Chapter 3

Deep Learning Models for Chest X-ray Image Segmentation

Abstract

Artificial Intelligence and CAD systems are becoming highly popular in medical diagnosis. The application of AI and deep learning in radiology is revolutionizing the healthcare sector with fast and accurate diagnosis. The Chest X-ray is one of the most significant radiological diagnostic methods being used for its easy availability, cost effectivity, and low radiation doses. The application of deep learning methods in chest X-rays has shown tremendous success in lesion detection. However, the chest-X ray contains a large non-region of interest in the form of the background that interrupts the AI system for accurate lesion detection. Towards the motive of solving the problem, this work proposes robust and accurate deep learning-based UNet and UNet + segmentation models to segment the region of interest, i.e., the lung region, and remove the background present in the X-ray images. Our model can successfully and accurately segment chest X-ray images. The UNet model performed best with an accuracy of 96.35% with dice coefficient and Jaccard index of 94.88% and 90.38%, respectively. Performing with high accuracy, dice, and Jaccard, our system proves its efficacy and robustness for efficiently segmenting the lung region to diagnose further numerous lung diseases, including COVID-19 and other pneumonia.

3.1 Introduction

In recent years, AI-CAD (Artificial Intelligence-based Computer-aided diagnosis) systems have been widely used in the medical field to diagnose multiple diseases [14]. The AI-CAD system shows accurate diagnoses with fast results and easily accessible methods [134]. The system is becoming popular and growing exponentially in developing countries where the lack of radiologists and rush in hospitals are major challenges. Several radiological assessments such as MRI, Ultrasound, CT scan, mammograms, and X-rays are widely used for tumor detection [98, 135], cardiovascular disease diagnosis [102, 103, 136], Plaque detection [137, 138], pneumonia detection [39, 139-141], Parkinson's disease diagnosis [142] and much more. Comparing the radiological techniques, MRI and CT are costly and have higher radiation doses, whereas X-rays are comparatively cost-effective, easily available, and have low radiation doses [143]. Also, the X-ray provides quick and instant results compared to CT and MRI. Therefore the advantages of X-rays over CT and MRI make them more popular for medical imaging, AI, and Computer-based diagnosis, especially for pneumonia detection [144]. Traditionally, Chest X-rays (CXRs) are widely used to diagnose pneumonia, tuberculosis, and other lung infections. The CXR has also been successfully used for AI-derived deep learning approaches based detection of several diseases such as COVID-19, tuberculosis, viral/bacterial pneumonia, lung tumors, and much more [145]. However, the CXR also contains a large area of non-region of interest that may interrupt the AI models to diagnose diseases accurately [146]. Therefore, it is highly recommended to separate the lung area from the whole CXR before implementing the AI-based classification or detection approach [50]. In this work, we have presented AI-based deep learning models: UNet and UNet+, for the automatic segmentation of CXR images to segment the lung region for further detection of COVID-19, pneumonia, and other diseases infecting the lung.

3.2 Related works

Previously, several researchers have introduced and implemented different methods to segment chest X-ray images. Many of them have reported significant outcomes. Candemir et al. [147] presented a nonrigid registration-based segmentation model. The system uses an image retrieval-driven patient-specific adaptive model to detect the lung boundaries in chest X-ray images. They demonstrated an accuracy of 95.4% on

the JSRT dataset and 94.1% on the Montgomery dataset. Ngo et al. [43] implemented the blend of a distance-regulated level set and deep belief system for the segmentation of CXR images taken from the JSRT dataset. Their model demonstrated an accuracy of 96.5%. Mittal et al. [45] applied an encoder decoder-based deep neural network for the segmentation of the CXR images taken from the JSRT and Montgomery datasets. Their model demonstrated an accuracy and Jaccard index of 98.73% and 95.10%, respectively. Hooda et al. [42] formulated a new CNN-based deep neural network to segment the CXRs from the JSRT dataset. They got an accuracy and Jaccard index of 98.92% and 95.88%, respectively. Saidy et al. [44] implemented an encoder-decoder approach to derive deep CNN and segmented the CXRs from the JSRT dataset. They got the test results with a Dice coefficient of 96%. Gaal et al. formulated a new neural network and implemented it to the JSRT dataset. Their model performed with a Dice coefficient of 97.5%. Zhang et al. [49] implemented a modified UNet model with dual encoders for the segmentation of the Shenzhen and Montgomery datasets. Their system demonstrated an accuracy of 98.04%. The model also revealed a Dice of 96.67% and an AUC of 0.98. Liu et al. [148] formulated a modified UNet by using pre-trained EfficientNetB4 as the encoder and LeakyReLU activation function in the decoder part. They tested the network on JSRT and Montgomery datasets separately. They achieved the accuracy, dice, and Jaccard of 98.55%, 97.92%, and 95.73% on JSRT and 98.94%, 97.82%, and 95.55% on Montgomery datasets, respectively. Chandra et al. [149] proposed a multistage superpixel classification-based method for disease localization and severity detection in CXR images. The method was tested on the Montgomery dataset. Their system performed with an average Jaccard Index of 82% and Pearson's correlation coefficient of 0.95. Chandra et al. [150] presented a context-aware-adaptive scan algorithm to scan and correct the artifacts present along inner lung boundaries in CXRs. The algorithm was tested on the Montgomery dataset with a significant average segmentation accuracy improvement of 2.5%.

3.3 Methodology

Figure 3.1 represents the step-wise schematic diagram showing the overall methodology opted in the study. First, we utilized the "Chest X-ray masks and labeled" dataset [151] for the training of the segmentation model that was UNet. We utilized the images for our experiment in a manner of 70:20:10 ratio. 70% for

training, 20% for validation, and 10% for testing the model. We implemented a five-fold cross-validation approach for model training. After training, we evaluated our model using testing accuracy, dice, Jaccard, AUC (Area-under-the-curve), and ROC (Receiver operating characteristic). Thereafter we applied the model for the segmentation of the "COVID-19 radiography database," [114] having a collection of >18000 CXR images to test the efficacy and feasibility of the model on larger databases.

