



Enhanced lung image segmentation using deep learning

Shilpa Gite^{1,2} · Abhinav Mishra¹ · Ketan Kotecha^{1,2}

Received: 10 August 2021 / Accepted: 30 October 2021 / Published online: 3 January 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

With the advances in technology, assistive medical systems are emerging with rapid growth and helping healthcare professionals. The proactive diagnosis of diseases with artificial intelligence (AI) and its aligned technologies has been an exciting research area in the last decade. Doctors usually detect tuberculosis (TB) by checking the lungs' X-rays. Classification using deep learning algorithms is successfully able to achieve accuracy almost similar to a doctor in detecting TB. It is found that the probability of detecting TB increases if classification algorithms are implemented on segmented lungs instead of the whole X-ray. The paper's novelty lies in detailed analysis and discussion of U-Net + + results and implementation of U-Net + + in lung segmentation using X-ray. A thorough comparison of U-Net + + with three other benchmark segmentation architectures and segmentation in diagnosing TB or other pulmonary lung diseases is also made in this paper. To the best of our knowledge, no prior research tried to implement U-Net + + for lung segmentation. Most of the papers did not even use segmentation before classification, which causes data leakage. Very few used segmentations before classification, but they only used U-Net, which U-Net + + can easily replace because accuracy and mean_iou of U-Net + + are greater than U-Net accuracy and mean_iou, discussed in results, which can minimize data leakage. The authors achieved more than 98% lung segmentation accuracy and mean_iou 0.95 using U-Net + +, and the efficacy of such comparative analysis is validated.

Keywords TB dataset · Data preprocessing · Segmentation models · Lung segmentation · Deep learning

1 Introduction

Tuberculosis (TB) is a primal disease that has probably affected humans since the dawn of humankind. TB is a contagious disease and caused by the bacteria mycobacterium tuberculosis. In 2019, TB was the cause of death of almost 1.4 million people around the globe. It is among the infectious diseases responsible for the death of millions of people [1]. Early diagnosis of TB is essential to save the patients from getting it fatal and life-threatening. TB can be diagnosed using computed tomography scans, magnetic resonance imaging, X-rays, etc. Analysis of X-ray is one of

the main tools for TB screening. Individuals suspected of TB need biological and clinical investigation before confirming TB diagnosis, and the medication is prescribed as per the guideline provided by World Health Organization (WHO). Regular screening is necessary for the early and correct diagnosis of TB. The chest is one of the primary tools because of its sensitivity and interpretation [2]. However, it is inevitable to eliminate the chances of intuitive inconsistencies in diagnosing disease from radiography [3, 4]. Chest X-ray of TB is often misdiagnosed with other diseases because of similar radiologic patterns [5, 6]. Misdiagnosis of TB can lead to wrong medication and worsening health conditions and cause other severe side effects. Hence, there is a need for correct lung diagnosis.

Lower-middle- and low-income countries face a scarcity of trained radiologists, especially in the country's rural areas. In these types of outlooks, large-scale screening of pulmonary TB by analyzing CXR images can be done using a computer-aided diagnosis (CAD) system. Recent advancements in GPUs and computer vision and the availability of large-scale chest X-ray-labeled datasets

✉ Shilpa Gite
shilpa.gite@sitpune.edu.in

¹ Computer Science Engineering Department, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India

² Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis International (Deemed University), Pune 412115, India

helped in successful image recognition. Convolutional neural networks can learn highly nonlinear functions and hierarchical visual features from appropriate training data, but acquiring datasets in the medical imaging domain as comprehensive annotation as ImageNet is challenging [7, 8]. It can be seen that countries around the globe are investing a significant part of their annual budget in the healthcare industry. However, it is still unable to fulfill society's aspirations [9]. Furthermore, the lack of health workers in the countries puts a significant workload on existing healthcare workers, resulting in fatigue and health issues [10]. Thus, the deep learning applications in the healthcare sector are getting massive attention in the recent decade.

Solutions based on machine learning (ML) and deep learning (DL) have been suggested for many medical applications, especially in diagnosing a brain tumor, lung nodules, pneumonia, breast cancer, etc. Deep learning, part of machine learning (ML), encourages image classification and segmentation results, hence widely adopted by the research community [11]. The cost associated with the X-ray imaging technique is low, and the abundance of data for deep learning techniques created a favorable condition for computer-aided diagnostic system development. The study shows that classification of lung images after segmentation techniques improves model accuracy [12]. Therefore, the theme of this paper is to make use of four popular segmentation techniques for the lung images. Significant contributions of this paper could be stated as

- Review of state-of-the-art techniques in lung segmentation problems.
- Implementation and analysis of four primary segmentation techniques, namely FCN, SegNet, U-Net, and U-Net + + .
- Result analysis of the above-implemented benchmark segmentation architectures and their comparison on different performance measures.
- Discussion of these segmentation techniques and their efficacy in TB diagnosis.

The rest of the paper is organized as follows: Sect. 2 states the existing literature review and lung classification techniques. In Sect. 3, the proposed comparison methodology is explained in a stepwise manner. Section 4 presents the results and a discussion based on the results generated by four segmentation techniques. Lastly, Sect. 5 concludes with the findings of this implementation and gives recommendations for future work.

2 Literature review

The literature review is divided into two sections, namely segmentation and classification.

2.1 Segmentation

Rehman et al. [12] generated lung segments from X-ray images using U-Net with mean_iou of 92.82. Shaoyong Guo et al. [39] proposed a novel automatic segmentation model using radiomics with a combination of handcrafted and automated features. Dice similarity coefficients of 89.42% are achieved on ILD database MedGIFT. Chen Zhou et al. [40] developed an automatic segmentation model by integrating (3D) V-Net and spatial transform network (STN) to segment pulmonary parenchyma in CT images and analyze texture and features from the segmented pulmonary parenchyma regions to assist the radiologist in COVID-19 diagnosis. Mizuho Nishio et al. [41] used U-Net architecture optimized via Bayesian optimization on Japanese and Montgomery and obtained DSC of 0.976 and 0.973 on respective datasets. Ferreira et al. [42] proposed a modified U-Net model for automatic detection of infection caused by COVID-19. Trained and evaluated on the CT database of the actual clinical case from Pedro Ernesto University Hospital of the state of Rio de Janeiro, this model achieved a dice value of 77.1% and an average specificity of 99.76%. Feidao cao [43] improves the traditional architecture of U-Net by introducing variational autoencoder (VAE) in each layer of decoder–encoder to improve the ability of the network to extract the features. The network was tested and trained on NIH and JRST datasets and achieved accuracy and F1 score of 0.9701, 0.9334 and 0.9750, 0.9578, respectively.

Advantages of segmentation:

1. Segmentation of the image is the most important medical imaging process. It extracts the ROI (region of interest) through an automatic process. Segmentation divides the image into an area based on a specific interest, like segmenting body organs/tissue.
2. Implementation of classification neural network algorithms on segmented radiological images can improve the segmentation accuracy significantly.
3. Segmentation can increase the computational cost, but it can significantly decrease the overall cost of disease diagnosis.

2.2 Classification

Detecting tuberculosis is an arduous job because of discrete manifestations such as cavities, small opacities, large

opacities, consolidation, and focal radiologists. Recent advancements in convolutional neural network architecture made it possible to detect lung infections, including TB and pneumonia, from chest X-ray images. Convolutional neural network (CNN) was used to detect tuberculosis from CXR images. Rehman et al. [12] generated lung segments from X-ray images using U-Net with a mean_iou of 92.82. They performed classification using pre-trained models like ResNet, Vgg-16, and Vgg-19 on segmented images with a test accuracy of 98.14%. Tahir et al. [13] used transfer learning to classify diseases that belong to the coronavirus family (SARS, MERS, and COVID-19) with sensitivity greater than 90%. Ali Narin et al. [14] used pre-trained models like ResNet-50, ResNet-101, ResNet-152, and Inception-ResNet V2 to classify between COVID-19, normal, viral pneumonia, and bacterial pneumonia. ResNet-50 performs the highest classification with a max accuracy of 99.17%. Pchhikara et al. [15] used some pre-trained models (ResNet, ImageNet, Xception Net, and Inception Net) to detect pneumonia from chest X-ray images.

Zak et al. [16] implemented TB classification using pre-trained Vgg-16, Vgg-19, ResNet-50, and inception V2 with accuracies 64%, 72%, and 81%, AUC 0.82, 0.76, and 0.87, and sensitivity 0.77, 0.68, and 0.77. Rohan et al. [17] proposed a model comprising three standard architectures: AlexNet, GoogLeNet, and ResNet, with an accuracy of 88.24% and AUC 0.93. Melendez et al. [18] classify tuberculosis using MIL, SVM, and MIL + AL with pixel-level AUC 0.855, 0.900, and 0.877 and case-level AUC 0.801, 0.878, 0.861. Rahul et al. [19] presented a convolutional neural network model comprising 7 Conv layers and 13 fully connected layers. The optimizer used was Adam, with a validation accuracy of 88.76% from chest X-ray images. Volkov et al. [22] introduced a CNN model with an accuracy of 86.6%. Dao et al. [23] classified the performance of the pre-trained models for the classification of TB using Shenzhen (CHN) and Montgomery (MC). UKloupes et al. [24] worked on the pre-trained model's performance on two datasets (CHN and MC) to distinguish between positive and negative, and the accuracy achieved is 80%.

