Proceedings of the 4th International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/80/2025.21240

Improved Unet Model for Lung X-ray Image Segmentation Based on AG Attention Mechanism with Min-ASPP Module Fused with GSConv Module

Zhuo Xu[1,a,*], Ningning Xu[2]

[1]*School of information science and technology, NanTong university, Jiangsu, Nantong, 226000, China*
[2]*Xuanchen central hospital, Anhui, Xuancheng, 242000, China*
a. 1791664335@qq.com
**corresponding author*

Abstract: In this paper, innovative improvements and optimizations are made on the basis of the traditional Unet base model, which mainly introduces the AG attention mechanism, Min-ASPP module, and incorporates the GSConv module to enhance the segmentation performance of lung X-ray images. During the training process, we recorded the loss change curves of the Unet model and the improved model respectively. The results show that the improved model converges significantly faster, while the loss values of the training and validation sets almost completely overlap, which indicates that the model proposed in this paper has stronger generalization ability. In the testing phase, we compare the evaluation indexes of Unet and the improved model in this paper. The results show that the improved model improves the segmentation accuracy by 9% over Unet, while the intersection-to-union ratio (IoU) is improved by 9.76%. The improvement of these metrics indicates that the model proposed in this paper has a superior performance in processing lung X-ray images. In terms of segmentation effect, compared with the traditional Unet, the improved model in this paper shows obvious advantages in the segmentation of lung structures, especially in the more accurate detail processing of edge contours. This result not only verifies the technical validity of the proposed method, but also provides a more reliable tool for practical medical image analysis. In summary, this paper brings new ideas and methods to the field of lung X-ray image analysis by optimizing the Unet model in multiple dimensions, which not only improves the segmentation accuracy, but also enhances the model's adaptability to different datasets. These research results will help to further promote the development of medical image processing technology and improve the efficiency and accuracy of clinical diagnosis.

Keywords: Lung X-ray Image, Deep Learning Network Algorithm, GSConv.

1. Introduction

Lung radiography is a widely used diagnostic imaging tool in clinical practice, which is mainly used to detect a variety of lung diseases, such as pneumonia, tuberculosis, and tumors [1]. With the aging of the population and the increase of environmental pollution, the incidence of lung diseases is gradually increasing, and there is an urgent need for efficient and accurate diagnostic tools. However,

DOI: 10.54254/2753-8818/80/2025.21240

traditional manual interpretation of X-ray images is not only time-consuming, but also susceptible to subjective factors, which may lead to misdiagnosis or missed diagnosis [2]. Therefore, how to improve the automation and accuracy of lung X-ray image analysis has become an important topic in medical imaging research.

In this context, image segmentation techniques have emerged. Image segmentation is to divide an image into multiple meaningful regions for subsequent analysis [3]. In lung X-ray images, the goal is usually to segment out the lung fields, the heart, and other relevant structures. This process is critical for physicians to quickly identify diseased areas. In recent years, with the development of computer vision and deep learning techniques, deep learning-based methods have shown great potential in the field of medical image segmentation.

Deep learning is a machine learning method that simulates the structure of the neural network of the human brain, which is used to extract features from data through multilayer nonlinear transformations, thus realizing the automated processing of complex tasks [4]. In medical image analysis, deep learning models are able to learn rich feature representations by training with large amounts of labeled data. This capability has led to the widespread use of deep learning in lung X-ray image segmentation.

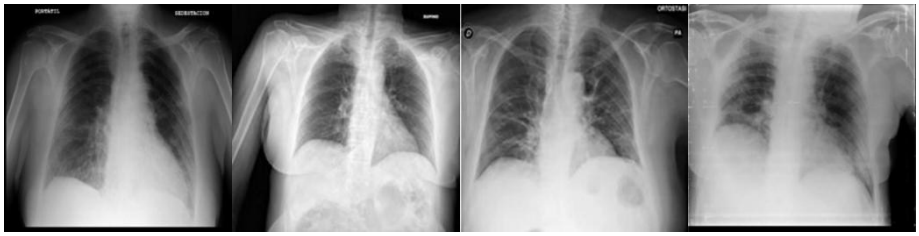
First, convolutional neural network (CNN) is a commonly used architecture in deep learning, which performs well in processing 2D images [5]. Through convolutional operations, CNNs can effectively extract local features, enabling the model to recognize the boundaries between different tissues or organs. For example, in lung X-ray images, healthy lung tissues and diseased regions can be automatically identified and separated by training the CNN. In addition, some improved networks, such as U-Net and SegNet, are specifically optimized for medical image segmentation, and they improve detail recovery by introducing mechanisms such as jump connections, which enable the model to better capture fine structures while maintaining global contextual information [6].