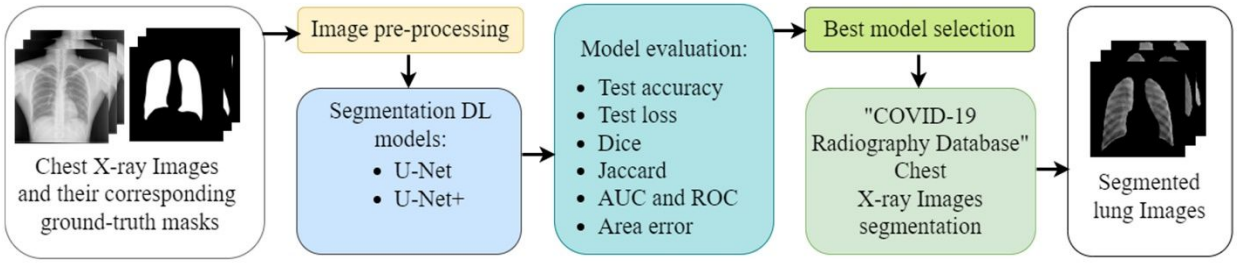

Figure 3.1: The step-wise schematic diagram showing the overall methodology.

3.4 Dataset utilized in the experiment

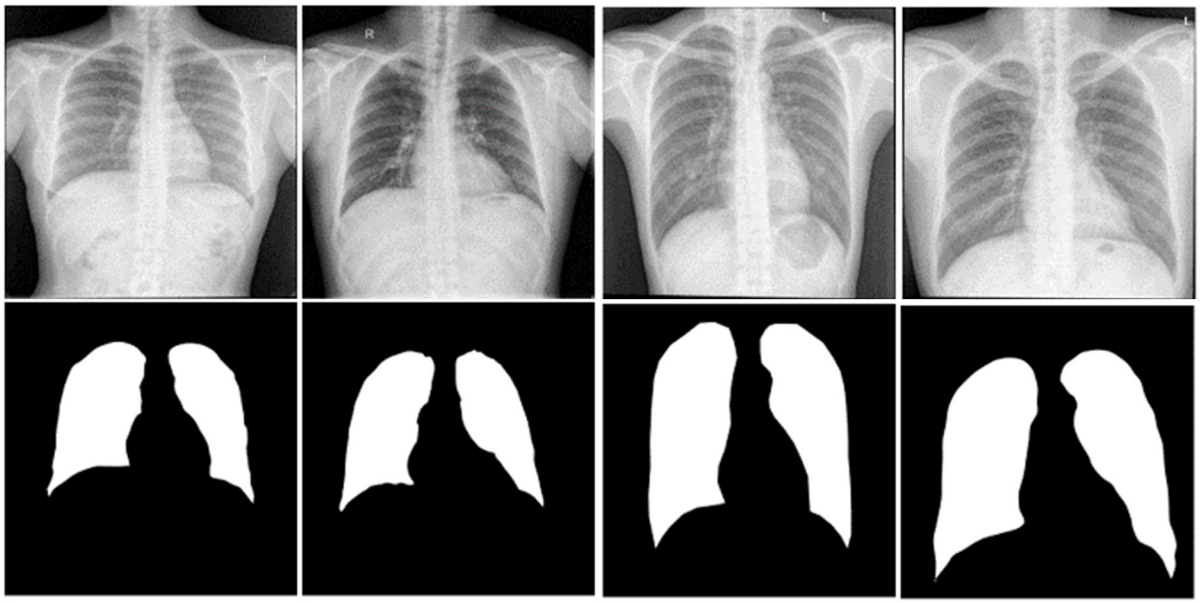

Figure 3.2: Sample CXR images (top row) and their ground truth masks (bottom row).

The Kaggle dataset named 'Chest X-ray Masks and Labels' [151] has been used to train the segmentation models in this work. The dataset contains 704 chest X-ray images and their corresponding ground truth masks. A team of expert radiologists annotated each mask. The data source is the National Library of Medicine, NIH, USA, and Shenzhen No.3 People's Hospital, GMC, Shenzhen, China. The dataset contains

360 normal chest X-rays and 344 infected chest X-ray images. Figure 3.2 shows sample chest X-ray images and their corresponding masks.

The second dataset utilized to test the performance of our model on larger datasets was the "COVID-19 radiography database". [38, 55] The dataset contains >18000 CXRs in COVID-19, pneumonia, and normal classes.

3.5 The architecture of segmentation networks

Two deep neural network models, namely UNet and UNet+, were applied for our first experimental phase, i.e., segmentation of chest X-ray images.

