BİTLİS EREN ÜNİVERSİTESİ
FEN BİLİMLERİ DERGİSİ

ISSN: 2147-3129 / e-ISSN: 2147-3188

**Article Type** : Research Article
**Received** : September 9, 2025
**Revised** : December 6, 2025
**Accepted** : December 9, 2025
**DOI** : 10.17798/bitlisfen.1780664

**Year** : 2025
**Volume** : 14
**Issue** : 4
**Pages** : 2694-2711

# PERFORMANCE EVALUATION OF DIFFERENT YOLO MODELS FOR LUNG NODULE DETECTION

**İbrahim ARUK** [1]

[1] *Kahramanmaras Istiklal University, Department of Software Engineering, Kahramanmaras, Türkiye, ibrahim.aruk@istiklal.edu.tr*

## ABSTRACT

Lung cancer is one of the leading causes of cancer-related deaths worldwide. The early diagnosis of this disease is critically important for the success of treatment. Computer-aided diagnosis systems and deep learning methods are widely used to ensure accuracy and speed in the automatic detection of lung nodules. In this study, the performance of medium models of four different YOLO architectures (YOLOv8, YOLOv9, YOLOv10, and YOLOv11) in lung nodule detection was comprehensively evaluated on the LUNA16 dataset. The models were compared using metrics such as precision, recall, F1-score, overall accuracy (mAP50, mAP50-95), and processing speed. The obtained results have shown that YOLOv8 offers high speed and accuracy, YOLOv10 provides the best sensitivity, and YOLOv11 excels in overall accuracy. To our knowledge, this study presents one of the first comprehensive comparisons of the latest YOLO architectures under fair experimental conditions. By systematically analyzing the relationships between performance metrics, this study fills a gap in the literature. Furthermore, our study demonstrates that deep learning-based YOLO models can be reliable and effective tools for the early diagnosis of lung cancer. The findings obtained are of a nature that will contribute to accurate and rapid diagnostic processes in clinical applications.

**Keywords:** Lung cancer, Deep learning, YOLO, Tumor detection, LUNA16, CT imaging.

## 1 INTRODUCTION

Cancer is a serious disease that negatively affects people's lives and quality of life. Among the types of cancer, lung cancer is a disease characterized by abnormal cells that multiply uncontrollably in the lung tissue [1]. According to the World Health Organization, this disease is the leading cause of cancer-related deaths [2]. In a study conducted by Siegel et al.,

it is predicted that there will be a total of 226,650 new lung cancer diagnoses in the United States in 2025, including both men and women. It is also estimated that a total of 124,730 people will die from lung cancer in 2025 [3]. The symptoms of lung cancer generally do not appear in the early stages. Due to this fact, treatment becomes very difficult, and there are very low chances of survival for the patient. In some instances, the symptoms appear after the disease reaches the terminal stage. Some patients do not show any symptoms at all [4].

The identification of nodules in the patient's lungs is a critical indicator of lung cancer. The size of the nodule gives substantial information about the patient's condition. Nodules between 3 mm and 30 mm are closely followed because they can either be benign or malignant. The course of the disease is positively influenced by early identification of these nodules. However, their relatively small size and resemblance to normal lung tissue make identification by specialists difficult. Even though methodologies like computed tomography may give detailed information about the size, location, and other characteristics of the nodules, the resemblance of nodules in computed tomography images to blood vessels and other tissues increases the possibility of misdiagnosis [4, 5].

According to [6], CT imaging can be regarded as the primary method for diagnosing lung diseases. Regular CT scans are conducted with the purpose of decreasing the mortality rate related to lung cancer in the risk group. In clinical practice, tumor characteristics in CT images are typically used by doctors to diagnose lung cancer [7, 8]. The only disadvantage of this approach is that huge amounts of data from CT scans result in serious complications in the diagnostic process. Frequent repeat examinations can even provoke fatigue in an experienced specialist, which might lead to misdiagnoses due to poor visibility of findings. Therefore, the development of a system that can automatically detect lung tumors in CT images is of a high priority to improve diagnostic accuracy and decrease the doctor's workload [9, 10].

The CAD systems help in the detection of lung nodules, thereby facilitating early diagnoses of lung cancers. These technologies enhance accuracy and efficiency in the processes related to the detection of lung nodules by reducing the workload on radiologists. In fact, CAD systems reduce the need to manually review CT images; this saves time and allows radiologists to focus on critical cases.

In recent years, advances in computer technology, machine learning, and imaging, with the increasing use of digital IT images, deep learning methods have been used more frequently in the solution of complex problems such as early detection of lung cancer. These particularly

attracted wide interest since they are beyond traditional methods' limitations of CAD systems [11]. Instead, CNN-based models have higher detection accuracy in CAD systems, helping radiologists to identify nodule candidates with increased precision. These methods enable even small nodules to be detected at an early stage and warn radiologists that detailed examination may be required, thus improving patient outcomes.

Deep learning-based object detection technologies form the basis of today's advanced CAD systems and are very important in highlighting nodules and abnormalities in CT images as a decision support mechanism. Since false detections and omissions in diagnostic processes carry critical risks, maximizing detection and classification performance in lung images is of great importance. In recent years, the YOLO algorithm family, which is widely used in general object detection, has also started to be widely used in the field of lung nodule detection thanks to its high accuracy and flexibility. Therefore, since the different versions and parameter setting of the YOLO algorithm may indicate considerable differences in terms of detection success, it is required to investigate the performances of these models in lung nodule detection in detail.

This study comprehensively evaluates the current approaches for deep learning-based object detection algorithms for lung nodule detection on the LUNA16 dataset. In particular, it compares the intermediate versions of the YOLOv8, YOLOv9, YOLOv10, and YOLOv11 models under the same experimental conditions. It also performs a detailed analysis of the effectiveness of different architectures in lung nodule detection. To this end, the study contributes to the literature by systematic evaluation of different YOLO models in terms of important metrics, namely accuracy, speed, and reliability, which provides objective and comparable performance data. These results give indications regarding which model is likely to be more advantageous. At the same time, such a study will serve as a good reference for further development of deep learning based diagnostic systems. Thus, this study provides information on the practical value of AI-based methods in the early diagnosis of lung cancer and on new contributions to the literature.

The organization of the paper is as follows: related studies on lung cancer detection are presented in the second section. The third section gives the material and methods used. The fourth section presents the experimental studies and their results obtained. Discussion and recommendations for the future are given in the fifth section. Finally, the last section contains the general evaluation and conclusions.

## 2 LITERATURE REVIEW

In recent years, the use of deep learning-based object detection algorithms has increased significantly for the early diagnosis and accurate classification of lung cancer. In particular, the YOLO architecture and its derivatives offer high accuracy rates in detecting lung nodules and cancer subtypes in medical images. This literature review examines the current applications of different YOLO-based approaches in lung cancer diagnosis.

