Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Lungs And Nodules Segmentation from Computed Tomography Scans Using Auxiliary U-Net

Abderrahim El yahyaouy[1], Abdelkader Hadjoudja[2], Abdelmajid EL Moutaouakkil[3], Rachida Latib[4], Youssef Omor[5], Hamza Retal[6]

[1,2]SETIME Laboratory, Faculty of Sciences, Ibn Tofail University, BP 133, 14000 Kenitra, Morocco.
[3]LAROSERI Laboratory, Faculty of Sciences, University of Chouaib Doukkali, El Jadida, Morocco.
[4-6]Radiology Department, National Institute of Oncology, Ibn Sina University Hospital, Rabat, Morocco.

E-mail: [1]abderrahim.elyahyaouy@uit.ac.ma, [2]abdelkader.hadjoudja@uit.ac.ma, [3]elmou01@yahoo.com, [4]rachidalatib@gmail.com, [5]omor.youssef3@gmail.com, [6]retal.hamza95@gmail.com

Abstract

Due to the increasing prevalence of lung cancer and the urgent need for artificial intelligence technologies to support medical fields, there is an immediate demand for highly accurate and dependable automated diagnostic models for the segmentation of chest CT scans. Accurate segmentation will reduce clinicians' manual labor and improve diagnostic results. This paper proposes a new segmentation model, which combines an improved U-Net architecture with a squeeze-and-excitation mechanism. The model can utilize the loss in a multilevel fashion, thanks to the auxiliary classification at the bottleneck layer. Furthermore, the squeeze-and-excitation block focuses on the key characteristics extracted from convolutional layers. Experiments on the LUNA16 dataset were carried out to train and evaluate the model. The proposed approach outperformed many other segmentation methods and was very effective in defining the shape of lung nodules more precisely. On the LUNA16 dataset, it achieved 98.42% DSC and 98.56% accuracy. It has been proven that the segmentation of lung nodules in chest CT scans using an auxiliary U-Net architecture together with a squeeze-and-excitation mechanism is highly efficient and cost-effective. High accuracy and Dice coefficient against the most advanced segmentation models demonstrate the possibility of clinical usage and provide accurate boundary detection, which is crucial for an accurate diagnosis and treatment. The proposed contraction and stimulation mechanism with an auxiliary U-Net architecture has achieved very effective yet inexpensive lung nodule segmentation on chest CT scans, compared to state-of-the-art segmentation models. Accurate identification of the boundaries is of paramount importance for diagnosis and treatment planning, and the high Dice coefficient and accuracy support its use in clinical settings.

Keywords: LUNA16, Lung Nodules, Segmentation, Auxiliary Loss, Squeeze and Excitation Block, U-Net.

DOI: https://doi.org/10.36548/jiip.2025.4.013
Received: 22.09.2025, received in revised form: 20.10.2025, accepted: 03.11.2025, published: 12.11.2025

1. Introduction

Lung cancer is one of the deadliest cancers, ranking second in cancer deaths among men and women, according to US cancer statistics [1], one of the earliest symptoms of lung cancer is lung nodules, which are detected by computed tomography (CT) scans.

Several solutions for diagnosing lung nodules on CT have been presented recently [2], including detecting lung nodules from radiographic images using artificial intelligence, and classifying them [3]; [4]. There are several datasets that can be used in research aimed at analyzing and classifying lung nodules for early cancer detection, such as the LIDC-IDRI dataset [5] and LUNA16 dataset [6]. These two datasets are used to detect lung nodules, other datasets are also available for studying the progression of lung nodules, such as the NSL [7] and NELSON datasets [8].

Despite the advantages of the LUNA16 dataset, it is very large in size (about 65 GB), because each image consists of 121 slices. For processing in the 3D domain, this requires very large resources Moreover, most research has focused on segmenting lung nodules without addressing the importance of segmenting the lungs. Since most clinics and hospitals still use two-dimensional imaging for diagnosis due to its low cost—both financially and computationally and time efficiency, this research proposes a novel model for segmenting both the lungs and lung nodules in the two-dimensional domain, achieving high performance, practicality, economy, and speed.

1.1 Available Resources

In this research, the researchers use Google Colab, which provides resources (15 GB GPU, 127 GB RAM, and 70 GB hard disk) to implement the proposed model. As known, architectures like UNET require very large resources, especially recent versions like UNET++ and others.

1.2 Limitation

Due to limited resources, a set of operations was performed that could be modified if sufficient resources were available:

1. Scaling images to 320x320
2. Using images in 2D instead of 3D, and limiting the middle slice only
3. Using separable convolution instead of traditional convolution. Not augmenting the dataset
4. The dataset was divided into 90% for training (799 images) and 10% for validation (89 images) due to the small number of images and the inability to augment the dataset

In this research, the researchers use accuracy and DSC to evaluate performance because accuracy alone is not sufficient to evaluate performance in cases of classes imbalance. DSC is the best metric to evaluate the performance of segmentation systems in case of class imbalance and is more sensitive to small details than other measures, so it is mainly used in medical systems.

2. Literature Review

This section delves deeply into several important studies in the field of lung nodule segmentation, which is a crucial step towards early lung cancer detection and improving diagnostic accuracy.

One prominent study in this domain, conducted on the LUNA16 dataset, utilized a 3D-Unet [9] for lung nodule segmentation. The 3D version of the U-Net architecture is particularly well-suited for volumetric data such as Computed Tomography (CT) scans, as it can process contextual information across all three axes (length, width, depth). This capability allows it to capture complex relationships between voxels within the nodules. The researchers in this study achieved a Dice Similarity Coefficient (DSC) of 7878%. The Dice coefficient is a widely used metric for evaluating segmentation accuracy, measuring the overlap between the predicted segmentation and the ground truth, with higher values indicating better performance.

In a related context, other researchers in study [10] classified and segmented lung nodules using the pre-trained Segment Anything Model (SAM), a foundation model known for its general object segmentation capabilities This study demonstrates the effectiveness of pre-trained models in specialized medical tasks, as they achieved a classification accuracy of 9671% and a segmentation DSC of 9708% on the same LUNA16 dataset, this particularly high segmentation performance indicates the immense potential of large, broadly trained deep learning models in medical segmentation tasks, even when applied to a narrower scope.

Furthermore, [11] aimed to develop a deep learning-based system for lung parenchyma nodule detection and segmentation, their proposed algorithm was split into two main stages: The first stage involved lung segmentation using the previously developed LungQuant algorithm to determine the Region of Interest (ROI) and reduce the volume of data requiring processing, the second stage focused on nodule segmentation itself, employing a specially designed and optimized Attention Res-Unet for this task Attention mechanisms are crucial in neural networks as they allow the model to selectively focus on the most important parts of an image, thereby enhancing segmentation accuracy Additionally, the study demonstrated the reliability of the proposed algorithm for clinical use by employing Grad-CAM, an Explainable AI (XAI) technique Grad-CAM enables clinicians and researchers to understand which parts of an image the model focuses on to make its decisions, increasing trust in the system, the average Dice Similarity Coefficient (DSC) for the LungQuant algorithm in the first phase was 90%, while the DSC scores for nodule segmentation reached 81% on test sets Moreover, the model's average sensitivity and specificity measures were 086 and 092, respectively, which are important metrics for evaluating the model's ability to detect true nodules and avoid false alarms.

In another study, [12] conducted research on the Far Eastern Memorial Hospital (FEMH) dataset using a UNET architecture. They achieved a DSC of 902% on this dataset. When the FEMH dataset was combined with the LUNA16 dataset, the segmentation performance improved, with the DSC reaching 926%. This improvement highlights the importance of data diversity and volume in enhancing a model's generalization capability and performance across different cases.