3.5.1 UNet architecture

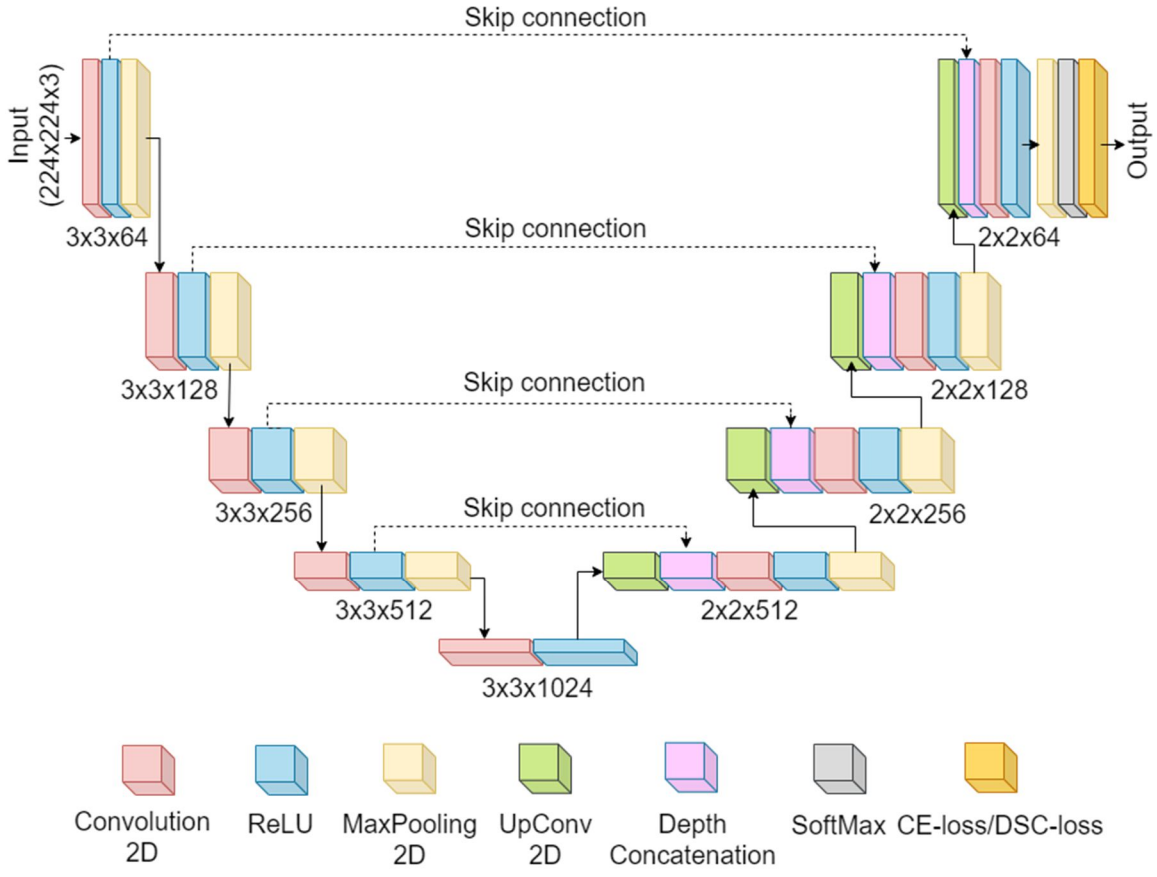

Figure 3.3: UNet architecture.

UNet is the most popular convolutional neural network for segmentation. It was proposed by Ronneberger et al. [152]. The network applies the idea of deconvolution, which was introduced by Zeiler et al. [153].

Figure 3.3 represents the UNet architecture. It consists of a blend of encoder-decoder stages [154]. The encoder encompasses a combination of convolutional layers followed by the ReLU and Maxpooling. The encoder has a 3 × 3 convolution with a MaxPooling that downsamples the images to the next stage and finally to the bridge network. The bridge network is at the bottom of the U-shaped network, connecting the encoder with the decoder. The bridge network has 3×3×1024 filters and a ReLU layer. Next to the bridge stage, the decoder functions by up-sampling the images. The decoder comprises up-convolution, convolution, ReLU, and MaxPooling layers. Each decoder stage has 2 × 2 convolutional filters. The spatial features from the encoder stage are delivered to the corresponding decoder stage through a skip connection. The spatial features are transferred from the encoder to the decoder. After the fourth decoder stage, the ADAM optimizer reduces the loss. Finally, an efficient classifier, Softmax, classifies the up-sampled features into two classes: the lung area and the background.

3.5.2 UNet+ architecture

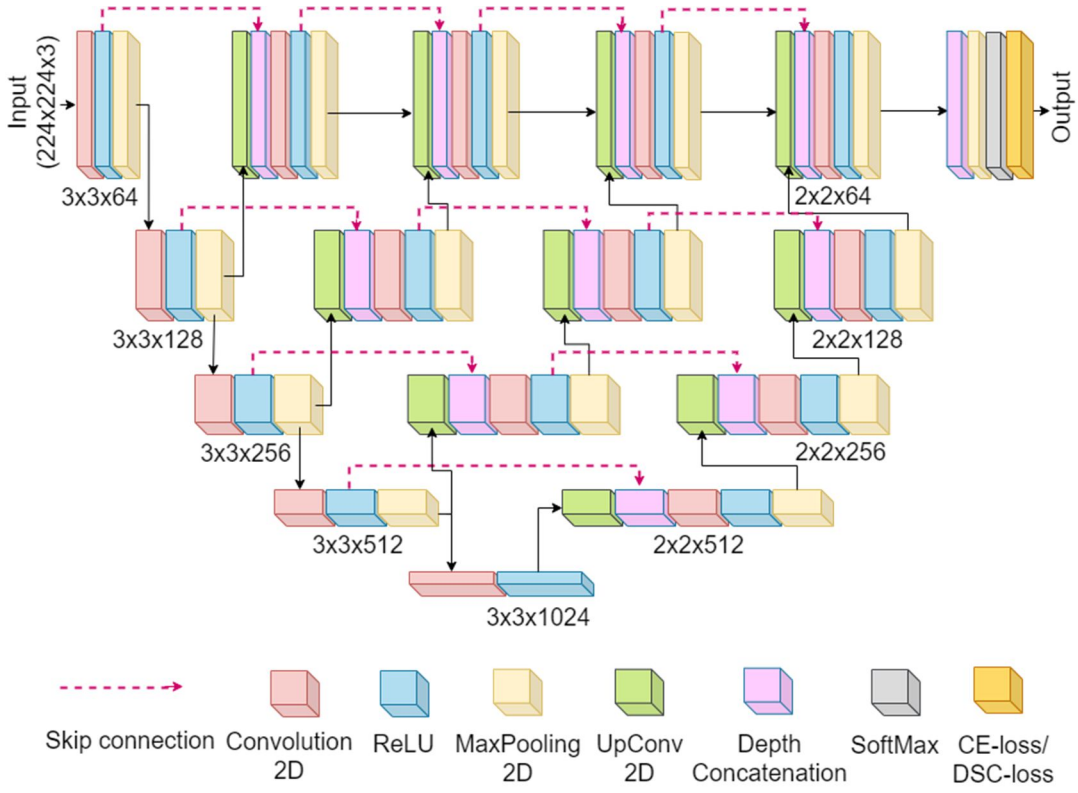

Figure 3.4: UNet+ architecture.

The UNet+ is a modified version of the UNet network. Figure 3.4 represents the UNet+ model. The UNet+ model differs from the original UNet by having a few intermediate encoder stages between compression and expansion. The first intermediate stage has three encoder stages, the second has two, and the third has one intermediate stage. The several intermediate up-sampling units with varying depths in the UNet+ model have overcome the limitation of optimal depth in the UNet encoder-decoder network. All intermediate up-sampling units are connected to the decoder stage with the exact resolution by reformed skip connections. Finally, after the fourth decoder stage, similar to the UNet, the ADAM optimizer reduces the loss, and the Softmax classifies the up-sampled features into two classes: the lung area and the background.