Four different CNN models (Vgg-16, Vgg-19, ResNet-50, and GoogLeNet) were explored, and the results generated by these models were analyzed (Yaakob et al. [25]). For TB detection, Ahsan et al. [26] presented a general pre-trained CNN, and accuracies achieved are 81.25% and 80%. Yadav et al. [27] transfer learning technique was used along with a deep learning framework which shows an accuracy of 99.98%. Ardila et al. [28] introduced a deep learning architecture to detect lung cancer and achieved an AUC of 94.4%. Thriach et al. [29] proposed a pre-trained Conv neural network for lung cancer classification with a mean accuracy of 74.43%, mean specificity of 74.96%, and mean sensitivity of

74.68%. Hua et al. [30] used CNN and DBN to classify lung cancer for achieving 82.2% and 73.4% of specificity and sensitivity. Islam et al. [31] presented the architecture, a combination of LSTM and CNN, to classify COVID-19 by analyzing chest X-ray and achieved AUC of 99.9%, specificity of 99.2%, sensitivity 99.3%, and an F1-score of 98.9%.

Based on the above extensive literature review, it is found that plenty of CAD architectures used ML and traditional DL architectures to detect TB and other infectious diseases and achieved accuracy up to 90% and more. In medical applications, robust and flexible algorithms or methods can increase the accuracy of the CAD system to diagnose TB from chest X-ray images and make the system reliable. The utilization of the latest and different architectures or assembling of benchmark algorithms can increase classification accuracy. Usually, the whole chest X-ray is used to diagnose lung disorder using convolutional neural networks. Although lungs are used to detect infectious diseases like TB and pneumonia, CXR also contains other regions of the chest cavity. Therefore, focusing only on lungs from CXR throughout training and classification can increase the accuracy significantly. Thus, to isolate lungs from X-ray images, segmentation techniques are used. So in this study, four benchmark segmentation techniques are explored, and their results are analyzed comprehensively. Based on the extensive literature review, along with detailed overview of the models and their accuracy in diagnosis of the lung related diseases as presented in Table 1.

3 Proposed methodology

This section contains comprehensive information about the dataset, preprocessing techniques, and segmentation models. Later in this section, the details about evaluation metrics like accuracy, dice coefficient, mean_iou, recall, specificity, sensitivity, and precision are discussed and are used to compare four models employed in this study. In this study, the authors explored four broadly used segmentation algorithms in multiple fields, including medical diagnoses, and tried to analyze which type of algorithm works well with limited medical images. The attention net has an additional mechanism that adds more parameters to the model, resulting in increased training time. It requires powerful graphical processing units (GPUs) to train, which is not very cost-effective. These are the few reasons why authors considered U-Net + + over attention net.

3.1 Dataset description

a. Montgomery County X-ray set:

Montgomery is a labeled dataset consisting of X-ray images with a frontal view. This dataset contains 138

X-ray images; 80 of these chest X-ray images contain no disease and 58 chest X-ray images show infection caused by TB. This dataset has been acquired by the Department of Health and Human Services of Montgomery County, MD, USA. Dataset also contains manually generated lung segment masks of every X-ray image of the dataset and is in DICOM format [36].

b. **Shenzhen Hospital X-ray set:**

Shenzhen dataset is a labeled dataset consisting of X-ray images with a frontal view. This dataset contains 662 X-ray images; 326 of which are regular chest X-rays and 336 of these X-ray images show the presence of infection caused by TB. The Hospital in Shenzhen, China, has collected this dataset in JPEG format [36].

3.2 Preprocessing

First, X-ray images of Montgomery and Shenzhen datasets were converted to PNG format because X-ray images of Montgomery and Shenzhen datasets are in DICOM and JPEG format, so it will be simple to train CNN models. Then, in the next level, standardization of images was performed by resizing because both datasets had different sizes. In this study, 512×512 pixel size was considered for FCN, SegNet, U-Net, U-Net + + as the input size for different convolutional architecture is different, and normalization of data was done using Z-score normalization using mean and standard deviation.

After preprocessing, the datasets were randomized and divided into 80% training and 20% test data, where 80% of the training data were used to provide experience to the segmentation models regarding lung segments. The remaining 20% of the data were used to evaluate the segmentation models.

3.3 Segmentation models

There are many other algorithms which are used for medical image segmentation, including DeepLab v1, DeepLab v2, DeepLab v3, DeepLab v3 + , 3D U-Net, V-Net, Res-U-Net, DenseUNet, H DenseUNet, GANs, SegAN, SCAN, PAN, AsynDGAN, etc. The authors have chosen four segmentation models, namely FCN, SegNet, U-Net, and U-Net + +, because of their features, for example, SegNet requires low memory for training and testing, FCN is fast and uses pixel-wise classification to produce segments, U-Net is effective with fewer data, and U-Net + + is a modified version of U-Net and thus utilizes attributes such as redesign skip connection and deep supervision to produce perfect segments. FCN is among the

first segmentation algorithms, so it is used as a benchmark in this paper for the other three algorithms (SegNet, U-Net, and U-Net + +). The authors also tried to track the improvement in segmentation algorithms and their performance in medical image segmentation, especially lung segmentation. There are very few papers that tried to track and analyze the improvement in segmentation algorithms comprehensively. These are among the most used semantic segmentation architectures shown by their Google Scholar citation scores presented in Fig. 1.

These four segmentation models are called “modified” in their architecture diagrams as per the finalized dataset and the problem statement.

a. **Fully convolutional networks (FCNs)**

Fully convolutional networks use CNNs for a pixel-to-pixel transformation, as shown in Fig. 2. However, unlike CNN, the weight and height of all intermediate layers feature maps are brought back to the original size through convolutional transpose in FCN, allowing localization and skip connections to be implemented to recover the fine spatial information lost downsampling [37].

FCN model architecture: FCNs use locally connected layers like convolution, upsampling, and pooling. Each layer is a 3-D array of size $h \times w \times d$ where h and w are spatial dimensions and d represents no channels. Dense layers are avoided in FCN, which means fewer parameters make the network faster and easier to train. The downsampling path of the model is responsible for the interpretation and extraction of the context, and upsampling is used to model for localization. FCNs employ skip connections to recover fine-grained information lost in downsampling. This improvement allows the model to have pixel-wise predictions.

b. **SegNet**

SegNet was developed by the University of Cambridge and primarily used for semantic segmentation [35]. This segmentation CNN model incorporates two

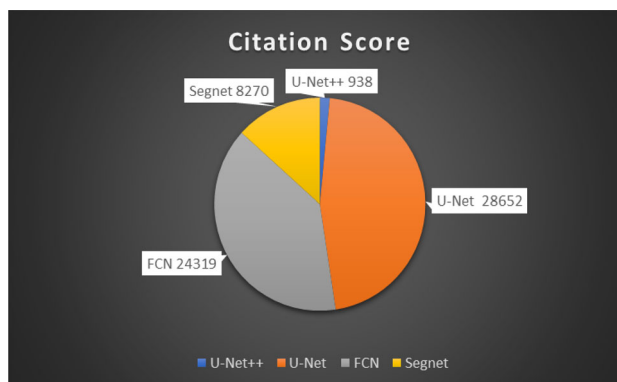


Fig. 1 Google scholar citation scores of segmentation models

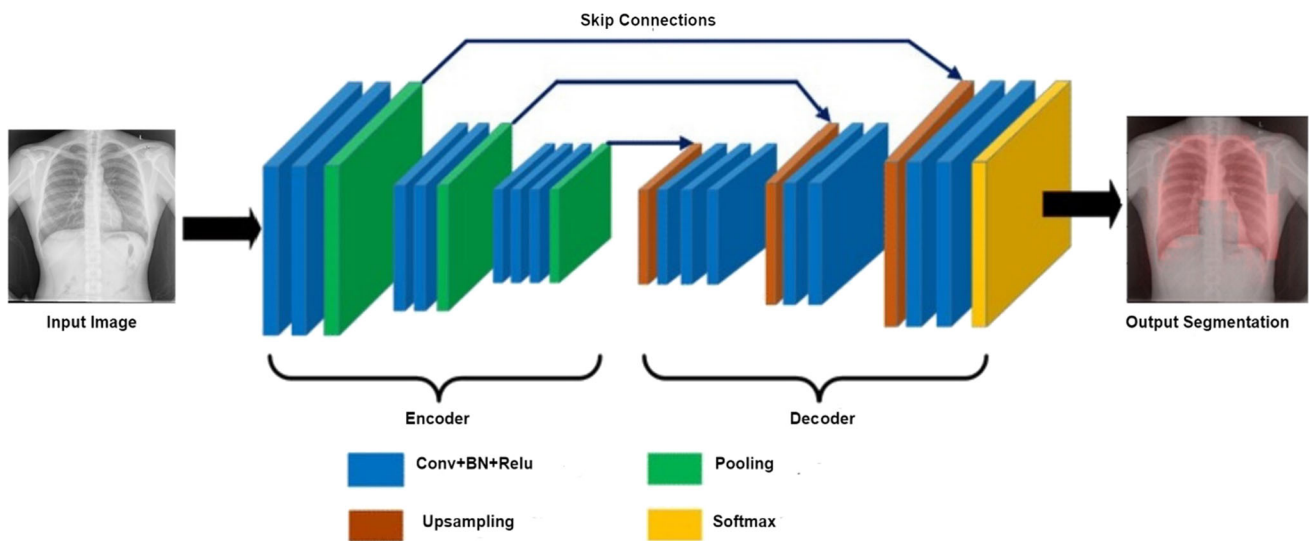


Fig. 2 Modified FCN architecture [37]

halves. The first half is an encoder, and the second half is a decoder followed by a pixel-wise classification. The network architecture of the encoder is identical to Vgg-16, and low-resolution encoder feature maps are converted into a full input feature map by the decoder network of the architecture for pixel-wise classification. The pooling indices of max-pooling layers are computed during downsampling to perform nonlinear upsampling, as shown in Fig. 3.

