POLYTECHNIQUE MONTRÉAL
TECHNOLOGICAL UNIVERSITY

Lung Cancer Detection in CT Images Using Region Proposal Network with Integrated Feature Maps from VGG16

Yadollah (Amir) Zamanidoost

Department of Computer and Software Engineering

Polytechnique Montreal, Montreal, QC H3T 1J4, Canada

Yadollah.zamanidoost@polymtl.ca

a.zamanidoost@gmail.com

August 2024

POLYTECHNIQUE MONTRÉAL
TECHNOLOGICAL UNIVERSITY

Contents

Introduction:

Lung cancer stands as one of the deadliest known diseases worldwide, comprising nearly two-thirds of all existing cancers [1]. The mortality rate of this type of cancer among all recognized tumours is 18% [2]. Studies indicate that early detection of lung cancer can improve treatment outcomes and increase patients' survival rates [3]-[5]. Among the available imaging modalities, computed tomography (CT) imaging is important in lung cancer detection and diagnosis [6],[7]. With increased access to CT equipment, physicians review a substantial volume of CT images daily. However, due to the prolonged time required for physicians to examine each CT scan, errors in cancer detection may occur due to fatigue or external factors, posing significant risks to patients [8]. Therefore, to reduce individual errors, computer-aided detection (CAD) systems have been developed to assist physicians in rapidly and accurately identifying tumours.

Lung nodule detection systems typically identify nodules in two stages. The first stage involves extracting candidate nodules to increase the system's sensitivity. Traditional methods for extracting remaining nodules used threshold-based or region-based algorithms, which perform poorly in extracting nodules with lower contrast than the surrounding tissue. The second stage involves removing false-positive candidate nodules to increase the system's precision. However, lung nodule detection approaches need more efficiency, such as lengthy processes, lack of end-to-end detectability, and challenges with larger datasets [9].

In previous years, object detection in medical images has not seen significant advancements due to hardware limitations in machine performance. However, with improved processor speeds and the introduction of deep learning techniques, various object detection methods in images have been widely presented in recent years. For example, Faster R-CNN [9], and Cascade R-CNN [10] are two-stage object detection techniques that accurately detect objects. Additionally, YOLO [11] and SSD [12] are one-stage methods that rapidly detect objects. In detecting lung nodules, the integration of region proposal networks in Faster R-CNN is utilized, leading to increased accuracy in detecting lung nodules, especially small ones.

Our method utilizes pretrained VGG16 with 5-group convolution as the main feature extraction network [13]. After a series of convolutions and pooling in VGG16, the size of the feature map of the last layer is reduced, which leads to limited performance in the ROIs detection of nodules. It has been observed that the utilization of small feature maps does not provide sufficient resolution to represent the features of nodules accurately. We can create a feature map that demonstrates the feature resolution of various sizes of nodules in the proposed method by combining the last three layers of VGG16.

The proposed feature map enters a region proposal network (RPN) and obtains a set of rectangular-shaped nodule proposals, each of which has a score in the output. The proposed network of the region consists of a fully convolutional network. In the false-positive reduction stage, the selected

proposed regions are input into a 2D deep convolutional neural network (2D DCNN) is designed to reduce false-positive nodules.

The remainder of this report is structured as follows: Section 2 explains the methodology. Section 3 presents the implementation of the method and its results. Section 4 provides a discussion of challenges and future work possibilities. Section 5 offers a conclusion.

Methodology:

Fig. 1 illustrates an automated pulmonary nodule detection system. This system takes three-dimensional CT scan images as input and outputs the position of nodules. The implementation of this system aims to achieve high sensitivity in nodule detection while reducing the average number of false positives per scan. The methodology employed is described in detail below.

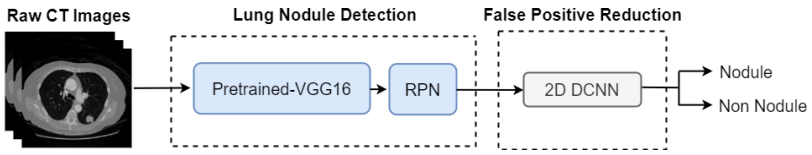


Fig. 1. Overall framework of an automatic pulmonary nodule detection system based on Pretrained-VGG16.

Feature extraction structure:

The fundamental architecture of the feature extractor is based on pretrained VGG16, comprising five convolutional groups. The upper layers encompass a semantic feature map and get more intricate details from nodules. Conversely, the lower layers offer heightened resolution but cannot extract finer intricacies from nodules. To harness the advantages of detailed features with enhanced resolution, we concatenate the upper and lower layers of VGG16 (Fig. 2). Using the proposed structure, we can select features from feature maps of Three different layers, which include features with high accuracy and resolution.

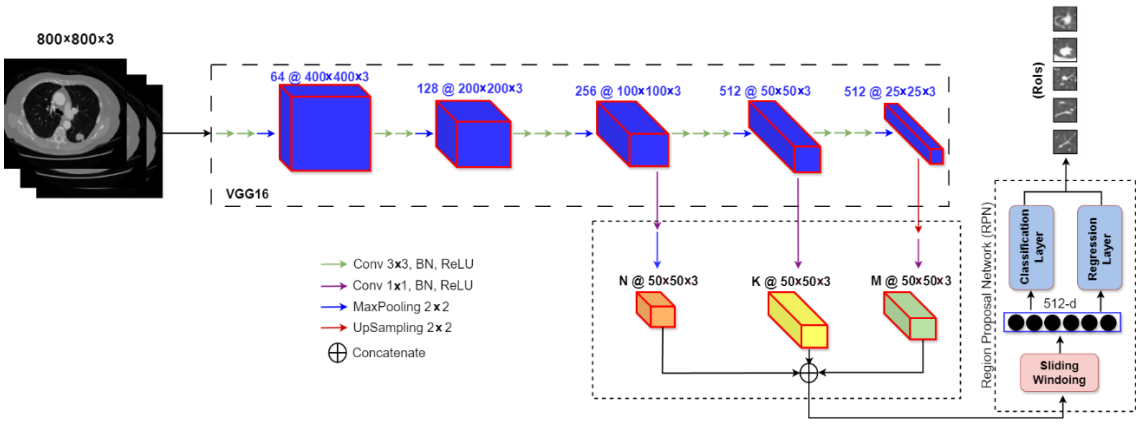


Fig. 2. Overall framework of lung cancer detection with concatenating layers of VGG16.

