Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Pulmonary Nodule Segmentation

Shenbagarajan A.[1], Abinaya M.[2], Bhuvanika S.[3], Rajakumari S.[4]

Artificial Intelligence and Data Science, Mepco Schlenk Engineering College, Anna University, Sivakasi, India.

Email: [1]shenbagarajan@mepcoeng.ac.in, [2]abi680774_bai25@mepcoeng.ac.in,
[3]bhuvanilakshmi1102_bai25@mepcoeng.ac.in, [4]rajakumariasha11_bai25@mepcoeng.ac.in

Abstract

Lung cancer remains one of the leading causes of death worldwide, mainly due to delayed diagnosis and treatment. Lung nodules must be properly and quickly recognized as benign or malignant to improve survival. This work presents a deep learning approach for segmentation using the Residual U-Net (ResiU-Net) architecture improved with Multiscale Aggregation (MSA) focusing on accurately identifying lung nodules from CT scans. A total of 1,190 labeled CT scan slices from the publicly available IQ-OTHNCCD dataset, including the benign, normal, and malignant classes, are used to test the model. To enhance feature representation and data balance, advanced preprocessing methods such as data augmentation and standardization were used. The new ResiU-Net architecture makes use of residual connections to overcome vanishing gradients and achieve contextual depth, while the MSA block enables the collection of global and fine-grained information required for the identification of small nodules (as small as 3 mm).

A rigorous 5-fold cross-validation technique was used to test the model's performance. The strength of segmentation was evaluated using metrics such as F1-Score, Accuracy, Dice Score, and Intersection over Union (IoU). The method demonstrated better detection accuracy and excellent generalization compared to baseline U-Net models with an F1-score of 0.8385 and a best IoU of 0.7971 and validation IoU of 0.7300. The model's ability to accurately recognize nodules of varying sizes and shapes has been established by the testing results. The goal of the project is to help design a clinical decision-support system that is accurate, automated, and affordable for monitoring and early-stage lung cancer diagnosis.

DOI: https://doi.org/10.36548/jaicn.2025.3.001
Received: 22.05.2025, received in revised form: 26.06.2025, accepted: 17.07.2025, published: 01.08.2025

Keywords: Pulmonary Nodule Segmentation, Deep Learning, U-Net, Resnet-152, Computed Tomography (CT) Images, Medical Image Processing, Convolutional Neural Networks (CNNs).

1. Introduction

Lung cancer is a malignant and prevalent disease. It needs to be diagnosed early enough so that it becomes more survivable. Pulmonary nodules are hard to find because they occur in extremely varied sizes, shapes, and positions and develop before lung cancer. Preparation for its treatment is the first extremely crucial task, i.e., properly segmenting such tumors from CT scans. Due to the poor resolution of the image and the noisy condition of the image, manual segmentation is practically impossible, time-consuming, and heterogeneous. Deep learning is a robust medical image analysis technique and also serves to effectively address detection, classification, and segmentation issues. U-Net is one of the broadly used benchmark models of biomedical image segmentation since it possesses an encoder-decoder structure, enabling it to learn low-level and high-level features. Residual networks (ResNets) are also a model that has been used extensively to enhance the performance of the model and provide effective training by preventing problems such as the vanishing gradient problem.

This paper proposes a new enhanced ResiU-Net model with a single MSA layer placed between the encoder and decoder to enable the easy representation of features for superior accuracy and efficiency in segmentation. The MSA layer is expected to enhance multi-scale feature aggregation and allow the model to better detect nodules of varying sizes and complexity. It was evaluated on a multi-patient database of CT scans in which patients varied according to nodule characteristics. This reflects enhanced performance, compared with the existing state-of-the-art, in segmentation quality and in accommodating nodules of different sizes and shapes, which is one of the primary lung cancer detection challenges.