[13] modified the U-Net structure by replacing the ReLU activation function with the Mish function, which is known for properties that can contribute to more stable training and improved performance Some other modifications were also implemented with the aim of

reducing computational resources used (such as memory and processing time), which is critically important for practical clinical applications. These modifications achieved a DSC of 8889% on the LUNA16 dataset.

Despite the significant progress demonstrated by these studies, the achieved performance has not yet reached the level required for widespread practical application in the clinical environment. One of the main challenges often overlooked in previous studies is the issue of class imbalance Lung nodules, by their nature, are relatively small compared to the total size of a CT image. This means that the number of voxels representing nodules (the positive class) is significantly smaller than the number of voxels representing the background or healthy tissue (the negative class). This substantial disparity leads to trained models being biased towards the more prevalent class (background), resulting in poor performance in segmenting small and critical nodules.

Therefore, in this research, the researchers present an innovative model based on the U-Net architecture to address these challenges, this model features multi-label classification, where the background, nodules, and lungs are treated as separate and distinct classes, providing a more accurate representation of each component in the image. To enhance the model's focus on important features and mitigate the effect of class imbalance, a squeeze and excitation mechanism will be incorporated, enabling the model to dynamically recalibrate the importance of feature channels, allowing it to concentrate on the most relevant areas within CT scans, such as subtle nodules. Additionally, an auxiliary loss in the bottleneck of the U-Net will be employed, this additional loss is added at intermediate layers of the network to provide supplementary gradient signals during training, which helps guide the learning process and improve feature representation in deeper layers. Consequently, this approach aims to reduce the imbalance between classes and enhance the model's ability to accurately segment nodules, even very small ones.

Table 1. Summary of Strengths and Weaknesses in Research Studies on Lung Nodule Segmentation

Study	Strengths	Weaknesses
Nam et al (2018)	Use of the 3D-Unet architecture, suitable for volumetric data (CT scans), and its ability to capture three-dimensional context	Relatively modest performance (DSC 7878%) compared to more recent studies
Asha & Bhavanishankar (nd)	Achieved very high performance (DSC 9708%) using a pre-trained foundation model (SAM), demonstrating the effectiveness of such models	The general model may lack the precise specialization needed to handle complex and rare medical cases, despite its high performance on the selected dataset
Zafaranchi et al (2024)	A two-stage approach that reduces the volume of data processed, use of an Attention mechanism to focus on important features, and employment of	Nodule segmentation accuracy (81%) is good but lower than some other mentioned studies, and the multi-stage approach may increase computational complexity

	Explainable AI (Grad-CAM) to increase clinical reliability	
Li et al (2024)	Demonstrated that data diversity and volume (by combining two datasets) significantly improve the model's performance and generalization capability (DSC 926%)	The improvement heavily relies on the availability of multiple large datasets, which may not always be accessible
Bhattacharyya et al (2023)	Modified the U-Net architecture to improve computational efficiency and reduce resource consumption, while achieving good performance (DSC 8889%)	Despite its efficiency, the performance did not reach the highest levels recorded in the field

As summarized in Table 1, the main contributions and inherent limitations of a selection of prominent studies in the field of lung nodule segmentation are outlined. The table highlights the diversity of methodologies employed While the study by [9] focused on leveraging the 3D nature of CT data using the 3D-Unet architecture, the study by [10] demonstrated the immense potential of large foundation models like SAM in achieving exceptional segmentation accuracy On the other hand, the study by [11] introduced a practical approach aimed at increasing clinical trust by integrating attention mechanisms and Explainable AI, albeit at the expense of achieving the highest possible accuracy, the study by [12] also highlighted a crucial factor: the importance of data, proving that increasing the diversity and volume of training data significantly enhances model performance Finally, the study by [13] addressed an important practical aspect, which is improving computational efficiency to make models more viable for use in clinical settings with limited resources.

Despite this notable progress and the multiple strengths highlighted by the research, they collectively reveal a fundamental research gap that has not been adequately addressed, this gap lies in the problem of class imbalance, which is inherent in CT lung scans, the nature of lung nodules means they occupy a very small number of voxels compared to the background or healthy lung tissue, which constitutes the vast majority of the data, this severe disparity leads to models being heavily biased towards the more prevalent class (the background), resulting in poor performance in accurately segmenting small nodules, which are often the most critical for early diagnosis, the neglect of this issue in most previous studies represents the main obstacle preventing these technologies from reaching the level of reliability required for widespread clinical application, this is precisely the gap that this research aims to address directly.

3. Proposed Methodology

The proposed architecture is an improved version of the traditional U-Net, incorporating two enhancements: Squeeze-and-Excitation (SE) blocks and an Auxiliary Classifier at the bottleneck, these additions aim to improve feature discrimination and mitigate class imbalance during segmentation.

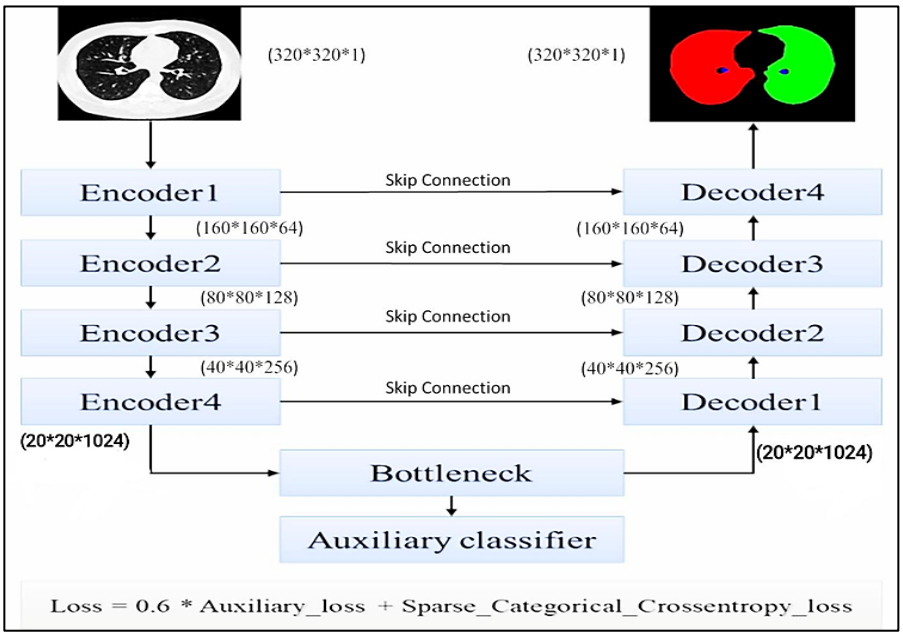

Figure 1. Overall Architecture of the Proposed Model

Each encoder and decoder block follows a consistent structure, As shown in Figure 2, the encoder block applies two sets of convolutions, batch normalization, and activation layers, followed by an SE block for emphasizing important feature maps, and concludes with max-pooling for downsampling. All encoders blocks are similar. All decoders blocks are similar, and Figure 2 shows the detailed structure of each block.