Wehbe et al. used the YOLOv8 algorithm, a single-stage object detection method, to detect and classify different subtypes of lung cancer in their study. In experiments conducted on an open-access dataset, the YOLOv8 model achieved an average mAP of 97.10%, and the small model variant reached a sensitivity rate of 96.10%. Subsequently, the features obtained from YOLO were subjected to dimension reduction. The developed neural network classified tumors, lymph nodes, and metastasis stages with 98% accuracy. Additionally, the model achieved a recall rate of 91.00% on an independent dataset [12]. Liu developed a method called STBi-YOLO to accurately detect lung nodules in CT images. This approach, based on YOLOv5, was compared with leading object detection models such as YOLOv3, YOLOv4, YOLOv5, and Faster R-CNN with SSD. Experimental results showed that STBi-YOLO achieved 96.10% accuracy and 93.30% recall in lung nodule detection [5].

Another study by Sharma et al. reports the use of deep learning-based YOLO models to detect lung nodules in computed tomography images. The accuracy and confidence scores of different versions of YOLO were compared: YOLOv7, YOLOv8, YOLOv9, and YOLOv11. These analyses revealed that the most successful results were achieved with YOLOv11 and YOLOv8 with a 93.33% accuracy rate, which had classified 14 out of 15 test images correctly. However, the YOLOv9 model reached an accuracy of 80%, while YOLOv7 showed the least success rate with 66.66% [2]. Mammeri et al. proposed a two-stage approach for the detection and classification of lung nodules in their joint study. In the first stage, the nodules have been detected in computed tomography images using YOLOv7, and the best result has been obtained as 81.28% mAP. In the second stage, nodules are classified as benign, suspicious, and malignant by using a VGG16-based transfer learning method [4].

In their joint study, Goel and Mishra developed a new method for optimizing network weights in conjunction with the YOLOv3 architecture for lung cancer detection. After preprocessing the LIDC-IDRI and LUNA16 datasets, the developed model achieved 92.00% accuracy on LUNA16 and 90.01% accuracy on LIDC-IDRI. The model classified lung cancer

cases with 91.80% sensitivity, 97.40% recall, and 94.60% F1-score metrics [9]. Finally, Choudhary et al. developed a method using the YOLO-NAS architecture to improve accuracy in lung cancer diagnosis. By combining neural network search algorithms with YOLO-NAS, a model was created that diagnoses lung cancer from medical images. The system, evaluated on the dataset, achieved 96.80% mAP [13]. To enhance the readability of the literature review, Table 1 provides a summary of the studies examined.

***Table 1. Summary of the literature review.***

| Author | Dataset | mAP (%) | Sensitivity (%) | Recall (%) | Precision (%) |
|---|---|---|---|---|---|
| Wehbe et al. [12]. | Lung PET-CT-DX | 97.10 | - | 94.50 | 96.10 |
| Liu [5]. | LUNA16 | 96.10 | - | 93.30 | - |
| Sharma et al. [2]. | 15 CT images | 93.33 | - | - | - |
| Mammeri et al. [4]. | LIDC-IDRI | 81.28 | - | 74.72 | 82.68 |
| Goel and Mishra [9]. | LIDC-IDRI | 92.00 | 91.80 | 97.40 | - |
| Choudhary et al. [13]. | Kaggle | 96.80 | - | - | - |

Within the scope of the reviewed literature, it is generally observed that YOLO-based models achieve successful results in lung nodule detection. However, as seen above, most studies have used older architectures such as YOLOv3, YOLOv5, and YOLOv7. Furthermore, comparisons were not made fairly because they were conducted with different hyperparameters. For example, Sharma et al. examined current models. However, they performed their tests on only 15 images. This indicates that the results cannot be generalized. Furthermore, a comprehensive comparison of current architectures such as YOLOv10 and YOLOv11 under the same experimental conditions on a standard and extensive dataset such as LUNA16 is not yet available in the literature. This study fills this gap by comparing the four most recent YOLO architectures under the same training and testing conditions, providing an objective guide on which model is more suitable for clinical applications.

# 3 MATERIAL AND METHODS

## 3.1 Dataset

In our study, evaluations were performed on the Lung Nodule Analysis 2016 (LUNA16) dataset. LUNA16 is one of the most widely used open-access databases for evaluating deep learning-based algorithms for lung nodule detection and CT images. LUNA16 is derived from the LIDC-IDRI dataset. It contains 1,186 lung nodules from a total of 888 patients. Researchers

can use LUNA16 to compare the performance of different algorithms and develop new methods. The LUNA16 dataset provides a rich and reliable resource for studies in the areas of false positive reduction and nodule detection [14-16].

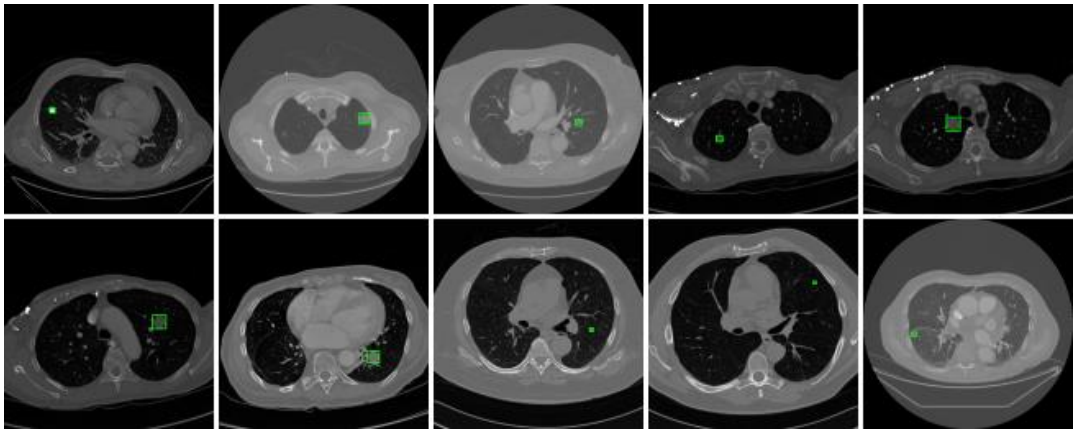

***Figure 1. Sample images from the LUNA16 dataset.***

2D axial sections were extracted from the LUNA16 dataset consisting of 3D CT images. Following this process, a data filtering step was applied to create the experimental dataset. The label files in the dataset were scanned. Sections where the label file was not empty were selected from these files. Sections without nodules were excluded from this dataset in order to focus the model training specifically on nodule localization and classification capabilities. Some examples of nodules labeled with boxes based on real coordinates in the images contained in the LUNA16 dataset are shown in Figure 1. Furthermore, Table 2 shows the number of images and labels in each section, along with the dataset being split into 70% train, 15% validation, and 15% test data.

