Comparative Analysis of Detection Transformers and YOLOv8 for Early Detection of Pulmonary Nodules

Victor Ferraz[1], **Marcelo Oliveira**[1], **Nilson Carvalho**[1], **Tarcísio Ferreira**[1]

[1]Instituto de Computação – Universidade Federal de Alagoas (UFAL)
57.072-970 – Maceió – AL – Brazil

{vmhf,oliveiramc,nsc,tlf}@ic.ufal.br

Abstract. *Lung cancer (LC) is the second most prevalent type of cancer worldwide and the deadliest, accounting for one in every five cancer-related deaths globally. The chances of survival for patients detected with this type of cancer increase considerably when the diagnosis is made early, with the 5-year survival rate reaching up to 70%. Radiologists perform LC diagnosis through Computed Tomography (CT) images, but such diagnosis is a complex and error-prone task. Through computer-aided tools, this diagnostic process can be automated, reducing time and effort for specialists, as well as improving confidence in the diagnosis. The objective of this work was to evaluate and compare the effectiveness of Convolutional Neural Network (CNN) and Transformer architectures in detecting small lung nodules ($\leq$15mm), where the guiding research question of this work was "What is the impact of the size of lung nodules on the detection accuracy of CNN and Transformer architectures?". The dataset used was based on the public database LUNA16, filtering the test set to include only sections with nodules smaller than 15mm. The models chosen for our comparisons were YOLOv8, a CNN considered state-of-the-art in object detection, and DEtection TRansformer (DETR), which combines the transformer architecture with a CNN layer, where we obtained results such as mAP50 = 0.70, Sensitivity = 0.91 and Λ = 0.85 for the DETR and mAP50 = 0.90, Sensitivity = 0.83 and Λ = 0.77 for the YOLOv8. We also assessed the impact of nodule size on the performance of both models, where the performance of YOLOv8 was impacted by the decrease in nodules size, while DETR continued to show satisfactory results regardless of how small the nodules were.*

1. Introduction

Global Cancer Statistics indicate that lung cancer is the second most prevalent type of cancer worldwide, with 2.2 million new cases, and the most lethal, with 1.8 million deaths reported in the year 2020, representing approximately one in every 10 (11.4%) cancer diagnoses and one in every 5 (18%) cancer deaths worldwide. Early diagnosis of lung cancer is essential to increase the chances of patient survival, with a 5-year survival rate reaching 70% when diagnosed at stage I [Blandin Knight et al. 2017]. Therefore, programs aimed at early detection of pulmonary nodules must be implemented to increase the survival chances of individuals affected by lung cancer.

Computed Tomography (CT) is the primary tool used by radiologists to detect lung nodules as it provides high-resolution 3D images with contrast, reflecting differences in intensity, texture, and shape of tumors. However, the process of diagnosing lung nodules, especially in the early stages (with up to 15mm in diameter size) is a challenge on CT images. Besides their small size, nodules may have low contrast at an early stage compared to the lung tissue and can be attached to other complex lung structures, increasing the difficulties [Lima et al. 2021].

Computer-aided diagnosis (CAD) systems are important tools that can provide support for radiologist's decision-making, functioning as a second opinion [Halder et al. 2020]. CAD tools can automate the diagnostic process, reducing the time and effort required for analysis and improving the task's reliability and repeatability [Ferreira et al. 2018, Choi and Choi 2013]. CAD systems involve both lesion localization in medical images (CADe) and lesion classification as malignant or benign (CADx), which brings various challenges to the field [Firmino et al. 2016].

Currently, Deep Learning (DL) techniques are the state-of-the-art in CAD applications for lung nodule detection, and several studies [McBee et al. 2018, Adams et al. 2021] have demonstrated the potential of DL in lung nodule detection, as well as the cost-effectiveness of using CAD systems in healthcare centers. Among DL architectures, Convolutional Neural Networks (CNNs) have emerged as the leading architecture in the medical informatics field due to exceptional results obtained in computer vision [Ravì et al. 2016]. However, despite the advances in CNN models, the number of false positives in lung nodule detection remains a challenge for implementing systems in medical clinics [Liang et al. 2021, Shaukat et al. 2019]. The complex boundaries of lung nodules and visual similarity to surrounding tissues make accurate detection of small lung nodules complex. Traditional CNN-based lung nodule detection models focus on extracting local features from neighboring pixels and ignore global contextual information, making detecting small lung nodules challenging [Li et al. 2022].

