

Lung tumor segmentation from CT images using 2.5D spatial data and 3D volumetric post-processing

Nguyen Duc Man¹, Ha Manh Toan², Le Manh Cuong³, Do Nang Toan^{2,1*}

¹International School, Duy Tan University, Da Nang, 550000, Vietnam

²Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, 10072, Vietnam

³K67A3 Computer and Information Science, VNU University of Science, Hanoi, 10072, Vietnam

Abstract

INTRODUCTION: Nowadays, lung cancer is known as one of the serious threats to human health and life globally. To deal with lung cancer, a problem is to accurately segment tumors from CT scans to help doctors in assessment and treatment planning.

OBJECTIVES: This study proposes a method for segmenting lung tumors from CT scan data based on exploiting spatial information between slices to feed into the deep learning model and applying 3D volumetric post-processing. In addition, we also provide visualization of 3D tumor reconstruction data.

METHODS: We incorporate spatially continuous data into the 2.5D input. In addition, we use a hybrid TransUNet deep learning model, and finally post-process 3D volumes. Our method allows the deep learning model to take advantage of spatial context information along the slices. TransUNet takes advantage of both CNN and Transformer. The 3D volume post-processing technique refines the results, removing discrete noisy blocks. This results in increased continuity and more accurate reconstruction in three-dimensional space.

RESULTS: The segmentation results were satisfactory. In case of 2.5D of 3 slices, results on NSCLC-Radiomics reached the Dice over slices of 0.891 ± 0.277 , the Dice over 3D volumes of 0.641 ± 0.235 , the IoU over slices of 0.875 ± 0.291 , the IoU over 3D volumes of 0.508 ± 0.228 , False Positives of 94, and False Negatives of 224. In case of 2.5D of 5 slices, results on NSCLC-Radiogenomics reached the Dice over slices of 0.951 ± 0.186 , the Dice over 3D volumes of 0.671 ± 0.231 , the IoU over slices of 0.941 ± 0.200 , the IoU over 3D volumes of 0.539 ± 0.223 , False Positives of 8, and False Negatives of 67.. In addition, the 3D tumor visualization showed good visualization capabilities.

CONCLUSION: The proposed method has proven to be effective in lung tumor segmentation. In this study, the elements of the proposed method have been analyzed comprehensively and comparatively. In the future, we expect to continue to conduct in-depth research and expand the experiment on patient data in Vietnam.

Keywords: lung cancer, 2.5D, CT scans, deep learning, 3D post-processing, tumor visualization

Received on 25 September 2026, accepted on 19 October 2026 published on 27 October 2026

Copyright © 2026 Nguyen Duc Man *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.11.10388

1. Introduction

Lung cancer is currently one of the major challenges for modern medicine and is the leading cause of death among cancers in the world. According to the GLOBOCAN 2022

statistical report, there are about 2.5 million new cases of lung cancer and nearly 1.8 million deaths from this disease, accounting for the highest cancer mortality rate globally [1]. Regarding lung cancer in Vietnam, the whole country

*Corresponding author. Email: dntoan@ioit.ac.vn

recorded 24,426 new cases (ranked 3rd after breast cancer and liver cancer) and 22,597 deaths (ranked 2nd after liver cancer) [2]. The mortality rate of lung cancer is almost equivalent to the number of new cases, showing that the low survival rate is often due to late detection or highly malignant pathological characteristics. This raises an urgent need for the development of early, accurate and highly applicable diagnostic methods.

In the process of early detection of lung cancer, chest CT images are an important supporting tool that helps doctors clearly observe the internal structure of the lungs and detect abnormalities. However, manually analyzing hundreds of CT slices from each patient is not only time-consuming but also subjective, depending on the experience and concentration of each doctor. This leads to the possibility of missing small lesions or tumors at an early stage [3].

Faced with this situation, automating the process of detecting and segmenting lung tumors from CT images has emerged as a potential solution. Remarkable advances in deep learning, especially in medical image processing, have opened up opportunities to build faster, more accurate and efficient diagnostic support systems. In the early stages, many studies still focused on improving image quality and manually extracting features. In 2016, Sangamithraa et al [4] proposed an image preprocessing pipeline using Median and Wiener filters, segmentation using K-Means and EK-Means, GLCM feature extraction and classification using backpropagation networks. In 2017, Kaucha et al [5] extracted features with entropy, correlation, energy, homogeneity from GLCM and trained SVM. In 2018, Patel et al [6] proposed to combine Local Energy-Based Shape Histogram and Sensitivity Analysis with histogram equalization, Otsu thresholding and shape and energy feature extraction, also in 2018, Perumal et al [7] combined Adaptive Histogram Equalization with Otsu segmentation and used level 5 discrete wavelet transform to identify ROI regions more accurately. The highlight of this study is to apply Enhanced Artificial Bee Colony Optimization algorithm to identify suspected tumor regions.

However, the above methods rely on manual feature extraction and traditional machine learning, which are difficult to scale to complex CT image data. Meanwhile, deep learning techniques are gradually replacing them thanks to their ability to automatically learn features. In 2017, Alakwaa et al [8] proposed a three-step pipeline including lung segmentation by thresholding, tumor detection by UNet, and malignancy classification by 3D CNN, also in 2017, Kuan et al. [9] improved the 3D convolutional neural network architecture with ResNet101 and handled data imbalance by balancing the weights in the loss function. In 2019, Hossain et al [10] developed a Dilated Hybrid-3D convolutional neural network architecture, combining 2D dilated convolution and 3D convolution.

Recent trends focus on UNet variants and deep supervision. In 2022, Farhanaz et al [11] combine discrete wavelet transform and Deeply Supervised MultiRes-Net for lung tumor segmentation, achieving an average Dice coefficient of up to 0.8472. In 2020, Kamal et al [12] develop a Recurrent 3D-DenseUnet network with ConvLSTM blocks

to help the model capture 3D spatial relationships better. In 2021, Dutande et al [13] propose a pipeline consisting of a SquExUNet model for segmentation and a 3D-NodNet for classification, combining 2D/3D inputs and advanced data augmentation. Recently in 2023, Le et al [14] introduced the RRcUnet 3D architecture, a variant of UNet 3D that combines Focal Loss and Generalized Dice Loss to handle imbalance. In 2025, Dhafer et al. [15] introduced a Squeeze-Excitation Dilated Attention-based Residual U-Net to address the diverse types and sizes of lung nodules and enhance lung nodule segmentation. After image preprocessing and extraction of regions of interest, the data were fed into the SEDARU-Net model, where the encoding part uses dilated convolutional layers. In 2026, YaXin et al. [16] presented CMS-UNet, an architecture using a two-branch encoder that combines the local segmentation capabilities of U-Net with the long-range dependency modeling capabilities of Mamba. In experiments, the authors verified results showing superior performance compared to other reference models.