SegNet Architecture: SegNet architecture can be divided into two halves, followed by a pixel-wise classification layer.

- c. Encoder network.
- d. Decoder network.
- e. Encoder:

It performs convolution with a filter to produce a set of feature maps. It has 13 convolutional layers that are

not fully connected, max-pooling layers, and these are used to achieve translation invariance. Combining it with subsampling leads to pixels governing large input feature maps. These methods achieve better classification accuracy and a reduction in the size of the feature maps. This is also responsible for the lossy image presentation with faded boundaries, which are unsuitable for image segmentation. The output should have the exact image resolution as the original image. This is achieved by using upsampling in the decoder. To achieve the exact image resolution in output as the input image, it is essential to store and capture the details of the edges in the encoder feature map before subsampling. SegNet accumulates only the max-pooling indices.

- f. Decoder:

For each encoder, the corresponding decoder input

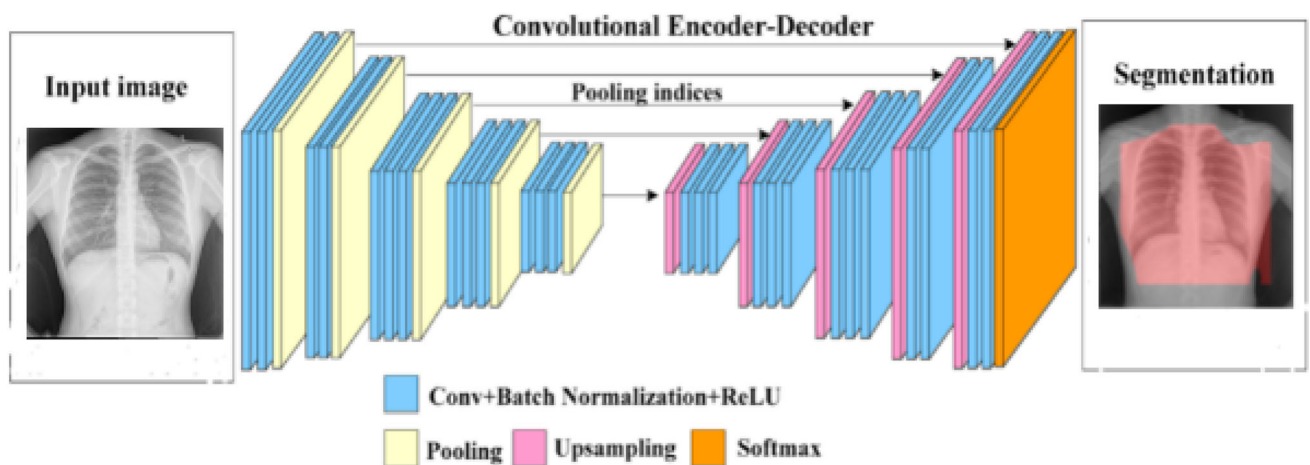


Fig. 3 Modified SegNet architecture [35]

Table 1 An overview of deep learning and segmentation techniques

Author	Models	Diseases	Datasets	Accuracy
Rehaman et al. [12]	CNN with transfer learning models like ResNet-101, ResNet-50, Vgg-19	Tuberculosis	Montgomery, Shenzhen, NLM, NIAID TB, RSNA, CXR, Belarus	98.14%
Tahir et al. [13]	CNN with transfer learning models like SqueezeNet, MobileNet, DenseNet	COVID-19	QU-covid-family	90%
Ali Narin et al. [14]	CNN with transfer learning Models like ResNet-152	COVID-19	Cohen's GitHub and LDOCTCXR	99.7%
p.chhikara et al. [15]	CNN with transfer learning models like Xception, inception	Pneumonia	Guangzhou women and children's medical center data[32]	90.1%
Zak M et al. [16]	CNN with transfer learning model like Vgg-16 and ResNet-50	Tuberculosis	Montgomery, Shenzhen	81%
Rohan et al. [17]	CNN with data augmentation	Tuberculosis	Montgomery, Shenzhen, Belarus	88.24%
J.Melendez et al. [18]	MIL,SVM,MIL + AL	Tuberculosis	Health center in Lusaka Zambia data	–
Rahul et al. [19]	CNN	Tuberculosis	Montgomery, Shenzhen	82.09%
A Rohilla et al. [20]	CNN with transfer learning	Tuberculosis	Montgomery, Shenzhen	80%
LGC.Evangelista et al. [21]	CNN with transfer learning	Tuberculosis	JSRT, Montgomery, Shenzhen	88.76%
F.pasa et al [22]	CNN	Tuberculosis	Montgomery, Shenzhen, Belarus	86.6%
Nguyen et al. [23]	CNN with transfer learning	Tuberculosis	Montgomery, Shenzhen	–
UK Lopes et al. [24]	CNN with transfer learning	Tuberculosis	Montgomery, Shenzhen	80%
S.S. Meray et al. [25]	CNN with transfer learning, model, like Vgg-16	Tuberculosis	Montgomery, Shenzhen, Indiana university	–
Ahsan et al [26]	CNN with transfer learning	Tuberculosis	Montgomery, Shenzhen	81.25%
O.Yadav et al.[27]	CNN with transfer learning	Tuberculosis	Montgomery, Shenzhen, NIH CXR	94.98%
Ardila et al. [28]	CNN	Lung cancer	LUNA, LIDC, NLST	–
A.Thirach [29]	CNN with transfer learning	Lung cancer	JSRT, NIH	74.43%
Hua et al. [30]	CNN and DBN	Lung cancer	LIDC-IDRI	82.2%
M.Z. et al. [31]	CNN + LSTM	COVID-19	Kaggle repository, NIH data, Mendeley	92%

feature maps are upsampled by memorized max-pooling indices from the corresponding encoder feature map and convolved with decoder filter banks to produce a dense feature map. The feature maps produced by decoders are of the same size and channels as their encoder inputs. The trainable classifier is fed with the higher-dimensional feature representation present at the final decoder as output. The classifier classifies each pixel and produces a channel image of probability at the output.

g. U-Net

The U-Net is a CNN architecture for solving segmentation problems in the biomedical field and other image transformation tasks. U-Net is more successful than other convolutional models in pixel-based image segmentation because it is very effective

with limited data. This unique model was developed by Olal Ronneberger et al. [33], as shown in Fig. 4.

U-Net model architecture: To segment biomedical images, the U-Net architecture has two paths. The first path is a contraction (also called encoder). The encoder captures context via a compact feature map. The encoder is a stack of max-pooling and convolution layers like Vgg-16. The other half of the architecture is a uniform expanding path (also known as a decoder) which is also the second path. The second path did the precise localization using transposed convolution. The encoder section is made of many contraction blocks. The encoder follows the classic architecture of ConvNet. The network uses a repeated implementation of two 3×3 convolutions (ReLU) and a 2×2 max-pooling operation with stride 2 for contraction. The