The Transformer is an encoder-decoder architecture that revolutionized the field of Natural Language Processing (NLP) by presenting a simpler structure without the need for convolutions, becoming state-of-the-art in translation tasks, and has since been used in various areas. The Vision Transformer (ViT) emerged as an adaptation of the Transformer for computer vision tasks, functioning in combination with traditional CNN architectures or completely replacing them. In the area of nodule detection using ViT, [Zhu et al. 2022] proposed an end-to-end architecture that utilizes a U-shaped residual network in combination with the attention mechanism and achieved 95% sensitivity in detecting pulmonary nodules with a significantly lower number of parameters compared

to CNN models while also reducing the number of false positives. [Niu and Wang 2022] proposed a 3D ViT-based region model to identify pulmonary nodules in a set of candidate regions. The proposed model achieved superior results (3% improvement) in nodule detection compared to state-of-the-art 3D CNN models.

Despite a wide range of studies on the use of CAD in the detection of pulmonary nodules, the early identification of these nodules, particularly those smaller than 15 millimeters (mm), is still a problem that needs further investigation.

In this context, the main goal of this work was to evaluate and compare the effectiveness of Convolutional Neural Network (CNN) and Transformer architectures in detecting small lung nodules ($\leq$15mm). The guiding research question of this work was "What is the impact of the size of lung nodules ($\leq$15mm) on the detection accuracy of Convolutional Neural Network (CNN) and Transformer architectures?"

2. Materials and Methods

The overview schema of this work is presented in Figure 1. The image dataset used for model training was generated from LUNA16, a 3D chest CT scans. The image dataset was preprocessing (Figure 1-A), and specific slices from each CT scan were selected to compose the training dataset; in addition to delimiting the bounding boxes of each nodule, section 2.1 outlines each of the preprocessing steps undertaken to obtain the image dataset for model training (Figure 1-B). The dataset was then split into training and test sets, with an 80/20 ratio. Following this, an analysis of nodule sizes was performed on the slices present in the validation set, keeping only the slices where the nodules diameters were less than 15mm (Figure 1-C), described in section 2.2. Subsequently, two models were trained: DETR and YOLOv8, undergoing a loop where both models were trained, the results evaluated, and improvements in the hyperparameters were made in order to enhance the performance of the models (Figure 1-D), section 2.3 elucidates the motivation behind the choice of the models used in this work and section 2.4 illustrates the configuration and execution of the training for both models. Finally, an analysis and comparison of the results (Figure 1-E) obtained using the metrics described in section 2.5 were conducted. In this work, the Python programming language (version 3.7) was utilized, along with the DETR [Carion et al. 2020] and YOLOv8 [Jocher et al. 2023] models, in addition to the Pytorch, Skimage, Scipy, Monai, and Ultralytics libraries.

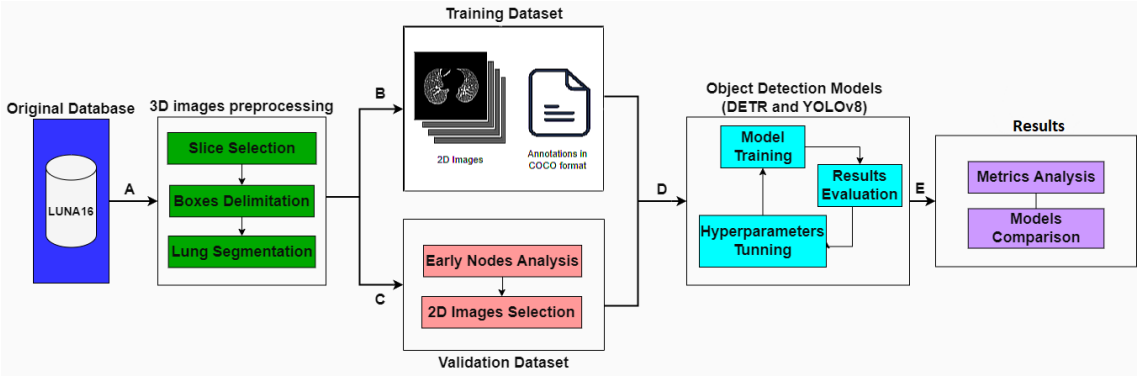

Figure 1. Methodology scheme applied in this work.

2.1. Database

The selected database for this study was LUNA16 (LUng Nodule Analysis 2016) [Setio et al. 2017], a publicly available dataset of computed tomography (CT) scans of the thoracic region from patients identified with lung nodules, containing 888 CT Scans and 1186 nodule annotations.