Despite the many studies that have been done, the problem of lung tumor segmentation on CT images still faces many difficulties. Many artificial intelligent models process each slice independently, resulting in the inability to take advantage of spatial information in depth. This makes it difficult for the model to learn continuous spatial features between slices, while this is an important factor in determining the tumor boundary and shape more accurately in three-dimensional space. Meanwhile, some other models, that use 3D volumes, can take advantage of spatial information better, but in return, they have high computational costs, high memory requirements, and longer training times, making it difficult to implement in practice, especially for systems with limited resources.

To overcome this problem, this study proposes a more optimal approach, using spatially continuous data combination into 2.5D input, a deep learning approach with a hybrid architecture of CNN and Transformer along with 3D volumetric post-processing. This method allows the deep learning model to not only take advantage of spatial context information along the slices, but also significantly improve the accuracy in segmenting tumor boundaries. The additional 3D volumetric postprocessing helps to refine the results, eliminate discrete false positive regions, increase the continuity and reproduce more accurate tumor shapes in three-dimensional space. In addition, we also perform 3D reconstruction visualization of the tumor for clearer observation and important potential in supporting doctors in assessing the disease condition.

In the following sections, the paper is organized as follows: section II is the proposed methodology, which describes how we prepare the data, the techniques applied, and the experimental setup; section III is the results and evaluation; and finally, the conclusions.

2. Proposed Method

2.1. Data Preparation

The study used the NSCLC-Radiomics dataset [17] and the NSCLC-Radiogenomics dataset [18] for experiments. The NSCLC-Radiomics dataset consists of 3D chest CT images of patients diagnosed with non-small cell lung cancer. Each patient has a series of slices in the standard DICOM format with a fixed size of 512x512 pixels, covering the entire chest.

The tumor regions in the images were manually annotated by experts as binary masks. The data in the NSCLC-Radiomics dataset was divided into three sets, including 260 patients for the training set, 60 patients for the validation set, and 40 patients for the test set. The data in the NSCLC-Radiogenomics dataset was divided into the training set of 107 patients, 15 patients for the validation set, and the test set of 15 patients. The test patients were not used for model selection. Statistics on the number of slices with and without tumors in each set are presented in Table 1.

Table 1. Number of slices in the datasets

Category	Number of patients	Number of slices containing tumor	Number of slices without tumor
Train set (NSCLC- Radiomics)	260	4,296	26,951
Validation set (NSCLC- Radiomics)	60	1,088	6,412
Test set (NSCLC- Radiomics)	40	848	3,630
Train set (NSCLC- Radiogenomics)	107	1,723	15,516
Validation set (NSCLC- Radiogenomics)	15	256	2,099
Test set (NSCLC- Radiogenomics)	15	278	2,293

During data processing from datasets, slices and corresponding segmentation masks are extracted and processed automatically through our pipeline. We will extract both CT image data and associated segmentation information from DICOM files. Each CT image is converted to Hounsfield Unit (HU) values based on two parameters RescaleSlope (RS) and RescaleIntercept (RI) according to the formula:

$$HU = PixelValue \times RS + RI \quad (1)$$

The slices are then thresholded to a range of -910 to 590, which is the predominant value range of lung and soft tissues. The pixel values of the image slices are then normalized to a range of 0 to 255. The tumor segmentation mask is extracted from the GTV-1 (Gross Tumor Volume) annotation region in the DICOM segmentation object.

To increase the focus of the deep learning model on the likely tumor-containing region in the image, we define a global square bounding box covering all lung regions of the dataset to crop the image and resize it to 256x256 without distortion.

The 2.5D image used to train the model will be concatenated by $2k+1$ consecutive slices in the slice sequence of a patient as illustrated in Figure 1.

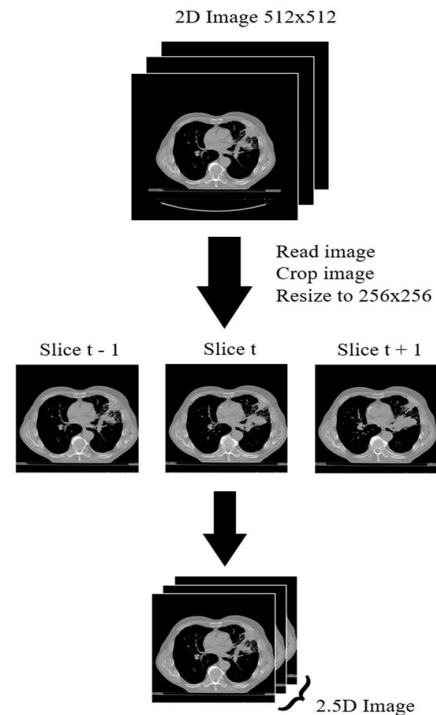


Figure 1. Process of making 2.5D images as input to neural network

Specifically, for the t^{th} slice in the axial axis, create an image $x_t \in \mathbb{R}^{256 \times 256 \times 3}$ by stacking $(t-k, \dots, t, \dots, t+k)$ in channel order, label y_t is the mask of the central slice t . At the beginning and end of the sequence, the missing channel is filled by repeating the slice.

2.2. CNN Architecture

In this study, we use the TransUNet network [19], which is a hybrid architecture between CNN and Transformer. This architecture is expected to exploit the advantages of both directions including the ability to capture global information of Transformer and the ability to preserve spatial details of CNN. Specifically, convolutional features facilitate the extraction of local spatial information from the tumor region, while the Transformer component enables the modeling of long-range spatial relationships within the lung slice. The Unet architecture, with its skip connections and decoder, also lays the foundation for preserving detailed spatial information that might be lost during feature encoding. This combination is expected to be well-suited for the task of lung tumor segmentation, given the significant diversity in tumor shapes and textural characteristics. The architecture is described in Figure 2.