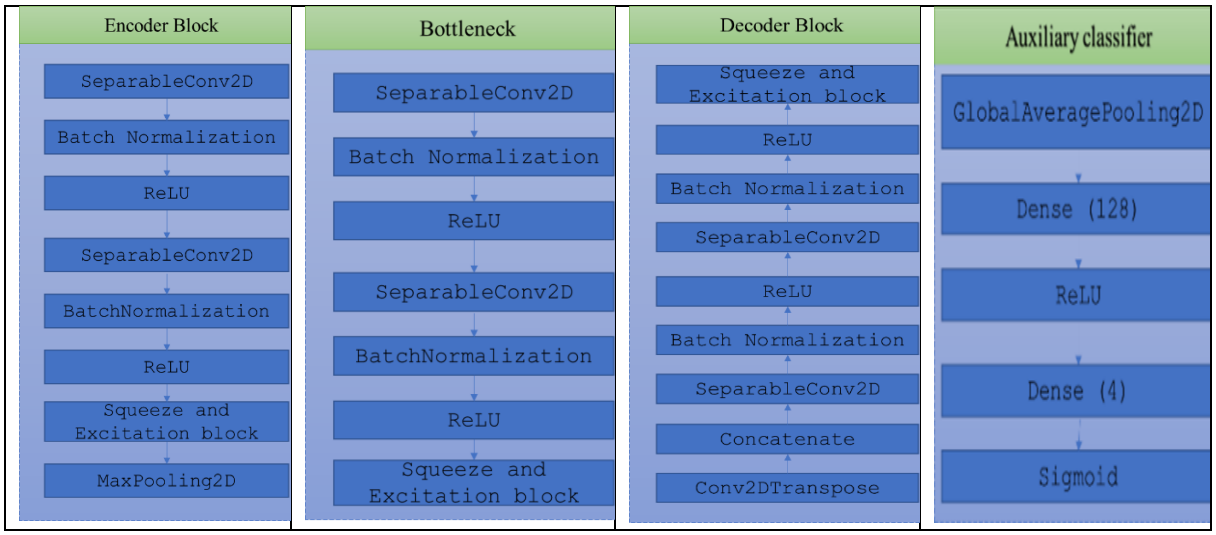

Figure 2. Detailed Structure of the Encoder Block with Squeeze-and-Excitation

Encoder Block: Consists of two convolutional layers with batch normalization and ReLU, followed by an SE block and max pooling [14].

Bottleneck Block: Similar to the encoder, but without down sampling, this block contains the Auxiliary Classifier, which computes a secondary loss to guide feature learning at this depth.

Decoder Block: Begins with upsampling, concatenates the encoder's corresponding features, and applies convolution and SE operations to reconstruct spatial details [15]

3.1 Squeeze and Excitation Block (SE)

The squeeze and excitation block is used to focus on important details in the image, especially small nodules. This block is less computationally complex than the attention block, as it focuses on improving the quality of existing important features rather than increasing the size of your features map. Figure3 shows the structure of this block.

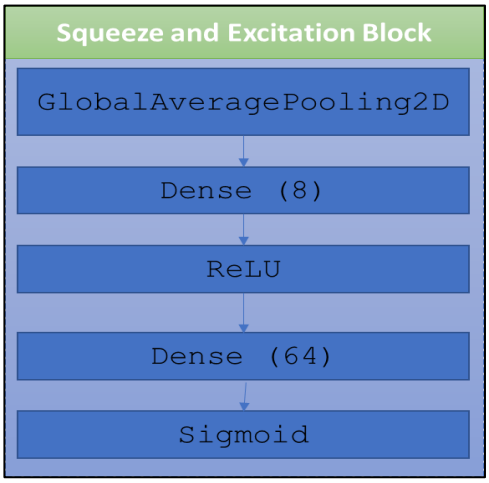

Figure 3. Structure of Squeeze and Excitation Block

- We can calculate the flops generated by each SE block using the formula:

$$Flops = H \times W \times C + 2 \times (C \times N_1 + N_1 \times N_2)$$

H: Height, W: Width, C: Channels number, N_1: number of first MLP (Multi-Layer Perceptron) layer of SE block, N_2: number of second MLP layer of SE block

If we calculate it for the first layer: H=320, W=320, C=64, N_1=8, N_2=64

$$\text{Flops} = 320 \times 320 \times 64 + 2 \times (64 \times 8 + 8 \times 64) = 6{,}555{,}648$$

- We can calculate the flops generated by traditional attention block using the formula:

$$Flops = 4 \times H \times W \times C \times C + H \times W \times C$$

$$\text{Flops} = 4 \times 320 \times 320 \times 64 \times 64 + 320 \times 320 \times 64 = 1{,}684{,}275{,}200$$

The number of flops produced by a traditional attention block is approximately 257 times that of the SE block.

Auxiliary Classifier: This classifier transforms the bottleneck feature map into a vector using global mean pooling and then passes it through two dense layers to predict multi-label outputs, the bottleneck loss is calculated by comparing these predictions to the ground-truth label vectors, which helps address class imbalance and speed up training Before applying this block, the image dataset must be concatenated with the label vectors to calculate the bottleneck

loss, this block is not responsible for segmenting a specific class in the image, but it helps in regularizing training and improving general features.

3.2 Dataset

The LUNA16 (Lung Nodule Analysis) dataset is one of the most important lung cancer datasets. This dataset contains 1,186 nodules across 888 CT scans Each image contains 121 slices, with each slice taken from a specific orientation to produce a 3D view of the chest. Figure 4 shows the middle slice of two images with the corresponding mask for each image from LUNA16 dataset.

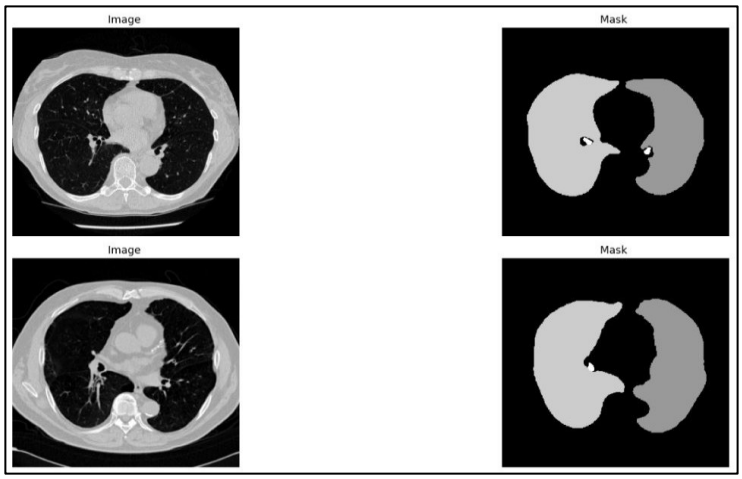

Figure 4. Samples of Dataset Images and Corresponding Masks

All nodules in the data set range in size from 3 to 30 mm, with nodules between 3 and10 mm accounting for 65% of the nodules. Some patients have nodules and others do not. Other patient information is not publicly available due to privacy concerns.

Each image in the dataset has a size of (121, 512, 512, 1) where 121 is the number of slices, 512*512 is the width and height in pixels, and 1 is the number of channels.

As for the masks, they contain 4 unique values: {0: background, 3: the left lung, 4: the right lung, 5: the nodules} Figure 4.

3.3 Preprocessing

Due to the large size of the dataset and limited resources, only the middle slice of each image, which has the largest cross-section of the chest and represents the anterior view of the chest, was retained. The images were saved in NPZ format and then resized to (320, 320, 1). No image augmentation operations, such as flipping or contrast adjustments, were performed due to resource limitations.

4. Results Discussion

The proposed model was trained on the Google Colab platform using RAM = 127 GB and GPU = 15 GB. The Adam optimizer was used with a starting error rate of (1e-3), a training

scheduler with a factor of 0.5, early stopping with a patience of 10, and a batch size of 8, for 50 iterations. Model parameters:

- Total params: 6,577,217 (2509 MB)
- Trainable params: 6,565,441 (2505 MB)
- Non-trainable params: 11,776 (4600 KB)

Training stopped at iteration 34, with a loss of (021), accuracy of (9856%), and Dice coefficient (9842%) on validation set, as shown in Figure (5).