In this work, we used 2D images due to the computational cost associated with processing 3D images and the fact that one of the analyzed models (YOLOv8) does not support 3D images. In constructing our 2D image database, it was essential to identify which slices contained information about the nodules. Since all slices from the same CT exam share the same properties (origin, spacing, direction matrix, directories, etc.), a class was designed to store information related to the same examination.

Then, our 2D database was created with 4,098 images (sections) containing one or more nodules. To adjust the brightness/contrast and highlight the lung region, the window/level values were adjusted to 1200/-400, respectively, as suggested by [Lima et al. 2020].

2.2. Dataset Organization

For the creation of our dataset, the database was split into training and testing sets, with 80% for training and 20% for testing. We mitigated data leakage by ensuring that images of the same nodule were not included in the training and testing sets. In total, the training set contains 3,255 images, while the testing set contains 832. The images were saved in their original resolution (512 x 512) with 8 bits.

Thereafter, an analysis of the nodules in the test set was performed, selecting only the slices of the CTs where the detected nodules had a diameter of less than 15mm, thus characterizing them as small nodules. This process involved separating all the sections belonging to the same CT scan in the test set and checking if the largest nodule in this set of sections has a diameter smaller than 15 mm. Thus, all the set of slices was added to the final version of the dataset, while those failing this check were removed. Consequently, out of the originally 832 images in the test set, only 587 remained, which were utilized for model validation. Furthermore, a subdivision was made in the dataset where three ranges (R1, R2 and R3) of nodule diameters (d) were selected (table 1).

Table 1. Table showing the disposition of images within the determined ranges.

Diameter (d)	Sections (Test Set)
R1: 15mm $\geq d >$ 10mm	172
R2: 10mm $\geq d >$ 5mm	385
R3: 5mm $\geq d$	30

2.3. Model Selection

In this work, we aimed to compare the results of two significantly distinct detection models. The first is a type of neural network that has been gaining popularity over the last years, known as Transformer, in this case, the *DEtection TRansformer* (DETR) [Carion et al. 2020]. The other chosen model was *You Only Look Once* (YOLO) [Jocher et al. 2023], a neural network maintained by Ultralytics that is well-known and

established in the field of object detection, currently in its eighth version, which was utilized in this study.

The DETR employs the Transformer architecture in conjunction with a CNN (Convolutional Neural Network) and simplifies the detection pipeline by not requiring custom layers, making it easily reproducible in any framework containing a standard CNN and Transformer classes [Carion et al. 2020].

The latest version of YOLO was chosen because, as mentioned earlier in this section, it is an acclaimed object detection model in the field and considered state-of-the-art. This model has demonstrated excellent results in various domains, and has shown considerable performance improvements compared to its previous versions [Jocher et al. 2023].

2.4. Model Training

The training for both models was conducted on a remote server running the Linux operating system (Ubuntu 20.04 LTS) and equipped with an NVIDIA Tesla T4 graphics card with 16GB of memory. The initial procedure for training involved creating a fork of the DETR repository on GitHub to make the necessary modifications and adapt the model for lung nodule detection. The number of layers was kept the same as in the original model. The first modification involved changing the number of object queries, which is the quantity of objects predicted per image where the number was decreased to 10 since the number of nodules in the same image is generally much smaller than 100 (default value). Other model hyperparameters were readjusted as training progressed.

One of the transformations performed by the model before training for better generalization involves randomly resizing the images, following a list of sizes. The maximum size for upscaling was set to 800 x 800. The batch size was set to 8. The model was then trained for 400 epochs. The learning rate drop value was also increased from 200 to 1000 due to the relatively small number of images in our database.

Continuing with training and utilizing the identified hyperparameters, we were able to train the model for 400 epochs, with each epoch taking an average of 10 minutes to complete, totaling approximately 66 hours of training.

The training of YOLOv8 was straightforward, as Ultralytics provides a Python package that greatly simplifies the training and model usage process. The training was set to run for 400 epochs, similar to DETR. However, an early stopping mechanism was implemented, whereby if no improvement was observed in training for the last 50 epochs, the training would be stopped. Consequently, the model had its training stopped after 126 epochs. The training for this model occurred relatively quickly compared to DETR, taking no more than 2 hours.

Due to the ease and speed of training, we decided to train a third YOLOv8 model, this time applying data augmentation to the training set. Rotations of 90, 180, and 270 degrees were applied to random images in the original set, thereby increasing the training set from 3,255 to 9,469 images. The training of this model was also conducted with the same configurations as the previous one, stopping after 77 epochs and taking approximately 80 minutes to complete.