***Table 2. Training, validation, and test partitions of the LUNA16 dataset.***

| Models | Images | Labels |
|---|---|---|
| Train (70%) | 830 | 830 |
| Validation (15%) | 178 | 178 |
| Test (15%) | 178 | 178 |
| Total (100%) | 1,186 | 1,186 |

## 3.2 Experimental Design and Hyperparameters

All experiments in this study were conducted on a computer running the Windows 11 operating system. The experiments used an Intel Core i7 14700K processor running at 2.10 GHz, 32 GB of DDR5 RAM, and an NVIDIA RTX 4060 graphics card. All evaluations were performed using the PyTorch framework, optimized with NVIDIA CUDA support. To ensure

that the models' performance was comparable and consistent, all models were evaluated using the same parameters during the training, validation, and testing phases.

Before training, pre-processing and data augmentation techniques were used to improve model robustness and avoid overfitting. All images were resized to 512 × 512 pixels according to the model's input dimensions. Pixel intensity values were normalized within the range [0, 1]. In addition, all images were augmented using data augmentation techniques such as mosaic data augmentation, random horizontal flipping, scaling, and color jittering. For consistency in the performance evaluation, these augmentation processes were disabled for the validation and testing phase.

We used the following hyperparameters for training models in a way that will ensure consistency and comparability between experiments. Each model was trained on images of size 512×512 pixels for 50 epochs. The batch size was 16, and data loading was done with four workers. We used a fixed learning rate of 0.001, with the Adam optimizer. Weight decay was set to 0.0005 to handle overfitting. We also used a warmup for five epochs at the beginning of training. In this way, the settings allowed us to train the models in a robust and reproducible manner.

## 3.3 YOLO Models

The YOLO family was first proposed by Redmon et al. in 2016. YOLO introduced an end-to-end approach to object detection, performing both classification and bounding box detection simultaneously [17]. YOLO models are important architectures in the field of deep learning for real-time object detection. Newer versions offer improved architectural structures and performance metrics, enhancing both the accuracy and speed of object detection. This section details the basic architectural features of the current models used in the study.

### 3.3.1 YOLOv8 Architecture

The C2f module forms the backbone of the YOLOv8 version, which was released in January 2023 and is an ELAN structure-inspired, faster implementation of the CSP architecture [18]. This model, whose overall architecture is presented in Figure 2, maintains detector learning with reduced computational loads. Moreover, the SPPF module further improves object detection performance on varying scales. In the neck of this model, the PAN-FPN architecture efficiently fuses multi-scale features. While the upper layers carry more information, the lower layers fortify the positional accuracy. The decoupled and anchor-free

header structure is one of the most significant innovations of YOLOv8, in which classification and detection are separated. Thus, objects can be detected via their center points without any predefined anchors, and their bounding box boundaries can be determined precisely [19].

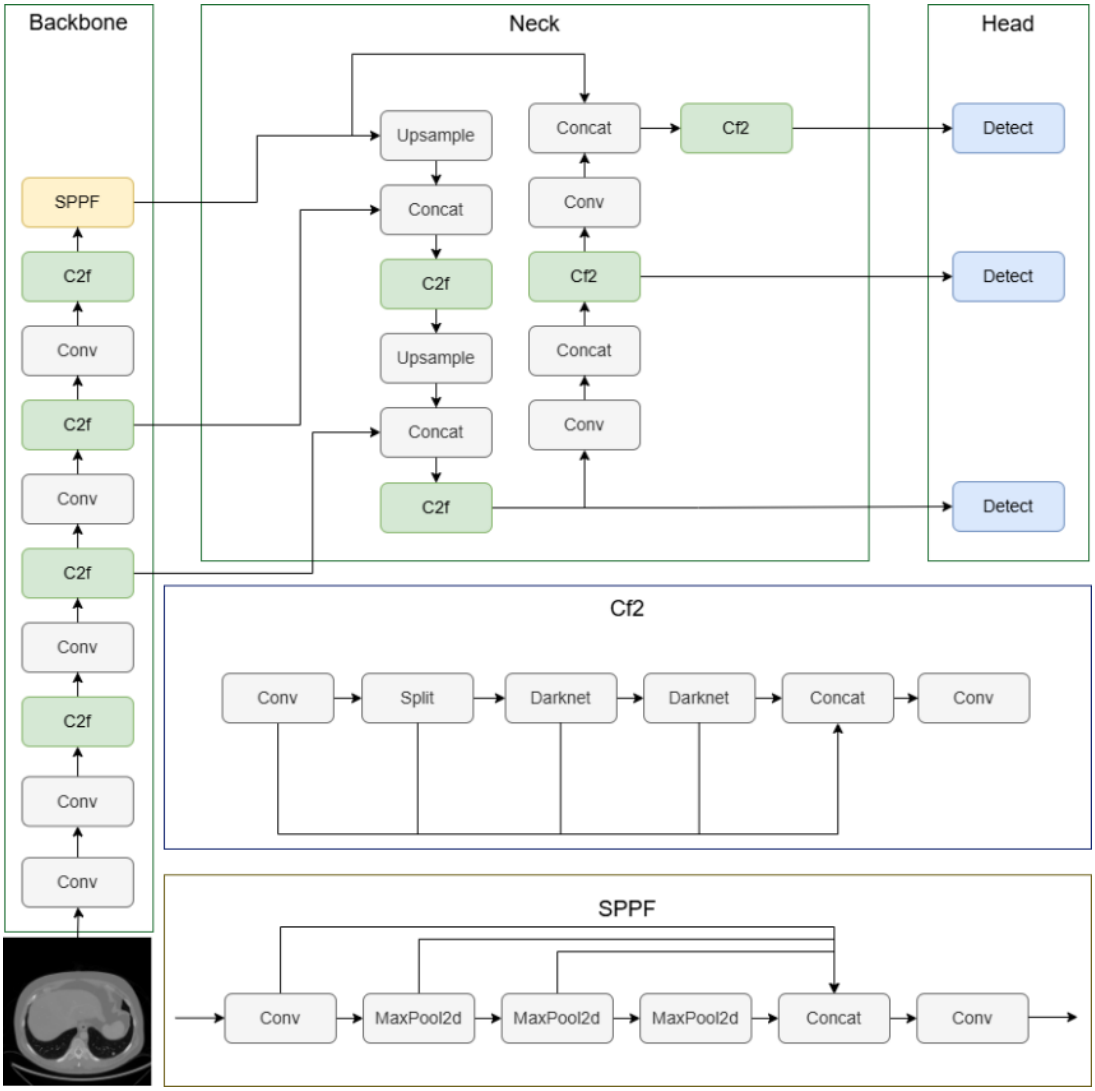


***Figure 2. YOLOv8 architecture [20].***

### 3.3.2 YOLOv9 Architecture

YOLOv9 represents a huge milestone in terms of both accuracy and efficiency compared to the previous YOLO series. Figure 3 portrays the basic architecture of this model, which is based on two most basic innovating ideas: PGI and GELAN. PGI helps in reducing information loss and error accumulation in the model training process. Then, more reliable gradients are generated and transmitted. This again results in more accurate updates of the model along with an increase in the overall detection performance. GELAN is an optimized structure in parameter usage for the traditional convolution. It used to control the performance of PGI. The best results obtained using GELAN architecture combined with the classic

hierarchical network structure of YOLO is known as RNELAN. Another new module used in YOLOv9 is named SPPELAN and is actually a combination of SPP with ELAN. SPPELAN and SPPF both used in YOLOv8 do have similar intentions about multi-scale contextual data gathering. YOLOv9 optimized the design of SPPELAN to deliver equivalent or even better performance with fewer parameters. Finally, the CBLinear and CBFuse layers make the model efficient. CBLinear accelerates feature extraction by directly applying linear transformations on the feature maps. The CBFuse generates a holistic representation by resizing and merging feature maps originating from different layers [21, 22].