We opted for a 3D input data approach instead of 2D input data. Due to the substantial computational demands associated with using the original 3D volume of the CT scan as input for

the nodule detection network, we resorted to employing axial slices as the input data. This was achieved by consolidating the primary CT scan slice housing the nodule with the adjacent upper and lower slices. 3D input data imparts a richer contextual backdrop and a more comprehensive portrayal of the nodule, aiding the model's differentiation between nodules and other structures or artefacts. Consequently, we extract the three contiguous slices for every axial slice in the CT images and convert this data into an 800 × 800 × 3 image.

Region Proposal Network (RPN):

An RPN aims to suggest potential nodule regions (called region-of-interest --- ROI). As shown in Fig. 3, RPN receives a feature map as an input and outputs a set of rectangular object maps, each of which has an object score.

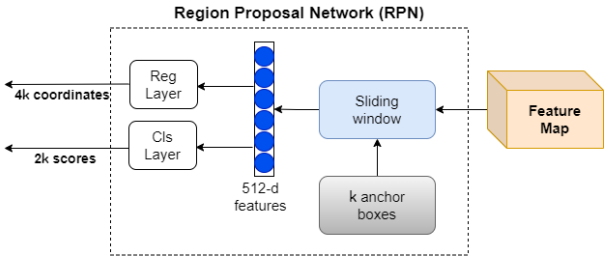


Fig. 3. The region proposal network (RPN) structure

In the RPN structure, a sliding window takes the feature map as input, obtains its convolution values using a 3*3 spatial window, and then maps each sliding window to a 512-dimensional feature. These features finally enter into two fully connected layers a box-regression layer (reg) and a box-classification layer (cls).

Simultaneously, K region proposals are predicted at each sliding window location. So, the (reg) layer has (4K) output that predicts the spatial coordinates of each box. The (cls) layer has 2K outputs that estimate each proposed region's nodule/non-nodule probability.

Each box is called an anchor in the centre of the sliding window. To train RPNs, a binary class label is assigned to each anchor. The anchor with the highest intersection over-union (IoU) overlaps with the ground truth box, or the anchor with an IoU overlap higher than 0.6 is assigned a positive label. On the other hand, if the overlap of the IoU with the ground truth box is less than 0.3, the anchor is assigned a negative label. Anchors that are neither positive nor negative do not contribute to the training goal.

The (reg) layer has 4K outputs. The x and y are the centres of the box, and the w and h are its width and height. For a region proposal (P) and a ground truth (G), these four parameters compute as follows:

$$\begin{cases} t_x = \frac{(G_z - P_z)}{P_w} \\ t_y = \frac{(G_y - P_y)}{P_{h0}} \\ t_w = log\frac{G_w}{P_w} \\ t_x = log\frac{G_h}{h} \end{cases} \quad (1)$$

where $P^i = (P_x^i, P_y^i, P_w^i, P_h^i)$ specifies the pixel coordinates of the center of proposal P^i and $G = (G_x, G_y, G_w, G_h)$ specifies the ground-truth bounding box.

In training the RPN for lung cancer detection, two key components are optimized: the classification (objectness) and regression (bounding box) tasks. The classification loss, L_{clc}, is computed using binary cross-entropy to determine whether a region proposal is foreground or background:

$$L_{clc}(P_i, P_i^*) = -\frac{1}{N}\sum_i [P_i^* \log(P_i) + (1 - P_i^*)\log(1 - P_i^*)] \quad (2)$$

where P_i represents the predicted probability, and P_i^* is the ground truth label. The regression loss, L_{reg}, applies a smooth L1 loss to refine the coordinates of the bounding boxes:

$$L_{reg}(t_i, t_i^*) = \frac{1}{N_{reg}}\sum_i P_i^*.smooth_{L1}(t_i - t_i^*) \quad (3)$$

where the smooth L1 function is defined as:

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & if\ |x| < 1, \\ |x| - 0.5 & Otherwise. \end{cases} \quad (4)$$

The combined loss function, L, is a weighted sum of these two losses:

$$L = \frac{1}{N_{clc}}\sum_i L_{clc}(P_i, P_i^*) + \lambda\frac{1}{N_{reg}}\log(P_i)\sum_i L_{reg}(t_i, t_i^*) \quad (5)$$

In the case of RPN model, its loss function encompasses both classification and regression components. When evaluating the loss for the regression prediction related to bounding boxes, the IoU metric comes into play. IoU quantifies the spatial overlap between the predicted bounding box and the ground-truth box, thereby assessing the precision of the regression prediction.

In this research, we employ the generalized intersection over union GIoU loss function instead of the IoU loss function to compute the loss for regression predictions. The equation is shown as follows:

$$\begin{cases} IoU = |\frac{B^{GT} \cap B^{Pred}}{B^{GT} \cup B^{Pred}}| \\ GIoU = IoU - \frac{|B - B^{GT} \cup B^{Pred}|}{|B|} \\ Loss_{GIoU} = 1 - GIoU \end{cases} \quad (6)$$

B represents the minimal bounding box that contains both B^{GT} (ground-truth box) and B^{Pred} (prediction box). When there is no intersection between the prediction box and the ground-truth box, the IoU is 0, and GIoU falls within the range of -1 to 0. GIoU equals 1 when the prediction box and the ground-truth box entirely overlap. Unlike the IoU loss function, the GIoU loss function addresses the issue of gradient optimization instability.

False Positive Reduction (FPR):

In the post-RPN stage of the lung cancer detection pipeline, a 2D Deep Convolutional Neural Network (2D DCNN) is employed to reduce false positives generated by the RPN. After the RPN proposes candidate regions, the 2D DCNN processes these regions to distinguish between true positives (actual lesions) and false positives (non-lesions or artifacts). This additional layer of classification helps to refine the detection results by filtering out irrelevant or incorrect region proposals, thereby improving the overall accuracy and reliability of the lung cancer detection model. The use of a 2D DCNN is particularly effective in capturing the spatial patterns and features unique to lung lesions, ensuring a robust reduction in false positives.

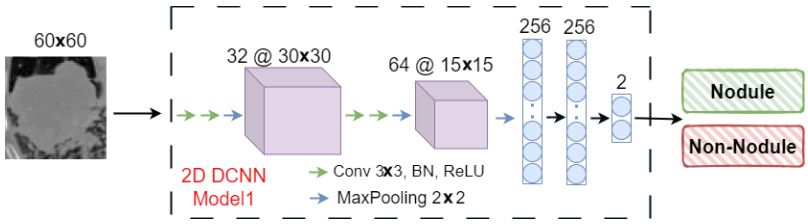


Fig. 4. The framework of false positive reduction model.