In addition to its technological superiority, the study advocates for ResiU-Net with MSA because it has great medical significance. Under the limitations of resource availability, the identification of patients at an early stage, made easier by the model through accurate segmentation, can also be supported by radiologists. In addition, the process of automated nodule segmentation precludes the risk of human bias and makes the diagnostic process uniform. Keeping in view the problems of cancer detection, the paper presents the integration of newer deep learning models into health systems. Early diagnosis and treatment of possible

lung cancer are worthy of further research to enhance patient outcome and decrease cancer death globally. The research points to the advantages of hybrid models like ResiU-Net and MSA.

2. Related Work

Recent developments in deep learning have had a profound impact on pulmonary nodule segmentation algorithms via CT imaging. Zhi et al. [1] conducted an extensive review of research on deep neural network models for segmentation, highlighting better performance by models such as U-Net and its variants due to high precision. Gao et al. [2] provided a systematic review that comprised current trends and issues in deep learning-based methods employed for pulmonary nodule detection and segmentation, stressing the trend towards more efficient end-to-end methods. Marinakis et al. [3] offered an extensive literature review of deep learning for classification, segmentation, and detection, establishing that multi-task fusion can lead to increased diagnostic accuracy. Shi et al. [4] compared multiscale residual U-Net with fuzzy C-means clustering and concluded that deep learning approaches far outstrip conventional clustering in feature preservation and boundary precision. Yadav et al. [5] proposed EDTNet, a novel attention-based transformer model with excellent spatial perception to improve segmentation accuracy.

Zhou et al. [6] proposed a cascaded multi-stage approach for low-resource environments with encouraging results in underdeveloped nations by reconciling performance and computational expense. Bhattacharjee et al. [7] explored transfer learning performance with pre-trained models fine-tuned for pulmonary nodule segmentation and demonstrated significant robustness and generalization on CT datasets. Lu et al. [8] also developed a domain-specialized deep network for nodule segmentation and demonstrated that domain-specialized architecture outperforms general models when designed with domain-specific features. Ni et al. [9] extended this by creating a two-stage multitask U-Net that segments and predicts malignancy risk in one go. This is representing a shift from single-task localization to multipurpose diagnosis models. Li et al. [10] provided a comprehensive review of deep learning application in pulmonary nodule detection and classification, including the future research and clinical potential of integrating segmentation with diagnostic pathways.

Overall, the literature presents a clear shift from older segmentation techniques to more recent deep learning algorithms that enhance accuracy, decrease false positives, and offer

greater clinical utility. Emerging trends include multitask learning, attention models, and transfer learning, aimed at providing more precise and interpretable outputs in lung imaging.

3. Proposed Work

3.1 Dataset Description

110 patients' CT scans totaling 1,190 images make up the IQ-OTH/NCCD Lung Cancer CT dataset from the National Center for Cancer Diseases and the Iraq Oncology Teaching Hospital. The information is available to the public and can be downloaded from Kaggle. It is divided into 40 malignant, 15 benign, and 55 normal classes. Initially saved in DICOM format, the images are now conveniently supplied in JPEG format. Using slices that were 1 mm thick, the CT scans were performed using a Siemens SOMATOM scanner. Images with contrast and brightness between 350 and 1,200 Hounsfield Units (HU) are suitable for imaging lung tissue. Table 1 presents an overview of the dataset and its performance.

Table 1. Dataset and Performance Summary

Dataset Name	**Optimizer**	**Training Accuracy**	**Validation Accuracy**	**Training F1-Score**
IQ-OTHNCCD	Adam (LR=0.0001)	0.9931	0.9906	0.8856
Validation F1-Score	Training IOU	Validation IOU	Training Loss	Validation Loss
0.8385	0.7971	0.7300	0.1134	0.1632

3.2 Data Preprocessing

This study uses the best preprocessed CT scans. The images are re-sampled to 256 × 256 pixels and pixel values are scaled to [0, 1]. Rotation (90° and 180°), horizontal and vertical flips, and cropping data augmentation operations are applied for model variability optimization and generalization. Cross-validation ground truth masks are also pre-prepared by radiologists to achieve maximum accuracy. An 80-20 data split ratio was employed to train the model,

resulting in a training loss of 0.10 and a validation loss of 0.22. The Google Colab Pro environment with Python 3.7.12, TensorFlow 2.1.0, Keras 2.3.1, CUDA 11.2 is employed to execute the model. It is trained for 100 epochs with batch size 8 and Adam optimizer learning rate 0.0001. Dice Loss with Binary Focal Loss as loss function was employed, and binary cross-entropy has also attempted. Themodel ensured convergence and performance in segmentation accuracy.