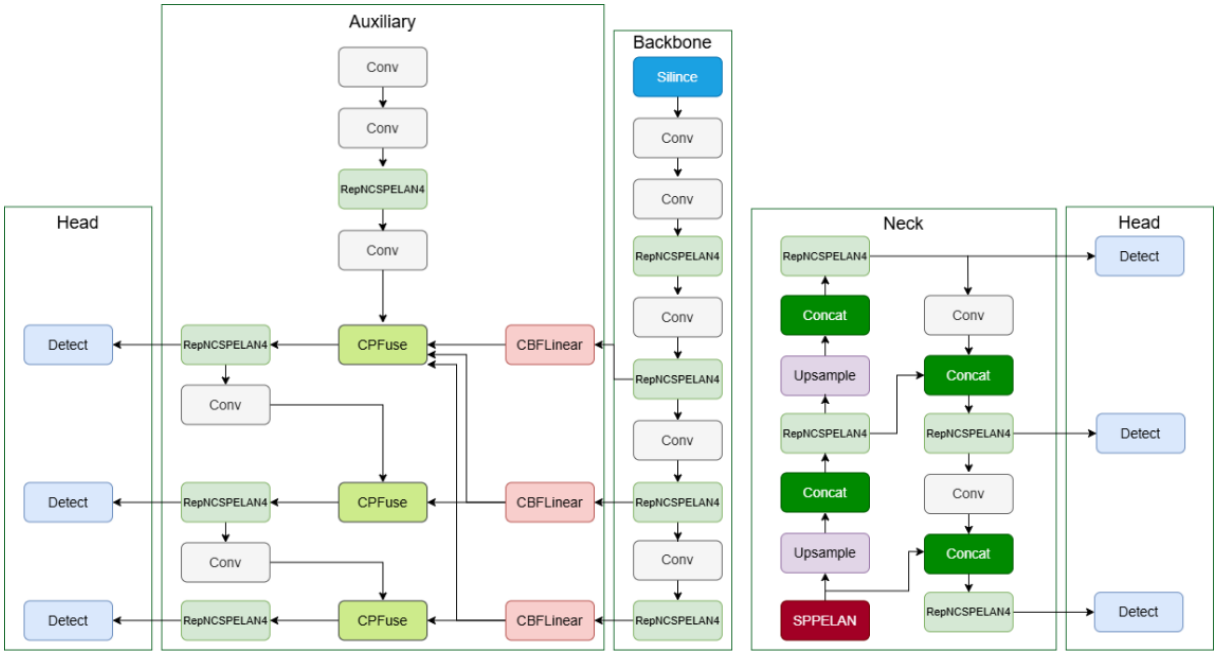


***Figure 3. YOLOv9 architecture [23].***

### 3.3.3 YOLOv10 Architecture

The basic architectural structure of YOLOv10, which was proposed by Tsinghua University in 2024, is shown in Figure 4. This network provides significant advances in real-time object detection. It adopts the NMS-free training paradigm, which removes the conventional Non-Maximum Suppression step to provide lower latency and faster results during inference. The model makes multiple predictions in training using the "Consistent Dual Assignments" method and selects the box with the highest IOU score. It has a strong, efficient process for multi-scale features because of the novel structures in the model, like Parallel Split-Attention modules and Compact Inverted Bottleneck blocks. A novel "one-to-one head" structure provides information richness with multiple boxes in training, while being more efficient since only this head is used during the time of inference. Internal ranking metrics and

designs of CIB are used to reduce box overlaps at different stages. Information flow is optimized by weight shortcuts and scaled residual connections [24, 25].