Fig. 4 shows the framework of false positive reduction model. The architecture comprises 6 layers, with (60× 60) patches serving as inputs. It entails a repetitive sequence of layers, each consisting of two convolutional layers with 32 and 64 kernel of size (3 × 3) and a max-pooling layer. The concluding segment encompasses two fully connected layers.

Experimental results and discussion:

Dataset:

This collection contains images from 422 non-small cell lung cancer (NSCLC) patients. For these patients, pre-treatment CT scans, manual delineation by a radiation oncologist of the 3D volume of the gross tumor volume, and clinical outcome data are available. The DICOM Radiotherapy Structure Sets (RTSTRUCT) in this data contain the annotation by a radiation oncologist. The code related to this section is located in the file ‘Data_Acquisition.ipynb’.

In this code, the annotations for each patient's CT scan are first extracted. The information, including the precise location and size of the cancerous nodule, is then saved in an Excel file (‘lung_nodule_data.xlsx’). Fig. 5 displays the information for each CT image in the file.

	lung_name	x-pos	y-pos	z-pos	dia
0	LUNG1-001	190.075427	272.766053	74	190.863360
1	LUNG1-002	182.106249	272.602799	72	99.672467
2	LUNG1-003	305.110428	253.777633	20	59.380757
3	LUNG1-004	316.562048	336.627413	83	66.841600
4	LUNG1-005	195.791061	261.483038	37	78.817810
..	...	...	...	...	...
417	LUNG1-418	316.902651	290.418092	53	108.590080
418	LUNG1-419	177.409936	221.029670	62	87.475200
419	LUNG1-420	310.701668	333.748897	71	39.541760
420	LUNG1-421	316.625274	228.135689	46	55.040000
421	LUNG1-422	228.858014	301.264879	60	43.735040

Fig. 5. The annotation of each CT scan image

Additionally, for the analysis of the available data, Fig. 6 illustrates the nodule density based on their diameter. This analysis shows that the dataset has the highest density of nodules with diameters ranging from 30 to 90 millimeters, with the smallest nodule being 10 millimeters and the largest nodule having a diameter of 260 millimeters.

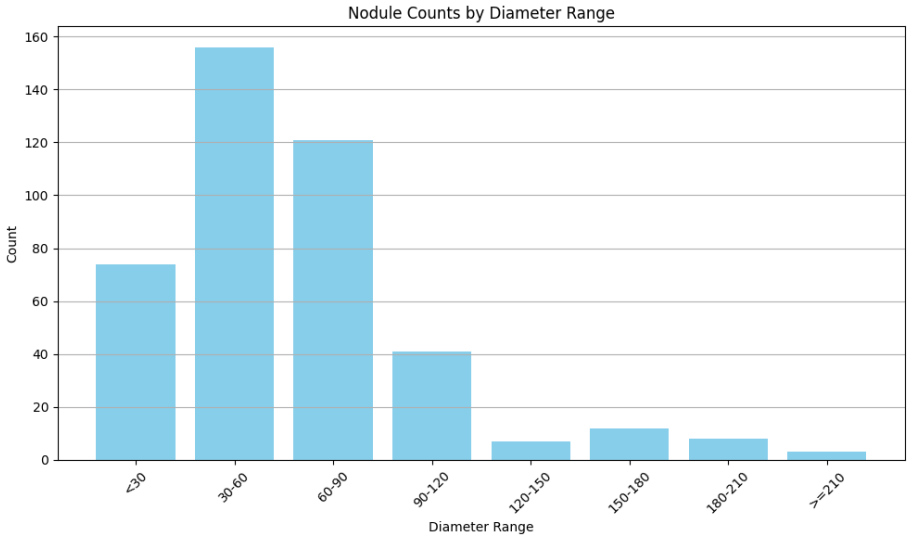


Fig. 6. The nodule density based on their diameter.

Furthermore, Fig. 7 indicates that most cancerous nodules are observable in slices ranging from 30 to 90.

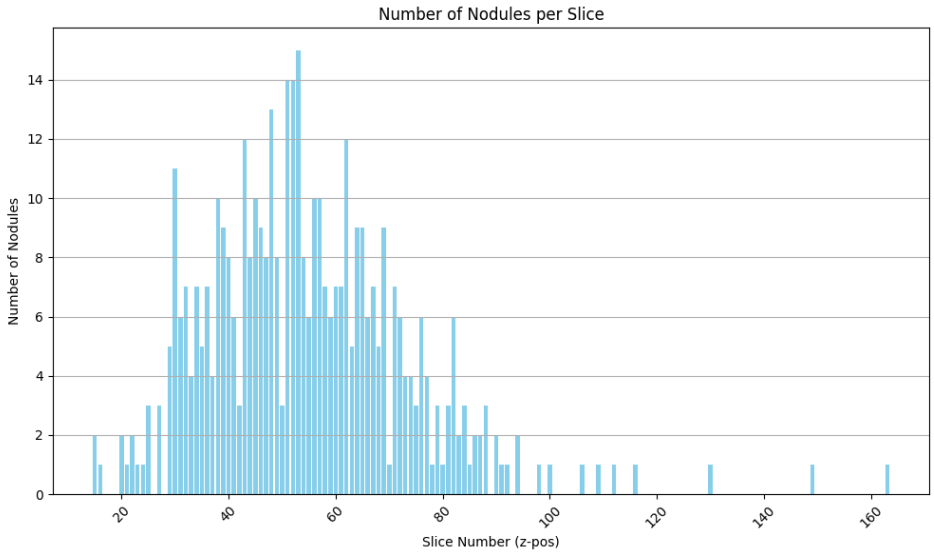


Fig. 7. Distribution of Cancerous Nodules Across Slices.

Subsequently, the slice containing the cancerous nodule, along with the preceding and following slices, is saved in a directory named "Dataset". Based on the information extracted from the dataset, Fig. 8 illustrates the precise location of the nodule.

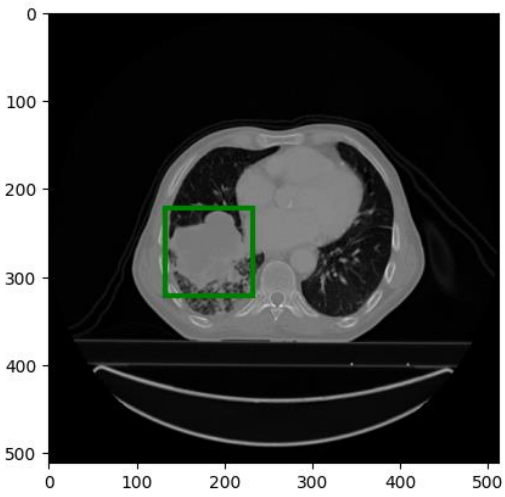


Fig. 8. Visualization of Nodule Location in Image.