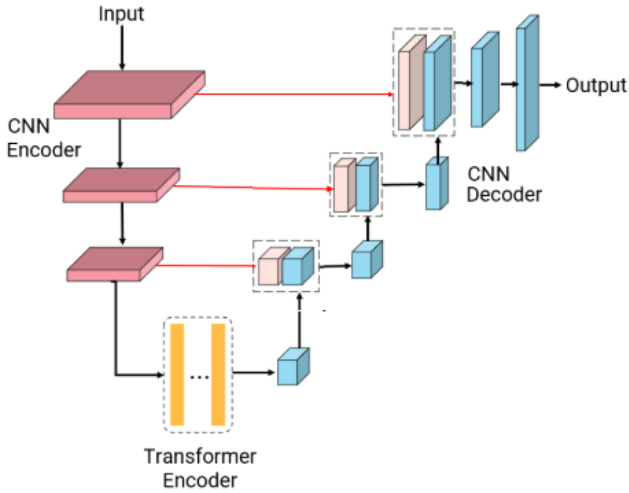


Figure 2. TransUNet Network Architecture Overview

The TransUNet architecture consists of three parts, namely CNN Encoder, Transformer Encoder, and CNN Decoder. To speed up training with large datasets, we customize the input image size of the network from 512 to 256, and to increase the representation learning ability, increase the hidden layer dimension in Transformer Encoder from 512 to 1024.

2.3. Experiment Setup

We conduct several experiments to evaluate the effectiveness of the method. The model training experiments are designed with three input configurations, including 2D input images, 2.5D-3, and 2.5D-5. 2.5D-3 is 2.5D with three consecutive slices. 2.5D-5 is 2.5D with five consecutive slices. This design aims to assess the impact of inter-slice spatial information. The 2D model serves as the baseline, while the 2.5D-3 and 2.5D-5 models incorporate varying levels of spatial contextual information. Subsequently, 3D volumetric post-processing is applied to the 2.5D results to evaluate the results before and after 3D post-processing. To increase the diversity of the training data, we apply data augmentation techniques with transformations such as horizontal flip, vertical flip, 90° rotation, free rotation in the range of -45° to 45° , slight scaling, image translation, and elastic transformation [20].

The model is implemented using the Pytorch library, an open-source deep learning library developed by Facebook AI Research [21]. The training process uses the Adam algorithm [22] with an initial learning rate of $1e-4$. To stabilize the model and avoid overfitting, we incorporate an early stopping strategy [23] if the validation loss does not improve after 5 consecutive epochs. In progress, if the validation loss does not decrease after 3 epochs, we reduce the learning rate by half. In our experiments, two datasets were used under a consistent configuration.

2.4. 3D Volumetric Post-Processing

The predicted binary masks are stored separately for each patient and arranged in the correct order. From this sequence of masks, we stack the slices in order to obtain the binary volume $\in \{0,1\}^{Z \times Y \times X}$. CT image data is typically anisotropic with slice thickness Δ_z being significantly different with pixel size Δ_x, Δ_y . Regarding voxel spacing, the NSCLC-Radiomics dataset has a spacing_z value of 3.0 mm. For spacing_x and spacing_y, the minimum value is 0.7207 mm and the maximum is 0.9770 mm. The NSCLC-Radiogenomics dataset has a spacing_z value of 2.0 mm. For spacing_x and spacing_y, the minimum value is 0.9366 mm and the maximum is 1.4496 mm. All parameters for the morphology are set in millimeters and then converted to the number of voxels on each axis to preserve the geometric meaning. Specifically, for a desired radius r (mm), the radius in voxels is:

$$r_x = \text{round}(\frac{r}{\Delta_x}), r_y = \text{round}(\frac{r}{\Delta_y}), r_z = \text{round}(\frac{r}{\Delta_z}) \quad (2)$$

In formula, the structuring element is a discrete ellipsoid of size $(2r_z + 1) \times (2r_y + 1) \times (2r_x + 1)$. This ensures that the 3mm expansion is actually 3mm in all directions.

First, the 3D volume is closed and opened with two ellipsoids of size r_{close} and r_{open} (usually, $r_{close} \approx 2 - 5 \text{ mm}$, $r_{open} \approx 1 - 3 \text{ mm}$). The closing method fills the narrow gaps and connects the adjacent regions, the opening

method removes small noises outside the boundary. Next, fill the small holes inside the predicted tumor volume. To reduce the discretization, the 3D connected components are separated and only those with volumes larger than a threshold τ (mm^3) are retained. In the implementation, this threshold is converted to the minimum voxel number $n_{min} = \left\lceil \frac{\tau}{\Delta_x \Delta_y \Delta_z} \right\rceil$, and then a small component removal is performed to remove the tiny positive parts. Depending on the clinical context, a component selection step can be applied after filtering: keep only the largest component or keep the top-k largest components. Here, we choose to keep only one largest component.

After volumetric post-processing, 3D mesh construction was performed using Marching Cubes [24] to visualize tumor shape and verify segmentation quality.

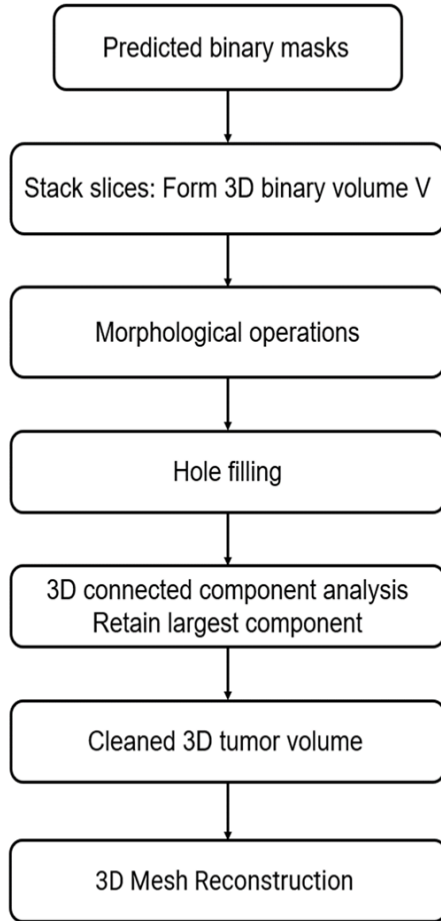


Figure 3. Our 3D volumetric post-processing pipeline

2.5. Evaluation Metrics

The performance of the model was evaluated by the quality in tumor location segmentation using two metrics, the Dice coefficient and IoU.