3.6 Experimental protocols

3.6.1 Cross-validation

A total of 704 CXR images and their 704 corresponding masks were used for the segmentation experiment. The K5 data partitioning method was implemented. The 5-fold cross-validation was done using 60%, i.e., 408 images for training, 20%, i.e., 148 images for validation, and 20%, i.e., 148 images for testing in each fold. After each fold's training and validation, testing was done on 148 new images that were not used in training or validation. The average test results for each fold were calculated to get the performance analysis, including the test accuracy and loss. Also, the mask was generated for images of the test set using each model developed by training on each fold's images. Next, all of the predicted masks from each fold's test images were compared with their ground truth masks to see how well they worked. This was done by generating the Dice, Jaccard, area error, Bland Altman plot, coefficient of correlation, and ROC.

3.6.2 Training parameters

Both the UNet and UNet+ models were trained for 100 epochs with a learning rate of 0.001, a dropout rate of 0.25, and a batch size of 4 images. The loss function used for training the model was the CE loss function, denoted by L_{ce} and mathematically represented as:

$$L_{CE} = [(y_i \times log a_i) + (1 - y_i) \times log(1 - a_i)] \qquad \text{(i)}$$

Here, y_i is the input GT label 1, (1-y_i) is GT label 0, a_i represents the Softmax classifier probability.

The entire experiment was carried out using Python 3.8. For training the network, we employed a workstation with an 8GB NVIDIA Quadro P4000 GPU (Graphics Processing Unit). The system had an Intel Core i7 8th Generation processor and 16GB of RAM.

3.7 Performance Evaluation Metrics

The performance of each network for image segmentation was evaluated on test data after the training and validation process. The following different matrices were utilized for the performance evaluation naming: accuracy, loss, Jaccard index, Dice coefficient, area error, and area-under-the-curve (AUC). The mathematical representations for the matrices are given in the equation below:

$$\text{Accuracy} = \frac{\text{TP+TN}}{(\text{TP} + \text{FN}) + (\text{FP} + \text{TN})} \qquad (3.1)$$

$$\text{Jaccard index} = \frac{\text{TP}}{(\text{TP} + \text{FN+FP})} \qquad (3.2)$$

$$\text{Dice Coefficient} = \frac{(2*\text{TP})}{(2*\text{TP} + \text{FN+FP})} \qquad (3.3)$$

3.8 Results

Figure 3.5 shows the masks generated by the UNet and UNet+ models and their comparison to ground truth masks. In addition, the comparative performances of both segmentation models are shown in Table 3.1. The values of the performance matrices are the average of results generated for test data of each fold and by each corresponding model from five folds. The UNet model performed with 96.35% accuracy, 0.15% test loss, 94.88% dice coefficient, 90.38% Jaccard index, 1.48 mm^2 area error, and 0.99 AUC with $p<0.001$. The UNet+ model performed with a test accuracy of 96.10%, a test loss of 0.17%, a dice coefficient of 92.35%, a Jaccard index of 86.07%, an area error of 2.63 mm^2, and an AUC of 0.98 with $p<0.001$.