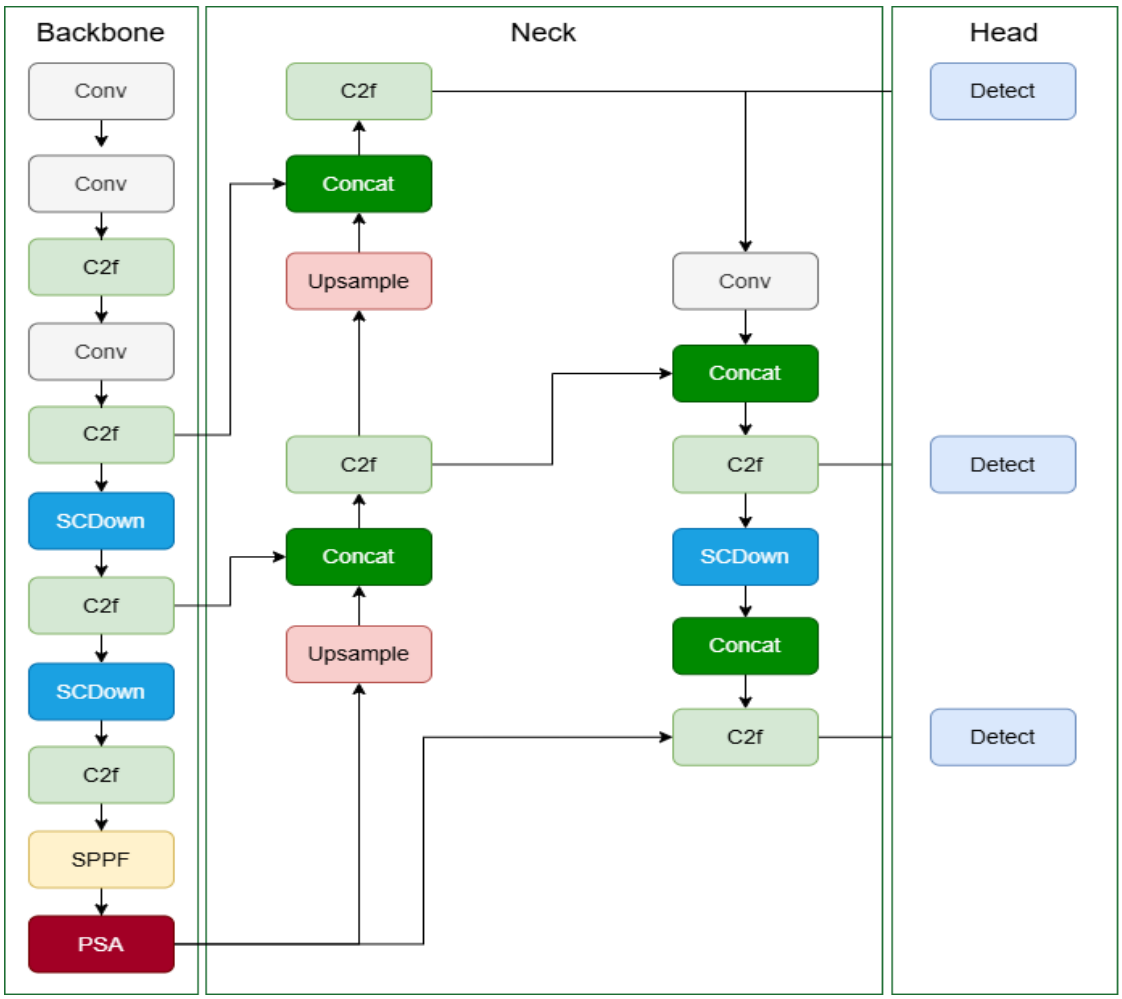


***Figure 4. YOLOv10 architecture [26].***

### 3.3.4 YOLOv11 Architecture

YOLOv11 signifies significant improvement over the YOLO family in terms of speed, accuracy, and feature extraction. The model architecture, as illustrated in Figure 5, is primarily made up of three parts: the backbone, neck, and head. In the backbone layer, several Conv blocks are adopted to extract multiscale features from the input image, where each block is composed of a Conv2D, BatchNorm2D, and SiLU activation function. Instead of C2f blocks utilized in YOLOv8, more efficient CSP implementations known as C3K2 blocks have been preferred. Besides, the SPPF block in the backbone allows for feature extraction on different scales with several maxpooling layers. The C2PSA block enhances accuracy by the spatial attention mechanism. Multiple Conv layers, C3K2 blocks, concat, and upsample operations in the neck are used to combine features at different scales and pass them to the head layer,

leveraging the advantages provided by the mechanism of C2PSA. Lastly, the head section estimates object classes, object scores, and bounding box coordinates with high accuracy, hence improving its general detection performance [27, 28].

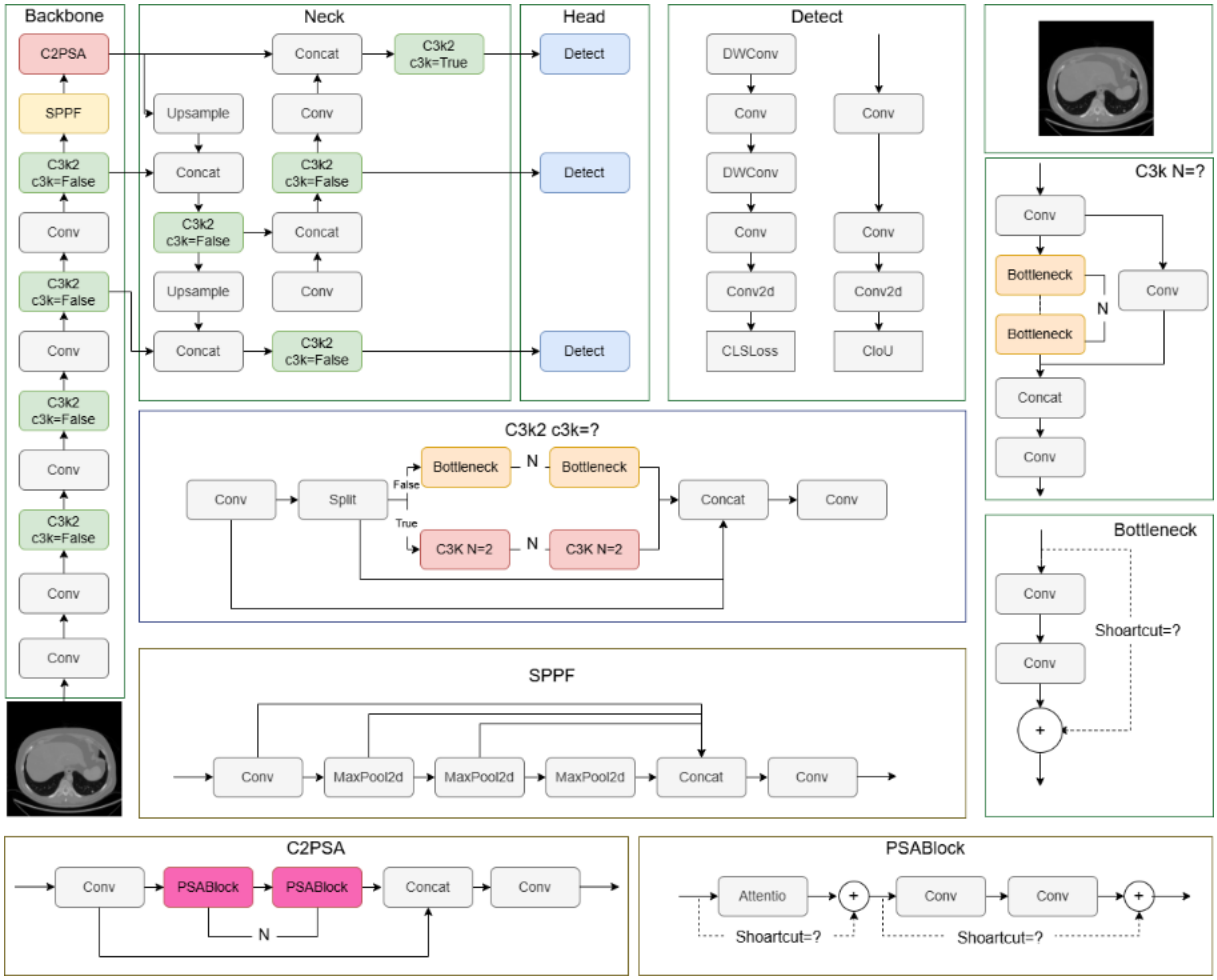


***Figure 5. YOLOv11 architecture [29].***

## 3.4 Performance Metrics

In our study, six different metrics were used to evaluate all models. These are precision, recall, F1-score, mAP50, mAP50-90, and speed. The metrics are evaluated based on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values. The evaluation of performance metrics was performed at the slice level using 2D axial images. The definitions of confusion matrix components depend on the Intersection over Union (IoU) threshold value. For the mAP50 metric in the evaluation, if the IoU between the predicted bounding box and the ground truth box is $\geq 0.5$, this detection is classified as a True Positive (TP). A False Positive (FP) is defined as a prediction with IoU $< 0.5$ or a detection in an area that is not actually a nodule. A False Negative (FN) occurs when a true nodule cannot be detected with an IoU of at least 0.5.