Second, deep learning models are more robust and adaptive than traditional methods. Since medical data tends to be noisy and inhomogeneous, deep learning models can be trained with large-scale datasets, thus improving their adaptability to image data generated by different patients and different devices. This adaptive ability makes them perform well in practical applications and achieve good segmentation results even when the data samples are limited or of poor quality [7].

In addition, with the development of migration learning and small amount of labeled data training techniques, pre-trained models can be utilized for effective migration and fine-tuning even in small sample scenarios with high labeling costs. This provides new solutions for clinical practice, enabling healthcare organizations to apply deep learning techniques more flexibly. In this paper, we improve and optimize the Unet base model based on the Unet model by adding AG attention mechanism, Min-ASPP module and fusing the GSConv module for segmentation of lung X-ray images.

2. Data set sources and data analysis

This dataset uses a private dataset, all data are collected and prepared by ourselves, the dataset contains two parts, the first part is the original images of lung X-ray images. The second part is the mask corresponding to the lung X-ray image. Some of the images are selected for presentation, as shown in Fig. 1.



DOI: 10.54254/2753-8818/80/2025.21240

Figure 1: Partial data.

3. Method

3.1. Unet

U-Net is a convolutional neural network architecture specially designed for medical image segmentation task.The design concept of U-Net is to efficiently capture the contextual information through encoder-decoder structure while maintaining the high resolution features for accurate segmentation.The model structure diagram of Unet is shown in Fig. 2.

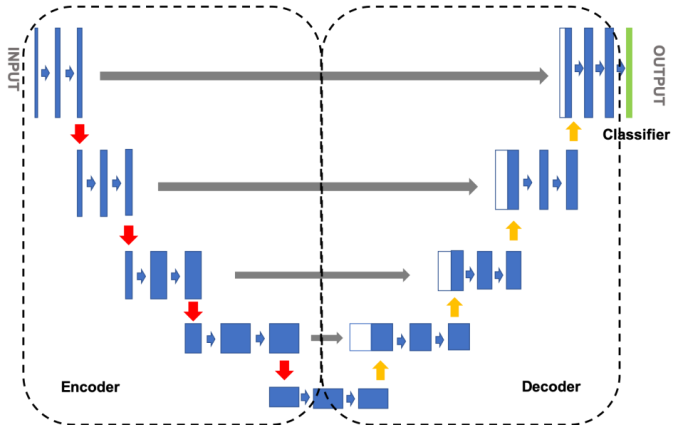


Figure 2: The model structure diagram of Unet.

U-Net consists of two parts: encoder and decoder. The encoder part usually consists of a series of convolutional and pooling layers, which are mainly responsible for extracting features from the image. In each layer, convolutional operations are used to extract local features, while pooling operations are used to reduce the spatial dimensions of the feature maps, thus gradually increasing the sensory field and capturing deeper information [8].

The decoder part then gradually restores the spatial resolution of the image through upsampling operations. Each upsampling step is followed by a splice (skip connection) with the feature map of the corresponding encoder stage. This skip connection allows the decoder to utilize the high-resolution features preserved in the encoder and thus localize and segment the target region more accurately.

Skip connections play a crucial role in U-Net. Since detailed information is often lost during the downsampling process, jump connections can pass high-resolution information directly to the decoding stage. This mechanism not only enhances the model's ability to capture details, but also improves the accuracy of the final output segmented image. As a result, U-Net performs well in dealing with complex shapes and boundaries, and is especially suitable for medical image analysis.

U-Net usually uses cross-entropy loss functions or Dice coefficients as optimization objectives, which can effectively measure the consistency between the prediction results and the real annotations.

DOI: 10.54254/2753-8818/80/2025.21240

In the training process, iterative learning through a large amount of labeled data allows the network to continuously optimize parameters for better adaptation to specific tasks [8].

3.2. AG Attention on Graphs

AG Attention on Graphs is an attention mechanism designed for graph data to enhance the performance of graph neural networks (GNNs). When dealing with graph-structured data in areas such as social networks, knowledge graphs and bioinformatics, traditional GNNs often have difficulty in effectively capturing the complex relationships and importance between nodes.AG Attention on Graphs introduces adaptive attention weights that enable the model to focus on different nodes and their neighborhoods, thus improving the accuracy and efficiency of analyzing graph data. The schematic diagram of AG Attention Mechanism is shown in Figure 3.