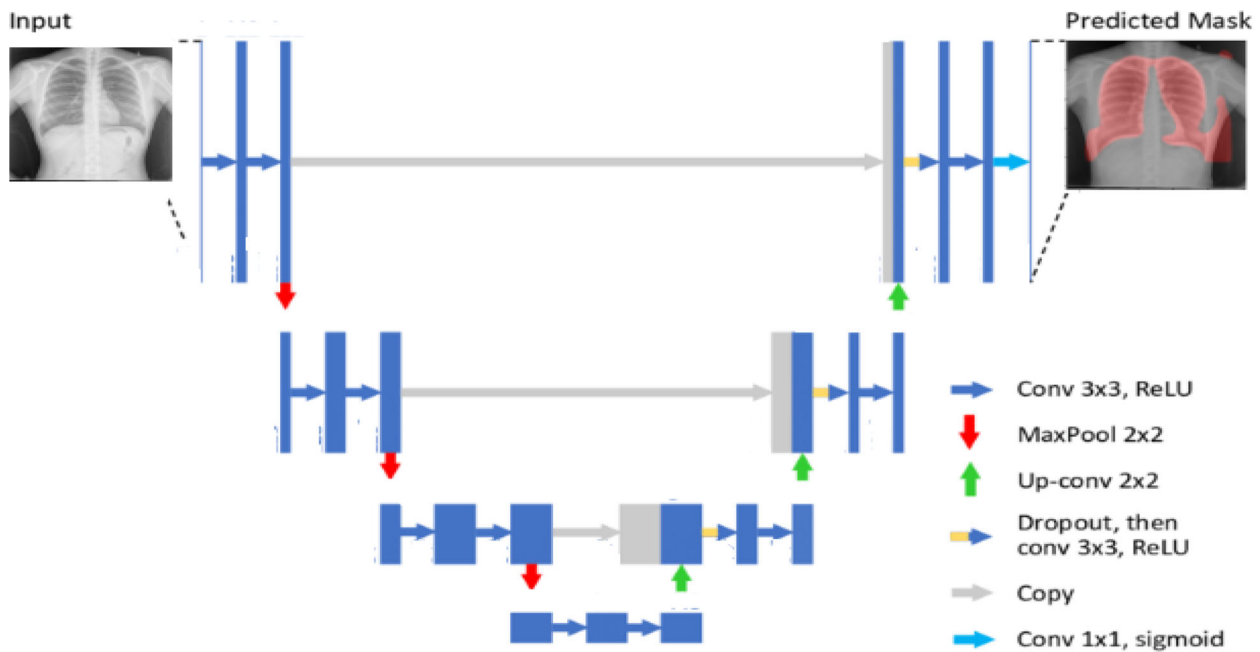


Fig. 4 Modified U-Net architecture [33]

numbers of feature channels double with every downsampling step. The expansive path contains a feature map's upsampling and 2×2 convolution ("up-convolution"), which halves the number of feature channels, concatenation with the corresponding feature map from skip connection and two 3×3 convolutions followed by ReLU. At the final layer, there is a 1×1 convolution used to map the component feature vector. In total, the network has 23 convolutional layers.

h. U-Net + +

The U-Net + + was proposed by Zhou et al. [34]. U-Net + + is the modified version of U-Net. U-Net + + uses the dense block ideas from DenseNet to improve U-Net. U-Net + + has three additional features to the original U-Net.

- i. Redesigned skip pathways.
- j. Dense skip connections.
- k. Deep supervision.

U-Net + + model architecture: U-Net + + architecture can be divided into three parts, as mentioned in the overview that distinguishes U-Net + + and U-Net.

1. Redesigned skip pathways: The U-Net + + consists of redesigned skip connections, as shown in Fig. 4. These are used to connect the semantic gap between encoder and decoder. The semantic gap of the feature map between encoder and decoder is reduced by skip connection Conv layers mentioned above. The direct connection of feature maps between encoder and decoder in U-Net results in semantic dissimilar feature

maps fusion. In U-Net + +, the output from the convolutional of the previous layers is concatenated with the output of the corresponding upsampled output of the low, dense block. This helps bring the feature maps of the encoders closer to the feature maps waiting in the corresponding decoder and helps in optimization quickly.

2. Dense skip connections: The U-Net + + has dense skip connections, as shown in Fig. 5, inspired by the DenseNet and the purpose of the dense skip connection to implement skip pathways between the encoder and decoder. This helps in improving the accuracy of segmentation and gradient flow. In addition, deep skip connection is responsible for accumulating prior feature maps and delivery to the right node due to dense convolution blocks along the skip pathways. This results in the generation of a full resolution feature at multiple semantic levels.
3. Deep supervision: Deep supervision in U-Net + +, as shown in red in Fig. 5, is implemented to adjust the model complexity to balance the speed and performance of the architecture. It is a must for an accurate CNN model to average the output from all segmentation branches.

After discussing the segmentation models above, the proposed methodology diagram (Fig. 6) gives a brief idea about the implementation flow adopted for this research. The dataset of chest X-ray is preprocessed first, and then the data split of 80–20% training and testing, respectively, takes place. Then, after applying four popular segmentation

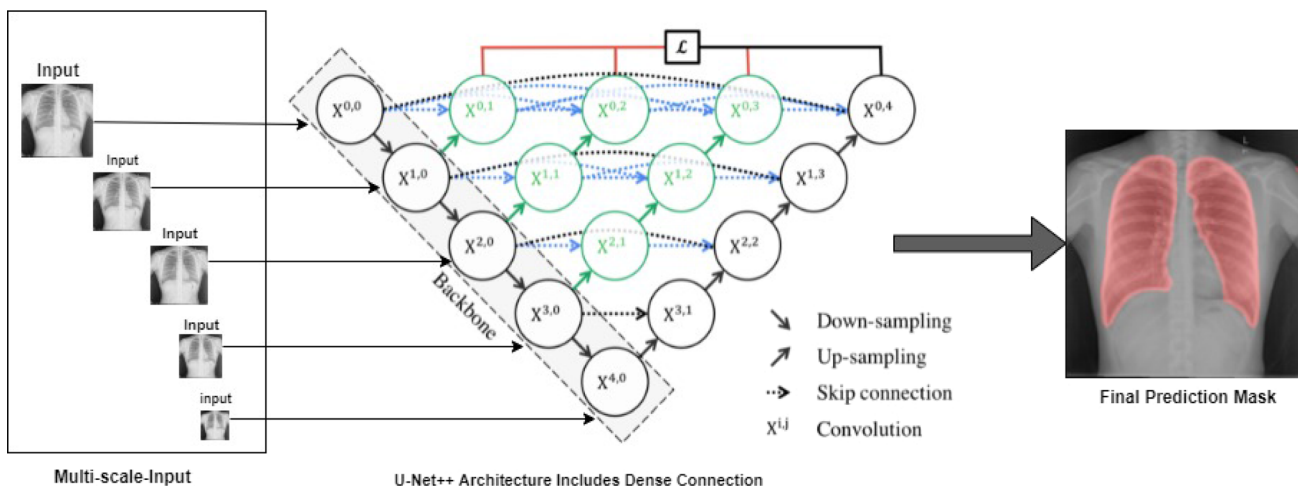


Fig. 5 Modified U-Net ++ architecture [34]

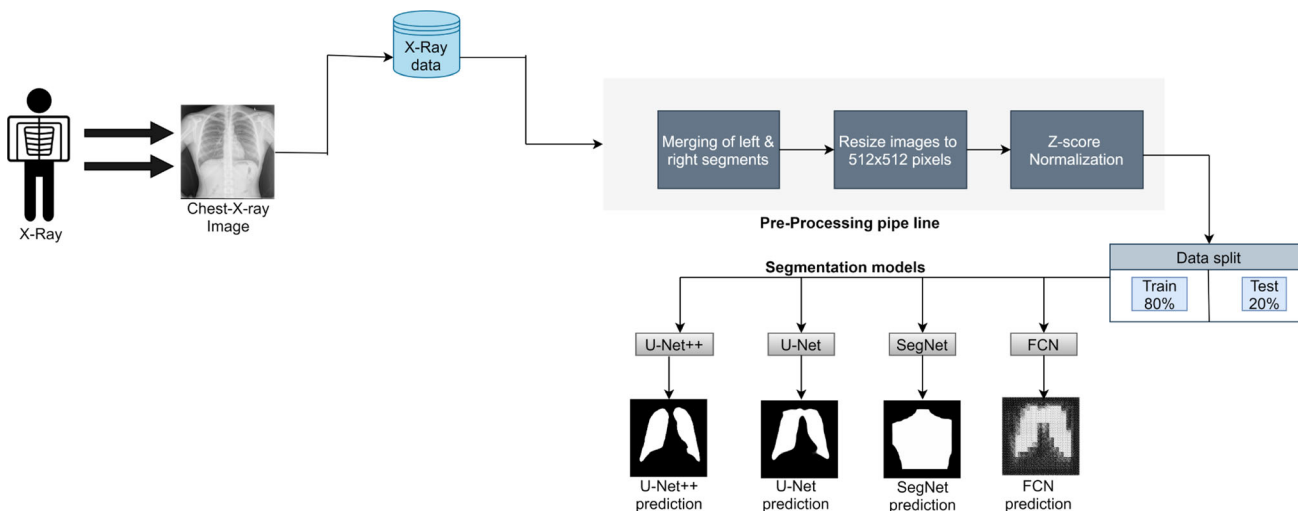


Fig. 6 Proposed methodology for comparative analysis of chest X-ray images

techniques on the image dataset, results are generated, compared, and analyzed in depth.

3.4 Performance matrix

Evaluation of models focuses on estimation of the performance of the model on unseen data. Thus, evaluating the performance of the neural network models for lung segmentation was done after finishing the training and validation phases, and models were compared using loss, accuracy, intersection over union (IoU), dice, sensitivity, specificity, specificity, recall, and precision.