3.3 Architecture Design

It is now suggested to segment pulmonary nodules using the appropriate ResiU-Net model, which includes a Multiscale Aggregation layer between the decoder and encoder. Although less computationally intensive, its model structure is more accurate. The overall strategy consists of the training strategy, model structure, and data preprocessing.

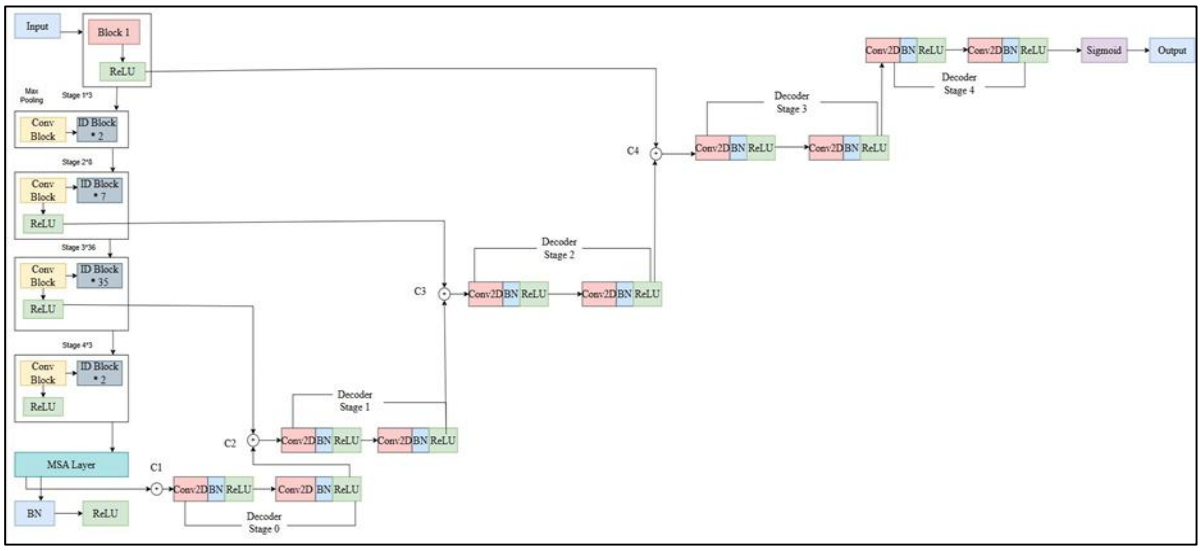

Figure 1. Architecture Design

Encoder: The Pre-trained ResNet152 model is utilized as the encoder, initializes its weights with ImageNet. This allows the network to extract hierarchical features like texture and edge patterns, which are needed for segmentation tasks.

- yc stage1 = G(M, {Sa}) + SW M (1)
- yid1 stage1 = G (yconv stage1, {Sa}) + yc stage1 (2)
- yid2 stage2 = G (yid1 stage1, {Sa}) + yid1 stage1 (3)
- yc stage2 = G (yid2 stage2, {Sb}) + Sv · yid2 stage2 (4)
- yid1 stage2 = G (yconv stage2, Sb) + yc stage2 (5)