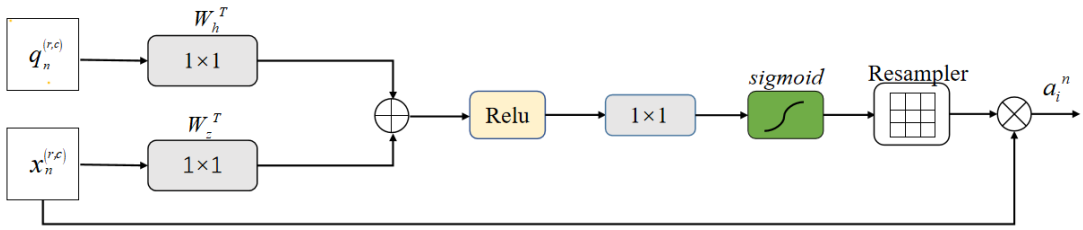


Figure 3: The schematic diagram of AG Attention Mechanism.

The core of the AG attention mechanism is to assign a weight to each node that reflects its importance. First, the model aggregates the features of the target node and its neighboring nodes, and then calculates the importance of each neighboring node to the target node to generate the attention weights. Finally, the updated feature representation is obtained by weighted summation of the features of neighboring nodes. This process allows the network to highlight key information while ignoring irrelevant or redundant information, thus improving the overall performance [9].

3.3. Min-ASPP module

The Min-ASPP (Minimum Atrous Spatial Pyramid Pooling) module is a structure designed to enhance the feature extraction capabilities of Convolutional Neural Networks (CNNs), especially for tasks such as image segmentation and target detection. The traditional ASPP (Atrous Spatial Pyramid Pooling) module enhances the model's sensitivity to multi-scale features by using multiple dilated convolutions to capture information at different scales. However, ASPP can be redundant in terms of computational overhead and parameter efficiency, resulting in high resource consumption during model training and inference. The Min-ASPP module optimizes the feature extraction process by simplifying the ASPP structure to more efficiently capture multi-scale information while reducing computational complexity. The structure of the Min-ASPP module is shown in Fig. 4.

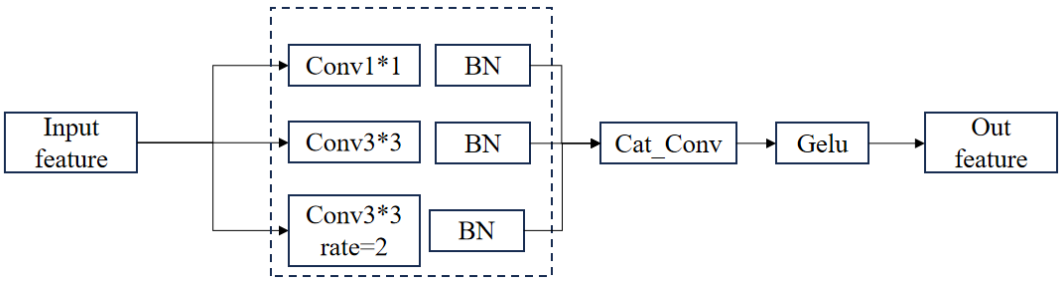


Figure 4: The structure of the Min-ASPP module.

DOI: 10.54254/2753-8818/80/2025.21240

The core of the design of Min-ASPP module is to utilize a small number of null convolution and global average pooling operations to achieve efficient multi-scale feature extraction. Specifically, the module first employs null convolutions with different expansion rates, which are able to efficiently capture different scale information in the image while keeping the sensory field unchanged. In addition, to enhance the capture of global contextual information, Min-ASPP introduces a global average pooling layer. By pooling the whole image, this layer is able to summarize important global features and provide support for subsequent feature fusion. Finally, all extracted features are spliced or weighted and summed to generate the final output. This process ensures that the model is able to synthesize information from different scales and contexts to improve the accuracy of the segmentation or detection task [10].

3.4. GSConv

The GSConv (Gated Spatial Convolution) module is an innovative architecture for improving the performance of Convolutional Neural Networks (CNNs), which is mainly applied to image processing and computer vision tasks.GSConv enhances the flexibility and accuracy of feature extraction by introducing a gating mechanism, thus improving the performance of the model in complex scenes. GSConv module's structure is shown in Figure 5.