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

In which, A is the prediction mask, and B is the ground truth. The Dice and IoU coefficients have values from 0 to 1, the closer to 1, the higher the match. The convention is that if the ground truth does not contain u and the prediction mask does not contain u, the Dice and IoU coefficients are 1. In addition to region-based metrics such as Dice and IoU, HD95 and ASSD are used to evaluate the results of tumor region segmentation. HD95 measures the Hausdorff distance at the 95th percentile, helping to mitigate the impact of extreme boundary outliers. ASSD measures the average symmetric surface distance, reflecting the average deviation of the segmentation boundary. The unit for both HD95 and ASSD is mm.

To assess practical applicability, the study provides data on inference time and resource usage. Runtime reflects real-world data processing capabilities, while memory requirements indicate hardware needs and suitability for resource-constrained systems. These metrics complement the evaluation of the model's deployment feasibility in clinical application environments.

3. Result and Evaluation

In this section, we present experimental results of the tumor segmentation model on the NSCLC-Radiomics dataset and the NSCLC-Radiogenomics. Experimental results focus on comparing the performance between the baseline model, which uses a single slice as input, and the proposed model with 2.5D approach. We use 2.5D-3, which uses three consecutive slices and 2.5D-5, which use five consecutive slices. The evaluation is mainly based on the Dice and IoU coefficients calculated on the 3D volumes of each patient, as these metrics more accurately reflect the clinical performance than the global slice metrics, which are heavily influenced by slices having no tumor.

We analyze tumor segmentation performance metrics for the NSCLC-Radiomics dataset without applying a 3D post-processing procedure. Details are described in Table 2. In this table, Dice 2D is dice over slices, Dice 3D is dice over 3D volumes, IoU 2D is IoU over slices, and IoU 3D is IoU over 3D volumes.

Table 2. Results for the NSCLC-Radiomics without the 3D volumetric post-processing

Category	2D	2.5D-3	2.5D-5
Dice 3D	0.580 $\pm$ 0.235	0.622 $\pm$ 0.233	0.625 $\pm$ 0.233
IoU 3D	0.458 $\pm$ 0.221	0.488 $\pm$ 0.225	0.485 $\pm$ 0.227
HD95	42.586 $\pm$ 37.137	28.080 $\pm$ 22.740	27.740 $\pm$ 18.920
ASSD	10.785 $\pm$ 6.835	9.769 $\pm$ 5.098	9.940 $\pm$ 6.050
Dice 2D	0.869 $\pm$ 0.307	0.874 $\pm$ 0.298	0.875 $\pm$ 0.298
IoU 2D	0.855 $\pm$ 0.317	0.858 $\pm$ 0.310	0.860 $\pm$ 0.310

The experimental results in Table 2 demonstrate a clear improvement in model performance when switching the input from 2D to 2.5D. The average 3D Dice score increased from 0.580 for the 2D input case to 0.622 for the 2.5D-3 case, approximately 7.24%, and 0.625 for the 2.5D-5 case, approximately 7.76%. The average 3D IoU increased from 0.458 for the 2D input case to 0.488 for the 2.5D-3 case, approximately 6.55%, and 0.485 for the 2.5D-5 case, approximately 5.9%. The average 2D Dice score increased from 0.869 for the 2D input case to 0.874 for the 2.5D-3 case and 0.875 for the 2.5D-5 case. The average 2D IoU increased from 0.855 for the 2D input case to 0.858 for the 2.5D-3 case and 0.860 for the 2.5D-5 case. The average HD95 improved from 42.586 for the 2D input case to 28.080 for the 2.5D-3 case and 27.740 for the 2.5D-5 case. The average ASSD

improved from 10.785 for the 2D input case to 9.769 for the 2.5D-3 case and 9.940 for the 2.5D-5 case.

Additionally, it is worth noting that the results show no significant difference in the metrics between the 2.5D-3 and 2.5D-5 approaches. This indicates that increasing the number of neighboring slices from 3 to 5 did not yield a noticeable improvement. This is likely because the additional slices contain redundant spatial information due to the tumor's size. Therefore, a preliminary conclusion is that the 2.5D-3 method can provide comparable performance with the 2.5D-5 method. To clarify further, we continue to analyze tumor segmentation performance metrics for the NSCLC-Radiogenomics dataset without applying a 3D post-processing procedure. Details are described in Table 3.

Table 3. Results for the NSCLC-Radiogenomics without the 3D volumetric post-processing

Category	2D	2.5D-3	2.5D-5
Dice 3D	0.567 $\pm$ 0.248	0.594 $\pm$ 0.206	0.618 $\pm$ 0.238
IoU 3D	0.432 $\pm$ 0.226	0.472 $\pm$ 0.182	0.482 $\pm$ 0.223
HD95	120.724 $\pm$ 40.680	105.108 $\pm$ 85.892	88.233 $\pm$ 81.567
ASSD	36.464 $\pm$ 19.538	32.652 $\pm$ 35.678	26.594 $\pm$ 32.350
Dice 2D	0.863 $\pm$ 0.324	0.896 $\pm$ 0.279	0.921 $\pm$ 0.257
IoU 2D	0.852 $\pm$ 0.331	0.885 $\pm$ 0.289	0.895 $\pm$ 0.278

Similar to Table 2, the experimental results in Table 3 demonstrate a clear improvement in model performance when switching the input from 2D to 2.5D. The average 3D Dice score increased from 0.567 for the 2D input case to 0.594 for the 2.5D-3 case, approximately 4.76%, and 0.618

for the 2.5D-5 case, approximately 8.99%. The average 3D IoU score increased from 0.432 for the 2D input case to 0.472 for the 2.5D-3 case, approximately 9.26%, and 0.482 for the 2.5D-5 case, approximately 11.57%. The average HD95 score improved from 120.724 for the 2D input case to

105.108 for the 2.5D-3 case and 88.233 for the 2.5D-5 case. The average ASSD score improved from 36.464 for the 2D input case to 32.652 for the 2.5D-3 case and 26.594 for the 2.5D-5 case. The average 2D Dice score increased from 0.863 for the 2D input case to 0.896 for the 2.5D-3 case, approximately 3.82%, and 0.921 for the 2.5D-5 case, approximately 6.72%. The average 2D IoU score increased from 0.852 for the 2D input case to 0.885 for the 2.5D-3 case, approximately 3.87%, and 0.895 for the 2.5D-5 case, approximately 5.05%. Furthermore, consistent with the previous observation, the results for the 2.5D-3 case do not differ significantly from those of the 2.5D-5 case, although the range of variation is wider than that observed with the NSCLC-Radiomics dataset.