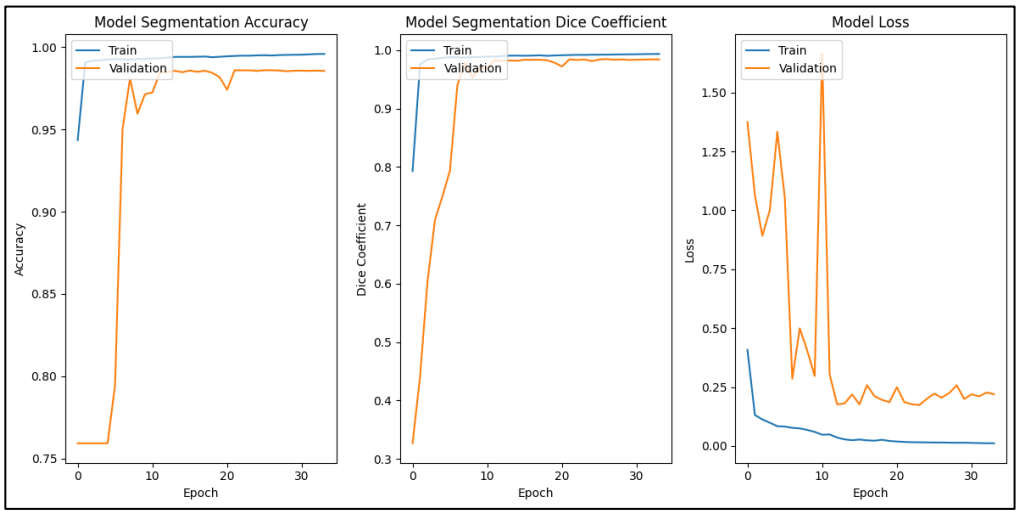

Figure 5. Training and Validation Accuracy/Loss Over Epochs (Seed 1)

Training Time 1779526 seconds

Predicting Time 34ms

Figure 5 presents the training curve of the proposed model, demonstrating that its performance metrics began to achieve promising values after only four iterations. This suggests an effective learning process and indicates that the network weights were initialized effectively. The model continued to optimize its parameters and reached its optimal performance around the 24th iteration. This likely represents the optimal point at which the model has captured the most significant features from the training data without starting to overfit. Notably, for the remaining training epochs beyond the 24th iteration, the model's performance remained stable. This stability indicates effective convergence, free from oscillations, divergences, or significant overfitting, which are common challenges in deep learning during prolonged training. The consistency in performance post-convergence underscores the robustness and reliability of both the training methodology and the model architecture. Figure 6 qualitatively evaluates the segmentation of the proposed model with a representative subset of its predictions. A full description of the segmentation process for one input slice is shown in every row or cluster of images in Figure 6. The input CT scan of the chest, which is the first frame in the sequence, is raw 3D medical image data, and our model processes this for segmentation. The ground truth can be found in the second figure. It is a highly annotated gold standard comparison, and its results explicitly illustrate the actual boundaries of these structures of interest as they compare to predictions by the model. The third row shows the mask resulting from applying the proposed model, demonstrating how the segmentation algorithm was able to identify and delineate

nodules and lung areas based on learned representations. A very important visualization is the fourth one: a composite view of the predicted mask overlaid on the original input CT image. This synthetic view provides a fast and interactive contrast between an anatomical context and the model's output. Some color coding is present on the overlay when these segmented structures are provided as input, which serves to highlight the differences. If there is any nodule in the slice, it is painted blue, and vice versa; the left lung is painted red and the right one green. Yellow is employed by the model in important areas to visualize and indicate any mismatches with the actual segmentation. This could either be false positives or false negatives, but instant feedback in terms of the distribution and existence of yellow areas highlights the shortcomings the model may have and points toward potential ways for improvement. Overall, Figure 6 qualitatively validates the quantitative performance metrics through its appearance and provides convincing proof that the model can indeed segment lung structures and nodules in a clinically relevant manner.

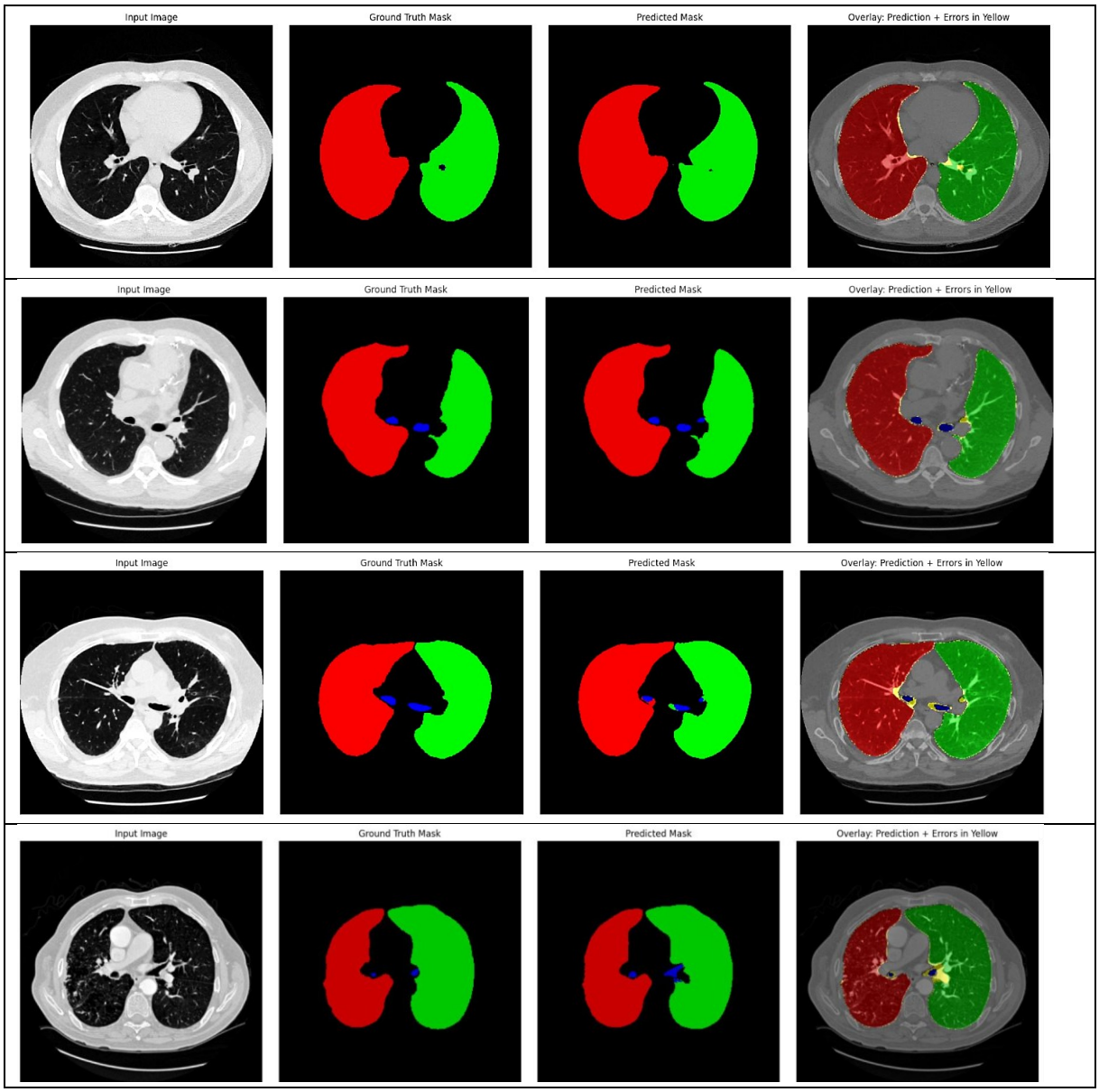

Figure 6. Example of Segmentation Results: CT Input, Ground Truth Mask, Predicted Mask, And Overlay

The high performance of the proposed model can be observed in both lung segmentation and nodule segmentation. To achieve higher performance, a larger dataset and higher resolution images can be used, but this requires more resources.

