International Journal of Scientific Engineering and Science
Volume 10, Issue 8, pp. 22-27, August 2026. ISSN (Online): 2456-7361

A Study on Deep Learning-Based Algorithms for Intelligent Detection and Growth Prediction of Lung Nodules

Zhao Zeda[1], Wan Nan[2]
[1]School of Medical Informatics, Wannan Medical College, Wuhu, Anhui, China,241002
[2]School of Clinical Medicine, Wannan Medical College, Wuhu, Anhui, China,241002

Abstract—*In response to the high incidence of lung nodules and the need for early-stage lung cancer screening, this project has developed an AI-assisted diagnostic system based on deep learning. The system employs an enhanced YOLOv8 model and incorporates a multi-head self-attention mechanism (MHSA) to improve detection of small nodules, achieving a precision of 87.6% and an mAP50 of 82.7% on the LUNA16 dataset and real-world hospital CT data—representing a significant improvement over the original model. Additionally, the system innovatively integrates the PredNet growth prediction network, which uses historical nodule images to predict volume change trends and risk levels, thereby achieving an integrated "detection–prediction–assessment" workflow. The system employs PyQt5 to develop its graphical user interface, supporting CT image upload, model switching, growth prediction, and risk analysis. It is easy to operate and can directly assist in clinical decision-making. Closely aligned with practical clinical needs, this solution aims to alleviate the burden on physicians during image review, reduce unnecessary radiation exposure, and optimize follow-up protocols. It holds practical significance in improving diagnostic efficiency, lowering healthcare costs, and advancing the implementation of smart healthcare, with broad market application prospects.*

Keywords— *Lung nodule, Deep Learning, YOLOv8, Multi-head self-attention (MHSA), PredNet*

I. Introduction

Given the high incidence of lung cancer and the persistently high rate of late-stage diagnosis, early screening for pulmonary nodules is key to improving prognosis. However, with the surge in the volume of CT scans and the low efficiency and high subjectivity of manual interpretation, there is an urgent clinical need for intelligent diagnostic tools. This project is based on the YOLOv8 object detection framework and incorporates a multi-head self-attention mechanism (MHSA) to enhance the ability to identify micro-nodules. It also integrates the PredNet growth prediction network to achieve automatic detection of lung nodules, prediction of growth trends, and risk stratification. The system was validated using the LUNA16 dataset and real-world hospital CT data, achieving a detection accuracy of 87.6%, with prediction results demonstrating clinical utility. The interactive interface, built using PyQt5, is designed to assist physicians in efficient diagnosis and treatment, reduce unnecessary radiation exposure, and promote personalized follow-up management, demonstrating clear prospects for clinical application.

II. Research Background

A. Introduction to Lung Nodules

A lung nodule is a focal, round or oval opacity on a lung imaging study with a diameter of ≤30 mm. It can be classified as a ground-glass nodule, a partially solid nodule, or a solid nodule; a solid component often indicates infiltrative growth. With the widespread adoption of low-dose CT screening, the detection rate of pulmonary nodules has risen significantly. A study involving 3,458 individuals undergoing health checkups showed a total detection rate of 41.01%, reaching as high as 61.47% in the 61－70 age group. Pulmonary nodules are a major indicator of early-stage lung cancer; the 5-year survival rate for early-stage diagnoses (such as Stage IA1) can reach 92%, whereas it drops significantly in advanced stages. However, current clinical practice primarily relies on manual interpretation of CT images, which is not only labor-intensive and inefficient but also prone to subjective interpretation, leading to missed diagnoses or misclassifications. At the same time, there is a lack of precise quantitative tools for nodule follow-up management, often requiring patients to undergo repeated follow-up examinations and exposing them to unnecessary radiation exposure and financial burdens. Therefore, the development of an intelligent, automated lung nodule detection and prediction system is of great significance for improving diagnostic and treatment efficiency and patient prognosis.

B. Background of the Problem

The *Journal of Anhui Medical College*, Vol. 24, No. 1 (Upper), 2025[1], published an analysis of lung nodule detection among 3,458 individuals who underwent physical examinations at a certain institution. The age distribution of the examinees is shown in Table 1:

TABLE I. Basic Information Form for Medical Examination Participants.