Next, we evaluate the improvement in result quality for cases 2.5D-3 and 2.5D-5 when applying 3D post-processing. We begin with the NSCLC-Radiomics dataset. Details are described in Table 4.

Table 4. 2.5D results for the NSCLC-Radiomics with the 3D volumetric post-processing

Category	2.5D-3	2.5D-5
Dice 3D	0.641 $\pm$ 0.235	0.632 $\pm$ 0.233
IoU 3D	0.508 $\pm$ 0.228	0.488 $\pm$ 0.228
HD95	23.977 $\pm$ 17.088	24.166 $\pm$ 19.051
ASSD	8.181 $\pm$ 6.503	8.591 $\pm$ 5.930
Dice 2D	0.891 $\pm$ 0.277	0.889 $\pm$ 0.281
IoU 2D	0.875 $\pm$ 0.291	0.874 $\pm$ 0.293

Table 4 demonstrates a marked improvement in the 2.5D-3 and 2.5D-5 results on the NSCLC-Radiomics dataset following the application of 3D post-processing. This improvement is evident across all metrics. Specifically, for the 2.5D-3 case, the mean 3D Dice score increased from 0.622 to 0.641 (an increase of approximately 3.05%), and the mean 3D IoU rose from 0.488 to 0.508 (approximately 4.1%). The mean 2D Dice score increased from 0.874 to 0.891 (approximately 1.95%), while the mean 2D IoU rose from 0.858 to 0.875 (approximately 1.98%). The mean HD95 improved from 28.080 to 23.977, and the mean ASD improved from 9.769 to 8.181. For the 2.5D-5 case, the mean 3D Dice score increased from 0.625 to 0.632 (approximately 1.12%), and the mean 3D IoU rose from 0.485 to 0.488 (approximately 0.62%). The mean HD95 improved from 27.740 to 24.166, and the mean ASD improved from 9.940 to 8.591. Additionally, the mean 2D Dice score increased from 0.875 to 0.889 (approximately 1.6%), and the mean 2D IoU rose from 0.860 to 0.874 (approximately 1.63%). These results clearly demonstrate the benefits of applying 3D post-processing.

Additionally, it is worth noting that the rate of improvement in the 2.5D-5 cases is generally lower than in the 2.5D-3 cases. Consequently, the final metrics for the 2.5D-3 cases are superior. This contrasts with the results observed prior to the application of 3D post-processing, although the difference is not significant. We continue to results for the NSCLC-Radiogenomics dataset with applying a 3D post-processing procedure. Details are described in Table 5.

Table 5. 2.5D results for the NSCLC-Radiogenomics with the 3D volumetric post-processing

Category	2.5D-3	2.5D-5
Dice 3D	0.668 $\pm$ 0.205	0.671 $\pm$ 0.231
IoU 3D	0.528 $\pm$ 0.179	0.539 $\pm$ 0.223
HD95	10.781 $\pm$ 18.159	12.590 $\pm$ 17.276
ASSD	6.501 $\pm$ 16.016	7.007 $\pm$ 15.461
Dice 2D	0.946 $\pm$ 0.191	0.951 $\pm$ 0.186
IoU 2D	0.935 $\pm$ 0.207	0.941 $\pm$ 0.200

Similar to Table 4, Table 5 also demonstrates improvements following the application of 3D post-processing. Specifically, for the 2.5D-5 case, the average 3D Dice score increased from 0.618 to 0.671 (approximately 8.58%), and the average 3D IoU rose from 0.482 to 0.539 (approximately 11.83%). The average HD95 improved from 88.233 to 12.590, and the average ASD improved from 26.594 to 7.007. The average 2D Dice score increased from 0.921 to 0.951 (approximately 3.26%), while the average 2D IoU rose from 0.895 to 0.941 (approximately 5.14%). For the 2.5D-3 case, the average 3D Dice score increased from 0.594 to 0.668 (approximately 12.46%), and the average 3D IoU rose from 0.472 to 0.528 (approximately 11.86%). The average HD95 improved from 105.108 to 10.781, and the average ASD improved from 32.652 to 6.501. The average 2D Dice score increased from 0.896 to 0.946 (approximately 5.58%), while the average 2D IoU rose from 0.885 to 0.935 (approximately 5.65%).

Furthermore, when comparing the results of the 2.5D-3 and 2.5D-5 cases, there is no clear superiority of one over the other. With 3D post-processing, while certain metrics for the 2.5D-3 case are higher than those for the 2.5D-5 case, other metrics show the opposite trend. This further reinforces the earlier observation that there is no distinct difference between the 2.5D-3 and 2.5D-5 results.

To provide a clearer assessment of the performance differences between the 2.5D-3 and 2.5D-5 configurations in this study, we performed a two-sided Wilcoxon signed-rank test on patient-specific outcomes in the test sets of the NSCLC-Radiomics and NSCLC-Radiogenomics datasets.

Four measures were considered: Dice 3D, IoU 3D, HD95, and ASSD. The significance level was set at $\alpha = 0.05$. In NSCLC-Radiomics, no statistically significant differences were observed between 2.5D-3 and 2.5D-5 for Dice 3D ($p=0.226$), IoU 3D ($p=0.249$), HD95 ($p=0.718$), and ASSD ($p=0.387$). Similarly, on NSCLC-Radiogenomics, no statistically significant differences were observed for Dice 3D ($p=1.000$), IoU 3D ($p=0.952$), HD95 ($p=0.588$), and ASSD ($p=0.639$). Therefore, the test results do not provide sufficient evidence to conclude that the segmentation performance of

the 2.5D-3 and 2.5D-5 configurations differs significantly in statistical terms.

To provide another perspective illustrating the effectiveness of the proposed method, we calculate the False Positive and False Negative values. Specifically, we will evaluate the measure of actual tumors in the slice but predicted no tumor, and actual no tumor in the slice but predicted tumors.