The original CT data is stored in DICOM format, typically consisting of 10-250 axial slices, each sized at 512*512 pixels. This dataset retains information regarding the position (coordX, coordY, coordZ). Consequently, this dataset is well-suited for training and evaluating the performance of my detection framework.

Evaluation Metrics:

The performance of networks is frequently described using a confusion matrix. Recall (Sensitivity), Precision, F1-score, and competition performance metric (CPM) are defined as follows:

$$Recall = \frac{TP}{TP + FN} \quad , \quad Precision = \frac{TP}{TP + FP} \quad , \quad F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$\begin{cases} CPM = \frac{1}{N} \sum_{i \in I} Recall_{fpr=i} \\ I = \{0.125, 0.25, 0.5, 1, 2, 4, 8\} \end{cases}$$

where the value of N is set at seven, the variable fpr represents the average number of false positives per scan, while $Recall_{fpr=i}$ signifies the recall rate associated with $fpr = i$.

Following the completion of model training and validation, the test images undergo the computation of average recall. The average recall (AR) involves IoU thresholds ranging from 0.5 to 0.95 in 100 region proposals.

Lung nodule detection:

In this project, a 3D image is used to input the feature extractor (Combined pretrained VGG16) network. Since using the original 3D volume of the CT scan as input to the nodule detection network has a very high computational cost, we use axial slices as input instead. Therefore, for each axial slice in CT images, we extract its two adjacent slices and then change it to an 800*800*3 image. The images are re-scaled for better resolution of nodules. This section converts the 512*512 CT scan images into 800*800 images with the cubic-interpolation method.

The original RPN network that uses pretrained VGG16-Net for feature extraction cannot extract the features of lung nodules with high accuracy, causing limited performance in detecting ROIs of nodules. To solve this problem, by concatenating the up-sampling of the last layer, the fourth layer, and the down-sampling of the third layer of VGG16 and tuning the number of kernels, we can obtain the best performance to detect the ROI of the lung nodules. Combining layers recovers more fine-grained features compared to the original feature map.

This study uses 422 cases. The CT images containing 422 lung nodules are utilized in this study. We consider 10% of images as validation data and use 80% of images as training data. We employ validation data to modify the parameters of the training model. In addition, we train the network end-to-end in the RPN stage by the Adam algorithm. We initialize the weight values of all VGG16 layers by pre-trained a model for ImagNet classification [14]. This model also uses six anchors of different sizes, including 8 * 8, 25*25, 38*38, 58*58, 85*85 and 120*120 for each sliding window. The network model trains in the 2 V100 GPUs environment, and the memory is 192G. Table.1 shows the parameter of the training model.

Table.1. Parameters for training RPN model.

Optimizer	Adam
Batch size	4
Learning rate	0.0001
Weight decay	0.00001

After training the model, when a CT image is input into the trained model, two values, 'anchor_deltas' and 'objectness_score', are output. By selecting the 'anchor_deltas' with the highest 'objectness_score', the most accurate Region of Interest (RoI) can be obtained.

Figure 9 shows examples of CT images that have been input into the trained model. In these images, the green box represents the ground truth bounding box, while the red box represents the predicted bounding box. Here, only the five bounding boxes with the highest 'objectness_score' have been selected as the predicted bounding boxes.

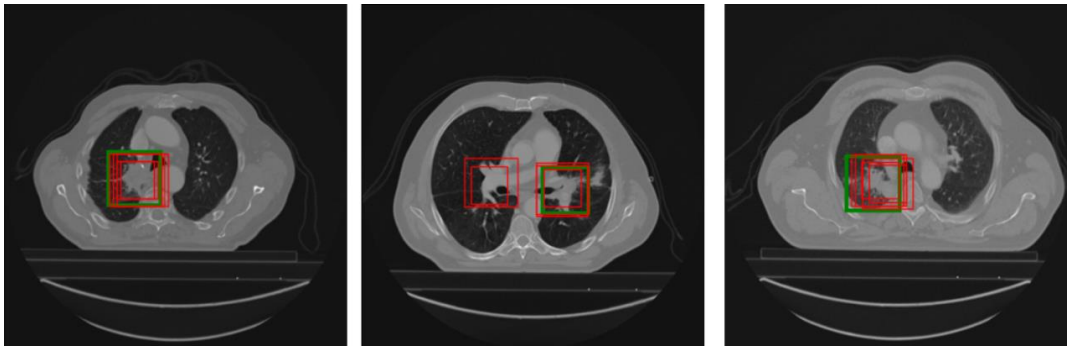


Fig. 9. Ground Truth and Predicted Bounding Boxes on CT Images with five highest score bounding box

Figure 10 shows examples of CT images where, using the top five predicted bounding boxes, the trained model failed to detect the cancerous nodules.

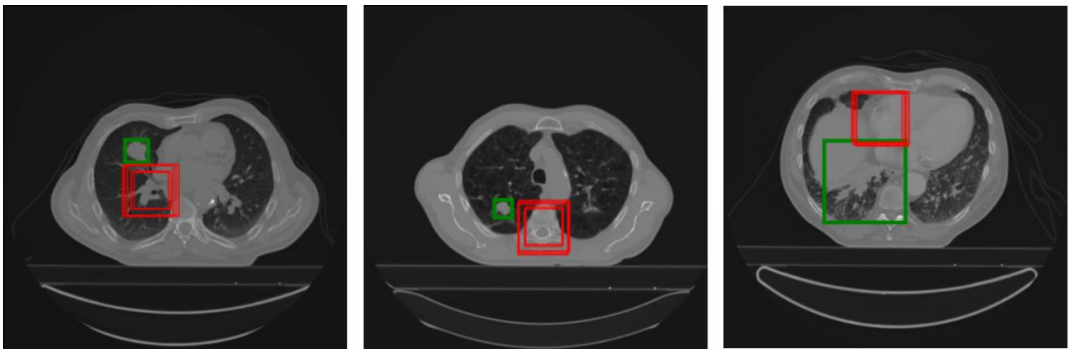


Fig. 10. Ground Truth and Predicted Bounding Boxes on CT Images with five highest score bounding box

Increasing the number of predicted bounding boxes can improve the likelihood of detecting the cancerous nodule; however, it also increases the number of false positives. Figure 11 demonstrates that when the number of predicted bounding boxes is increased from five to ten, the cancerous nodule is successfully detected by the model.