Table 2. Comparative Performance of Different Models on LUNA16 Dataset

Study / Model	Dataset Used	DSC (%)	Accuracy (%)	Key Techniques
Chang-Mo et al (2018)	LUNA16	7878	–	3D U-Net
Bhattacharyya et al (2023)	LUNA16	8889	–	Bi-directional U-Net, Mish Activation
Zafaranchi et al (2024)	LungQuant + Custom	8100	–	Attention Res-UNet
Li et al (2024)	FEMH + LUNA16	9260	–	U-Net Fusion
Authors (Proposed)	LUNA16 (2D slices)	9842	9856	SE Blocks + Auxiliary Loss + U-Net

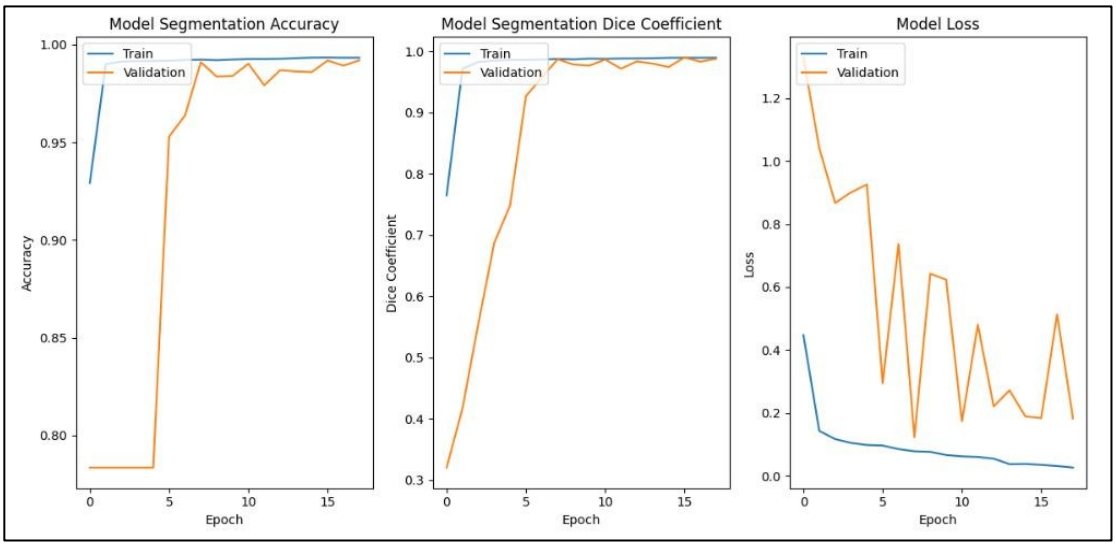

Figure 7. Model Performance Curves for the Robustness Test Run (Seed 2)

Training and validation accuracy curves are slowly increasing smoothly and converging to each other, meaning the model is learning well both from training and generalizing well to new data without overfitting. Meanwhile, the loss is decreasing during both training and validation. This implies steady training (ST), where the learning matures in epochs and the error of the model is also minimized. Figure 8 demonstrates the visual accuracy of lung and nodule segmentation by the model. As can be observed, the predicted mask of the model and the ground truth mask had a very high overlap. The model could delineate the boundaries of the target with high accuracy, which was evident from the overlay image created by applying the predicted mask over the original image. This reconciles with its high qualitative performance.

For more robustness and validation that the model is indeed strong and that the results are not a consequence of some coincidental split in the data, this experiment has been repeated with a different split of the dataset, i.e., with another seed. This gave the following outstanding results: an overall accuracy of 0.09919, a Dice Similarity Coefficient of 0.09877, and finally, the loss value reached a minimum of 0.003. The whole training process converged in 1,569,684 seconds at epoch #18, while the prediction of one image took only 0.034 ms. These figures confirm the high and stable performance of the model. The low prediction time underlines its efficiency and suitability for a clinical environment that requires rapid responses.

Note that other types of loss functions, including iou_loss, hybrid_loss, weighted_loss, and Tversky_loss, were also experimented with during the model formation process. Those experiments proved that such kinds of functions are not good enough or perform worse than the one we selected.