Precision tells us about the exactitude of the positives predicted by the model. TP is computed as the ratio of positive predictions against the overall number of positive predictions that a model makes. A higher precision reflects a lower number of FPs. This metric is represented with the equation below:

$$Precision = \frac{TP}{(TP+FP)} \tag{1}$$

Recall measures the capability of a model to detect all the positive examples correctly. It can be computed as the ratio of true positive predictions versus the total number of true positive examples in the dataset. A high recall value indicates fewer FNs. The recall metric can be computed by using the following equation:

$$Recall = \frac{TP}{(TP+FN)} \tag{2}$$

The F1-score is a harmonization metric that balances precision and recall values. The F1-score is expressed by the following equation:

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{(Precision+Recall)} \tag{3}$$

mAP is one of the most comprehensive metrics utilized in assessing the general performance of object detection models across all different categories. Average precision for each category is calculated and the average taken in order to get the overall performance value. The equation for the mAP value is given as follows [30]:

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{4}$$

where C is the total number of classes present in the dataset. A higher value of mAP represents better model performance [31]. In the case of mAP50, a detection is said to be true positive if IoU is greater than 0.5. The mAP50-95 considers thresholds of 0.5 to 0.95 with an increment of 0.05. Lastly, the speed metric is another performance metric that indicates the time in milliseconds taken by the model to process an image on test data. This metric will be considered where speed is needed.

## 4 EXPERIMENTAL RESULTS

The results obtained from the evaluations performed on the LUNA16 dataset are presented in Table 3. When Table 3 is evaluated, the YOLOv8 model stands out with the highest precision rate of 81.24% among the algorithms tested in this study. This high value means that the model provides reliable results for the vast majority of the nodules it detects by keeping the FP rate low. However, the recall value remaining at 65.83% reveals that the model is inclined

to miss some real nodules. The YOLOv8 model has shown a balanced performance with an F1-score of 72.73%, and generally, it is 76.30% accurate in mAP50 and 38.62% in mAP50-95. These metrics delineate the success of the model at different thresholds. For speed, the YOLOv8 model has the fastest running time at 9.61 ms/img, which provides its advantage in practical applications. The YOLOv8 exhibited high accuracy with fast performance to gain an important advantage in minimizing the risk of false alarms during clinical applications.

***Table 3. Performance metric results of YOLO models evaluated with the LUNA16 dataset.***

| Models | Precision (%) | Recall (%) | F1-score (%) | mAP50 (%) | mAP50-95 (%) | Speed (ms/img) |
|---|---|---|---|---|---|---|
| YOLOv8 | 81.24 | 65.83 | 72.73 | 76.30 | 38.62 | 9.61 |
| YOLOv9 | 78.92 | 71.66 | 75.12 | 76.35 | 39.60 | 9.84 |
| YOLOv10 | 80.05 | 73.60 | 76.69 | 77.04 | 40.07 | 10.29 |
| YOLOv11 | 80.78 | 70.00 | 75.00 | 77.11 | 40.93 | 10.01 |

The YOLOv9 model shows a significant increase in recall success compared to the previous version and stands out with a recall rate of 71.66%. However, its precision value of 78.92% lags slightly behind YOLOV8. This indicates an increase in the margin of error in some of the nodules it detects. The F1-score is 75.12%, which is higher than YOLOV8, indicating that the model has improved its ability to both accurately detect and find real nodules. The mAP50 metric has increased to 76.35% and the mAP50-95 value to 39.60%. These results show that the model provides more consistent detections at different thresholds. The speed performance is reasonable at 9.84 ms/img. In visual analysis, YOLOV9 is seen to be more stable in detecting nodules in some difficult cases and capable of detecting in a wider area. In terms of overall accuracy and sensitivity, YOLOv9 stands out with its balanced structure.

The YOLOv10 model, which was evaluated in third place, showed the highest recall rate among the tested algorithms; it was also the most successful model to find real nodules, 73.60%. Precision was similar to that of YOLOv8 and YOLOv11, with 80.05%. Its best balance was achieved with an F1-score of 76.69%. It provides optimal results regarding the correct detection rate and possibility to find real nodules. It gave a slight advantage in overall accuracy with mAP50 values of 77.04% and mAP50-95 values of 40.07%. In terms of speed performance, this model is slightly slower compared to other models, operating at 10.30 ms/img. However, this difference is acceptable considering such model sensitivity. In visual detections, it is outstanding due to its clear marking of the presence of nodules in suspicious areas and achieving quite high confidence scores.

Finally, among the models evaluated, the YOLOv11 model is distinguished as the most up-to-date algorithm in particular for general accuracy metrics. While it performs at a similar level compared to the previous models by giving a precision of 80.78% and a recall of 70.00%, it achieves the highest value with an mAP50 of 77.11% and an mAP50-95 ratio of 40.93%. Thus, in this way, the model makes the most consistent and reliable detections at different thresholds. The F1-score is 75.00%, which is a pretty balanced result in comparison with previous models. According to the visual analyses, it was observed that YOLOv11 offers very confident results in detecting nodules, especially in so-called low-contrast and problematic cases. Considering speed, it has a performance of 10.01 ms/img, which is basically sufficient in practical applications. Given the fine-grained detection capability and multi-metric stability of YOLOv11, it indeed represents a model that one can consider applying in real-world scenarios.