Age	Total	Male	Women
≤40 years old	1398	973（69.60）	435（30.40）
41 to 50 years old	605	381（62.98）	224（37.02）
51 to 60 years old	668	351（52.54）	317（47.46）
61 to 70 years old	364	165（45.33）	199（54.67）
≥71 years old	423	231（54.61）	192（45.39）
Total	3458	2101（60.76）	1357（39.24）

The results of the physical examination are shown in the table below:

TABLE II. Physical Examination Results Table.

Age	Group	Number of People	Has lung nodules	Detection Rate (%)
≤40 years old	1	1398	427	30.54
41 to 50 years old	2	605	233	38.51
51 to 60 years old	3	668	305	45.66
61 to 70 years old	4	364	193	53.02
≥71 years old	5	423	260	61.47
Total		3458	1418	41.01

Incidence of Lung Nodules: Among 3,458 participants in a health screening program, 1,418 were found to have lung nodules, resulting in an overall detection rate of 41.01%. Overall Detection Rates of Lung Nodules by Age Group: The detection rate of lung nodules gradually increases with age. The detection rate for the ≤40 age group was 30.54%, while the highest rate was observed in the 61–70 age group at 61.47%.

The high prevalence of lung nodules is alarming. You may already have lung nodules without even realizing it! This clearly illustrates the necessity of creating a method for the adjunctive treatment and detection of lung nodules!

C. Introduction to Deep Learning

Especially in areas like image identification and object recognition, deep learning, and convolutional neural networks (CNNs) in particular, have made enormous progress. In 2014, R-CNN first applied CNNs to object detection, and subsequent models such as Fast R-CNN and Faster R-CNN gradually optimized both speed and accuracy[2]. The YOLO series, proposed in 2016, treats detection as a regression problem and enables end-to-end real-time detection. Among these, YOLOv8 achieves a good balance between lightweight design, detection speed, and accuracy, making it particularly suitable for deployment on medical embedded devices. However, lung nodules are tiny and have blurred boundaries, making them easily confused with normal structures such as blood vessels; conventional feature extraction networks struggle to fully capture key information. To address this, the Multi-Head Self-Attention (MHSA) mechanism was introduced[3]. By simultaneously focusing on features in different subspaces, it enhances the model’s ability to capture long-range dependencies and fine-grained details, thereby effectively improving detection performance. Additionally, prediction networks based on encoder-decoder architectures—such as PredNet, which draws inspiration from the U-Net design—can infer future trends of targets based on historical images, providing a viable technical approach for predicting the growth of pulmonary nodules.

D. Current State of Research

Internationally, the release of public datasets such as LUNA16 has driven the rapid development of lung nodule detection algorithms. Y. Su et al. improved Faster R-CNN, achieving a detection accuracy of 91.2%; Cai et al. optimized Mask R-CNN using ResNet50 as the backbone, achieving an mAP of 88.2% on LUNA16[9-11]. However, these models have a large number of parameters and slow detection speeds, making them difficult to deploy on portable devices. Domestic researchers are also actively pursuing this field, but most work remains focused on static detection, with few systematic studies integrating growth prediction. In recent years, the European Respiratory Society proposed the concept of “growth entropy,” which uses AI to calculate nodule volume growth rates and morphological complexity to predict malignancy up to six months in advance with a sensitivity of 89%. This marks a shift in growth prediction research from “whether it grows” to “how it grows”. However, there is currently no integrated system that combines an enhanced YOLOv8 with PredNet to simultaneously achieve detection and growth prediction for clinical implementation. Building on the aforementioned research, this project aims to introduce MHSA to optimize YOLOv8 and combine it with PredNet to construct a comprehensive “detection – prediction – evaluation” solution, thereby filling a gap in this field[12-14].