True positive (TP)	False positive (FP)
True negative (TN)	False negative (FN)

The equations were used to calculate loss, accuracy, IoU, dice, sensitivity, and specificity:

$$\text{Accuracy} = (TP + TN) / ((TP + FN) + (FP + TN)). \tag{1}$$

$$\text{IoU(Jaccard Index)} = (TP) / (TP + FN + FP). \tag{2}$$

$$\text{Precision} = (TP) / (TP + FP). \tag{3}$$

$$\text{Specificity} = (TN) / (FP + TN). \tag{4}$$

$$\begin{aligned} \text{Dice coefficient(F1 - score)} \\ = (2 * TP) / (2 * TP + FN + FP). \end{aligned} \tag{5}$$

$$\text{Sensitivity} = (TP) / (TP + FN). \tag{6}$$

$$\text{Recall} = TP / (TP + FN). \tag{7}$$

4 Results and discussion

In this study, four neural network architectures of lung segmentation are evaluated on the Montgomery and Shenzhen datasets. This study uses a dataset of 704 images taken from these two datasets to check the model performance.

4.1 Segmentation techniques results an FCN

FCN is trained and validated on the datasets, and then the overall performance is evaluated using the test set. Best performing training and validation results are stated in Table 2.

Their results are discussed separately: Sect. 4.1 presents the results generated by four segmentation techniques separately, and 4.2 presents the comparison among them.

The performance of the FCN in the study is represented in Figs. 7, 8, 9, and 10. Accuracy is the way to measure how often an algorithm or architecture classifies positive and negative. The specificity is the measure of true positives identified by the model, and the sensitivity is the measure of true negatives identified by the model. FCN shows low validation dice_coefficient and mean_iou, which is shown in Figs. 22A and 23A, and Fig. 24B shows low accuracy as well. Low values of parameters show that the FCN is not suitable for organ segmentation which is a valuable research finding.

As seen and discussed above, segmented lungs generated by FCN are not satisfactory, which shows that FCN is unsuitable for medical application, which can be convinced from Fig. 11.

4.2 SegNet

Two standard publicly available datasets such as Montgomery and Shenzhen are used to generate training and testing results. Best training and validation results are stated in Table 3

The response of the SegNet Model is reported in Figs. 12, 13, 14, and 15. It can be seen from Figs. 12A and 13A that the values of validation mean_iou and dice_coefficient are higher than training mean_iou and

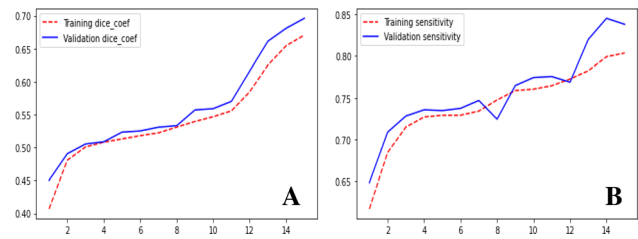


Fig. 7 Representation of training and validation coefficients (A), representation of training and validation sensitivity (B)

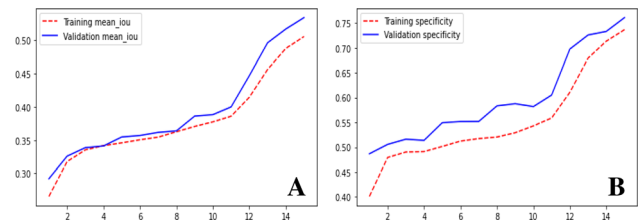


Fig. 8 Representation of training and validation mean_iou (A), representation of training and validation specificity (B)

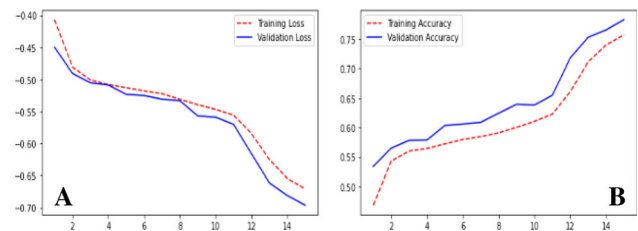


Fig. 9 Representation of training and validation loss (A), representation of training and validation accuracy (B)

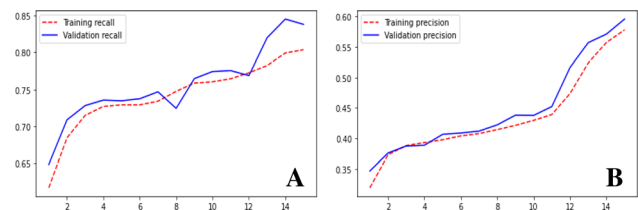


Fig. 10 Representation between training and validation recall (A), representation between training and validation precision (B)

dice_coefficient. Still, their values are 0.7914 and 0.6558. These values show that the model is not preferable for medical segmentation, and Fig. 14A and B shows that

Table 2 FCN best performing training and validation results

Model	Loss	Dice	Specificity	Mean_iou	Sensitivity	Recall	Precision	Accuracy
FCN training	– 0.6670	0.6670	0.7313	0.5018	0.8029	0.8029	0.5735	0.7536
FCN validation	– 0.6962	0.6962	0.7609	0.5343	0.8380	0.8380	0.5958	0.7832
difference	– 0.0292	0.0292	0.0296	0.0325	0.0351	0.0351	0.0223	0.03

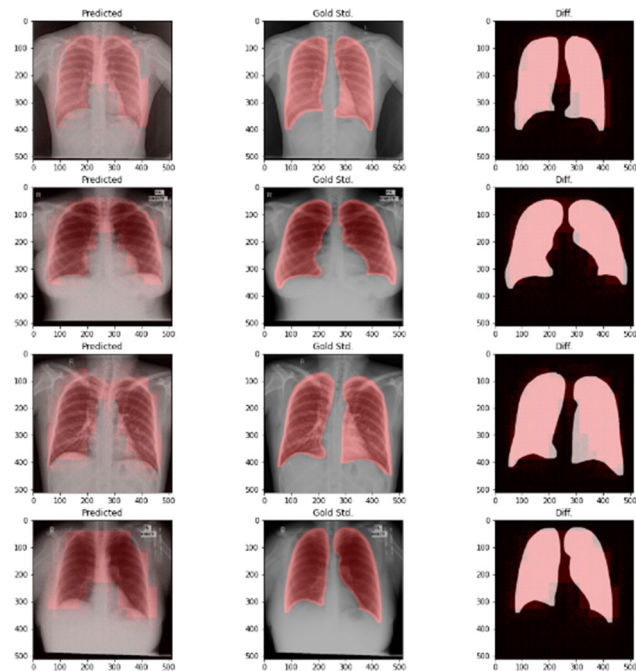


Fig. 11 Segmented lungs predicted by FCN and ground truth comparison

there is a big difference in values of validation loss, accuracy, and corresponding training values, which shows that the SegNet model is not suitable for a dataset with small size.

As presented and discussed above, segmented lungs generated by SegNet are not satisfactory, showing that SegNet is not suitable for medical application, as shown in Fig. 16.

Table 3 SegNet best performing training and validation result

Model	Loss	Dice	Specificity	Mean_iou	Sensitivity	Recall	Precision	Accuracy
SegNet training	– 0.6841	0.6841	0.6509	0.5219	0.9258	0.9258	0.5462	0.7358
SegNet validation	– 0.7914	0.7914	0.7995	0.6558	0.9684	0.9698	0.6701	0.8409
Difference	– 0.1073	0.1073	0.1486	0.1339	0.0426	0.044	0.1239	0.1051

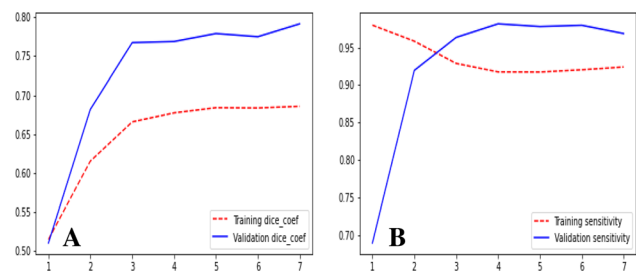


Fig. 12 Representation of training and validation coefficients (A), representation of training and validation sensitivity (B)

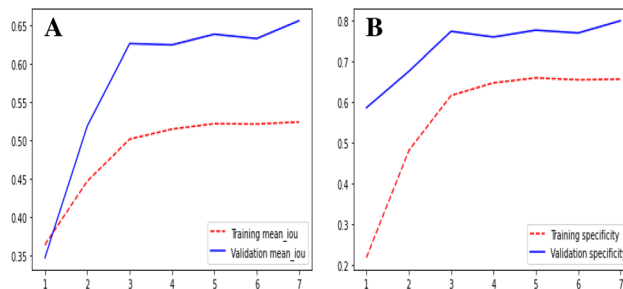


Fig. 13 Representation of training and validation mean_iou (A), representation of training and validation specificity (B)

4.3 U-Net

Originally, U-Net was trained on training data, validated, and evaluated on test datasets. As a result, the U-Net’s best performing training and validation parameters are stated in Table 4.

Dice coefficient is a measure of overlaps between two sets here; these two sets are ground truth masks and predicted masks. Sensitivity is the measure of ground truth. It can be seen from Fig. 17A, B that dice coefficient is increased with every epoch; maximum training dice coefficient and sensitivity are 0.8776, 0.8756. Validation dice coefficient and sensitivity are 0.9217 and 0.8904. It can be seen from the sensitivity results that U-Net does image localization by predicting the image pixel by pixel.