- yid7 stage2 = G (yid6 stage2, {Sb}) + yid6 stage2 (6)
- yc stage3 = G (yid7 stage2, {Sc}) + Su · yid7 stage2 (7)
- yid1 stage3 = G (yconv stage3, {Sc}) + yc stage3 (8)
- yid35 stage3 = G (yid34 stage3, {Sc}) + yid34 stage3 (9)
- yc stage4 = G (yid35 stage3, {Sd}) + St · yid35 stage3 (10)
- yid1 stage4 = G (yconv stage4, {Sd}) + yconv stage4 (11)
- yid2 stage4 = G (yid1 stage4, {Sd}) + yid1 stage4 (12)

MF and MSA Layer: The MF1 Block is capable of capturing features of any size by means of convolutional layers with different dilation rates (1, 3, 5, and 7). Through this, can learn more detailed information from an input image. Feature maps from all four convolutional layers are concatenated and batch-normalized for stable training. The features are then merged into an even finer representation in a 1×1 convolutional layer. There is also a skip connection (residual path) to maintain the input features, allowing gradients to propagate the during training, keeping the model stable, and enabling it to function effectively. The middle MSA layer, which sits between the encoder and decoder, uses a stack of convolutional layers of different sizes (1×1, 3×3, 5×5, and 7×7) to combine features at multiple scales. To normalize activations, batch normalization is used at every scale. Additionally, the block contains learnable attention weights in the form of distinct 1x1 convolutional layers, which enable the model to determine which salient features to focus on at what scale. To establish equivalence among feature maps of varied scales, downsampling compresses large feature maps while upsampling expands small feature maps to larger sizes in an effort to promote uniformity. Features of varied scales are concatenated and pooled to form the ultimate representative shape. Output is also enhanced through the use of a 1×1 convolutional layer and bypassing input to a residual connection to allow the model to extract meaningful information from the input.

Decoder: Decoder constructs segmentation masks from transposed convolutions for upsampled feature maps. Skip connections between the decoder and encoder retain spatial information and enhance prediction accuracy as well.

- C1 = γstage4 + γconv stage4 (13)
- C2 = γdecoder stage0 + γvconv stage3 (14)
- C3 = γdecoder stage1 + γconv stage2 (15)
- C4 = γdecoder stage2 + γblock1 (16)

Skip Connections: Decoder-to-encoder large-size skip connections provide feature persistence during upsampling, while mini skip connections in the network encoder address vanishing gradients.

3.4 Fine-Tuning

In the optimisation maxim of peak performance, even pretrained encoder weights are frozen during the initial training sessions to allow early decoder learning, Only the initialization is varied after that to freeze the encoder encoder, which is further fine-tuned through additional training where encoder adaptation is varied and segmentation accuracy is enhanced. This is at no cost to the pretrained models feature extraction ability but reduces training time.

3.5 Loss Function and Optimization

The model proposed here uses Dice Loss and Binary Focal Loss combined for better precision in segmentation along with class imbalance minimization.

- **Dice Loss:** Provides overlapping prediction mask and ground truth estimation, hence accuracy increases.
- **Binary Focal Loss:** Provides extra focus to hard-to-classify regions and allows more penalty on the wrong predictions.

Adam optimizer with learning rate of 0.0001 is employed for better convergence with less computation complexity.

4. Results and Discussion

The ResiU-Net model with an additional MSA layer was tested on the IQ-OTHNCCD Lung Cancer Dataset to evaluate the pulmonary nodule segmentation performance. The addition of multi-scale information was significantly improved by the additional MSA layer, considering the nodule segmentation performance as an independent variable of nodule size and complexity. The model performed best during the training and validation stages, achieving a precision of 0.9931, an F1-score of 0.8856, an IoU of 0.7971, and a loss of 0.1134 at the training step. It achieved a precision of 0.9906, an F1-score of 0.8385, an IoU of 0.7300, and a loss of 0.1632 at the validation step. This provided the model with stability to complete the MSA layer, resulting in more precise object segmentation.