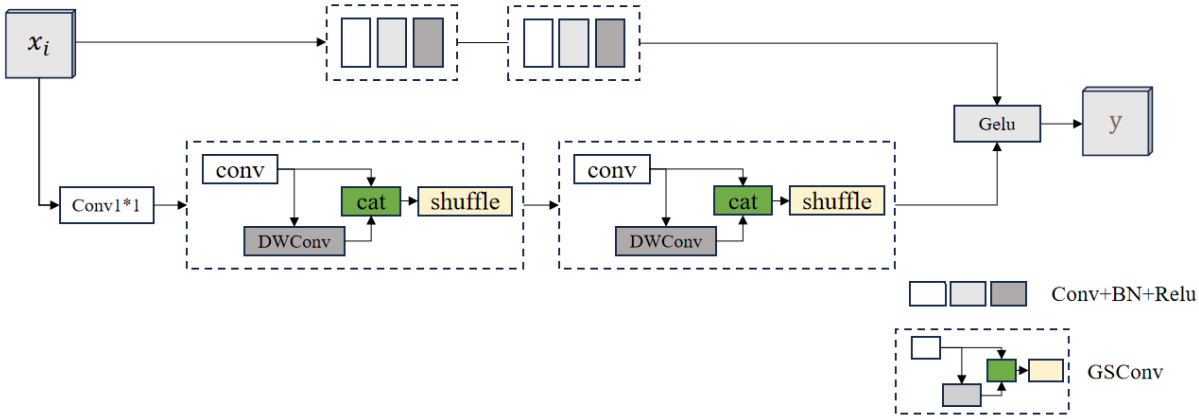


Figure 5: GSConv module's structure.

The core idea of GSConv is to dynamically adjust the information flow in the convolution operation through the gating mechanism. While traditional convolutional operations usually have fixed weights, GSConv introduces a gating unit that enables each convolutional kernel to adaptively and selectively activate or inhibit certain features when processing input features. This mechanism is based on the analysis of the input feature maps, which can effectively capture spatial context information, thus improving the model's ability to recognize different scenes and objects.

The GSConv module is composed of two main parts: the standard convolutional layer and the gating layer. The standard convolutional layer is responsible for extracting the basic features, while the gating layer generates a weight matrix based on the input features to adjust the importance of each feature. Specifically, the gating layer typically uses a sigmoid activation function that limits the output to between 0 and 1, thus creating an "attention" mechanism. In this way, the model can automatically decide which features to emphasize and which to ignore when performing convolutional operations.

DOI: 10.54254/2753-8818/80/2025.21240

3.5. Improved Unet model based on AG attention mechanism with Min-ASPP module fused with GSConv module

The U-Net model after fusing the AG attention mechanism, Min-ASP module and GSConv has several significant advantages. First, the model can better balance the global and local information and improve the segmentation accuracy when dealing with complex images; second, by dynamically adjusting the importance of features, it reduces the unnecessary information interference and helps to reduce the risk of overfitting; finally, the multi-scale feature fusion makes the model more robust and flexible when facing objects of different sizes. The schematic diagram of the model is shown in Fig. 6.

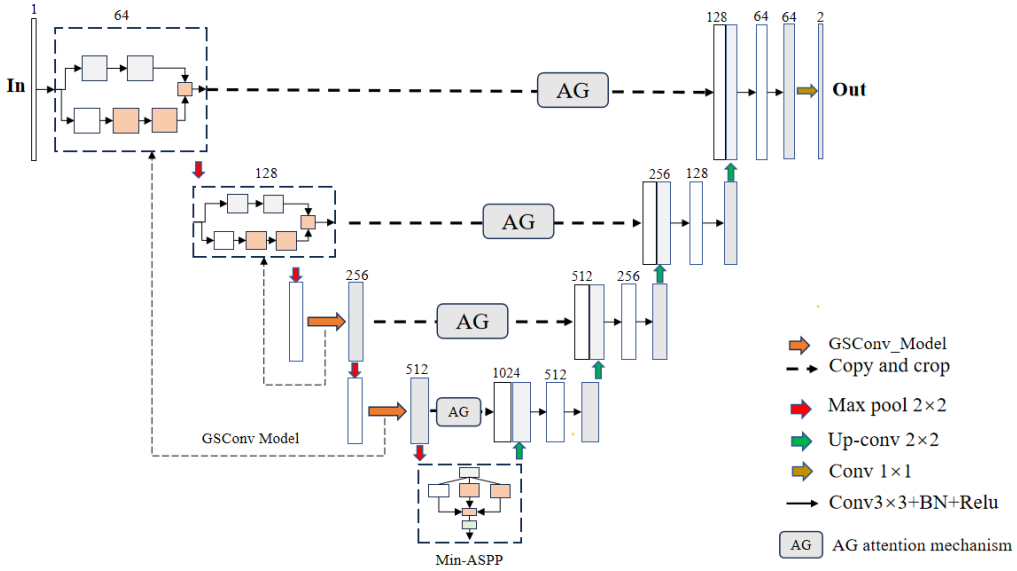


Figure 6: GSConv module's structure.