Loss predicts error in the model and accurately predicts how well the model is performed; as can be seen in Fig. 18A, B, loss decreases and accuracy increases with every epoch, and Fig. 19A, B shows that both mean_iou and specificity increase with every epoch and maximum

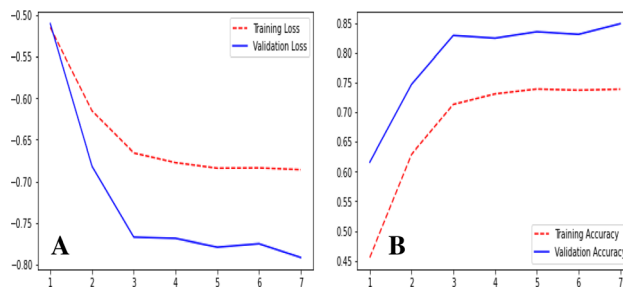


Fig. 14 Comparison of training and validation loss (A), comparison of training and validation accuracy (B)

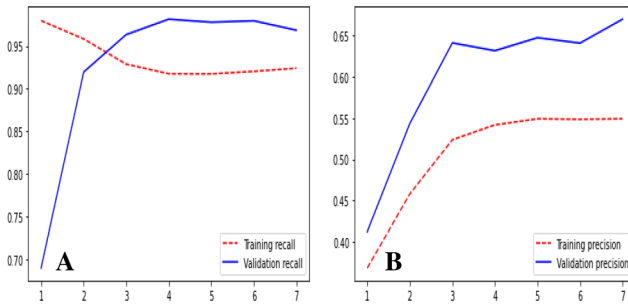


Fig. 15 Representation of training and validation recall (A), representation of training and validation precision (B)

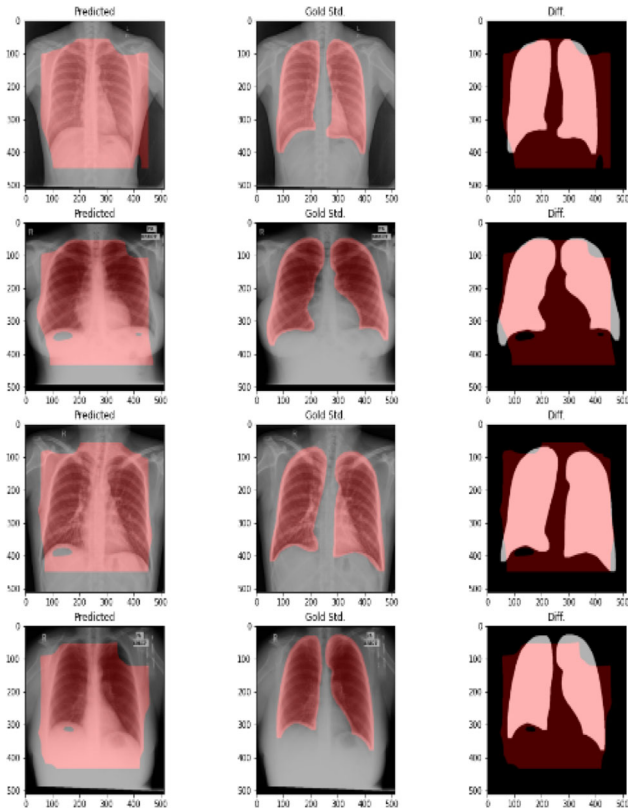


Fig. 16 Segmented lungs predicted by SegNet and ground truth comparison

Table 4 U-Net best performing training and validation results

Model	Loss	Dice	Specificity	Mean_iou	Sensitivity	Recall	Precision	Accuracy
U-Net training	− 0.8779	0.8779	0.9493	0.7852	0.8756	0.8756	0.8896	0.9256
U-Net validation	− 0.9217	0.9217	0.9837	0.8572	0.8904	0.8904	0.9584	0.9555
Difference	− 0.0438	0.0438	0.0344	0.072	0.0148	0.0148	0.0688	0.0299

validation values are 0.85 and 0.98, which shows that U-Net is doing an excellent job in creating segmented lungs as seen in Fig. 20.

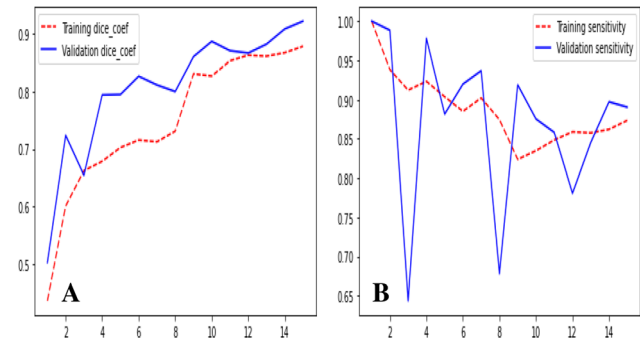


Fig. 17 Representation of training and validation coefficients (A), representation of training and validation sensitivity (B)

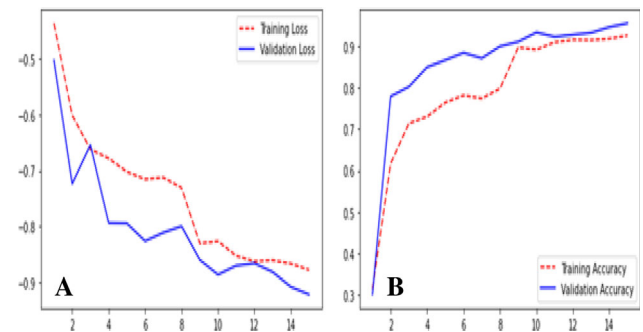


Fig. 18 Representation of training and validation loss (A), representation of training and validation accuracy (B)

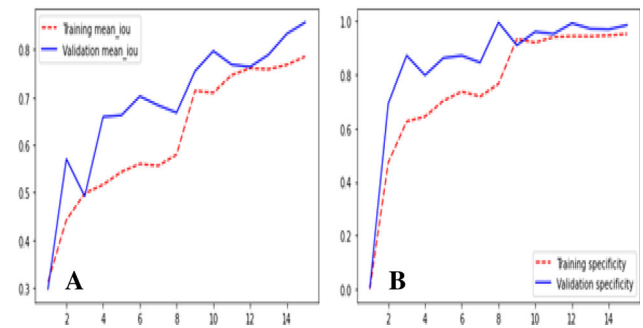


Fig. 19 Comparison of training and validation mean_iou (A), comparison of training and validation specificity (B)

Figure 21 shows segmented lungs predicted by U-Net, corresponding ground truth (Gold std), and the difference between ground truth and segmented lungs generated by

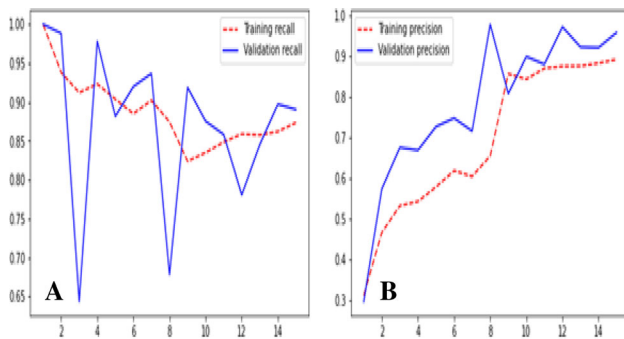


Fig. 20 Comparison between training and validation (A), comparison between training and validation precision (B)

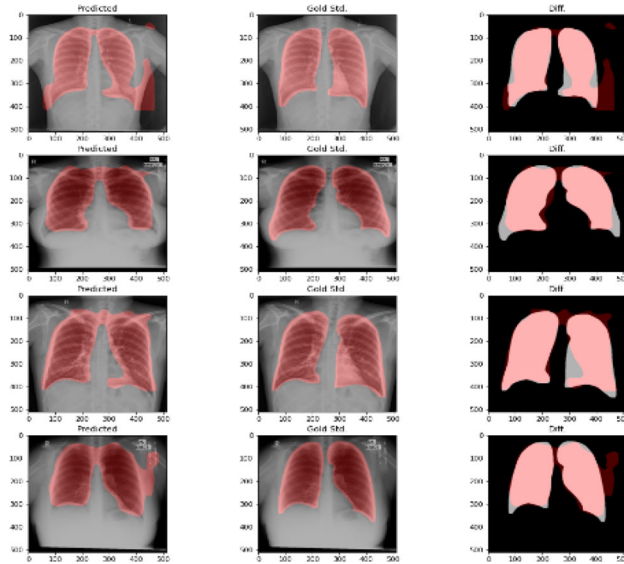


Fig. 21 Segmented lungs predicted by U-Net and ground truth comparison

trained U-Net. It can be seen from Fig. 11 that U-Net trained on Montgomery and Shenzhen datasets can segment the lung areas of X-ray images of the database very reliably and correctly.

4.4 U-Net + +

To train the U-Net + +, training dataset was utilized, and performance of the validation set is enhanced by using hyperparameters tuning, and the test dataset is used to evaluate the overall performance. Best performing training and validation results are stated in Table 5.