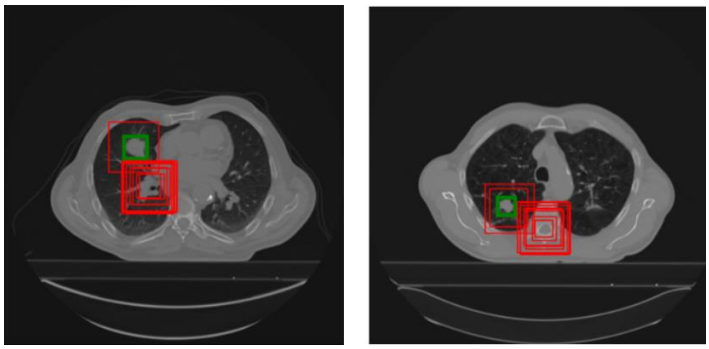


Fig. 11. Improvement in Cancerous Nodule Detection with Increased Number of Predicted Bounding Boxes

As the number of predicted bounding boxes increases, the sensitivity or recall increases. Figure 12 presents the graphs of precision, recall, and F1-score for the test data as the number of predicted bounding boxes increases. These results were obtained with IoU set to 0.5. This chart illustrates that the recall rate increases with the number of predicted bounding boxes, while the average precision at a fixed IoU remains unchanged. This is because, as the number of predicted bounding boxes increases, both the true positives (TP) and false positives (FP) rise simultaneously.

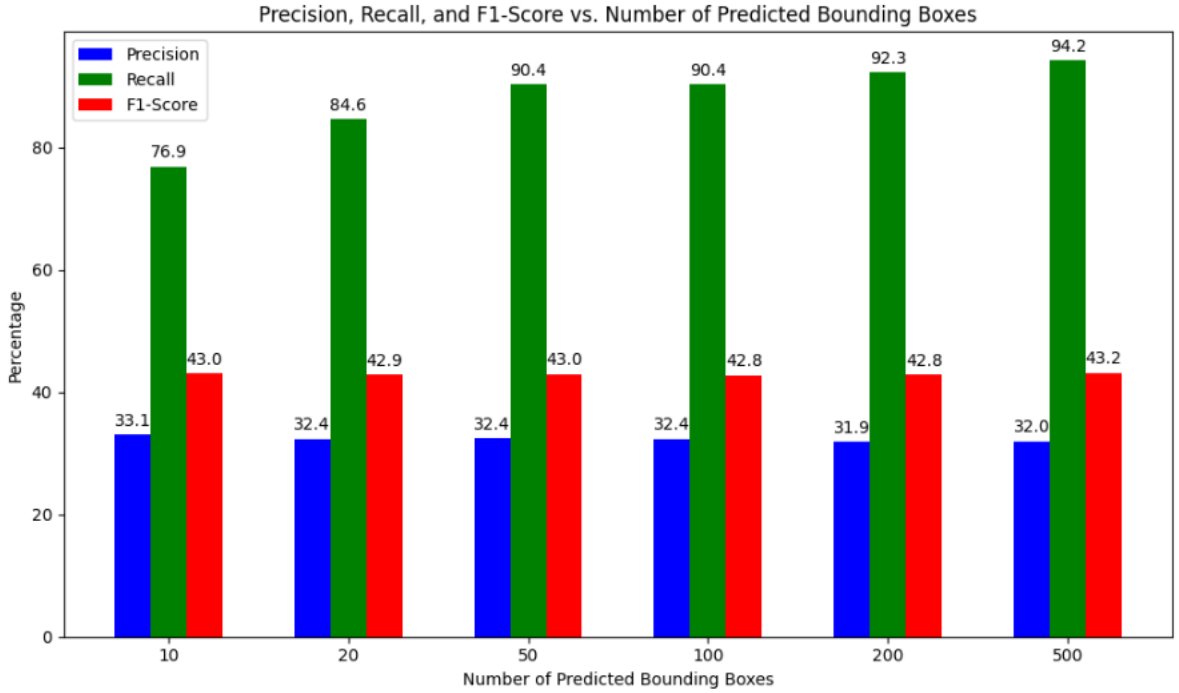


Fig. 12. Precision, Recall, and F1-Score (IoU = 0.5)

As the IoU value increases, the number of true positive nodules decreases, leading to a reduction in precision, recall, and F1-score. Fig. 13 shows the impact of increasing the IoU value on precision, recall, and F1-score. These results were obtained with the number of predicted bounding boxes set to 100. This figure indicates that the average recall (AR) is 59.3%

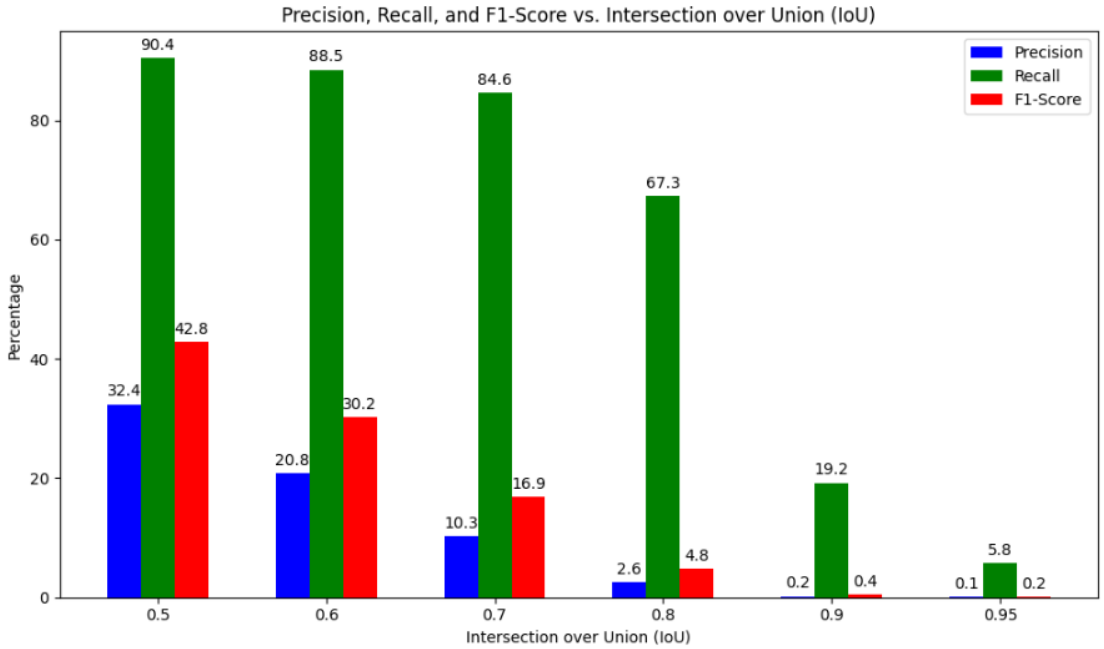
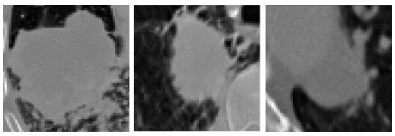