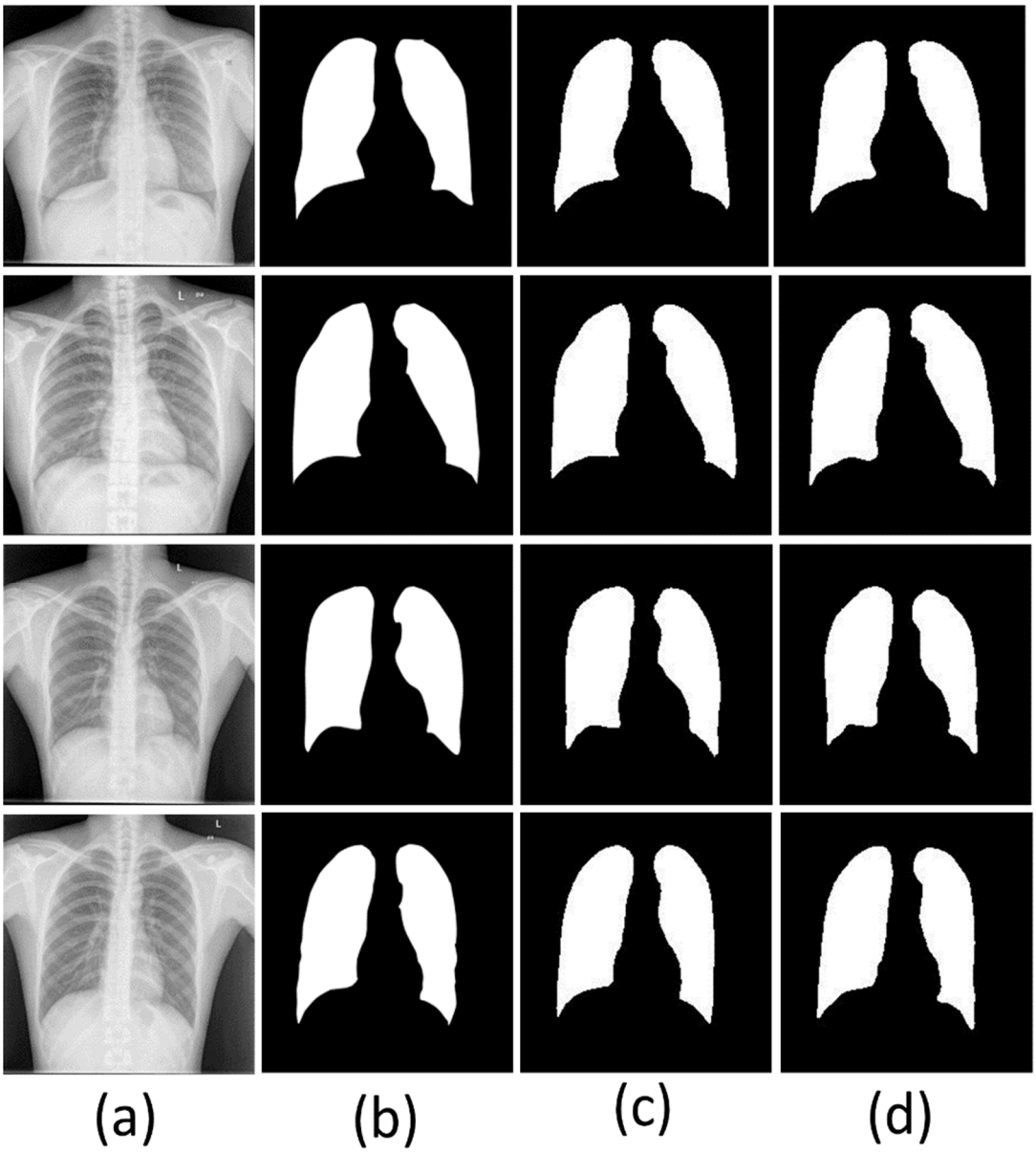

Figure 3.5: Comparison of results by two segmentation models: (a) Original CXR images, (b) Ground truth masks, (c) Masks generated by UNet model, (d) Masks generated by UNet+ model.

Table 3.1: Comparative performance of UNet and UNet+ model.

Model	Test accuracy (%)	Test loss	Dice (%)	Jaccard (%)	Area error (mm^2)	AUC (p-value)
UNet	96.35	0.15	94.88	90.38	1.48	0.99 ($p<0.001$)
UNet+	96.10	0.17	92.35	86.07	2.63	0.98 ($p<0.001$)

3.8.1 Cumulative Frequency Curves for Dice and Jaccard

The dice coefficient, or F1-score, and the Jaccard index, or IoU (area of overlap), are the most important metrics to evaluate the segmentation. The Dice coefficient is double the area of overlap between AI (predicted mask) and GT (ground truth mask) divided by the total number of pixels in both images. The Jaccard index is the area of overlap between AI and GT divided by the area of union between AI and GT. The Dice and Jaccard are very similar, and both are positively correlated with each other. Figure 3.6 shows the Cumulative Frequency Curves of Dice and Jaccard for both the UNet and UNet+ models. For the UNet model, 80% of scans had Dice and Jaccard >0.96 and >0.93, respectively, whereas, for the UNet+ model, 80% of scans had Dice and Jaccard are >0.95 and >0.91, respectively. Thus, the UNet model showed better performance in terms of Dice and Jaccard than the UNet+ model.

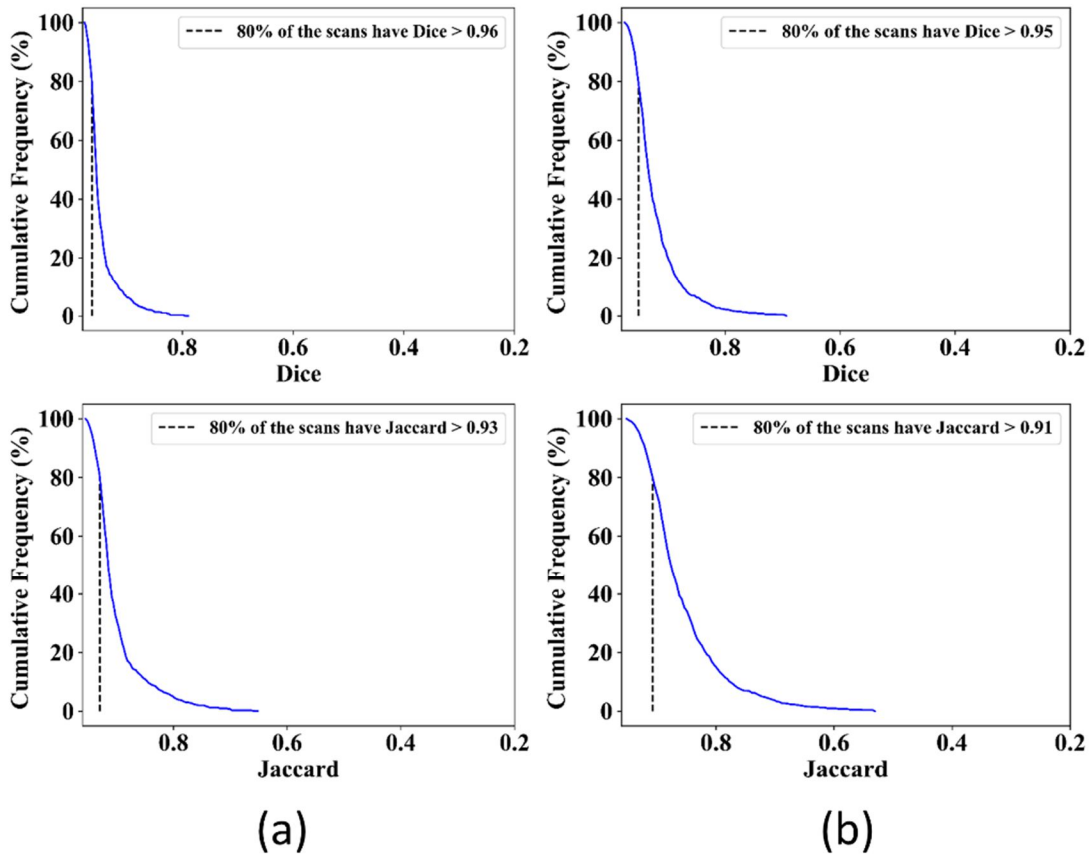

Figure 3.6: Cumulative Frequency Curves showing Dice (top) and Jaccard (bottom) for (a): UNet model, (b): UNet+ model.

3.8.2 Receiver Operating Curve and AUC analysis

The Receiver Operating Curve (ROC) is the graphical plot of sensitivity against the (1- specificity). A higher AUC indicates better performance. Figure 3.7 shows the ROC and AUC for the UNet and UNet+ models. The AUC performance by the UNet was 0.99, whereas the UNet+ was 0.98. Thus, the UNet model shows a better ROC curve with a higher AUC value by 1% than the UNet+ model.