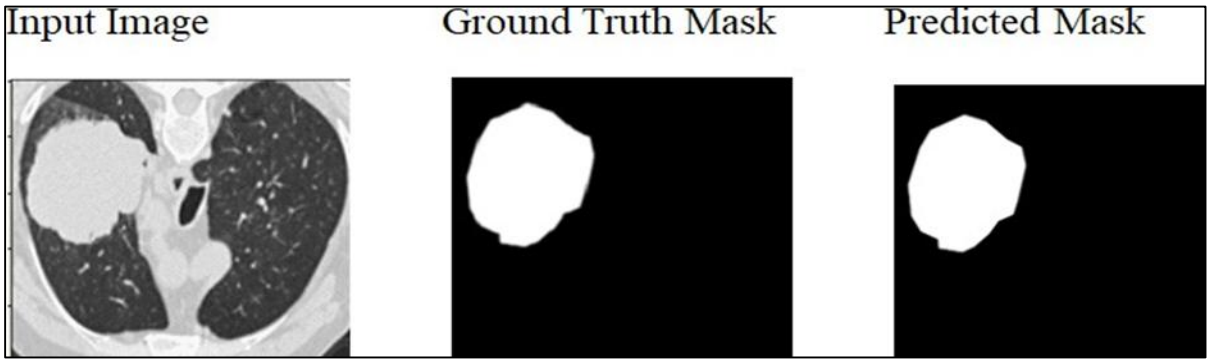

Figure 2. Segmentation Result (a)Input Test Image (b)Ground truth of Test Image (c)Predicted Test Image

These outcomes in Figure 2 further verify that the model developed here has achieved improved generalization capability to novel data and is performing more precise segmentation. Its ability to generalize effectively within the training set, not to mention the validation set, is a sign of excellence and accuracy. It is already being implemented in lung cancer early detection and for designing treatment and advanced computer-automated nodule segmentation methods.

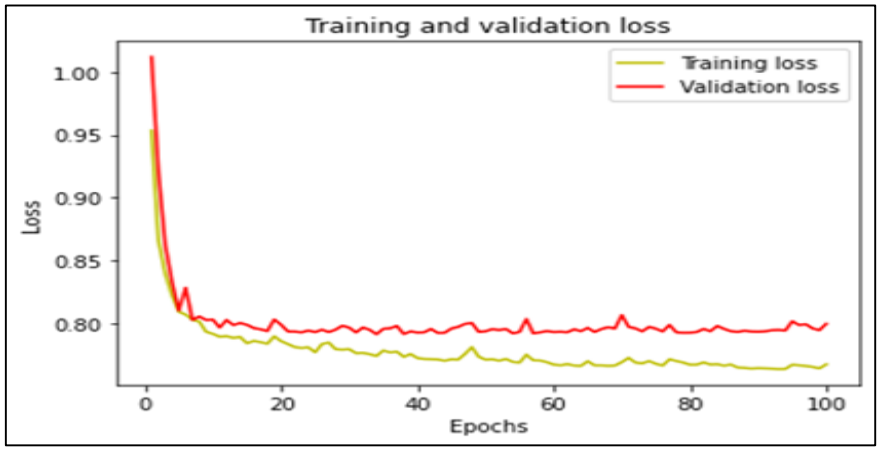

Figure 3. Training and Validation Loss

The pattern of training loss and validation loss across the epochs is shown in figure 3. Showing that the model is learning properly, both losses initially were very high but significantly reduced during the initial phase. The training loss continues to reduce during the training phase, showing that the model is mapping the training data correctly. On the validation loss side, however, the validation loss initially decreases before beginning to fluctuate or drop slightly. Either overfitting or randomness in the validation set is the reason for these oscillations. The validation loss is stable nevertheless, which indicates that good generalized features have been learned by the model and the model has generalized well while training.

