaging 37 (8) (2018) 1865–1876.
[25] R. Rashid, M.U. Akram, T. Hassan, Fully convolutional neural network for lungs segmentation from chest X-rays, in: International Conference Image Analysis and Recognition, Springer, 2018, pp. 71–80.
[26] F. Munawar, S. Azmat, T. Iqbal, C. Grönlund, H. Ali, Segmentation of lungs in chest X-ray image using generative adversarial networks, IEEE Access 8 (2020) 153535–153545.
[27] C. Chen, Q. Dou, H. Chen, P.-A. Heng, Semantic-Aware Generative Adversarial Nets for Unsupervised Domain Adaptation in Chest X-Ray Segmentation, Cham, 2018: Springer International Publishing, in Machine Learning in Medical Imaging, pp. 143-151.
[28] O. Gomez, P. Mesejo, O. Ibáñez, A. Valsecchi, O. Cordon, Deep architectures for high-resolution multi-organ chest X-ray image segmentation, Neural Comput. Appl. (2019) 1–15.
[29] J.C. Souza, J.O.B. Diniz, J.L. Ferreira, G.L.F. da Silva, A.C. Silva, A.C. de Paiva, An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks, Comput. Methods Programs Biomed. 177 (2019) 285–296.

[30] J.A. Onofrey et al., Generalizable multi-site training and testing of deep neural networks using image normalization, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, IEEE, pp. 348-351.
[31] C. Yan, B. Gong, Y. Wei, Y. Gao, Deep multi-view enhancement hashing for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 43 (4) (2020) 1445–1451.
[32] C. Yan, Z. Li, Y. Zhang, Y. Liu, X. Ji, Y. Zhang, Depth image denoising using nuclear norm and learning graph model, ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 16 (4) (2020) 1–17.
[33] C. Yan, T. Teng, Y. Liu, Y. Zhang, H. Wang, X. Ji, Precise no-reference image quality evaluation based on distortion identification, ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 17 (3s) (2021) 1–21.
[34] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[35] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: ICML, 2010.
[36] C. Yan, et al., Task-adaptive attention for image captioning, IEEE Trans. Circuits Syst. Video Technol. (2021).
[37] Y. Wen, M. F. Rahman, H. Xu, T.-L. B. Tseng, Recent Advances and Trends of Predictive Maintenance from Data-driven Machine Prognostics Perspective, Measurement (2021) 110276.
[38] C. Yan, et al., Age-invariant face recognition by multi-feature fusion and decomposition with self-attention, ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 18 (1s) (2021) 1–18.
[39] T. de Moor, A. Rodriguez-Ruiz, A. G. Mérida, R. Mann, J. Teuwen, Automated lesion detection and segmentation in digital mammography using a u-net deep learning network, in: 14th International Workshop on Breast Imaging (IWBI 2018), 2018, vol. 10718, International Society for Optics and Photonics, p. 1071805.
[40] H. Shaziya, K. Shyamala, R. Zaheer, Automatic lung segmentation on thoracic CT scans using U-Net convolutional network, in: 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, IEEE, pp. 0643-0647.
[41] H.T. Le, H.-T.-T. Pham, Brain tumour segmentation using U-Net based fully convolutional networks and extremely randomized trees, Vietnam J. Sci. Technol. Eng. 60 (3) (2018) 19–25.
[42] M.F. Rahman, T.-L.B. Tseng, M. Pokojovy, W. Qian, B. Totada, An automatic approach to lung region segmentation in chest x-ray images using adapted U-Net architecture, in: Proc. of SPIE Vol, vol. 11595, pp. 115953I.
[43] J.Y. Gil, R. Kimmel, Efficient dilation, erosion, opening, and closing algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 24 (12) (2002) 1606–1617.
[44] J.-M. Park, C.G. Looney, H.-C. Chen, Fast connected component labeling algorithm using a divide and conquer technique, Comput. Their Appl. 4 (2000) 4–7.
[45] M.F. Rahman, J. Wu, B. Tseng, Automated Fiber Extraction From SEM Images With Application to Quality Control of Fiber-Reinforced Composites Manufacturing, in: International Manufacturing Science and Engineering Conference, 2018, vol. 51371, American Society of Mechanical Engineers, p. V003T02A037.
[46] J. Bertels, et al., Optimizing the Dice score and Jaccard index for medical image segmentation: Theory and practice, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 92–100.
[47] H. Chung, H. Ko, S.J. Jeon, K.-H. Yoon, J. Lee, Automatic lung segmentation with juxta-pleural nodule identification using active contour model and bayesian approach, IEEE J. Translat. Eng. Health Med. 6 (2018) 1–13.