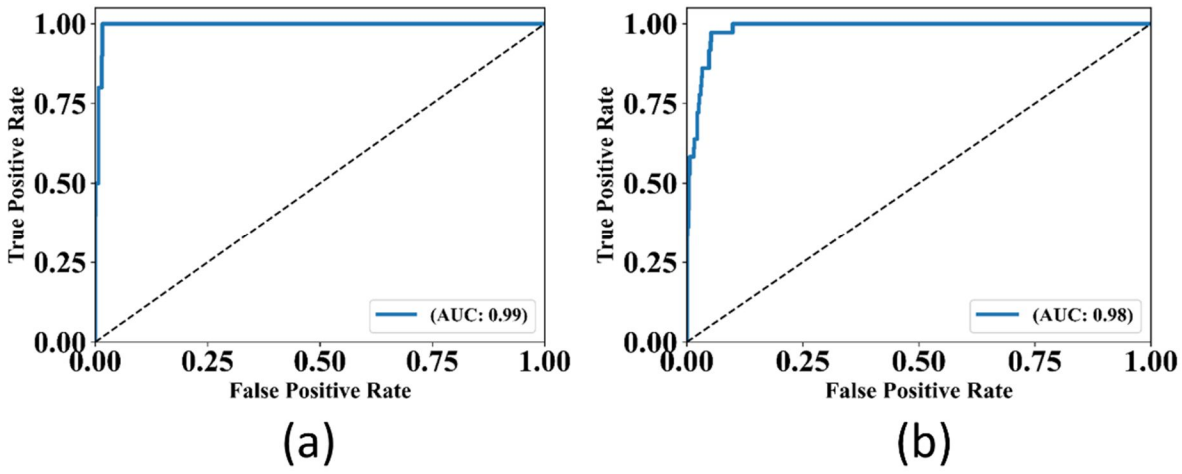

Figure 3.7: AUC and ROC Curves for (a): UNet model, (b): UNet+ model (p<0.0001).

3.8.3 Correlation analysis between AI and GT

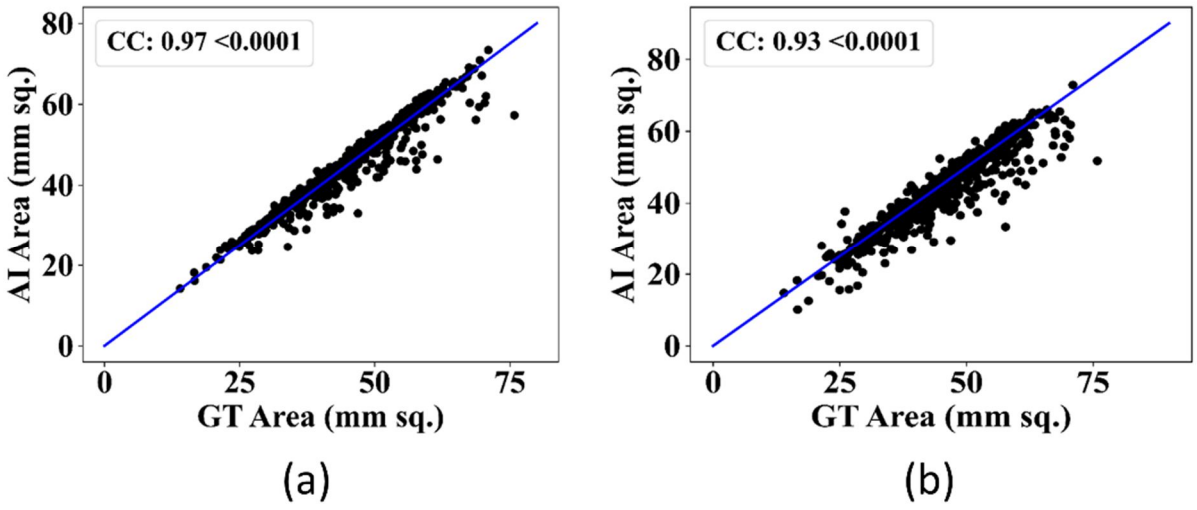

Figure 3.8: CC for GT and AI-estimated area for (a): UNet model, (b): UNet+ model.

The regression curve is a prevailing method to find a correlation between two measures. The Correlation coefficient (CC) signifies the relationship between the two measures. The higher CC value denotes a better model performance. Figure 3.8 shows the CC between AI-estimated and GT area for both models, i.e., UNet and UNet+. The CC value for the UNet model was 0.97, whereas the CC value for the UNet+ model was 0.93. The UNet model showed better performance by 0.04 CC than the UNet+ model.

3.8.4 Bland-Altman Plot for AI and GT area

The Bland-Altman plot denotes the difference between the AI and GT areas along the y-axis and the mean of AI and GT areas along the x-axis. The less the mean and SD (standard deviation) values show, the better the performance. Figure 3.9 shows the Bland-Altman plots for AI-estimated and GT area for both the UNet and UNet+ models. The mean and SD values for UNet were 0.08 mm^2 and 2.68 mm^2, respectively. In contrast, the mean and SD values for the UNet+ model were 1.60 mm^2 and 3.78 mm^2, respectively. So, the UNet model performs better than UNet+ by 1.52 mm^2 and 1.1 mm^2 in terms of mean and SD, respectively.

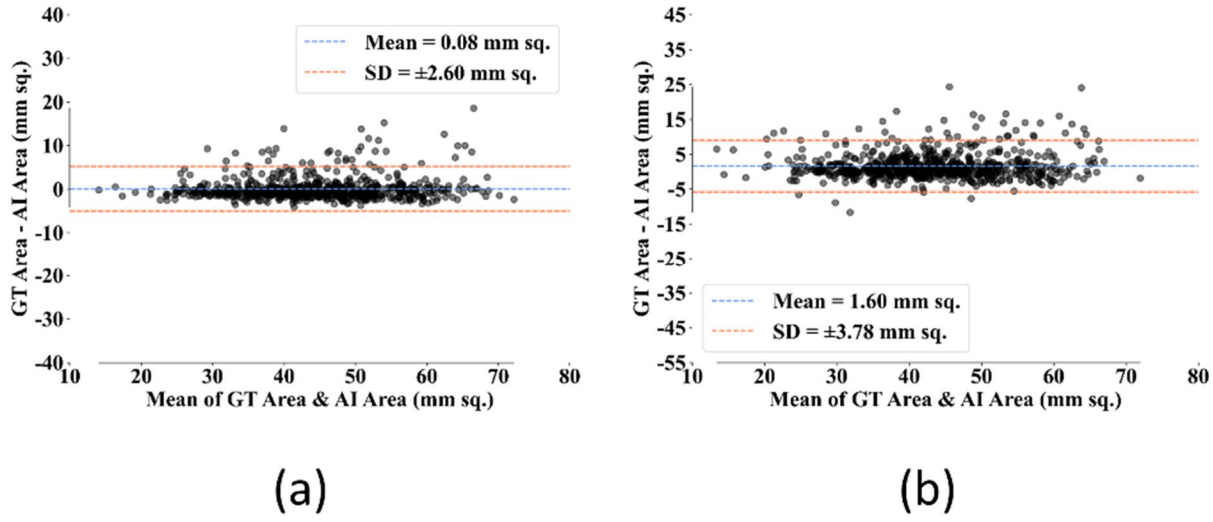

Figure 3.9: Bland-Altman plots for AI-estimated and GT area for (a): UNet model, (b): UNet+ model.