2.5. Metrics and Evaluation

We evaluated the models using various statistical metrics, which includes: sensitivity (Eq. 1), precision (Eq. 2), f1-score (Eq. 3), mAP50, AUC-FROC and Λ.

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$F1score = \frac{2 * precision * recall}{precision + recall} \tag{3}$$

Where TP is the number of true-positive samples, FN is the number of false-negative samples and FP is the number of false-positive samples. Average Precision (AP) is a numerical value that aids in comparing different object detection models. The general definition of Average Precision is the area under the precision-recall curve. The COCO evaluation method employs a 101-point interpolation for AP calculation along with the mean over ten IoU (Intersection Over Union) thresholds. AP@[.5:.95] corresponds to an average of AP for IoU ranging from 0.5 to 0.95 with intervals of 0.05. Finally, mAP50 represents the same as AP@[0.5].

The Free-Response Receiver Operating Characteristic (FROC) Curve is defined as the relation between the Sensitivity and the False Positive Rate (FPR) of a detection model, where we can calculate the Area Under Curve (AUC) score, which is an excellent way to describe the effectiveness of an object detection model, with a higher AUC indicating a better-performing model.

Although the area under the FROC curve summarizes the performance of the FROC system for all decision thresholds, in some instances the area under the FROC curve might be considered a suboptimal or, worse, a potentially misleading summary index of the overall performance of the system [Bandos et al. 2009].

To address this issue, [Bandos et al. 2009] proposed a method to enhance this metric by adding an augmentation of the FROC curve up to a specified FPR value and a new curve representing a naive model that would only attempt to guess the values, called guessed curve. Thus, the new metric would be the area under the augmented FROC curve and above the guessed curve, with this value denoted by Λ.

3. Results and Discussion

3.1. Training Results

Figure 4 presents the results of the DETR models for class error and mAP50 obtained after training for 400 epochs for both datasets. Our DETR model is represented by the orange curve. The class error values displayed in the image are not exact due to using fewer predictions per image (10) than the original model (100). However, the curve is useful for observing that the class error is decreasing throughout the training. The second metric shown is the mAP50, which refers to the average precision using an IoU of 0.5, where we can observe that the curve, exhibits a slight upward trend until the last recorded epochs,

indicating that with a few more training epochs, the bounding boxes could become even more precise. We can observe the training results of the best YOLOv8 model in Figure 3, behaving similarly to the DETR graphs, however, achieving higher mAP50 values within a relatively smaller epoch range. To compare the results more directly, Table 2 contains the metrics described in 2.5 for all three trained models. Thus, it is easy to observe that concerning model precision (mAP50), YOLOv8 stands out considerably compared to DETR, where the best YOLOv8 model achieved mAP50 = 0.896. Another interesting comparison can be made by observing equation 3, which relates sensitivity and precision, serving as a more reliable general indicator of the model's effectiveness, where the results of YOLOv8 stand out once again.

3.2. Evaluation Results

For a more precise validation of the models, we will utilize the FROC curve, as described in Section 2.5, to provide a more convincing comparison measure than mAP. Using the Monai Python package, it was possible to plot the FROC curves for the five models. The first observation is the AUC values of the models, where DETR clearly stands out, and the sensitivity score values are also superior, as can be seen in Table 2. However, paying closer attention to the metric of the DETR model, something slightly inconsistent is noted, where the AUC of the model is almost twice as high as the best YOLOv8 model, while the false positive rate is significantly higher. Consequently, we can understand what the authors of [Bandos et al. 2009] meant by stating that the AUC value of the FROC curve can be a misleading value. To address this, a script was created to plot the guessed curve and the augmented FROC curve following the equations described by the aforementioned authors, truncating at the value of FPS = 1 for all models. The resulting graphs can be observed in Figure 2, where the Λ metric seems to reflect more concisely what has been observed from the previous metrics.

For the final comparison between the models, we will look at the values of TP, FP, FN, Sensitivity, FROC AUC, and Λ for all trained models in Table 2. Using Λ as the primary metric for model comparison, we observe that the best model was the DETR, obtaining a $\Lambda = 0.847$, where an ideal model has a $\Lambda = 1$. An important observation is to note which model had fewer false negatives, as in a medical context, this is a very significant type of errors. In this case, we observe that out of a total of 595 nodules present in the 587 sections of the validation set, the DETR model once again performed the best, having the lowest occurrence of FN among the five models but presenting a high value of FP compared with the YOLOv8, which is a tradeoff that can be studied further in subsequent works.

Table 2. Table displaying values of some of the metrics described in Section 2.5, with the optimal value for each metric highlighted.