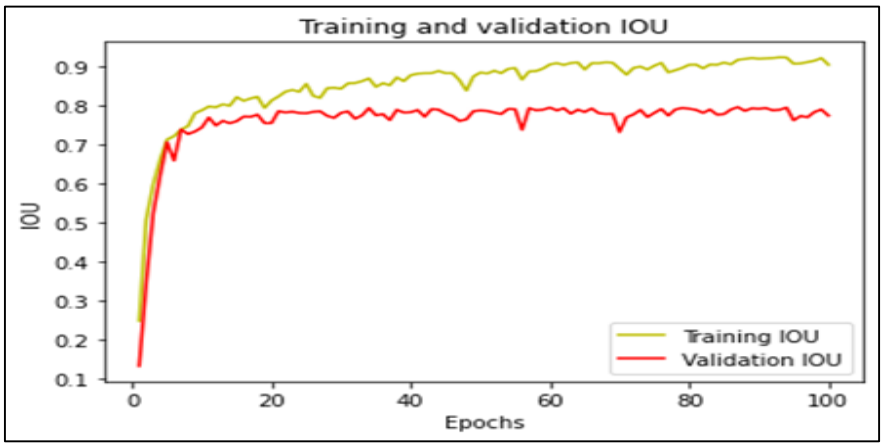

Figure 4. Training and Validation – IOU Curve

In the figure 4 shown, Intersection over Union (IoU) measures on training and validation data were monitored over epochs. In the early epochs, IoU of both classes increases rapidly, which is a sign that the model has learned successfully in the early stages of learning. The training IoU continues to rise through later epochs, indicating that the model is properly and consistently segmenting the training images. The validation IoU rises significantly in the early epochs and then reaches a plateau, slightly lower than the training IoU, indicating that the model is generalizing very well and is providing consistent good performance on unseen data.

$$IOU = \frac{|A \cap B|}{|A \cup B|}$$

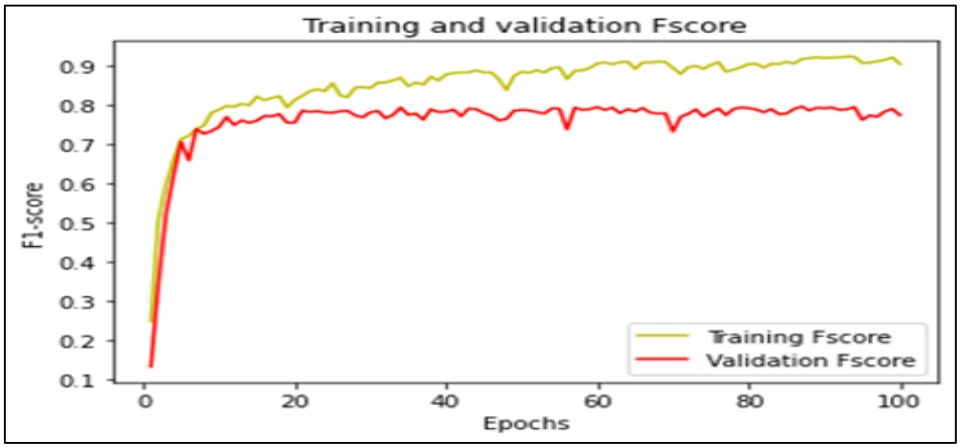

Figure 5. Training and Validation – F1 Curve

The graph 5 shows that the F-score was increasing for both the training and validation sets with an increase in the training epochs. Both were increasing very rapidly in the initial

phase, which suggests that the model is learning quickly and optimizing during training. The training F-score continued to increase steadily and reached a healthy, stable F-score, indicating it will perform well overall on the training set. The validation F-score also increased with training but remained just below the training F-score, indicating mild overfitting has occurred. The validation performance is at a level that suggests the model has generalized very well and performed commendably with unseen cases. An 80:20 training-validation split was used to create robust system validation. Five-fold cross-validation was also employed. Rotation, flipping, and cropping were performed to enhance generalization. The Adam optimizer (learning rate = 0.0001), batch size = 8, Binary Focal Loss, and Dice Loss were all used during training. Accuracy, IoU, Dice Score, and F1-Score were among the measures employed. Overfitting was mitigated through learning rate scheduling and early stopping. The overall validation scores achieved were IoU = 0.7300 and F1-score = 0.8385