3.8.5 Cumulative distribution curves for area error between AI and GT

The area error is one of the other metrics used to determine the model's performance. The area error is the difference between the area of AI and GT in mm^2. The area error is calculated by converting the area of the

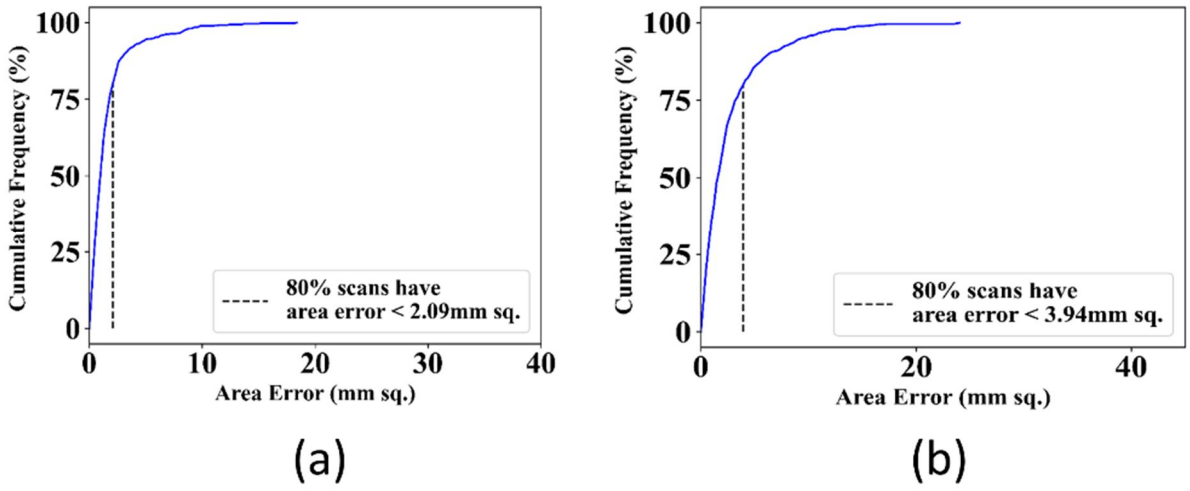

Figure 3.10: Cumulative distribution curves for area error between GT and AI-estimated masks by (a): UNet model, (b): UNet+ model.

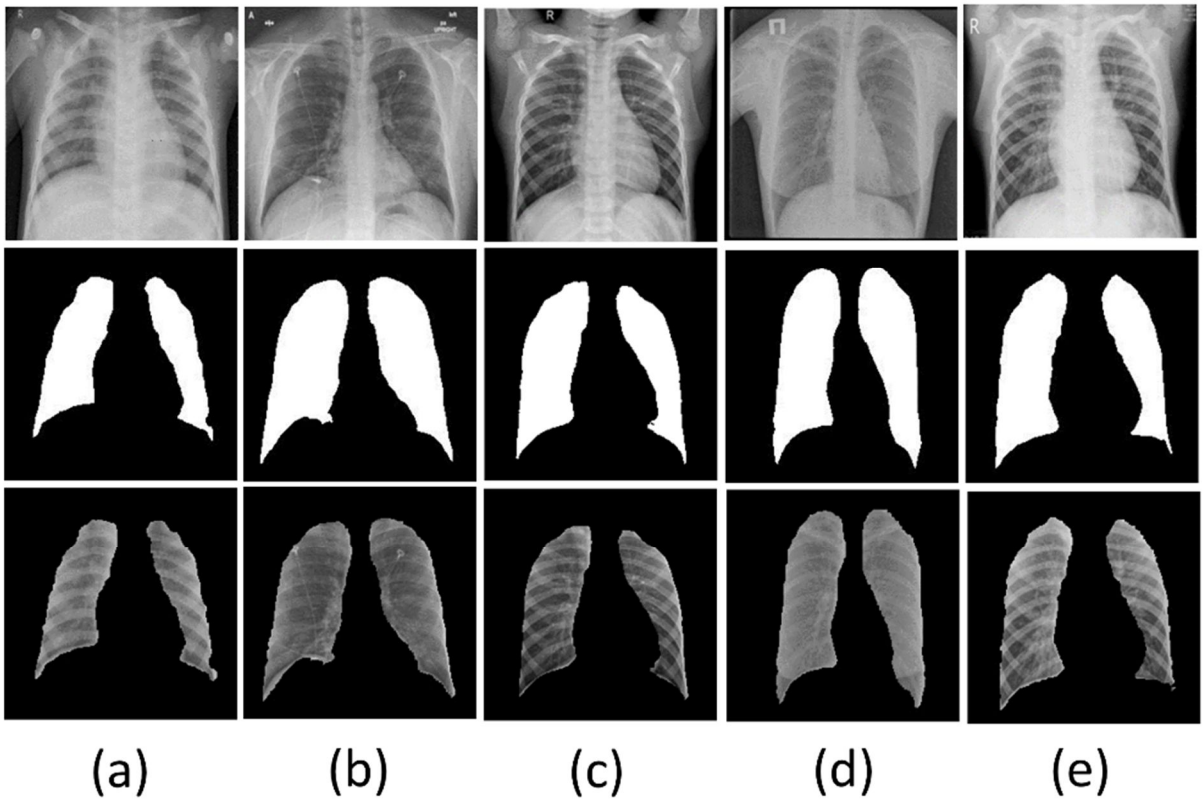

Figure 3.11: Example of images from classes (a): Bacterial Pneumonia, (b): COVID-19, (c): Normal, (d): Tuberculosis, (e): Viral Pneumonia; top row: original chest X-ray images, middle row: UNet generated corresponding masks, bottom row: final segmented lung images.

predicted and ground truth mask from pixel to mm dimensions and applying a resolution factor of 0.0625 mm to a pixel. A Lower error denotes better performance. Figure 3.10 shows the cumulative distribution

curves for area error between GT and AI-estimated masks for both the UNet and UNet+ models. 80% of scans had an area error < 2.09 mm^2 for the UNet model, whereas 80% of scans had an area error < 3.94 mm^2 for the UNet+ model. Therefore, the UNet model performed better with less area error of 1.85 mm^2 than the UNet+ model.