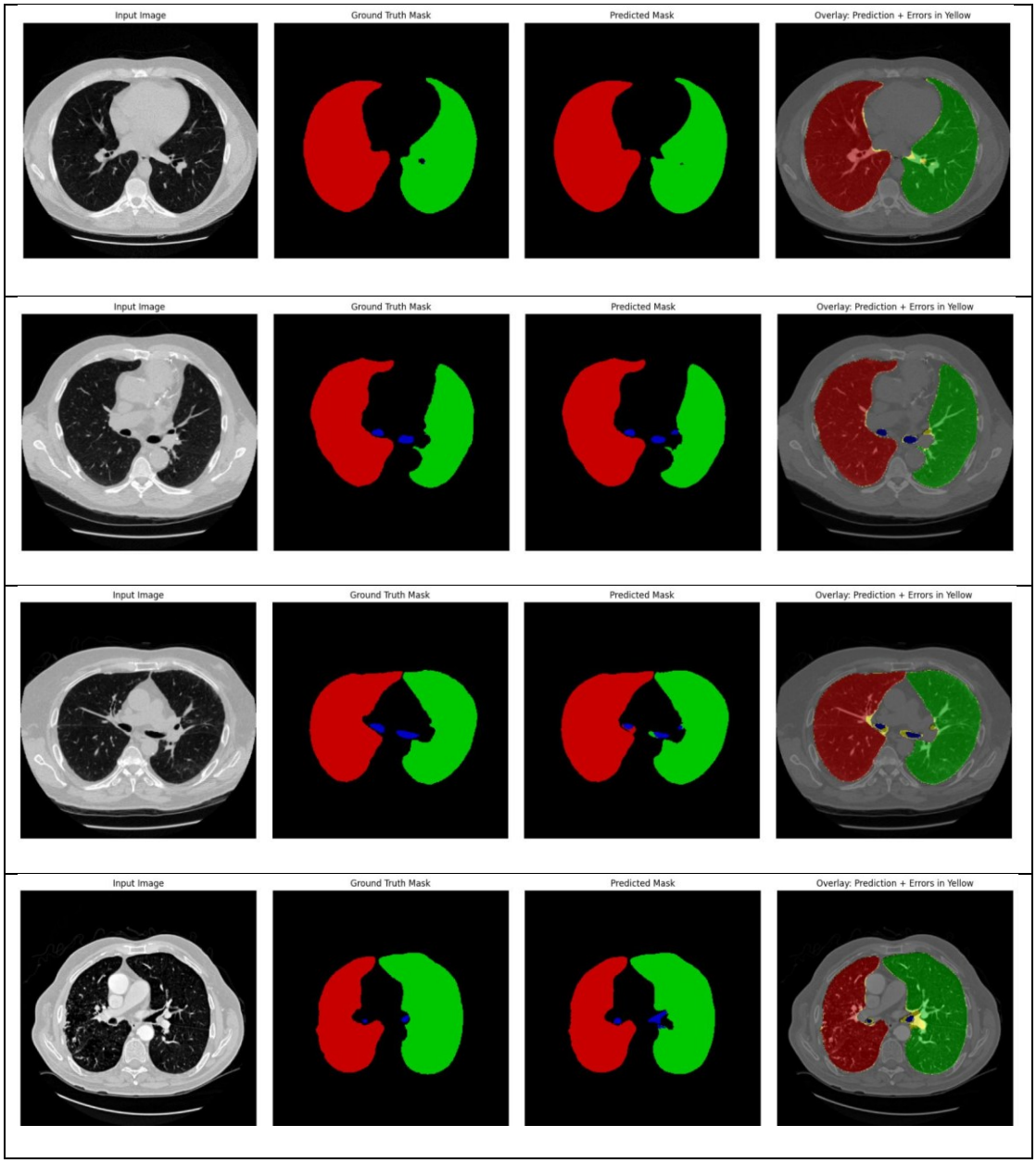

Figure 8. Example Segmentation Results from the Robustness Test Run

4.1 Seed3 Train Curves

1. Train time 139373 s
2. Test time 43ms
3. Dice 9871
4. Accuracy 9897

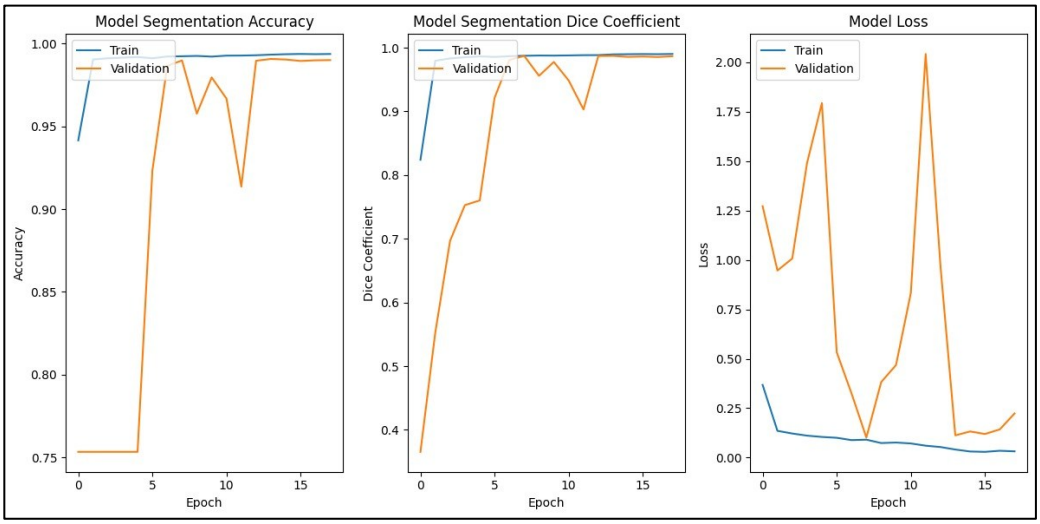

Figure 9. Model Performance Curves for the Robustness Test Run for (Seed 3)

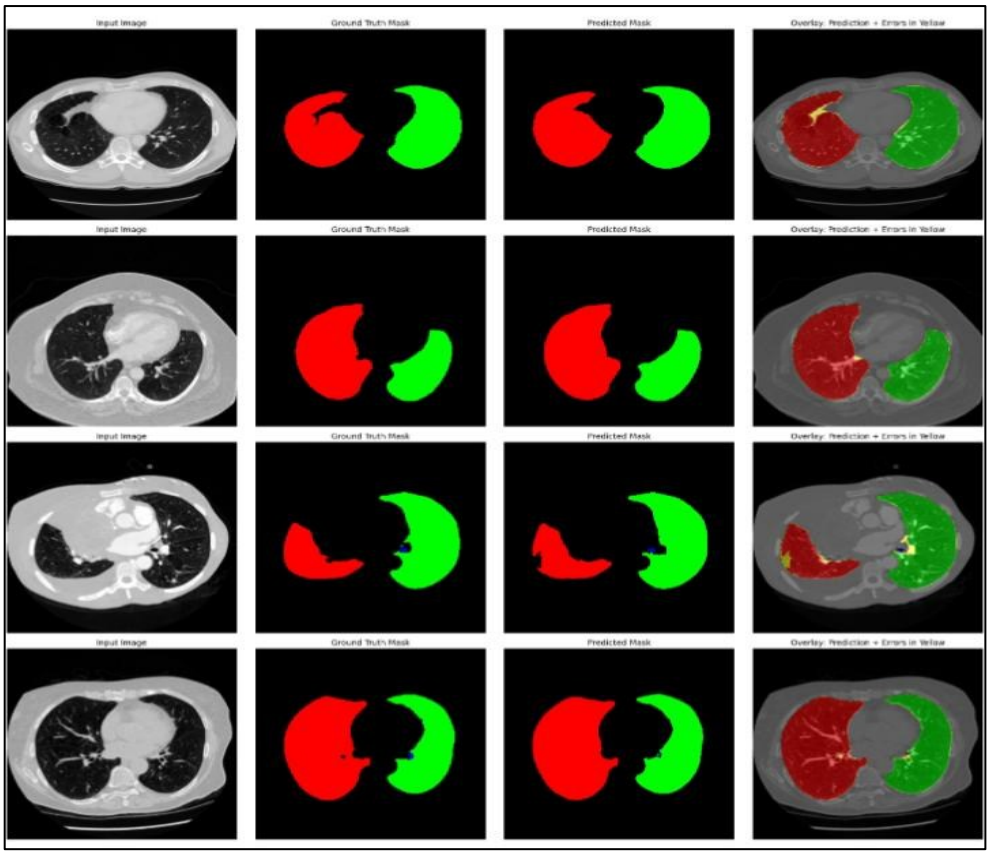

Figure 10. Example Segmentation Results from the Robustness Test Run for Seed 3

Table 3. The Change in Learning Rate and GPU Consumption Over Epochs for the Training Process of the Proposed Model

Epoch	Learning Rate	Batch Size	GPU
Seed1			
1-12	0001	8	0-137
13-21	00005	8	137-138
22-29	000025	8	138
30-34	0000125	8	138
Seed2			
1-13	0001	8	0-138
14-19	00005	8	138
Seed3			
1-13	0001	8	0-138
14-18	00005	8	138

Table 3 shows the change in learning rate and GPU consumption over epochs for the training process of the proposed model. From the table 3 it is noted that the learning rate varies from 0001 to 0000125, and the GPU usage varies from 0 to 138 GB. This consumption is constant across three different seeds although some seeds take fewer iterations than others.

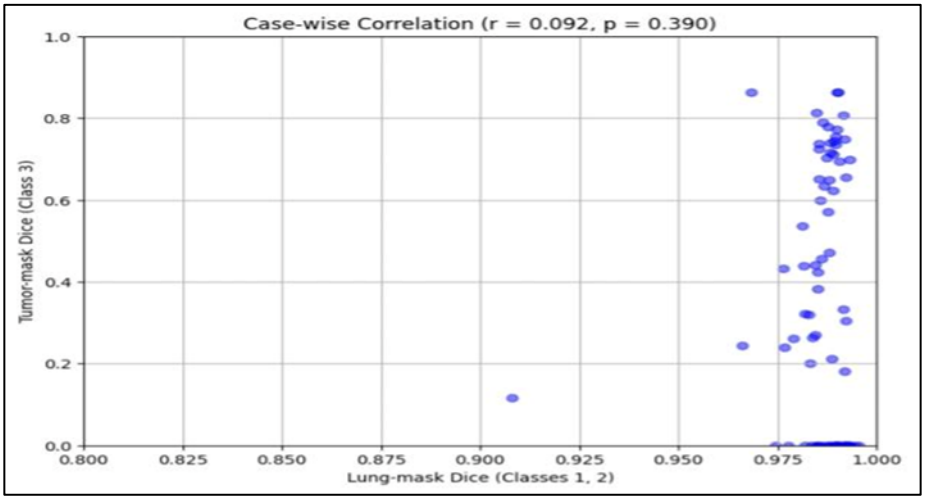

Figure 11. The Correlation Between Lung-mask Dice (classes 1, 2: X-axis, 08-10) and Nodules-Mask Dice (class 3: Y-axis, 00-10)

Figure 11 shows the correlation between lung-mask Dice (classes 1, 2: X-axis, 08-10) and Nodules-mask Dice (class 3: Y-axis, 00-10), the points are scattered, with a concentration in the upper right (Lung Dice >09, Nodules Dice >06) and some low points (Nodules Dice ~02-04)

R = 0092 (weakly positive), this reflects that lung Dice does not directly affect nodule Dice, but using lungs as targets decreases image class unbalancing and negative size bias (mean difference ≈ -60 pixels for nodules), with wide LoA (eg, [-258, +140]) indicating under-segmentation and better performance for lungs (slight bias, wider LoA due to their large size). Figure 12 shows the Bland-Altman for Class 3 (Nodules), with the difference (Y-axis, -500 to 100) versus the mean (X-axis, 0-400 pixels).