5. Conclusion

In the present work, we introduce ResiU-Net and Multiscale Aggregation layer to obtain better segmentation and computational efficiency in pulmonary nodule segmentation. It is computationally efficient with respect to managing the multi-scale information, improving spatial feature extraction, and maximizing the nodular segmentation with specific features. Our experimental verification demonstrates that the proposed method provides better results across three challenging datasets with nodules of different sizes and complexities. The MSA layer facilitates smooth fusion of the multiscale information. The model works much better on all the datasets suggesting its improvement in generalizability. Computer-aided segmentation of pulmonary nodules based on systems reduces significantly the dependence on visual observation, decreases the chances of human errors and increases the trustworthiness of the diagnosis. Better and precise segmentation is required for Pretreatment Planning and Early Diagnosis of Lung Cancer, which is critical for achieving best patient outcomes. The model can also be implemented in a clinical work setting in real time. More optimization of the model will be achieved in future study based on a large database and in clinical implementation. The development here shows another use of deep learning in medicine so that lung cancer can be diagnosed more quickly and more accurately.

References

[1] Zhi, Lijia, Wujun Jiang, Shaomin Zhang, and Tao Zhou. "Deep neural network pulmonary nodule segmentation methods for CT images: Literature review and experimental comparisons." Computers in Biology and Medicine 164 (2023): 107321.

[2] Gao, Chuan, Linyu Wu, Wei Wu, Yichao Huang, Xinyue Wang, Zhichao Sun, Maosheng Xu, and Chen Gao. "Deep learning in pulmonary nodule detection and segmentation: a systematic review." European radiology 35, no. 1 (2025): 255-266.

[3] Marinakis, Ioannis, Konstantinos Karampidis, and Giorgos Papadourakis. "Pulmonary nodule detection, segmentation and classification using deep learning: a comprehensive literature review." BioMedInformatics 4, no. 3 (2024): 2043-2106.

[4] Shi, Jianshe, Yuguang Ye, Daxin Zhu, Lianta Su, Yifeng Huang, and Jianlong Huang. "Comparative analysis of pulmonary nodules segmentation using multiscale residual U-Net and fuzzy C-means clustering." Computer Methods and Programs in Biomedicine 209 (2021): 106332.

[5] Yadav, Dhirendra Prasad, Bhisham Sharma, Julian L. Webber, Abolfazl Mehbodniya, and Shivank Chauhan. "EDTNet: A spatial aware attention-based transformer for the pulmonary nodule segmentation." PloS one 19, no. 11 (2024): e0311080.

[6] Zhou, Zhixun, Fangfang Gou, Yanlin Tan, and Jia Wu. "A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in developing countries." IEEE Journal of Biomedical and Health Informatics 26, no. 11 (2022): 5619-5630.

[7] Bhattacharjee, Ananya, R. Murugan, Tripti Goel, and Seyedali Mirjalili. "Pulmonary nodule segmentation framework based on fine-tuned and pretrained deep neural network using CT images." IEEE Transactions on Radiation and Plasma Medical Sciences 7, no. 4 (2023): 394-409.

[8] Lu, Dechuan, Junfeng Chu, Rongrong Zhao, Yuanpeng Zhang, and Guangyu Tian. "A novel deep learning network and its application for pulmonary nodule segmentation." Computational Intelligence and Neuroscience 2022, no. 1 (2022): 7124902.

[9] Ni, Yangfan, Zhe Xie, Dezhong Zheng, Yuanyuan Yang, and Weidong Wang. "Two-stage multitask U-Net construction for pulmonary nodule segmentation and malignancy risk prediction." Quantitative Imaging in Medicine and Surgery 12, no. 1 (2022): 292.

[10] Li, Rui, Chuda Xiao, Yongzhi Huang, Haseeb Hassan, and Bingding Huang. "Deep learning applications in computed tomography images for pulmonary nodule detection and diagnosis: A review." Diagnostics 12, no. 2 (2022): 298.