Model	mAP50	f1-Score	TP	FP	FN	Sensitivity	FROC AUC	Λ
DETR	0.703	0.772	**542**	156	**53**	**0.910**	**0.222**	**0.847**
YOLOv8	0.844	0.805	487	84	108	0.818	0.100	0.761
YOLOv8 - Aug.	**0.896**	**0.848**	492	**69**	103	0.827	0.085	0.774

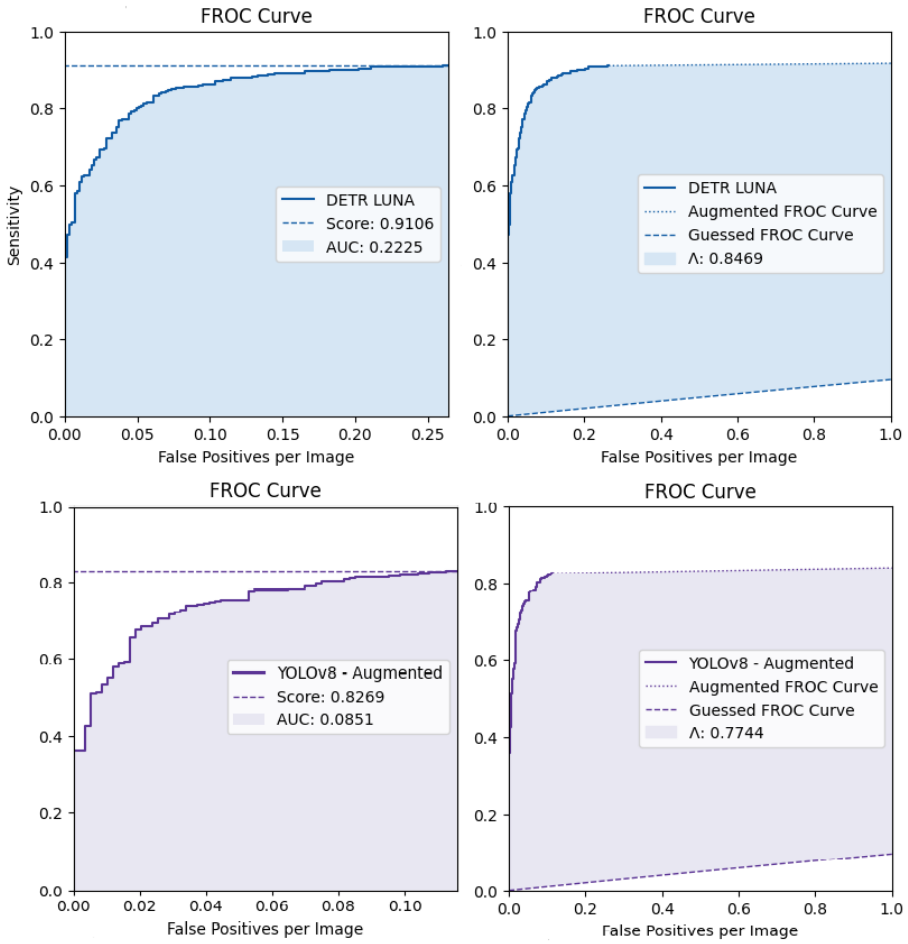

Figure 2. Plots of the FROC Curves and the respective AUCs and Λ for the best DETR and YOLOv8 models.