Figures 22A and 23A show that the validation dice coefficient and validation mean_iou are more significant than the training dice coefficient and training mean_iou. Validation dice coefficient and mean_iou are 0.9796, 0.9598. It shows that lung segmentation generated by U-Net + + and ground truth are almost similar. Figures 22B and 23B show that specificity increases steadily

with every epoch. Validation values of specificity and sensitivity are greater than training values. Value received during validation are 0.9932 and 0.9753, which shows that the redesigned skip pathway between the encoder and decoder Subpaths bridges the semantic path between the encoder and decoder which increased the optimization of the U-Net++. Figures 24 and 25 represented that validation accuracy and precision are near one and loss declines with every epoch, which indicates that U-Net + + is pretty accurate in generating segmented lungs.

Figure 26 represents the segmented lung segments generated by U-Net + +, ground truth, and the difference between ground truth and lung generated by U-Net + +. It can be noticed from Fig. 24a that U-Net + + generated pretty accurate lung segments.

4.5 Comparison Among models

In the section, the best training results of the four models are presented with various performance measures. Tables 6 and 7 show the training and validation results, respectively, and gives a fair idea about the superiority of U-Net + + implementation results.

The performance of the models used in this study is compared in this section. The authors trained four segmentation models on Shenzhen and Montgomery datasets and generated the results for all models: accuracy, precision, sensitivity, specificity, recall, precision, mean_iou, and dice_coefficient. Tables 6 and 7 compare the deep learning models trained and evaluated on the datasets. As segmented images play a vital role in the perfect diagnosis of the disease, Table 6 presents the performance for image segmentation. The algorithm that scores the best results in this study is U-Net + +. It can create segmented images with dice_coefficient 0.9796, mean_iou 0.9598, and accuracy 0.9874. The U-Net also scores acceptable results in this study and can segment lungs of dice_coefficient 0.9217, mean_iou 0.8572, and accuracy 0.9555, but SegNet and score are not satisfactory. All models in this study are based on encoder followed by decoder-type architecture, but U-Net + + performance has shown its best results because of redesigned skip pathways, dense skip connection, and deep supervision. So U-Net + + is the best performing model for chest X-ray images.

5 Conclusion and future scope

Segmentation is an important step to reduce the chance of data leakage and forces the classification architecture to focus only on essential areas and helps improve classification accuracy. The application of the segmentation technique has proven to be very helpful in the real world.

Table 5 U-Net + + best performing training and validation results

Model	Loss	Dice	Specificity	Mean_iou	Sensitivity	Recall	Precision	Accuracy
U-Net + + training	- 0.9630	0.9630	0.9858	0.9293	0.9585	0.9585	0.9174	0.9771
U-Net + + validation	- 0.9796	0.9796	0.9932	0.9598	0.9753	0.9838	0.9685	0.9874
Difference	- 0.0166	0.0166	0.0074	-0.0305	0.0168	0.028	0.0511	0.0103

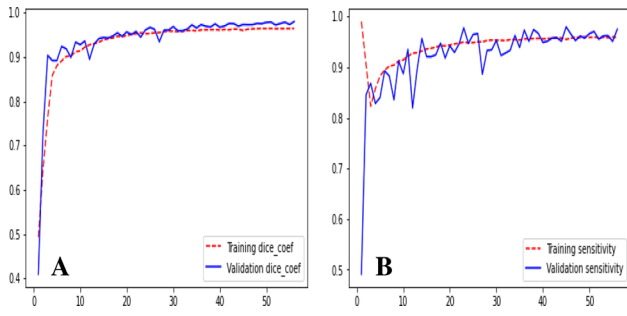


Fig. 22 Representation of training and validation coefficients (A), representation of training and validation sensitivity (B)

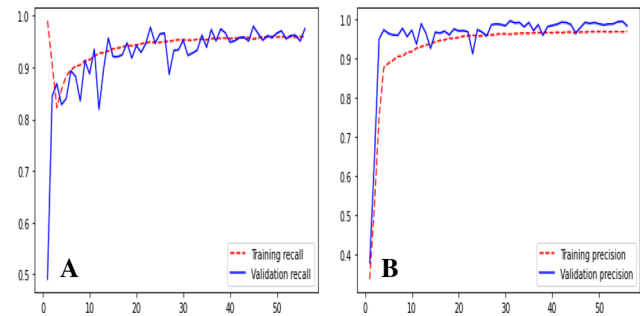


Fig. 25 Representation of training and validation recall (A), representation training and validation precision (B)

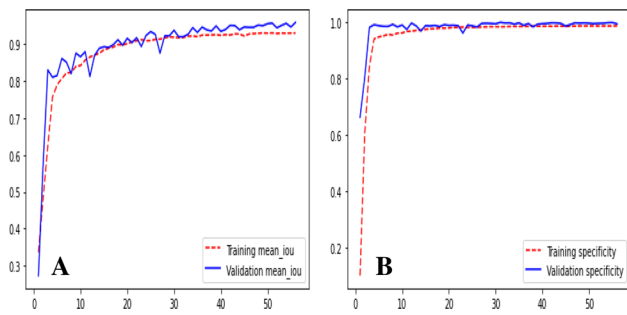


Fig. 23 Representation of training and validation mean_iou (A), representation of training and validation specificity (B)

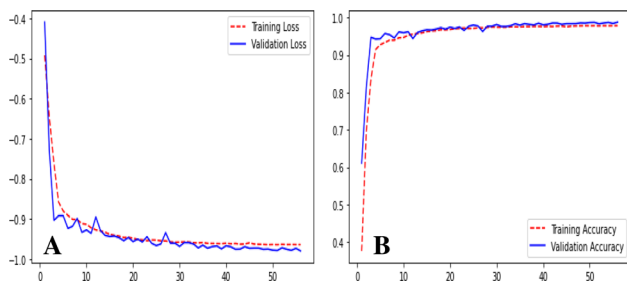


Fig. 24 Representation of training and validation loss (A), representation of training and validation accuracy (B)

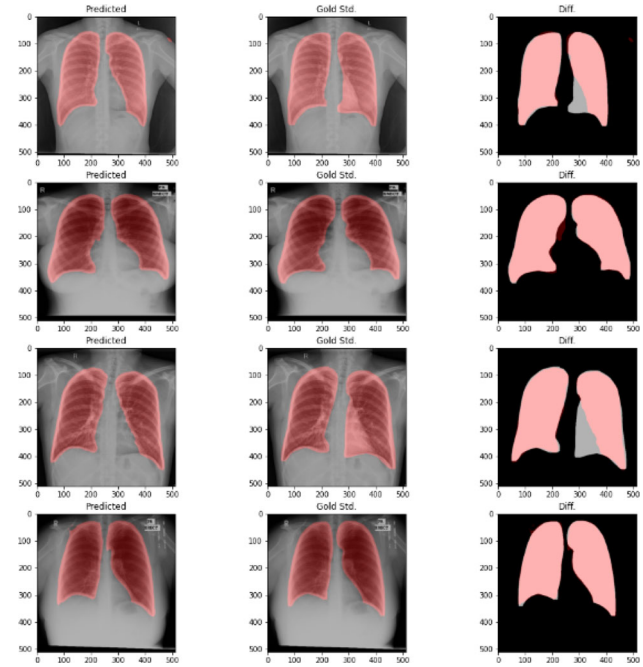


Fig. 26 Segmented lungs predicted by U-Net + + and ground truth comparison

The existing literature review section papers implemented machine learning and deep learning techniques for lung segmentation and got encouraging results. However, to the best of our knowledge, not even a single paper discussed U-Net + + and compared other segmentation techniques for lung segmentation. In this study, we studied four benchmark neural network architectures: U-Net, FCN,

SegNet, and U-Net + +, and the performance of these architectures is thoroughly explored and studied in this paper. The results generated by FCN, U-Net, SegNet, and U-Net + + are evaluated on Shenzhen and Montgomery datasets. Comparison between the results of four architectures shows that the U-Net + + surpasses other architectures by a considerable margin and achieved 98% accuracy because of its state-of-the-art architecture. FCN

Table 6 Best training results of models

Model	Loss	Dice	Specificity	Mean_iou	Sensitivity	Recall	Precision	Accuracy
FCN	– 0.6670	0.6670	0.7313	0.5018	0.8029	0.8029	0.5735	0.7536
SegNet	– 0.6841	0.6841	0.6509	0.5219	0.9258	0.9258	0.5462	0.7358
U-Net	– 0.8779	0.8779	0.9493	0.7852	0.8756	0.8756	0.8896	0.9256
U-Net + +	– 0.9630	0.9630	0.9858	0.9293	0.9585	0.9585	0.9685	0.9771

Table 7 Best validation results of models

Model	Loss	Dice	Specificity	Mean_iou	Sensitivity	Recall	Precision	Accuracy
FCN	– 0.6962	0.6962	0.5343	0.8380	0.8380	0.7609	0.5958	0.7832
SegNet	– 0.7914	0.7914	0.6558	0.9684	0.9684	0.7997	0.6701	0.8490
U-Net	– 0.9217	0.9217	0.8572	0.8904	0.8904	0.9837	0.9584	0.9555
U-Net + +	– 0.9796	0.9796	0.9598	0.9753	0.9838	0.9932	0.9685	0.9874

did not achieve satisfactory results with 78% and hence was not encouraged to explore further studies based on image segmentation.