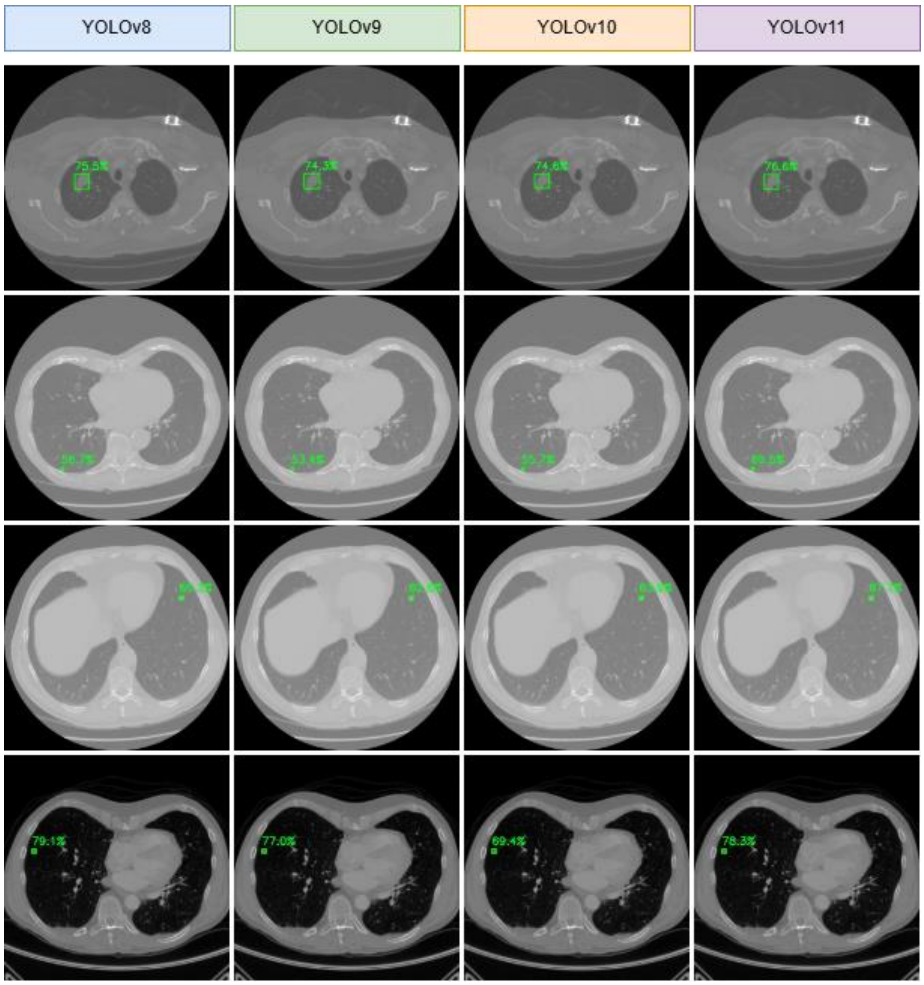


***Figure 6. Sample detection images generated using YOLO models on the LUNA16 test dataset.***

Figure 6 gives a visual overview of the detection success of models over diverse types of nodules. In these examples, all models were able to localize the nodules. However, confidence scores showed significant differences, especially for challenging images. For example, in Figure 6, the second row shows a small nodule close to the wall of the lung. In this image, which is a challenging case because of low contrast and similarity to the background, YOLOv11 outperformed YOLOv9 by 53.4% and YOLOv8 by 58.7%, with a high confidence score of 69.5%. This visual observation justifies the quantitative evidence in Table 3, where YOLOv11 achieved the highest mAP50-95. Conversely, in the fourth row, which is a high-contrast example, all models performed very well with high confidence at about 77% to 79%. Additionally, no false positives, meaning the background being misclassified as a nodule, were observed in these examples, which further confirms the high precision rates that have been detected for the models.

## 5 DISCUSSION AND FUTURE DIRECTIONS

A relationship between architectural complexity and detection performance is demonstrated by the findings of this study. The YOLOv8 architecture yielded the highest processing speed with a value of 9.61 ms/img. However, it reflects a weakness in the detection of nodules compared to newer versions, where sensitivity is well below at a value of 65.83%. This result again insinuates that while the C2f backbone promotes efficiency, it may struggle with feature preservation in challenging low-contrast examples. This weakness has been overcome by architectural innovation in the subsequent models. Adopting PSA modules in the YOLOv10 architecture seems to have resolved the previously unaddressed issue of a missed detection rate. The highest sensitivity performance was thus realized at a value of 73.60%. This was achieved at the expense of speed. Finally, the highest mAP50-95 value was realized by the YOLOv11 model due to the use of C3K2 blocks and the addition of C2PSA, which promote an enhanced ability to distinguish nodules from backgrounds when compared to the YOLOv8 and YOLOv9 architectures. These successes are not devoid of weaknesses in the architectures. A general weakness among all the tested 2D models is the loss of volumetric context. Given that all predictions have been made on individual slices, there is no way the models can leverage the nodule 3D continuity. This gives rise to lower confidence on slices with very small nodule appearance. Second, in as much as precision is high within the YOLOv8 model, false negative values stand in the way of utilizing it as a primary scanning tool without secondary verification.

Training deep learning models on large datasets such as LUNA16 is computationally expensive. All performance metrics in this study are therefore based on a single training run using a fixed random seed and standard training-validation-test split, so that results reported here are reproducible by other researchers. It does not account, however, for statistical variance due to random data shuffling. Further studies focused on clinical validation should be done with k-fold cross-validation to achieve mean and standard deviation values for a robust statistical evaluation.

In addition, research should be directed to rendering model architectures more efficient and generalizable. More comprehensive diagnostic systems can be developed by integrating data from several imaging modalities and clinical data of patients. To improve the variety and quality of the data, large, balanced datasets of different institutions and populations should be used. Engineering approaches to simplify models and optimize hardware for real-time applications will further open the way for clinical integration. Finally, the integration of developed algorithms into clinical decision support systems should be tested in collaboration with physicians, user-friendly interfaces should be developed, and issues like ethics and security should be addressed. In this regard, a new generation of computer-assisted diagnostic systems using deep learning-based methods for the early diagnosis of lung cancer can be developed.

## 6 CONCLUSION

This study demonstrates the potential of deep learning-based object detection algorithms in the detection of lung nodules in clinical applications, supported by numerical data. Comparisons carried out on the LUNA16 dataset have shown that the YOLOv8 model has the fastest and most accurate results, with 81.24% precision and a speed of 9.61 ms/img. The YOLOv10 model demonstrated the highest recall rate of detecting true nodules, with a recall rate of 73.60%. For accuracy in multiple metrics, YOLOv11 achieved the highest rates, at 77.11% for mAP50-95 and 40.93% for mAP50-95. Based on these findings, it can be concluded that automatic diagnosis systems have huge potential to bring about both increased accuracy and speed in lung cancer screening processes. The comprehensive and comparable performance data provided by the study are indicative for future studies. In the future, optimizing model architectures, integrating data from different imaging sources, and developing systems compatible with clinical infrastructures will further increase diagnostic accuracy. Thus, deep learning-based approaches can bring about a transformation in healthcare services for the early diagnosis of lung cancer.

## Conflict of Interest Statement

There is no conflict of interest between the authors.

## Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

## Artificial Intelligence (AI) Contribution Statement

This manuscript was entirely written, edited, analyzed, and prepared without the assistance of any artificial intelligence (AI) tools. All content, including text, data analysis, and figures, was solely generated by the authors.

## REFERENCES


[1] T. Ozcan, A. N. Toprak, I. Aruk, O. Sahin, and I. Ozcan, "Applications of deep learning techniques in healthcare systems: A review," *Journal of Clinical Practice & Research,* vol. 46, no. 5, 2024.

[2] D. K. Sharma, M. K. Pal, and A. K. Singh, "A comparative analysis of YOLO models for efficient lung tumor detection using CT images," *Health and Technology,* 2025/06/19 2025, doi: 10.1007/s12553-025-00989-1.

[3] R. L. Siegel, T. B. Kratzer, A. N. Giaquinto, H. Sung, and A. Jemal, "Cancer statistics, 2025," *Ca,* vol. 75, no. 1, p. 10, 2025, doi: https://doi.org/10.3322/caac.21871.

[4] S. Mammeri, M. Amroune, M.-Y. Haouam, I. Bendib, and A. Corrêa Silva, "Early detection and diagnosis of lung cancer using YOLO v7, and transfer learning," *Multimedia Tools and Applications,* vol. 83, no. 10, pp. 30965-30980, 2024/03/01 2024, doi: 10.1007/s11042-023-16864-y.

[5] K. Liu, "Stbi-yolo: A real-time object detection method for lung nodule recognition," *IEEE Access,* vol. 10, pp. 75385-75394, 2022.