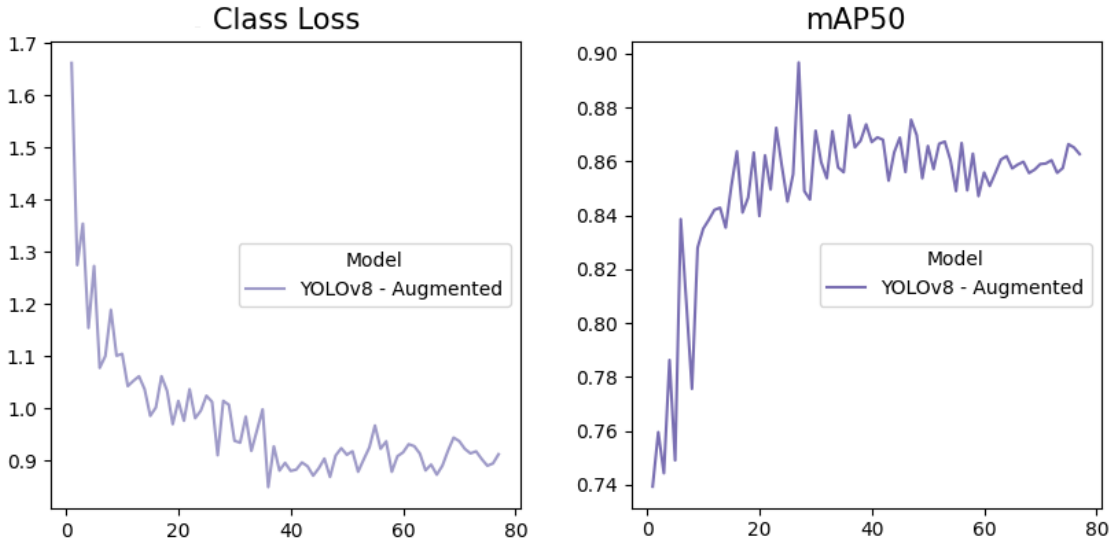

Figure 3. Plots of the best YOLOv8 model training logs.

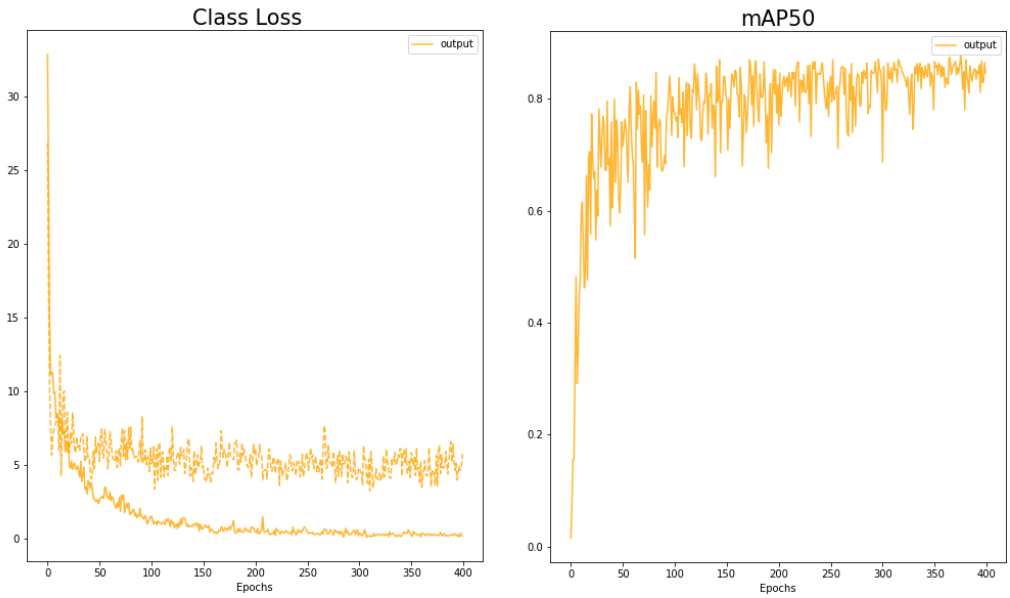

Figure 4. Plots of the DETR models training logs.

3.3. Inferences and Discussion

To perform inferences on the images, DETR allows us to load the generic model from the TorchHub repository and update it with the weights obtained during training. Each inference returns ten queries with ten predictions, but we retain only predictions with confidence above 0.8. For YOLOv8, the process is simplified using the Ultralytics Python package, where the model is automatically saved at the end of training, and this model can be loaded directly for inference, setting the confidence above 0.8, similar to DETR. Some examples of inferences were saved, where inferences made by the models are shown in Figure 5. The blue bounding boxes represent the model's inference, while the red boxes show the ground truth values.