The future work could be implementing other respiratory problems [e.g., chronic obstructive pulmonary diseases (COPD), pneumonia, etc.] using chest X-rays. Advanced feature extraction techniques with machine learning algorithms and the ensemble model localization scheme can be used to further downstream analysis, detect lung abnormality, and visualize explainable artificial intelligence (XAI) Grad-CAM.

Data availability It is confirmed by the authors that data supporting this research finding are present within the article, and the publicly available datasets used in this study are Montgomery County X-ray Set and Shenzhen Hospital X-ray Set [36].

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

- Global Tuberculosis Report (2019) WHO. World Health Organization, Switzerland
- World Health Organization (2016) Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches. World Health Organization, Geneva
- Brady AP (2017) Error and discrepancy in radiology: Inevitable or avoidable? *Insights Image* 8(1):171–182
- Degnan AJ, Ghobadi EH, Hardy P, Krupinski E, Scali EP, Stratchko L, Ulano A, Walker E, Wasnik AP, Auffermann WF (2019) Perceptual and interpretive error in diagnostic radiology—causes and potential solutions. *Acad Radiol* 26(6):833–845
- van Cleeff M, Kivihya-Ndugga L, Meme H, Odhiambo J, Klauser P (2005) The role and performance of chest X-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in Nairobi, Kenya. *BMC Infect Dis* 5(1):111
- Graham S, Gupta KD, Hidvegi JR, Hanson R, Kosiuk J, Zahrani KA, Menzies D (2002) Chest radiograph abnormalities associated with tuberculosis: reproducibility and yield of active cases. *Int J Tuberc Lung Dis* 6(2):137–142
- Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Noguees I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning. *IEEE Trans Med Imag* 35(5):1285–1298
- Ravishankar H et al (2016) Understanding the mechanisms of deep transfer learning for medical images. In: Carneiro G (ed) *Deep learning and data labeling for medical applications. DAL-MIA, LABELS (lecture notes in computer science)*, vol 10008. Springer, Cham. https://doi.org/10.1007/978-3-319-46976-8_20
- Zhou L, Ampon-Wireko S, Asante Antwi H et al (2020) An empirical study on the determinants of health care expenses in emerging economies. *BMC Health Serv Res* 20:774. <https://doi.org/10.1186/s12913-020-05414-z>
- Joseph B, Joseph M (2016) The health of the healthcare workers. *Indian J Occup Environ Med* 20(2):71–72. <https://doi.org/10.4103/0019-5278.197518>
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, pp. 1097–1105
- Rahman T, Khandakar A, Kadir M, Islam K, Islam K, Mazhar R, Hamid T, Islam M, Kashem S, Ayari M, Chowdhury M (2020) Reliable tuberculosis detection using chest x-ray with deep learning, segmentation, and visualization. *IEEE Access* 8:191586–191601. <https://doi.org/10.1109/ACCESS.2020.3031384>
- Tahir A, Qiblawey Y, Khandakar A, Rahman T, Khurshid U, Musharavati F, Islam MT, Kiranyaz S, Chowdhury MEH (2020) Coronavirus: comparing COVID-19, SARS and MERS in the eyes of AI. [Online]. Available: <http://arxiv.org/abs/2005.11524>
- Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *arXiv* 2020, <http://arxiv.org/abs/2003.10849>
- Chhikara P, Singh P, Gupta P, Bhatia T (2020) Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays. In: Jain L, Virvou M, Piuri V, Balas V (eds) *Advances in bioinformatics, multimedia, and electronics circuits and signals (advances in intelligent systems and computing)*, vol 1064. Springer, Singapore. https://doi.org/10.1007/978-981-15-0339-9_13
- Zak M, Krzyzak A (2020) Classification of lung diseases using deep learning models. In: Krzhizhanovskaya VV (ed)

- Computational science—ICCS 2020. ICCS 2020. Lecture notes in computer science, vol 12139. Springer, Cham. https://doi.org/10.1007/978-3-030-50420-5_47
17. Hooda R, Mittal A, Sofat S, Automated TB (2019) classification using an ensemble of deep architectures. *Multimed Tools Appl* 78:31515–31532
 18. Melendez J, Sánchez CI, Philipsen RH, Maduskar P, van Ginneken B (2014) 'Multiple-instance learning for computer-aided detection of tuberculosis. *Proc. SPIE*, vol. 9035
 19. Hooda R, Sofat S, Kaur S, Mittal A, Meriaudeau F (2017) Deep-learning: a potential method for tuberculosis detection using chest radiography. In: *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Kuching, Malaysia, pp. 497–502
 20. Rohilla A, Hooda R, Mittal A (2017) Tb detection in chest radiograph using deep learning architecture. *Proc ICETETSM* 6(8):1073–1085
 21. Evangelista LGC, Guedes EB (2018) Computer-aided tuberculosis detection from chest X-ray images with convolutional neural networks. In: *Proc. Anais do XV Encontro Nacional de Inteligência Artif. e Computacional (ENIAC)*, pp. 518–527
 22. Pasa F, Volkov V, Pfeiffer F, Cremers D, Pfeiffer D (2019) Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. *Sci Rep* 9(1):1–9
 23. Nguyen QH, Nguyen BP, Dao SD, Unnikrishnan B, Dhingra R, Ravichandran SR, Satpathy S, Raja PN, Chua MCH (2019) Deep learning models for tuberculosis detection from chest X-ray images. In: *Proc. 26th Int. Conf. Telecommun. (ICT)*, pp. 381–386
 24. Lopes UK, Valiati JF (2017) Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Comput Biol Med* 89:135–143
 25. Meraj SS, Yaakob R, Azman A, Rum SN, Shahrel A, Nazri A, Zakaria NF (2019) Detection of pulmonary tuberculosis manifestation in chest X-rays using different convolutional neural network (CNN) models. *Int J Eng Adv Technol (IJEAT)* 9(1):2270–2275
 26. Ahsan M, Gomes R, Denton A (2019) Application of a convolutional neural network using transfer learning for tuberculosis detection. In: *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, pp. 427–433
 27. Yadav O, Passi K, Jain CK (2018) Using deep learning to classify X-ray images of potential tuberculosis patients. In: *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, pp. 2368–2375
 28. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G et al (2019) End-to-end lung cancer screening with deep three-dimensional learning on low-dose chest computed tomography. *Nat Med* 25:954–961
 29. Ausawalaithong W, Thirach A, Marukatat S, Wilaiprasitporn T (2018) Automatic lung cancer prediction from chest X-ray images using the deep learning approach. In: *Proceedings of the 2018 11th biomedical engineering international conference (BMEi-CON)*, Chiang Mai, Thailand, pp. 21–24
 30. Hua KL, Hsu CH, Hidayati SC, Cheng WH, Chen YJ (2015) Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Oncotargets Ther* 8:2015–2022
 31. Islam MZ, Islam MM, Asraf A (2020) A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Inf Med Unlocked* 20:100412
 32. Kermany D, Goldbaum M (2018) Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley Data* 2
 33. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) *Medical image computing and computer-assisted intervention—MICCAI 2015*. MICCAI 2015. Lecture notes in computer science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
 34. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J (2018) UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov D (ed) *Deep learning in medical image analysis and multimodal learning for clinical decision support*. DLMIA 2018, ML-CDS 2018. Lecture notes in computer science, vol 11045. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_1
 35. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
 36. Jaeger S, Candemir S, Antani S, Wang Y-X, Pu-Xuan Lu (2014) Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg* 4(6):475–477. <https://doi.org/10.3978/j.issn.2223-4292.2014.11.20>
 37. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. *CVPR*. <https://doi.org/10.1109/CVPR.2015.7298965>
 38. Guo R, Passi K, Jain CK (2020) Tuberculosis diagnostics and localization in chest X-rays via deep learning models. *Front Artif Intell* 3:583427. <https://doi.org/10.3389/fraci.2020.583427>
 39. Ting P, Guo S, Zhang X, Zhao L (2019) Automatic lung segmentation based on texture and deep features of HRCT images with interstitial lung disease. *Biomed Res Int*. <https://doi.org/10.1155/2019/2045432>
 40. Zhao C, Xu Y, He Z, Tang J, Zhang Y, Han J, Shi Y, Zhou W (2021) Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images. *Pattern Recognit* 119:108071. <https://doi.org/10.1016/j.patcog.2021.108071>
 41. Monshi MM, Poon J, Chung V, Monshi FM (2021) CovidXrayNet: optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. *Comput Biol Med*. <https://doi.org/10.1016/j.compbmed.2021.104375>
 42. Diniz JOB, Quintanilha DBP, Santos Neto AC et al (2021) Segmentation and quantification of COVID-19 infections in CT using pulmonary vessels extraction and deep learning. *Multimed Tools Appl* 80:29367–29399. <https://doi.org/10.1007/s11042-021-11153-y>
 43. Cao F, Zhao H (2021) Automatic lung segmentation algorithm on chest X-ray images based on fusion variational auto-encoder and three-terminal attention mechanism. *Symmetry* 13(5):814. <https://doi.org/10.3390/sym13050814>

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