Fig. 13. Precision, Recall, and F1-Score with 100 Predicted Bounding Boxes

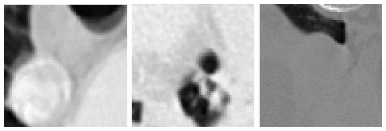
As shown in Figure 13, increasing the intersection over union (IoU) leads to an increase in the number of false positive nodules, which results in a decrease in average precision. By eliminating false positive nodules, we can improve precision and F1-score. Therefore, to reduce false positive nodules, we employ a 2D DCNN model.

False positive reduction stage:

We obtain the patches of size 60*60 by extracting various slices to identify potential nodules. For data augmentation, we use the RPN model from the previous stage. When a CT image is input into the RPN model, the RoIs with an IoU greater than 0.7 are labeled as positive patches, while RoIs with an IoU of zero are labeled as negative patches. To maintain a balance between positive and negative data, we use an equal number of positive and negative patches. Fig. 14 shows examples of positive and negative patches.



(a)



(b)

Fig. 14. Examples of Positive and Negative Patches: (a) Positive Patches; (b) Negative Patches.

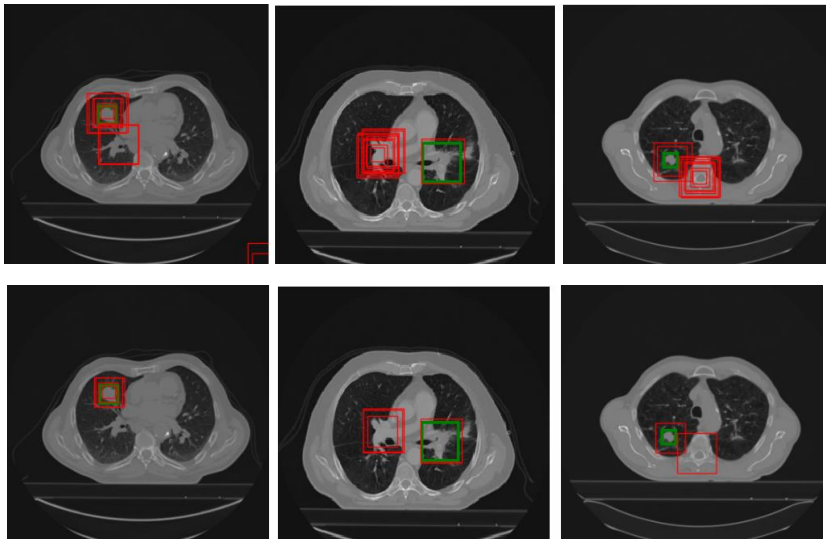


Fig. 15. Predicted Regions of Interest after RPN and FPR Stages: (First line) RPN stage; (second line) FPR stage.

In the false positive reduction (FPR) stage, The Data containing 1739 patches are utilized in this study. We consider 10% of patches as validation data and use 90% of patches as training data. Furthermore, the network model employs the Adam optimizer, with the learning rate and batch size set to 0.001 and 16, respectively.

After training the FPR model, the predicted RoIs from the RPN model are input into the trained FPR model. Fig. 15 illustrates the predicted RoIs after the RPN and FPR stages. As shown in the figure, the number of false positive nodules decreases.

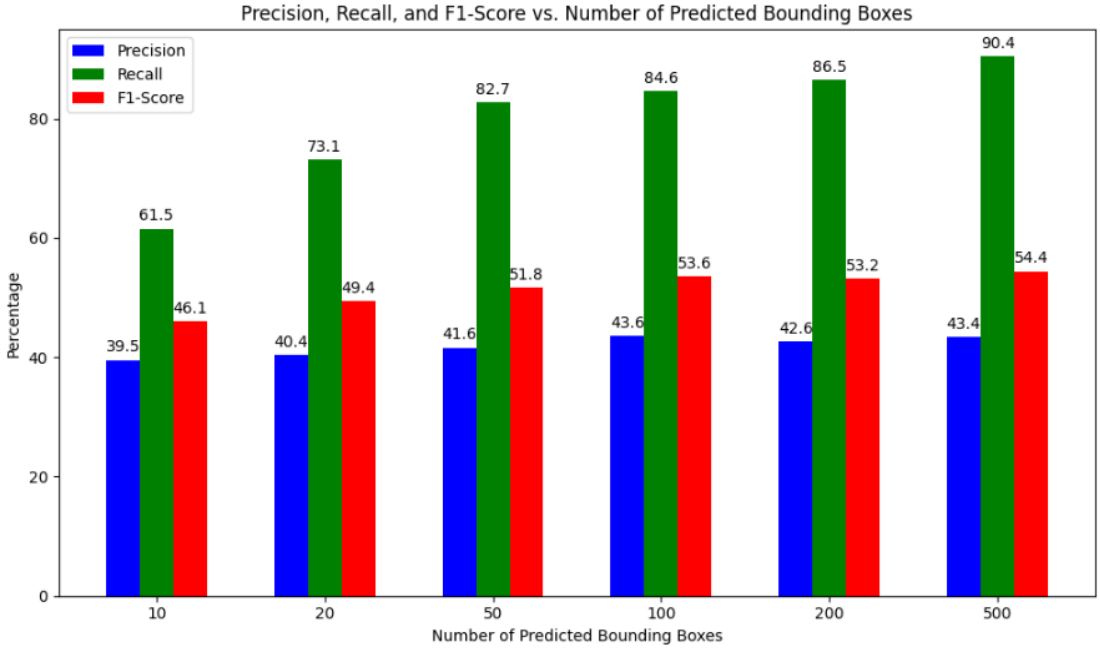


Fig. 16. Precision, Recall, and F1-Score (IoU = 0.5) after FPR stage.