[6] J. H. Lee *et al.*, "Deep learning to optimize candidate selection for lung cancer CT screening: advancing the 2021 USPSTF recommendations," *Radiology,* vol. 305, no. 1, pp. 209-218, 2022.

[7] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," *IEEE transactions on medical imaging,* vol. 25, no. 4, pp. 385-405, 2006.

[8] S. Makaju, P. Prasad, A. Alsadoon, A. Singh, and A. Elchouemi, "Lung cancer detection using CT scan images," *Procedia Computer Science,* vol. 125, pp. 107-114, 2018.

[9] L. Goel and S. Mishra, "A hybrid of modified YOLOv3 with BBO/EE optimizer for lung cancer detection," *Multimedia Tools and Applications,* vol. 83, no. 17, pp. 52219-52251, 2024.

[10] B. Ozdemir, E. Aslan, and I. Pacal, "Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection," *IEEE Access,* 2025.

[11] I. Aruk, I. Pacal, and A. N. Toprak, "A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification," *Expert Systems with Applications,* p. 127721, 2025.

[12] A. Wehbe, S. Dellepiane, and I. Minetti, "Enhanced Lung Cancer Detection and TNM Staging Using YOLOv8 and TNMClassifier: An Integrated Deep Learning Approach for CT Imaging," *IEEE Access,* vol. 12, pp. 141414-141424, 2024, doi: 10.1109/ACCESS.2024.3462629.

[13] C. Choudhary, Anurag, J. Thakur, Urvashi, H. Bhardwaj, and V. Kaushik, "Detection of Lung Cancer with YOLO-NAS: A Novel Approach for Enhanced Diagnostic Accuracy," in *Intelligent Communication, Control and Devices*, Singapore, A. Kumar, R. K. Pachauri, R. Mishra, and P. Kuchhal, Eds., 2025// 2025: Springer Nature Singapore, pp. 37-49.

[14] C.-Y. Lin *et al.*, "Development of a modified 3D region proposal network for lung nodule detection in computed tomography scans: a secondary analysis of lung nodule datasets," *Cancer Imaging,* vol. 24, no. 1, p. 40, 2024/03/20 2024, doi: 10.1186/s40644-024-00683-x.

[15] J. Wang *et al.*, "Preparing CT imaging datasets for deep learning in lung nodule analysis: Insights from four well-known datasets," *Heliyon,* vol. 9, no. 6, 2023, doi: 10.1016/j.heliyon.2023.e17104.

[16] A. A. A. Setio *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Medical Image Analysis,* vol. 42, pp. 1-13, 2017/12/01/ 2017, doi: https://doi.org/10.1016/j.media.2017.06.015.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[18] G. Jocher *et al.*, "ultralytics/yolov5: v6. 1-tensorrt, tensorflow edge tpu and openvino export and inference," *Zenodo,* 2022.

[19] F. Solimani *et al.*, "Optimizing tomato plant phenotyping detection: Boosting YOLOv8 architecture to tackle data complexity," *Computers and Electronics in Agriculture,* vol. 218, p. 108728, 2024/03/01/ 2024, doi: https://doi.org/10.1016/j.compag.2024.108728.

[20] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, "A Review on YOLOv8 and Its Advancements," in *Data Intelligence and Cognitive Informatics*, Singapore, I. J. Jacob, S. Piramuthu, and P. Falkowski-Gilski, Eds., 2024// 2024: Springer Nature Singapore, pp. 529-545.

[21] J. Anandakrishnan, A. K. Sangaiah, H. Darmawan, N. K. Son, Y. B. Lin, and M. J. F. Alenazi, "Precise Spatial Prediction of Rice Seedlings From Large-Scale Airborne Remote Sensing Data Using Optimized Li-YOLOv9," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 18, pp. 2226-2238, 2025, doi: 10.1109/JSTARS.2024.3505964.

[22] R. Sapkota, Z. Meng, M. Churuvija, X. Du, Z. Ma, and M. Karkee, "Comprehensive performance evaluation of yolov12, yolov11, yolov10, yolov9 and yolov8 on detecting and counting fruitlet in complex orchard environments," *arXiv preprint arXiv:2407.12040,* 2024.

[23] Y. Huang *et al.*, "YOLO-IAPs: A Rapid Detection Method for Invasive Alien Plants in the Wild Based on Improved YOLOv9," *Agriculture*, vol. 14, no. 12, doi: 10.3390/agriculture14122201.

[24] M. Hussain and R. Khanam, "In-Depth Review of YOLOv1 to YOLOv10 Variants for Enhanced Photovoltaic Defect Detection," *Solar*, vol. 4, no. 3, pp. 351-386doi: 10.3390/solar4030016.

[25] T.-N. Nguyen, T. D. Tran, and P. V. Cuong, "Segmentation of Concrete Surface Cracks Using DeeplabV3+ Architecture," in *Proceedings of the Third International Conference on Sustainable Civil Engineering and Architecture*, Singapore, J. N. Reddy, C. M. Wang, V. H. Luong, and A. T. Le, Eds., 2024// 2024: Springer Nature Singapore, pp. 1531-1538, doi: https://doi.org/10.1007/978-981-99-7434-4.

[26] H. Cheng and F. Kang, "Corrosion Detection and Grading Method for Hydraulic Metal Structures Based on an Improved YOLOv10 Sequential Architecture," *Applied Sciences*, vol. 14, no. 24, doi: 10.3390/app142412009.

[27] A. F. Rasheed and M. Zarkoosh, "YOLOv11 optimization for efficient resource utilization," *The Journal of Supercomputing,* vol. 81, no. 9, p. 1085, 2025/06/27 2025, doi: 10.1007/s11227-025-07520-3.

[28] L.-h. He, Y.-z. Zhou, L. Liu, W. Cao, and J.-h. Ma, "Research on object detection and recognition in remote sensing images based on YOLOv11," *Scientific Reports,* vol. 15, no. 1, p. 14032, 2025/04/23 2025, doi: 10.1038/s41598-025-96314-x.

[29] L.-h. He, Y.-z. Zhou, L. Liu, Y.-q. Zhang, and J.-h. Ma, "Research on the directional bounding box algorithm of YOLOv11 in tailings pond identification," *Measurement,* p. 117674, 2025.

[30] T. Q. Nguyen, H. L. Tran, T. K. Tran, H. P. Phan-Nguyen, and T. H. Nguyen, "FA-YOLOv9: Improved YOLOv9 Based on Feature Attention Block," in *2024 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, 15-16 Aug. 2024 2024, pp. 1-6, doi: 10.1109/MAPR63514.2024.10661057.

[31] A. A. Alsabei, T. M. Alsubait, and H. H. Alhakami, "Enhancing Crowd Safety at Hajj: Real-Time Detection of Abnormal Behavior Using YOLOv9," *IEEE Access,* vol. 13, pp. 37748-37761, 2025, doi: 10.1109/ACCESS.2025.3545256.