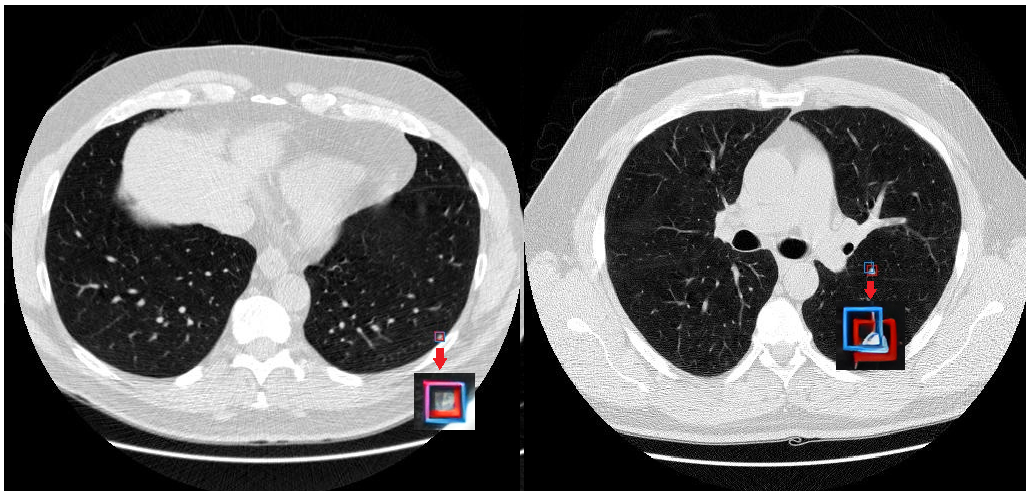

Figure 5. Model inferences (in blue) alongside ground truth values (in red). YOLOv8 inferences on the left and the DETR ones on the right, with a zoom in on the nodules for better visualization.

The detections obtained by all models showed good results in terms of bounding box localization and size, closely approximating the values marked by radiologists. The best-performing model in detecting small nodules observed in this study (DETR) occupies 1.9 GB of GPU memory, and each inference takes 0.6 seconds to execute (1.4 seconds to perform the inference and plot the result). Therefore, we understand that the model could be used on a standard computer with a simple GPU.

At the end of Section 2.1, it was mentioned that the test dataset was divided into three categories: nodules between 15 and 10 mm in diameter (R1), 10 to 5 mm (R2), and smaller than 5 mm (R3). It is important to note that our test set contained only 30 images in R3, which makes the detection of these nodules extremely challenging. Besides the minimal number of examples present in the dataset, a CT scan slice with such a nodule would present very little information (pixels), thus hindering the generalization of the models. To address the question posed at the end of Section 1, we will examine the values of mAP50, Sensitivity, and Λ for both models in ranges R1, R2, and R3 (Table 3). At first glance, we can observe that the results follow the same logic as those presented in Section 3.2, where YOLOv8 performs better in precision, while DETR stands out in the other two. We can clearly see the impact of decreasing nodule size on YOLOv8, where all metrics decrease as nodules become smaller, while DETR manages to remain relatively stable across metrics, regardless of the range being evaluated. This is especially evident when directly comparing the results of R1 (larger nodules, more images) and R3 (smaller nodules, less images), where YOLOv8 experiences a sharp drop in mAP50 and a considerable decrease in the other two metrics, whereas DETR not only appears unaffected negatively but even improves in mAP50 and sensitivity, remaining nearly unchanged in Λ.

Table 3. Table displaying values of mAP50, Sensitivy and Λ,for both models in the three diameters ranges, with the optimal value for each metric highlighted.

Model	mAP 50			Sensitivity			Λ		
	R1	**R2**	**R3**	**R1**	**R2**	**R3**	**R1**	**R2**	R3
YOLOv8 - Aug.	**0.907**	**0.902**	**0.727**	0.829	0.830	0.766	0.785	0.775	0.662
DETR	0.663	0.729	0.665	**0.858**	**0.851**	**0.900**	**0.801**	**0.790**	**0.799**

To evaluate the performance of our models with larger nodules, we conducted an evaluation using the test dataset before filtering out small nodules, i.e., with nodules larger and smaller than 15 mm. YOLOv8 achieved an mAP50 of 0.907, similar to the metric in R1, and a Sensitivity of 0.847, higher than all obtained in all ranges, which reinforces the difficulty the model faces when reducing the size of nodules. On the other hand, DETR achieved an mAP50 of 0.866, significantly superior to all ranges, and a Sensitivity of 0.899, similar to R3.

After all these analyses, it becomes clear to note that YOLOv8 excels in terms of precision, indicating that the model has a low rate of false positives (FP), while DETR stands out in sensitivity, presenting a low rate of false negatives (FN). Both types of errors are relevant in the context of nodule detection; therefore, it would be up to the professional to choose which model to apply for each case, depending on which type of error would be less impactful.

Object detection models using the ViT architecture are relatively new and pose some challenges to the research community. In the field of medical images, this challenge is particularly heightened by the scarcity of medical images, especially annotated ones, compared to other domains. In this work, we demonstrated that a lung nodule detection model in CT images using a Transformer architecture achieved satisfactory results in terms of mAP, sensitivity, and F1-score. It also outperformed the YOLOv8 model in the analysis based on the FROC curves and exhibited bounding boxes in the inferences with satisfactory alignment with the expected values. Transformer models, including DETR, require a considerably high training time to achieve the best results, making it challenging to experiment with various hyperparameter options—a clear disadvantage compared to YOLOv8, which offers a more simplified and faster training process.