Table 6. Results of cases on the test set in terms of False Positive and False Negative

Category	NSCLC-Radiomics (FP/FN)	NSCLC-Radiogenomics (FP/FN)
2D	166/250	228/65
2.5D-3	166/215	145/56
2.5D-3 + 3D volumetric post-processing	94/224	14/66
2.5D-5	156/229	125/49
2.5D-5 + 3D volumetric post-processing	95/232	8/67

Table 6 clearly illustrates the changes in False Positive and False Negative values across the test cases. Applying the 2.5D approach resulted in clear improvements in these metrics. For the NSCLC-Radiomics dataset, compared to the 2D case, the number of False Negatives dropped from 250 to 215 for the 2.5D-3 case and to 229 for the 2.5D-5 case. For the NSCLC-Radiogenomics dataset, False Positives decreased from 228 to 145 for the 2.5D-3 case and to 125 for the 2.5D-5 case. Furthermore, the application of 3D volumetric post-processing led to a significant reduction in False Positives. For the 2.5D-3 case on the NSCLC-Radiomics dataset, False Positives fell from 166 to 94. For the 2.5D-5 case, they dropped from 156 to 95. For the NSCLC-Radiogenomics dataset, False Positives decreased from 145 to 14 for the 2.5D-3 case and from 125 to 8 for the 2.5D-5 case.

Table 6 shows that 3D post-processing slightly increases the false-negative rate, and retaining only the largest connected component can discard multifocal lesions. On the one hand, the results suggest that 3D post-processing may

create a trade-off in processing performance. Removing small connected components reduces the predicted areas of noise or false positives. However, this process may also slightly increase false negatives, especially for small lesions or those isolated from the main lesion. On the other hand, the strategy of retaining only the largest connected component assumes that the target area is primarily composed of a single continuous structure. This assumption may not be suitable for multifocal lesions. However, it may still be relevant in clinical practice where identifying the location and presence of the lesion is crucial. Often, support systems play a role in providing a warning signal to help physicians quickly notice and thoroughly examine the case. Therefore, in subsequent practical implementation, the application of the technique also needs to be customized and adjusted to ensure convenience in supporting doctors and medical professionals.

Next, we examine the inference time and resource usage for the cases. The results were evaluated on both CPU and GPU hardware platforms. Details are presented in Table 7.

Table 7. Inference time and resource usage

Category	2D	2.5D-3
Inference time (CPU)	9.23 samples/s	9.01 samples/s
Inference time (GPU)	174.63 samples/s	172.81 samples/s
Resource usage (CPU)	1971.57 MB	2277.64 MB
Resource usage (GPU)	1477.57 MB RAM, 773.34 MB GPU memory	1511.51 MB RAM, 777.44 MB GPU memory

Table 7 presents runtime and resource usage data for the study. Given the differing processing methods across configurations, inference time is calculated based on the number of samples. For instance, in the 2D case, a sample consists of a single image containing one slice. For the 2.5D-3 case, a sample comprises an image with three slices. For the 2.5D-5 case, a sample consists of an image with five slices. Since processing time is essentially determined by the network's execution time, the 2D, 2.5D-3, and 2.5D-5 configurations exhibit roughly similar processing times, regardless of whether a CPU or GPU is used. Overall, processing speeds are approximately 9 samples/s on the CPU and 171 samples/s on the GPU. Additionally, the 3D volumetric post-processing time is approximately 7.049 ± 5.379 seconds per scan. Memory usage exceeds 2000 MB, with allocation varying based on the hardware (CPU or GPU). In a real-world clinical context, these memory and runtime requirements are suitable for personal computers during the testing phase, prior to deploying a GPU server to handle high-volume workloads.

In addition to the quantitative results, we also demonstrate the effectiveness of the proposed method through visualization. Some comparison images between 2D and 2.5D cases are illustrated in Figure 4.

Figure 4 presents the detailed prediction results in images on some slices when comparing the 2D and 2.5D cases. The first line is Ground Truth, the second line is the 2D case segmentation result, the third line is the 2.5D-3 case segmentation result, the fourth line is the 2.5D-3 case segmentation result with 3D volumetric post-processing, the fifth line is the 2.5D-5 case segmentation result, and the sixth line is the 2.5D-5 case segmentation result with 3D volumetric post-processing. We can see an improvement in 2.5D results compared to 2D, and an improvement in 2.5D results with post-processing compared to without. Accordingly, the best results achieved with Dice were 0.793, 0.893, and 0.9.