Fig. 16 presents the average precision, recall, and F1-score at an IoU of 0.5 after the FPR stage. The results indicate a 9.5% increase in average precision, while recall has decreased by 7.8%. At this stage, the reduction in false positives (FP), which leads to an increase in average precision, also results in the removal of a number of true positives (TP), causing a decline in recall.

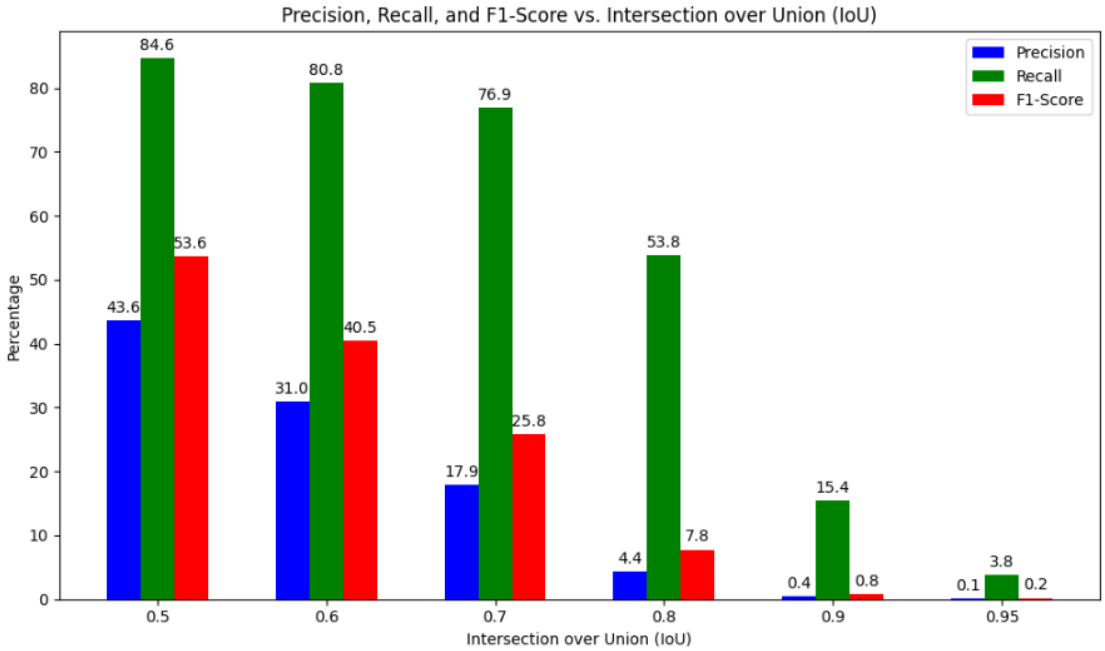


Fig. 17. Precision, Recall, and F1-Score with 100 Predicted Bounding Boxes after FPR Stage

Fig. 17 shows the average precision, recall, and F1-score with 100 Predicted Bounding Boxes after the FPR stage. This figure shows that the average precision at IoU = 0.5 (AP@0.5) is 43.6%, which represents an 11.2% increase compared to the previous stage. Additionally, the average recall (AR) at this stage is 52.55%.

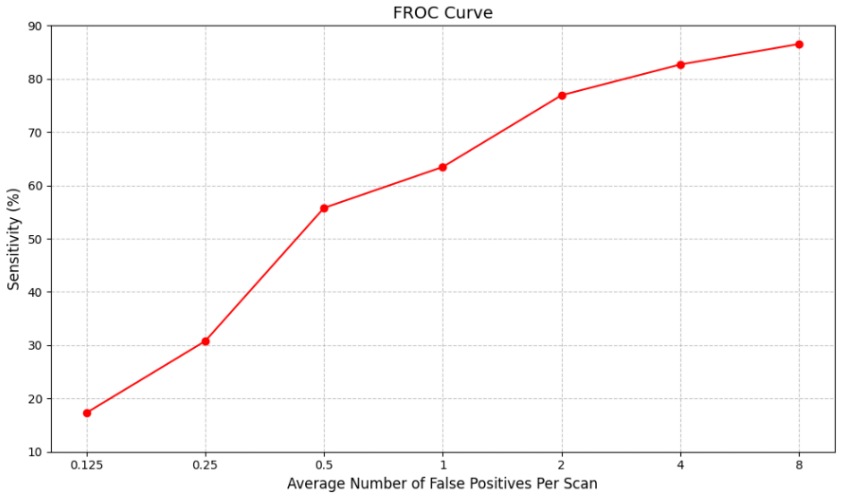


Fig. 18. FROC Curve for Proposed Model.

Fig. 18 displays the FROC curves of VGG16-RPN+FPR model. Sensitivities are calculated at specific false positive rates (FPRs) per patient, including 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPRs. The CPM for the system is derived by averaging sensitivities at these specific points. Our detection system achieves the CPM score of 0.593. The sensitivities at 0.125, 0.25, 0.5, 1, 2, 4, and 8 FPs/scan are 0.185, 0.308, 0.558, 0.638, 0.769, 0.827, and 0.865, respectively.

The proposed method for detecting potential nodules achieves a sensitivity of 90.4% and an average precision of 53.1%. On average, there are 18.2 candidates per scan.

Discussion of challenges and future work:

Lung cancer detection faces several significant challenges that impact its effectiveness and reliability. The variability in nodule appearance, such as differences in size, shape, and texture, complicates detection, while low contrast in CT images and noise add further difficulties. Imbalanced datasets, where cancerous samples are far fewer than healthy ones, often lead to biased models with high false positive or false negative rates, which can result in unnecessary procedures or missed diagnoses. Additionally, the interpretability of deep learning models remains a concern, as black-box models are hard for clinicians to trust. Integrating these AI tools into clinical workflows is complicated by the need for regulatory approval and compatibility with existing systems. Ethical issues, such as data privacy and potential biases, also pose significant challenges. Moreover, maintaining model accuracy over time requires continuous learning and adaptation to evolving medical knowledge and data drift. Addressing these challenges is essential to improve early detection and ultimately save lives.

One of the key challenges in lung cancer detection is the significant variability in nodule sizes. In our dataset, the nodule sizes range from 10 to 250 millimeters, presenting a complex problem for accurate detection. The Region Proposal Network (RPN) model, which we utilize for nodule detection, requires the careful setting of anchor boxes to accommodate this wide range of sizes. However, the current anchor settings may not effectively capture nodules at both ends of this size spectrum, leading to suboptimal detection performance. This variability underscores the difficulty

in designing a model that is both sensitive to small nodules and robust enough to identify larger ones.