4. Result

After cropping the data, image enhancement is performed on the image using histogram equalization. In terms of hardware configuration, this paper uses a 3090 graphics card with 32G of RAM, the epoch is set to 50, the batch size is set to 16, and the training, validation, and test sets are divided according to the ratio of 40%, 40%, and 30%. The segmentation effect of the model is evaluated using loss, Accuracy and Miou. The loss variation curves for the training and validation sets of the Unet model are shown in Fig. 7, and the loss variation curves for the training and validation sets of the model in this paper are shown in Fig. 8.

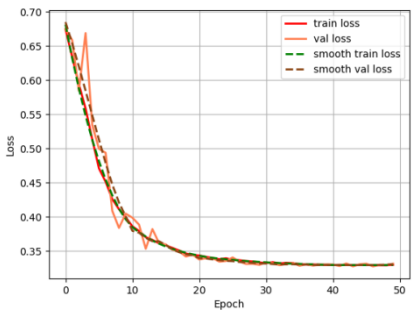


Figure 7: The loss variation curves for the training and validation sets of the Unet model.

DOI: 10.54254/2753-8818/80/2025.21240

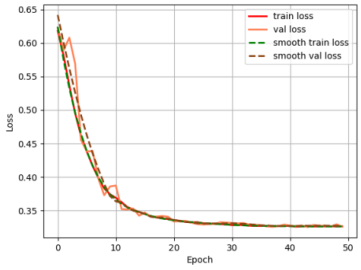


Figure 8: The loss variation curves for the training and validation sets of the model in this paper.

Output the loss change curves of Unet and our model during the training process respectively, it can be seen that the model proposed in this paper is faster in convergence, and the loss values of the training set and validation set of our model almost overlap, indicating that the model has a stronger generalization ability.

In the testing part, the evaluation indexes of two models, Unet and our model, are outputted respectively, and our model is 9% higher than Unet in segmentation accuracy, and 9.76 higher than Unet in iou, and the segmentation results of Unet and GSConv Unet are outputted as shown in Fig. 9, with the first column as the gold standard, the second column as the segmentation results of Unet, and the third column as the segmentation results of our model. The first column is the gold standard, the second column is the segmentation result of Unet, and the third column is the segmentation result of this model.

Table 1: Modelling assessment.