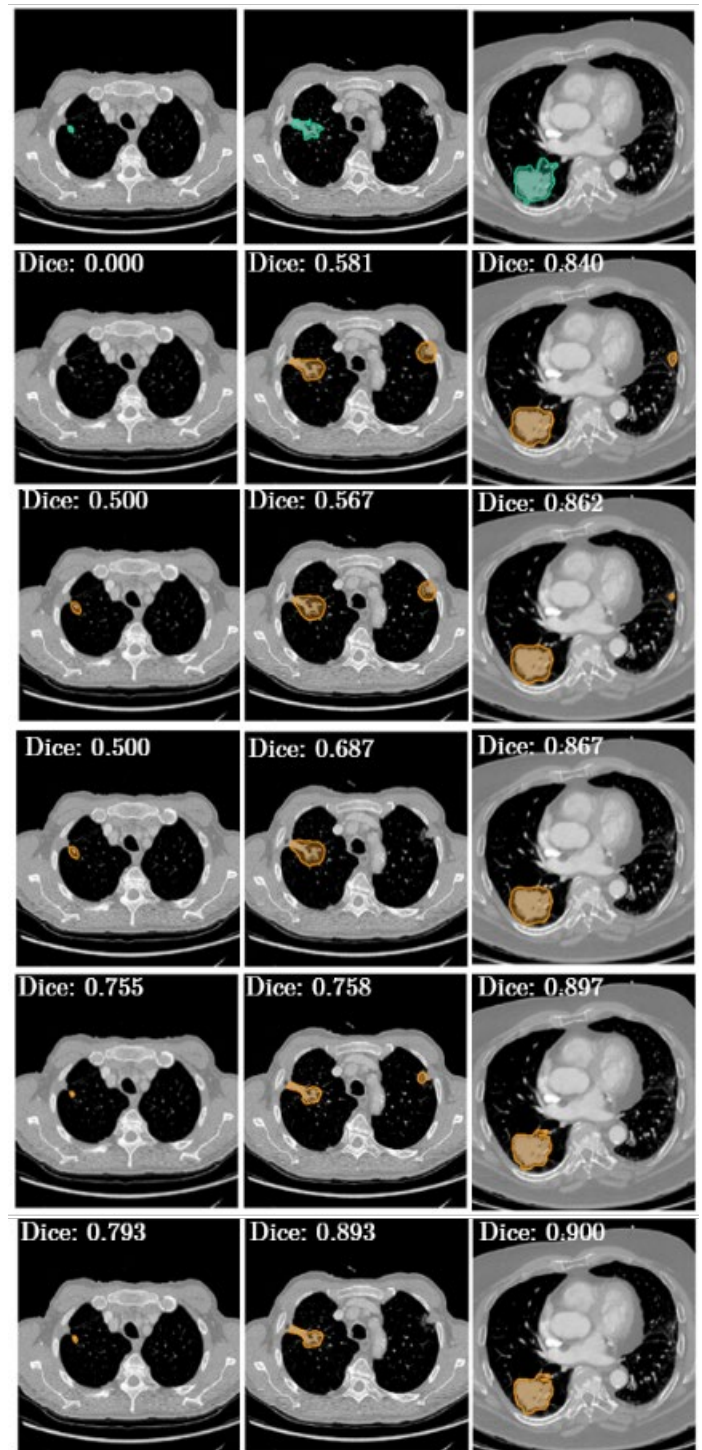


Figure 4. Some comparison images between 2D and 2.5D cases

Next, we also visually illustrate the effectiveness of applying the 3D volumetric post-processing technique for 2.5D-5 results from a 3D perspective. The details are shown in Figure 5.

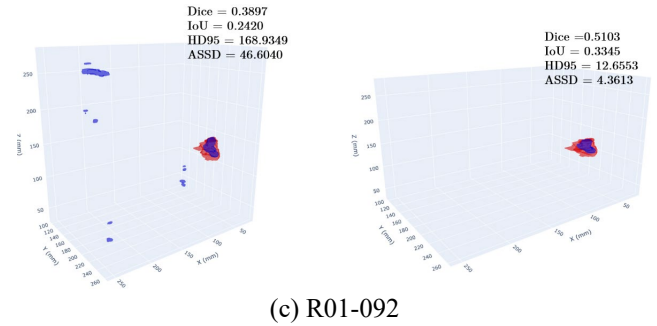
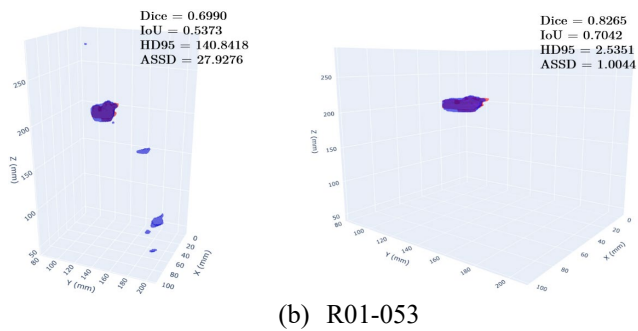
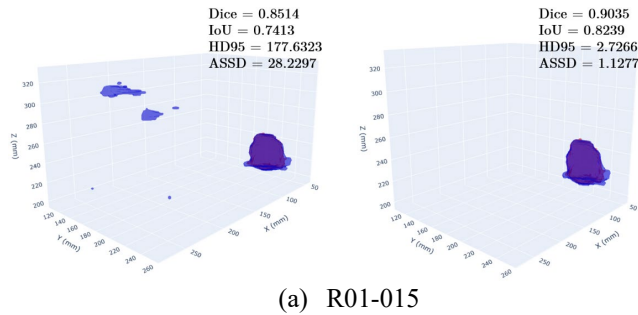


Figure 5. Some 3D perspective images illustrate the results with and without 3D volumetric post-processing

Figure 5 illustrates the 3D reconstruction results of the segmented tumor on samples R01-015, R01-092 and R01-092. Without 3D volumetric post-processing, the reconstructed 3D volume, although it fits relatively well to the ground truth, still has small, discrete structures or rough edges, leading to suboptimal Dice and IoU values. After applying 3D volumetric post-processing, the many small areas are removed, the tumor edges become smoother, and the spatial continuity is significantly improved.

Overall, the illustrations confirm the presented quantitative results. The 2.5D case provides a clear improvement compared to the 2D case by taking advantage of the continuity between slices, while the 3D volumetric post-processing step continues to improve the segmentation quality, both reducing noise and increasing the spatial consistency of the tumor.

We also evaluated the obtained results with the published results of Kamal et al [12] with Recurrent3D-DenseUNet method and Farheen et al [11] with Deeply Supervised MultiResUnet method. The data were analyzed with Dice average over slices, False Positive and False Negative. Details are shown in Table 8.

Table 8. Comparison of our results with previous studies

Category	False Positive	False Negative	Dice average over slices
Kamal et al [12]	321	331	0.7228
Farheen et al [11]	123	324	0.8472
Our proposed method (2.5D-3)	94	224	0.891
Our proposed method (2.5D-5)	95	232	0.889

From the results in Table 8, it can be seen that our study has achieved relatively better results than Kamal et al and

Farheen et al. In particular, in terms of segmentation measure with Dice average, we achieved 0.891 in case of 2.5D-3 and

0.889 in case of 2.5D-5, in which Kamal et al achieved 0.7228 and Farheen et al achieved 0.8472. In terms of the two indices False Negative and False Positive, related to the actual tumor in the slice but the prediction of no tumor, and the actual tumor in the slice but the prediction of tumor, we also achieved better results. This also partly proves our research hypothesis, with the combination of many factors of spatial information between slices, the use of a deep learning approach, and 3D volume post-processing.

4. Conclusion

In this study, we proposed a lung tumor segmentation method based on exploiting spatial information between slices, using a deep learning approach with the TransUNet model combining CNN and Transformer, along with an important step of 3D volumetric post-processing. Experimental results on the NSCLC-Radiomics dataset and the NSCLC-Radiogenomics dataset show that the model achieves high segmentation efficiency, and the resulting illustrations also show improvements in the ability to remove small noise blocks. In addition, we also illustrate the 3D image reconstruction of the tumor, which promises to help visualize clearly and support doctors in assessing the disease status.