4. Conclusion

In this study, we sought to evaluate the effectiveness of models based on CNN (YOLOv8) and Transformer (DETR) architectures in detecting small lung nodules ($\leq$15mm) and compare their results, where our best DETR model achieved values of mAP50 = 0.703, Sensitivity = 0.91 and Λ = 0.847, results considered satisfactory, especially when comparing with the results obtained by YOLOv8, where it falls behind only in mAP50. The detected nodule bounding boxes also showed good alignment with ground-truth values. The inferences made in this work were done using a Jupyter notebook, but the models can easily be ported to be integrated into a more robust software pipeline or even added to a mobile application, which opens up various possibilities for its use. Considering the results achieved in this work, we can assert that a model with the Transformer architecture is capable of achieving similar results to a CNN model in the detection of small lung nodules, even outperforming YOLOv8. In terms of the impact of nodule size on model performance, it was observed that DETR is not greatly affected by decreasing size, achieving good results across all proposed diameter ranges, whereas YOLOv8 performance drops as nodule size decreases. We also concluded that both models have their limitations, with DETR exhibiting more cases of false positives and YOLOv8 showing more false negatives. This is something that should be taken into consideration by professionals when choosing the most suitable model for each case.

This work was supported by Alagoas Research Foundation (FAPEAL).

References

Adams, S. J., Henderson, R. D., Yi, X., and Babyn, P. (2021). Artificial intelligence solutions for analysis of x-ray images. *Canadian Association of Radiologists Journal*, 72(1):60–72.

Bandos, A. I., Rockette, H. E., Song, T., and Gur, D. (2009). Area under the free-response roc curve (froc) and a related summary index. *Biometrics*, 65(1):247–256.

Blandin Knight, S., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., and Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open biology*, 7(9):170070.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer.

Choi, W.-J. and Choi, T.-S. (2013). Automated pulmonary nodule detection system in computed tomography images: A hierarchical block classification approach. *Entropy*, 15(2):507–523.

Ferreira, J. R., Oliveira, M. C., and de Azevedo-Marques, P. M. (2018). Characterization of pulmonary nodules based on features of margin sharpness and texture. *Journal of digital imaging*, 31:451–463.

Firmino, M., Angelo, G., Morais, H., Dantas, M. R., and Valentim, R. (2016). Computer-aided detection (cade) and diagnosis (cadx) system for lung cancer with likelihood of malignancy. *Biomedical engineering online*, 15(1):1–17.

Halder, A., Dey, D., and Sadhu, A. K. (2020). Lung nodule detection from feature engineering to deep learning in thoracic ct images: a comprehensive review. *Journal of digital imaging*, 33(3):655–677.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). YOLO by Ultralytics.

Li, S., Wu, C., and Xiong, N. (2022). Hybrid architecture based on cnn and transformer for strip steel surface defect classification. *Electronics*, 11(8):1200.

Liang, J., Ye, G., Guo, J., Huang, Q., and Zhang, S. (2021). Reducing false-positives in lung nodules detection using balanced datasets. *Frontiers in Public Health*, 9:671070.

Lima, L., Vieira, T., Costa, E., Azevedo-Marques, P., and Oliveira, M. (2020). Using support vector machine and features selection on classification of early lung nodules. In *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 60–71.

Lima, L. L., Ferreira Junior, J. R., and Oliveira, M. C. (2021). Toward classifying small lung nodules with hyperparameter optimization of convolutional neural networks. *Computational Intelligence*, 37(4):1599–1618.

McBee, M. P., Awan, O. A., Colucci, A. T., Ghobadi, C. W., Kadom, N., Kansagra, A. P., Tridandapani, S., and Auffermann, W. F. (2018). Deep learning in radiology. *Academic radiology*, 25(11):1472–1480.

Niu, C. and Wang, G. (2022). Unsupervised contrastive learning based transformer for lung nodule detection. *Physics in Medicine & Biology*, 67(20):204001.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G.-Z. (2016). Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1):4–21.

Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., Van Den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M. E., Geurts, B., et al. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Medical image analysis*, 42:1–13.

Shaukat, F., Raja, G., and Frangi, A. F. (2019). Computer-aided detection of lung nodules: a review. *Journal of Medical Imaging*, 6(2):020901–020901.

Zhu, X., Wang, X., Shi, Y., Ren, S., and Wang, W. (2022). Channel-wise attention mechanism in the 3d convolutional network for lung nodule detection. *Electronics*, 11(10):1600.