Model	Accuracy	Miou
Unet	81.46%	77.15
GSConv Unet	90.47%	86.91

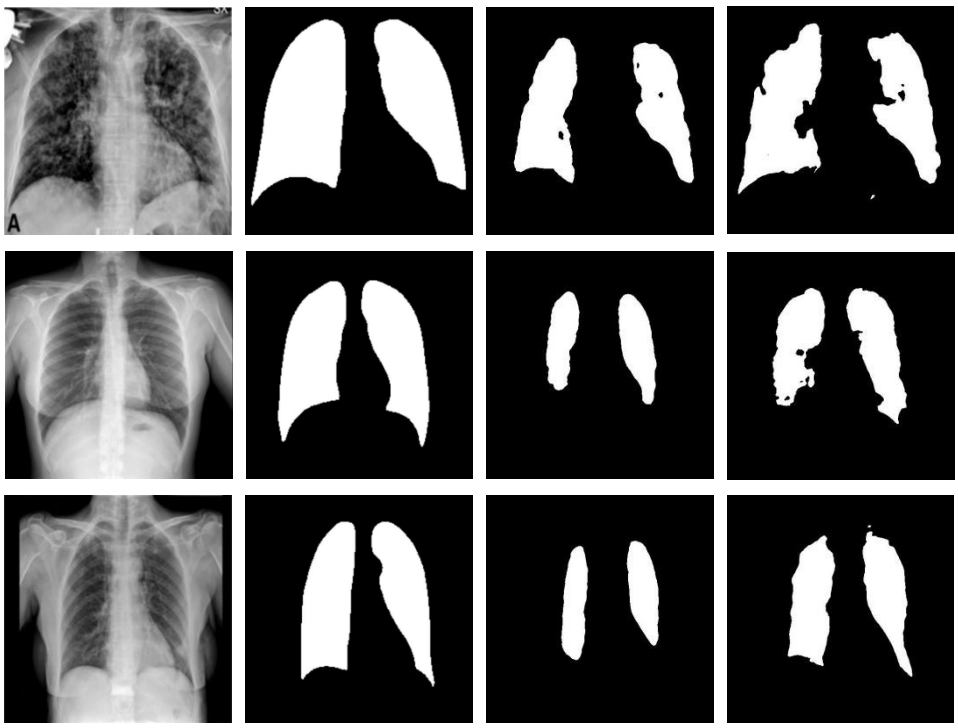


Figure 9: Test Set Output Results.

DOI: 10.54254/2753-8818/80/2025.21240

From the segmentation results of Unet and the model in this paper, compared with Unet, the model in this paper has better segmentation of the lungs, especially in the segmentation effect of the edge contour.

5. Conclusion

In this study, we have made several improvements and optimizations on the classical U-Net base model, especially introducing the AG (Adaptive Gated) attention mechanism, Min-ASPP (Minimum Adaptive Spatial Pyramid Pooling) module, and incorporating the GSConv (Gated Spatial Convolution) module to improve the segmentation performance of lung X-ray images.

During the training process, we recorded the loss variation curves of the U-Net model and the improved model in this paper, respectively. By observing these curves, it can be clearly seen that the model proposed in this paper converges significantly faster than the traditional U-Net.This fast convergence not only reflects the improved model's high efficiency for feature learning, but also indicates that it is able to better capture the important information in the data during the training phase. In addition, the loss values of the model in the training and validation sets almost coincide, which further indicates that the model has stronger generalization ability and can effectively avoid overfitting, thus showing better stability in practical applications.

During the testing phase, we conducted a comprehensive evaluation index comparison between U-Net and the improved model in this paper. The results show that the model proposed in this paper outperforms U-Net by 9% in segmentation accuracy and improves 9.76% in IoU (Intersection over Union) metrics. The improvement of this series of indexes not only proves the advantage of the new model in accuracy, but also reflects its good ability to grasp the details of the complex lung structure.

By comparing the segmentation results of U-Net and the improved model in this paper, it can be clearly seen that the improved model performs much better in the segmentation effect of the lung region. Especially in the processing of the edge contour, our model can more accurately capture the subtle changes of the lung structure, which is crucial for medical image analysis. Accurate segmentation not only helps doctors to make subsequent diagnosis, but also provides a reliable basis for clinical decision-making.

References

[1] *Sulaiman, Adel, et al. "A convolutional neural network architecture for segmentation of lung diseases using chest X-ray images." Diagnostics 13.9 (2023): 1651.*

[2] *Ullah, Ihsan, et al. "A deep learning based dual encoder–decoder framework for anatomical structure segmentation in chest X-ray images." Scientific Reports 13.1 (2023): 791.*

[3] *de Almeida, Pedro Aurélio Coelho, and Díbio Leandro Borges. "A deep unsupervised saliency model for lung segmentation in chest X-ray images." Biomedical Signal Processing and Control 86 (2023): 105334.*

[4] *Arvind, S., et al. "Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays." Results in Engineering 17 (2023): 100929.*

[5] *Öztürk, Şaban, and Tolga Çukur. "Focal modulation network for lung segmentation in chest X-ray images." Turkish Journal of Electrical Engineering and Computer Sciences 31.6 (2023): 1006-1020.*

[6] *Ghimire, Samip, and Santosh Subedi. "Estimating Lung Volume Capacity from X-ray Images Using Deep Learning." Quantum Beam Science 8.2 (2024): 11.*

[7] *Gaggion, Nicolás, et al. "CheXmask: a large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images." Scientific Data 11.1 (2024): 511.*

[8] *Brioso, Ricardo Coimbra, et al. "Semi-supervised multi-structure segmentation in chest X-ray imaging." 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2023.*

[9] *Chen, Lingdong, et al. "Development of lung segmentation method in x-ray images of children based on TransResUNet." Frontiers in radiology 3 (2023): 1190745.*

[10] *Iqbal, Ahmed, Muhammad Usman, and Zohair Ahmed. "Tuberculosis chest X-ray detection using CNN-based hybrid segmentation and classification approach." Biomedical Signal Processing and Control 84 (2023): 104667.*