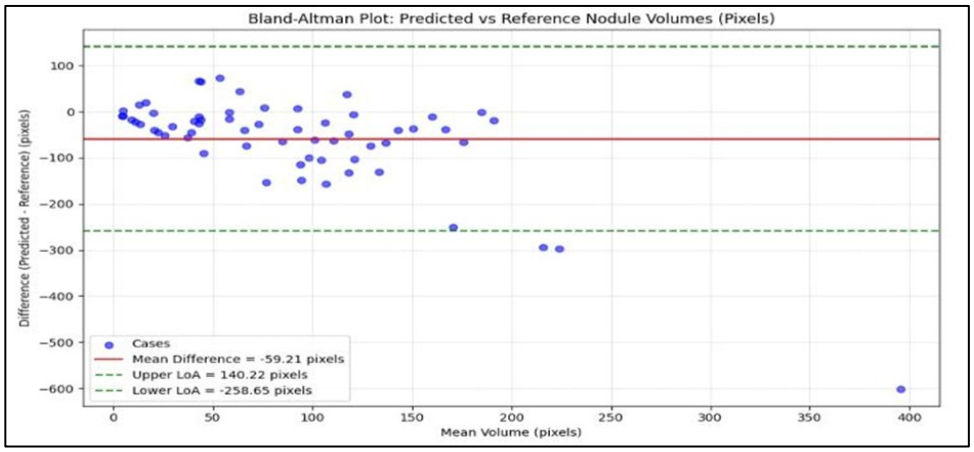

Figure 12. Bland-Altman for Class 3 (Nodules), the Difference (Y-axis, -500 to 100) Versus the Mean (X-axis, 0-400 Pixels

The points are scattered, with most near 0 and some significantly negative; mean difference = -5921 pixels, LoA = [-25865, +14022]. There is a strong negative bias (30-50% of the size of a small nodule) with an asymmetric LoA (greater spread in the negative). This is because of the nodule edges being classified as background due to their small size and variable number (0-2), causing under-segmentation in cases of multiple or ambiguous nodules. Figure 13 Bland-Altman for Class 1 (Right Lung), Shows the difference (Y-axis, -20,000 to 0) versus the mean (X-axis, 4,000-20,000 pixels).

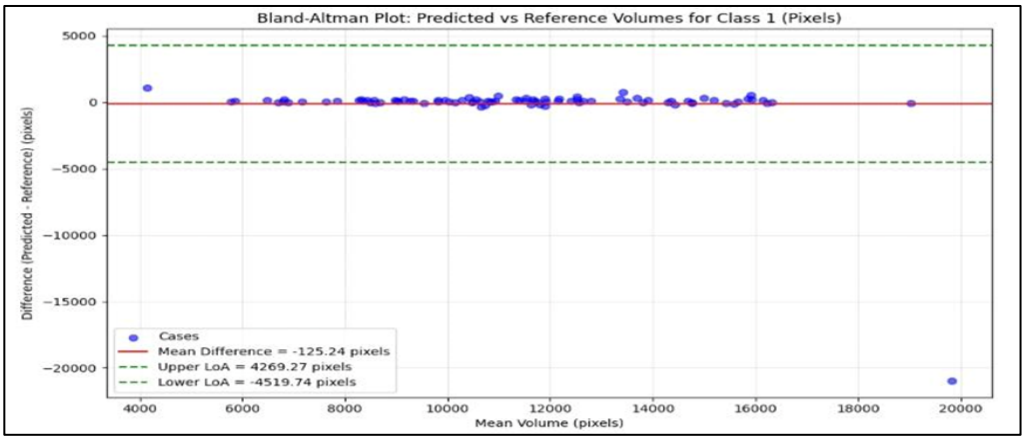

Figure 13. Bland-Altman for Class 1 (Right Lung), Shows the Difference (Y-axis, -20,000 to 0) Versus the Mean (X-axis, 4,000-20,000 Pixels)

Points are near 0, with negative spread; mean difference = -12524 pixels, LoA = [-425374, +420327] ,slight negative bias (<1% of lung volume) with wide LoA (high spread in negative), this is due to the lung margins being classified as background, especially in cases of nodules that affect segmentation, but the large size makes the effect less pronounced. Figure 14 shows the nBland-Altman for Class 2 (Left Lung), illustrating the difference (predicted – reference: Y-axis, -15,000 to 0) versus the mean (mean volume: X-axis, 5,000 to 22,500 pixels).

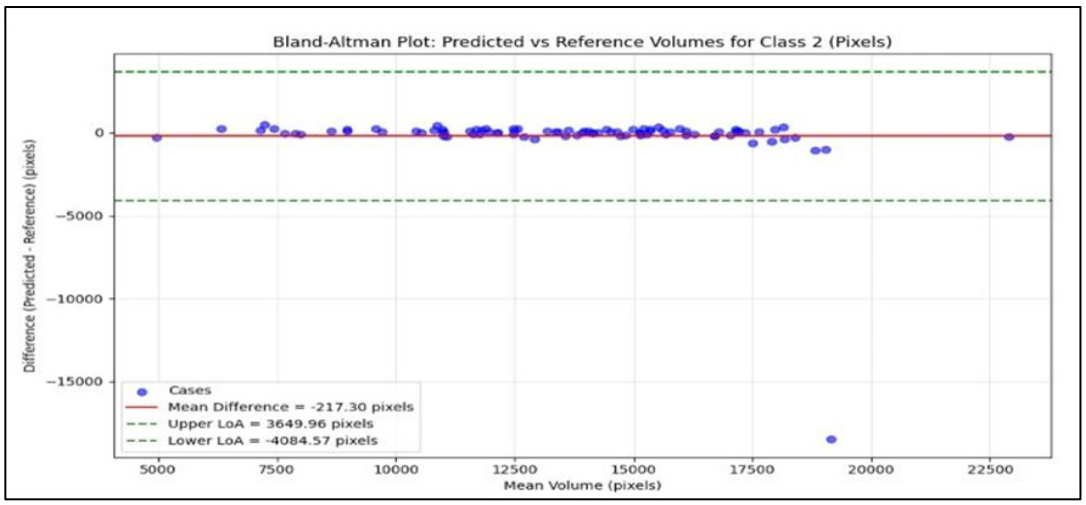

Figure 14. Bland-Altman for Class 2 (Left Lung), Shows the Difference (Predicted – Reference: Y-axis, -15,000 to 0) Versus the Mean (Mean Volume: X-axis, 5,000 to 22,500 Pixels)

Points near 0, with negative spread; mean difference = -21730 pixels, LoA = [-78457, +34996], slight negative bias (<2% of lung volume) with asymmetric LoA (greater spread in negative); this is due to the lung margins being classified as background, especially in cases of internal nodules causing confusion, but the large lung volume reduces the effect

1. The relationship between slice thickness, number of nodules, lung volume, and bias.

2. Slice thickness (05-5 mm): Determines the vertical depth of each of the 121 slices in a single case Large lungs (10,000-20,000 pixels) are easy to segment, while small nodules (20-400 pixels) are affected by partial volume effects (tissue merging) at high thicknesses.

3. Number of nodules (0-2+): Variation in number increases the complexity of segmentation in the middle slice (2D) only.

4. Lung volume: Large, which reduces the effect of slice thickness, but their edges are classified as background.

5. Bias: Negative bias (-60 pixels for nodules) is due to the edges of nodules being classified as background due to partial volume effects, especially with multiple nodules, which reduces the volume and increases LoA contrast Lungs suffer from a slight bias (-125 to -217 pixels) due to fuzzy edges.