To address this challenge, future work will focus on optimizing the anchor list using metaheuristic optimization algorithms, such as the advance parameter setting free harmony search (PSF-HS) [15]. By leveraging this approach, the anchor boxes can be more precisely tuned to detect nodules across various sizes, improving the RPN model's overall accuracy. This optimization method holds promise in enhancing the detection capability of the model, ensuring that nodules of all sizes are accurately identified. Ultimately, this approach aims to reduce false negatives and improve early detection rates, which are critical for successful lung cancer treatment and patient outcomes.

Another significant challenge in lung cancer detection lies in obtaining an efficient feature map from the feature extractor structure. The ability to capture both local and global features is crucial for accurately detecting nodules, as these features offer complementary insights into the nodule's characteristics. Traditional Convolutional Neural Networks (CNNs) excel at extracting local features, such as edges and textures, but they fall short in capturing long-term dependencies and global context. This limitation can lead to suboptimal detection, particularly when nodules have complex patterns that require understanding beyond local neighborhoods.

To overcome this challenge, one promising solution is to incorporate Vision Transformers, specifically the Shifted Window (Swin) Transformer [16], into the feature extraction process. Swin Transformers are adept at capturing global features due to their self-attention mechanism, which models long-range dependencies across the image. However, while Swin Transformers excel at global feature extraction, they may lack the fine-grained localization that CNNs naturally provide, which is critical for precise nodule detection. Therefore, relying solely on Swin Transformers could result in a trade-off between global context and local accuracy.

To address these shortcomings, future work will explore the combination of CNNs and Swin Transformers to create a hybrid feature extraction architecture. By integrating the local feature maps generated by CNNs with the global feature maps provided by Swin Transformers, this approach aims to leverage the strengths of both models. The hybrid architecture would enable the detection system to maintain high sensitivity to local details while also understanding the broader context, improving the accuracy and robustness of lung nodule detection across a range of scenarios. This fusion of CNN and transformer features has the potential to significantly enhance the performance of lung cancer detection models, making them more reliable and effective in clinical applications.

Conclusion:

In conclusion, this study presents an effective method for lung cancer detection using CT images by integrating a Region Proposal Network (RPN) with feature maps derived from the pretrained VGG16 model. The key innovation lies in combining the last three layers of VGG16 to enhance the resolution and detail of the feature maps, thereby improving the accuracy of detecting nodules of various sizes. The RPN efficiently generates region proposals that are further refined through a false positive reduction stage using a 2D Deep Convolutional Neural Network (DCNN). The

experimental results demonstrate that the proposed method significantly enhances detection performance, with improvements in both accuracy and stability.

This approach addresses the limitations of existing methods, such as inadequate feature resolution and high false positive rates. By utilizing a more detailed and integrated feature extraction process and incorporating advanced loss functions like the Generalized Intersection over Union (GIoU), the method achieves more reliable and precise nodule detection. The use of a false positive reduction stage further enhances the robustness of the system, making it a valuable tool for early lung cancer detection.

Addressing the challenges in lung cancer detection requires innovative approaches to improve model performance across diverse scenarios. The significant variability in nodule sizes, necessitates the optimization of anchor settings in the RPN model, potentially through metaheuristic algorithms like advanced PSF-HS, to enhance detection accuracy across all size ranges. Additionally, the integration of Vision Transformers, particularly the Swin Transformer, into the feature extraction process offers a promising solution to capture global features, though it may compromise fine-grained localization. To overcome this, future work will explore a hybrid architecture that combines the strengths of CNNs and Swin Transformers, enabling a more comprehensive and precise feature map for lung nodule detection. By tackling these challenges, we aim to improve early detection rates, reduce false negatives, and ultimately enhance clinical outcomes for lung cancer patients.

References:

[1] M. Sudhamani et al., “Techniques for detection of solitary pulmonary nodules in human lung and their classifications-a survey,” International Journal on Cybernetics & Informatics (IJCI), vol. 4, no. 1, p. 27, 2015.

[2] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: a cancer journal for clinicians, vol. 71, no. 3, pp. 209–249, 2021.

[3] Z. Zhou, F. Gou, Y. Tan, and J. Wu, “A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in developing countries,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 11, pp. 5619–5630, 2022.

[4] H. Huang, R. Wu, Y. Li, and C. Peng, “Self-supervised transfer learning based on domain adaptation for benign-malignant lung nodule classification on thoracic ct,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 8, pp. 3860–3871, 2022.

[5] H. Mkindu, L. Wu, and Y. Zhao, “Lung nodule detection of ct images based on combining 3d-cnn and squeeze-and-excitation networks,” Multimedia Tools and Applications, vol. 82, no. 17, pp. 25 747–25 760, 2023.

[6] S. A. Agnes, J. Anitha, and A. A. Solomon, “Two-stage lung nodule detection framework using enhanced unet and convolutional lstm networks in ct images,” Computers in Biology and Medicine, vol. 149, p.106059, 2022.

[7] L. Zhu, H. Zhu, S. Yang, P. Wang, and H. Huang, “Pulmonary nodule detection based on hierarchical-split hrnet and feature pyramid network with atrous convolution,” Biomedical Signal Processing and Control, vol. 85, p. 105024, 2023.

[8] D. Zhao, Y. Liu, H. Yin, and Z. Wang, “An attentive and adaptive 3d cnn for automatic pulmonary nodule detection in ct image,” Expert Systems with Applications, vol. 211, p. 118672, 2023.

[9] J. Xu, H. Ren, S. Cai, and X. Zhang, “An improved faster r-cnn algorithm for assisted detection of lung nodules,” Computers In Biology And Medicine, vol. 153, p. 106470, 2023.

[10] Z. Cai and N. Vasconcelos, “Cascade r-cnn: Delving into high quality object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6154–6162, 2018.

[11] X. Han, J. Chang, and K. Wang, “You only look once: unified, real-time object detection,” Procedia Computer Science, vol. 183, no. 1, pp.61–72, 2021.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp. 21–37. Springer, 2016.

[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” International journal of computer vision, vol. 115, no. 3, pp. 211–252, 2015.

[15] Y.-W. Jeong, S.-M. Park, Z. W. Geem, and K.-B. Sim, “Advanced parameter-setting-free harmony search algorithm,” Applied Sciences, vol. 10, no. 7, p. 2586, 2020.

[16] Liu, Z.; Lin, Y. T.; Cao, Y.; Hu, H.; Guo, B. N. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022, 2021.