Several limitations of the study need to be considered. First, the test set was not very large, with 40 patients on the NSCLC-Radiomics set and 15 patients on the NSCLC-Radiogenomics set. Therefore, the evaluation results may not fully reflect the generalizability of the method and require larger trials in the future. Second, the single-annotation ground truth may contain subjective biases of the evaluator regarding unclear lesion areas. This could lead to some cases being considered false positives or false negatives related to annotation uncertainty. This issue requires more detailed verification in practical implementation. Finally, comparisons with previous studies such as Kamal et al. and Farheen et al. rely on published figures that may use different data splits. Therefore, the comparative results should be considered as a reference for performance rather than a completely direct comparison. These are valuable factors that will be considered in our further research.

In the future, our research direction will continue to delve into deep learning methods as well as 3D techniques. In addition, we will also aim to expand the experiment on patient data in Vietnam to verify the research results as well as bring the research results to reality in the medical industry. With the current positive results, we hope that the research can open up potential practical applications in lung cancer prevention in Vietnam.

References

- [1] World Health Organization. World Fact Sheet. 2022. [Online]. Available: <https://geo.iarc.who.int/media/globocan/factsheets/population/s/900-world-fact-sheet.pdf> [accessed: 2025 September 23].
- [2] World Health Organization. Viet Nam Fact Sheet. 2022. [Online]. Available: <https://geo.iarc.who.int/media/globocan/factsheets/population/s/704-viet-nam-fact-sheet.pdf> [accessed: 2025 September 23].
- [3] Ming S., Yang W., Cui S., Huang S., Gong X.. Consistency of radiologists in identifying pulmonary nodules based on low-dose computed tomography. *Journal of thoracic disease*. 2019;11(7):2973.
- [4] Sangamithraa P., Govindaraju S.. Lung tumour detection and classification using EK-Mean clustering. In: 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). 2016. p. 2201-2206.
- [5] Kaucha D.P., Prasad P.W.C., Alsadoon A., Elchouemi A., Sreedharan S.. Early detection of lung cancer using SVM classifier in biomedical image processing. In: 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). 2017. p. 3143-3148.
- [6] Patel T., Nayak V.. Hybrid Approach for Feature Extraction of Lung Cancer Detection. In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). 2018. p. 1431-1433.
- [7] Perumal S., Thambusamy V.. Lung cancer detection and classification on CT scan images using Enhanced Artificial Bee Colony Optimization. *International Journal of Engineering & Technology*. 2018;7:74.
- [8] Alakwaa W., Nassef M., Badr A.. Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN). *International Journal of Advanced Computer Science and Applications*. 2017;8.
- [9] Kuan K., Ravaut M., Manek G., Chen H., Lin J., Nazir B., Chen C., Howe T.C., Zeng Z., Chandrasekhar V.. Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge, arXiv preprint arXiv:1705.09435. 2017.
- [10] Hossain S., Najeeb S., Shahriyar A., Abdullah Z.R., Ariful Haque M.. A Pipeline for Lung Tumor Detection and Segmentation from CT Scans Using Dilated Convolutional Neural Networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019. p. 1348-1352.
- [11] Farheen F., Shamil M.S., Ibtehaiz N., Rahman M.. Revisiting segmentation of lung tumors from CT images. *Computers in Biology and Medicine*. 2022;144:105385.
- [12] Kamal U., Rafi A.M., Hoque R., Wu J., Hasan M.K.. Lung Cancer Tumor Region Segmentation Using Recurrent 3D-DenseUNet, arXiv preprint arXiv:1812.01951. 2020.
- [13] Dutande P., Baid U., Talbar S.. LNCDS: A 2D-3D cascaded CNN approach for lung nodule classification, detection and segmentation. *Biomedical Signal Processing and Control*. 2021;67:102527.
- [14] Le V., Saut O.. RRe-UNet 3D for lung tumor segmentation from CT scans of Non-Small Cell Lung Cancer patients. In: 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2023. p. 2308-2317.
- [15] Alhajim D., Ansari-Asl K., Akbarizadeh G., Soorki M.N.. Improved lung nodule segmentation with a squeeze excitation dilated attention based residual UNet. *Sci Rep*. 2025; 15(1):3770.
- [16] YaXin Z., Cheng J., RenZheng X., Zheng L., Qing Z., DaTong C., DongBin X., Yue S., GongQi L., MingHui L.. CMS-UNet: A CNN-Mamba Synergistic Dual-Encoder u-net for lung tumor segmentation in CT images. *Biomedical Signal Processing and Control*. 2026; 116:109451.
- [17] Aerts H. J. W. L., Wee L., Rios Velazquez E., Leijenaar R. T. H., Parmar C., Grossmann P., Carvalho S., Bussink J., Monshouwer R., Haibe-Kains B., Rietveld D., Hoebbers F., Rietbergen M. M., Leemans C. R., Dekker A., Quackenbush

- J., Gillies R. J., Lambin P. Data From NSCLC-Radiomics (version 4) [Data set]. The Cancer Imaging Archive. 2014.
- [18] Bakr S., Gevaert O., Echegaray S., Ayers K., Zhou M., Shafiq M., Zheng H., Zhang W., Leung A., Kadoch M., Shrager J., Quon A., Rubin D., Plevritis S., Napel S.. Data for NSCLC Radiogenomics Collection (Version 4) [Dataset]. The Cancer Imaging Archive. 2017.
- [19] Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y., Lu L., Yuille A.L., Zhou Y.. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306. 2021.
- [20] Castro E., Cardoso J.S., Pereira J.C.. Elastic deformations for data augmentation in breast cancer mass detection. In: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE. 2018. p. 230-234.
- [21] Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., Killeen T., Lin Z., Gimelshein N., Antiga L., others. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems. 2019;32.
- [22] Kingma D.P., Ba J.. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR).. 2015.
- [23] Prechelt L.. Automatic early stopping using cross validation: quantifying the criteria. Neural networks. 1998;11(4):761-767.
- [24] Lorensen W.E., Cline H.E.. Marching cubes: A high resolution 3D surface construction algorithm. ACM SIGGRAPH Computer Graphics. 1987;21(4):163-169.