6. 2D model: Relies on the middle slice only, which limits the capture of nodules across the 121 slices, and increasing bias.

4.2 Comparative Performance Evaluation

This proposed model demonstrates a significant leap forward in lung nodule segmentation, as evidenced by its superior performance when compared to existing architectures Specifically, this approach achieves remarkable results with a Dice Similarity Coefficient (DSC) of 9842% and an accuracy of 9856% on the LUNA16 dataset, this performance notably surpasses that of previous state-of-the-art methods, such as the 3D U-Net by [14] which reported a DSC of 7878%, the Bi-directional U-Net with Mish activation by [13] at 8889%, and even more advanced fusion models like [12] that achieved 9260% by combining FEMH and LUNA16 datasets.

A key distinction of this methodology lies in Its efficiency and design philosophy Unlike many contemporary studies that rely on computationally intensive 3D U-Nets or complex, multi-stage architectures—such as the Attention Res-Unet used by [11] which had an 8100% DSC—this model achieves these superior results using a lightweight 2D framework.This design choice significantly reduces computational overhead and memory requirements, making the model more practical for real-world applications.

The enhanced performance stems from the integration of two critical mechanisms: Squeeze-and-Excitation (SE) Blocks and an auxiliary loss mechanism SE Blocks dynamically recalibrate channel-wise feature responses, allowing the network to emphasize more informative features and suppress less relevant ones, this is particularly beneficial in medical imaging where subtle abnormalities like small nodules need to be highlighted Concurrently, the auxiliary loss, strategically applied within the U-Net architecture, provides additional gradient signals during training, this not only stabilizes the learning process but also directly aids in addressing the pervasive issue of class imbalance, where the small volume of nodules (positive class) is heavily outnumbered by the background (negative class) By providing supplementary supervision, the auxiliary loss ensures that the model learns robust feature representations for both dominant and minority classes.

Moreover, the model's ability to accurately segment both the overall lung structures and the intricate nodules using only a single 2D slice is a testament to its exceptional efficiency. This single-pass, 2D processing capability translates into faster inference times, which is crucial for high-throughput clinical environments. It simplifies the deployment process, as it can be readily integrated into existing Picture Archiving and Communication Systems (PACS) that typically display medical images in 2D slices. This streamlined approach, combined with the high accuracy achieved, underscores the model’s robust suitability for practical, real-world clinical deployment, paving the way for more rapid and precise early lung cancer detection.

5. Conclusion

During the research process, a high-accuracy segmentation model was developed and a new auxiliary unit was added to the U-Net architecture. To give more weight to informative features, the squeeze-and-excitation method dynamically recalculated feature responses. An auxiliary loss was added at the bottleneck to alleviate the class imbalance problem and further improve learning. The effectiveness of the model proposed herein receives full confirmation on the widely used dataset LUNA16. Its accuracy was as high as 98.56%; if the Dice Similarity Coefficient is 98.42%, its performance is considered satisfactory. These results show how well the approach presented here works to detect lung nodules and structures, a crucial clinical

testing step that requires accurate boundary definition. Firstly, the robustness and reliability of the model for various real-world clinical scenarios could be further improved by the addition of a larger and more diverse data set in addition to LUNA16. Second, though this study is focused on a very efficient 2D framework, the core principle of the proposed model, particularly the auxiliary unit, has great potential for application in the 3D domain. The extension of the model to directly process volumetric data in 3D may capture comprehensive contextual information and interslice relationships that may lead to finer segmentation details. However, such a transition will require very high computational resources, thus a very careful evaluation is required. Lastly, training on and using a range of datasets will be essential to further increase the model's ability to handle small nodules and irregular boundaries.

Statement and Declarations

Competing Interests

All authors declare that they have no conflicts of interest

Funding

No funds, grants, or other support was received

References

[1] Siegel, Rebecca L., Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. "Cancer Statistics, 2022." CA: a cancer journal for clinicians 72, no. 1 (2022). 7-33.

[2] Gu, Yu, Jingqian Chi, Jiaqi Liu, Lidong Yang, Baohua Zhang, Dahua Yu, Ying Zhao, and Xiaoqi Lu. "A Survey of Computer-Aided Diagnosis of Lung Nodules from CT Scans Using Deep Learning." Computers in biology and medicine 137 (2021): 104806.

[3] Hua, Kai-Lung, Che-Hao Hsu, Shintami Chusnul Hidayati, Wen-Huang Cheng, and Yu-Jen Chen. "Computer-Aided Classification of Lung Nodules on Computed Tomography Images via Deep Learning Technique." OncoTargets and therapy (2015): 2015-2022.

[4] Zheng, Sunyi, Ludo J. Cornelissen, Xiaonan Cui, Xueping Jing, Raymond NJ Veldhuis, Matthijs Oudkerk, and Peter MA van Ooijen. "Deep Convolutional Neural Networks for Multiplanar Lung Nodule Detection: Improvement in Small Nodule Identification." Medical physics 48, no. 2 (2021): 733-744.

[5] Samuel, G “The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans” Medical Physics, (2011): 38, 915-931.

[6] Setio, Arnaud Arindra Adiyoso, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas Van Den Bogaard, Piergiorgio Cerello, Hao Chen et al. "Validation, Comparison, and Combination of Algorithms for Automatic Detection Of Pulmonary Nodules in Computed Tomography Images: The LUNA16 Challenge." Medical image analysis 42 (2017): 1-13.

[7] National Lung Screening Trial Research Team. "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening." New England Journal of Medicine 365, no. 5 (2011): 395-409.

[8] Zhao, Ying Ru, Xueqian Xie, Harry J. De Koning, Willem P. Mali, Rozemarijn Vliegenthart, and Matthijs Oudkerk. "NELSON Lung Cancer Screening Study." Cancer Imaging 11, no. 1A (2011): S79-84.

[9] Nam, Chang-Mo, Jihang Kim, and Kyong Joon Lee. "Lung Nodule Segmentation with Convolutional Neural Network Trained by Simple Diameter Information." (2018).

[10] Asha, V., and K. Bhavanishankar. "Advanced Lung Nodule Segmentation and Classification for Early Detection of Lung Cancer using SAM and Transfer Learning." https://arxiv.org/abs/2501.00586

[11] Zafaranchi, Arman, Francesca Lizzi, Alessandra Retico, Camilla Scapicchio, and Maria Fantacci. "Explainability Applied to a Deep-Learning Based Algorithm for Lung Nodule Segmentation." In Proceedings of the 1st International Conference on Explainable AI for Neural and Symbolic Methods-EXPLAINS, vol. 1, Scitepress, 2024, 132-138.

[12] Li, Ai-Hsien Adam, Cristian Daniel Aruperes, Yen-Jun Lai, Ting-Ying Chien, Yen-Ling Chiu, and Chien-Lung Chan. "Lung Nodule Analysis in CT Images: Deep Learning for Segmentation and Measurement." In Proceedings of the 2024 8th International Conference on Medical and Health Informatics, 2024, 13-17.

[13] Bhattacharyya, Debnath, N. Thirupathi Rao, Eali Stephen Neal Joshua, and Yu-Chen Hu. "A Bi-Directional Deep Learning Architecture for Lung Nodule Semantic Segmentation." The Visual Computer 39, no. 11 (2023): 5245-5261.

[14] Li, Yuemeng, and Yang Fan. "DeepSEED: 3D Squeeze-and-Excitation Encoder-Decoder Convolutional Neural Networks for Pulmonary Nodule Detection." In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, 1866-1869.

[15] Dou, Qi, Hao Chen, Lequan Yu, Jing Qin, and Pheng-Ann Heng. "Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection." IEEE Transactions on Biomedical Engineering 64, no. 7 (2016): 1558-1567.