3.8.6 Segmentation of the classification dataset

The overall results analysis claims that the UNet model performed better than the UNet+ model in each parameter on our dataset. Therefore, we selected the UNet model for the further segmentation of our classification data. Figure 3.11 shows the sample of segmented CXR images from the five class classification data by the UNet model.

3.8.7 Benchmarking

Table 3.2 shows a comparison of our segmentation model to the existing state-of-the-art segmentation methods. Hooda et al. [42] applied a novel deep CNN on the JSRT CXR dataset and achieved an accuracy of 98.92% with a Jaccard index of 95.88%. Ngo et al. [43] applied a combination of Distance Regularized Level Set and Deep Belief Network to segment the JSRT dataset and achieved an accuracy of 96.5%. Saidy et al. [44] also utilized the JSRT dataset for an encoder-decoder-based segmentation model development and achieved a dice coefficient of 96% on the test dataset. Mittal et al. [45] utilized the combination of JSRT and Montgomery CXR datasets for an encoder-decoder-based segmentation model and achieved an accuracy of 98.73% and a Jaccard index of 95.10%. Reamarron et al. [47] applied the total variation-based active contour method for the segmentation and used a combination of the JSRT and Montgomery datasets. The model achieved a dice of 89%. Gaal et al. [51] developed a novel segmentation method and applied it to the JSRT dataset. They got a dice coefficient of 97.5%. Munawar et al. [48] utilized three datasets: JSRT, Montgomery, and Shenzhen, for the training of the Generative Adversarial Network and achieved a dice coefficient of 97.4%. Zhang et al. [49] applied the Dual Encoder Fusion UNet model on a combination of Montgomery and Shenzhen datasets and achieved an accuracy of 98.04% with dice and an AUC of 96.67% and 0.98, respectively. Teixeira et al. [50] applied the UNet model on the combination of five datasets,

namely Cohen, JSRT, Montgomery, Shenzhen, and a private dataset. They achieved a dice coefficient of 98.2%. Souza et al. [46] applied a combination of AlexNet and ResNet-based CNN segmentation models on the Montgomery dataset and achieved the accuracy, dice, and Jaccard of 96.67%, 93.56%, and 88.07%, respectively.

Table 3.2: Benchmarking table showing comparison of proposed and existing segmentation models.

Author & Year	Dataset (chest X-ray)	Technique	Accuracy	Dice	Jaccard	AUC
Hooda et al. (2018) [42]	JSRT	New deep CNN	98.92%	N.A.	95.88%	N.A.
Ngo et al. (2015) [43]	JSRT	DRLS (Distance Regularized Level Set) + DBN(Deep Belief Network)	96.5%	N.A.	N.A.	N.A.
Saidy et al. (2018) [44]	JSRT	Encoder-decoder neural network	N.A.	96%	N.A.	N.A.
Mittal et al. (2018) [45]	JSRT+Montgomery	Encoder-decoder neural network	98.73%	N.A.	95.10%	N.A.
Reamarron et al. (2020) [47]	JSRT+Montgomery	TVAC(Total Variation-based Active Contour)	N.A.	89%	N.A.	N.A.
Gaal et al. (2020) [51]	JSRT	New deep CNN	N.A.	97.5%	N.A.	N.A.
Munawar et al. (2020) [48]	JSRT+Montgomery+ Shenzhen	GAN (Generative Adversarial Networks)	N.A.	97.4%	N.A.	N.A.
Zhang et al. (2021) [49]	Montgomery+ Shenzhen	DEFUNet(Dual Encoder Fusion UNet)	98.04%	96.67%	N.A.	0.98
Teixeira et al. (2021) [50]	Cohen v7labs+ JSRT+Montgomery+ Shenzhen+Private	UNet	N.A.	98.2%	N.A.	N.A.
Souza et al. (2019) [46]	Montgomery	AlexNet+ResNet based CNN	96.97%	93.56%	88.07%	N.A.
Proposed [155]	Chest X-Ray Masks and Labels (kaggle dataset)	UNet	96.35%	94.88%	90.38	0.99

In the proposed segmentation method, we utilized a Kaggle dataset naming: Chest X-ray Masks and Labels. The dataset contains 704 CXR images and their corresponding masks. We applied to the UNet network for training. The model performed with a test accuracy, Dice, Jaccard, and AUC of 96.35%,

94.88%, 90.38%, and 0.99, respectively. Our model performed best in terms of AUC score. In addition, most of the other works utilized JSRT or Montgomery datasets with a deficient number of images, such as 247 and 138, respectively, which may also be why some of them have higher accuracy than us. However, we have used a large number of images that make our model more stable and robust.

3.9 Summary

Chest X-rays are one of the most important and popular radiological analyses for the detection of several diseases, especially COVID-19 or other pneumonia. The applications of AI and deep learning methods using CXR for lesion detection have shown very significant results. However, the background or non-region of interest present in the CXRs may misguide the AI system in detecting the lesion. In this work, we have presented AI-based deep learning UNet and UNet+ models for the segmentation of CXR images to segment the lung region from the X-ray so that an accurate diagnosis of diseases could be made. Our UNet model performed best and achieved an accuracy of 96.35% with a dice coefficient and Jaccard index of 94.88% and 90.38%, respectively. The high-accuracy dice and Jaccard demonstrate the model's efficacy for the efficient segmentation of the lung region and potential in the field of AI and CAD-based diagnosis.