III. Research Process

A. Dataset Selection

This project uses the publicly available LUNA16 (Lung Nodule Analysis 2016) dataset, which is based on a preprocessed subset of the LIDC/IDRI (National Cancer Institute’s Lung CT Database). It contains 888 lung CT images and annotation files, comprising a total of 1,186 positive nodules, including 272 nodules measuring 3 – 5 mm, 633 nodules measuring 5 – 10 mm, and 281 nodules larger than 10 mm.

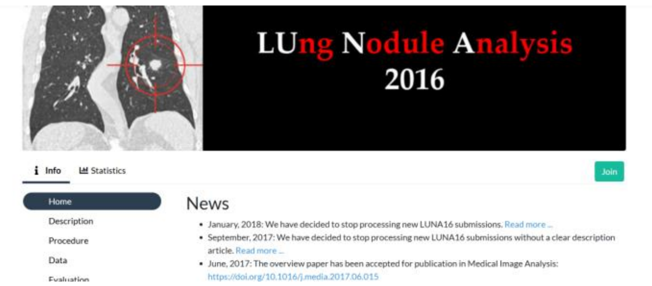

Fig. 1. luna16 Official Website Home Page.

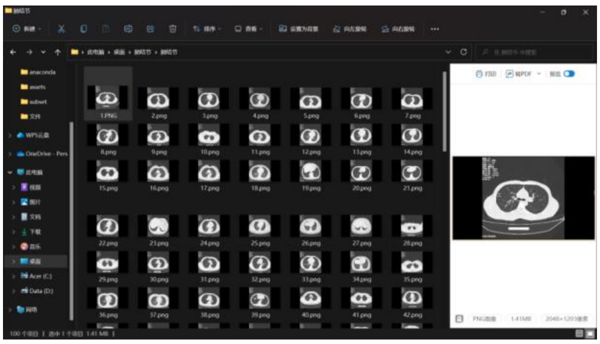

Fig. 2. Real Lung CT Images.

In addition, we have obtained real-world lung CT image

International Journal of Scientific Engineering and Science
Volume 10, Issue 8, pp. 22-27, August 2026. ISSN (Online): 2456-7361

data from local top-tier hospitals. This will significantly enhance the model's robustness, specificity, and other performance metrics, making the model more reliable. As a result, the trained model can be rapidly applied in clinical practice — advantages that public datasets simply cannot match.

B. Data Set Preprocessing

Since the LUNA16 dataset is in MHD and RAW formats and cannot be used directly for model training, the dataset must be preprocessed. The flowchart is shown below:

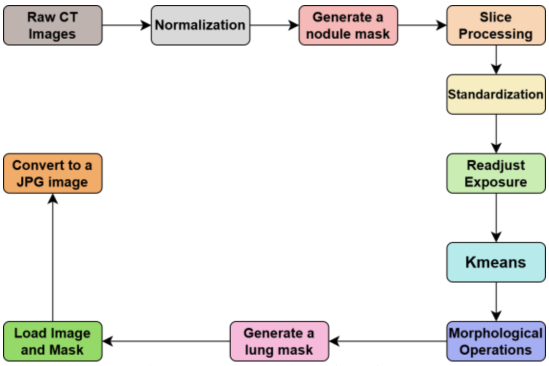

Fig. 3. Preprocessing Flowchart.

Based on the LUNA16 dataset format and actual training requirements, this project has designed a comprehensive preprocessing workflow. Since the raw data is in MHD and RAW formats, which cannot be directly input into the YOLOv8 model, the CT values must first be windowed and thresholded (-1000 HU to 400 HU) to exclude irrelevant tissues such as bone and skin. The pixel values are then linearly mapped to the [0,1] interval via minimum-maximum normalization to eliminate dimensional differences between different devices. Next, based on the nodule center's world coordinates, pixel spacing, and origin information provided in the annotation file, these coordinates are converted to voxel coordinates. The voxel coverage area is calculated using the nodule diameter to generate a binary mask of the same size as the original image, where the nodule location is marked as 1 and the rest as 0. After annotation, the 3D volume data is sliced layer by layer along the Z-axis to extract 2D cross-sectional images, preparing them for subsequent 2D model training. Subsequently, the sliced images were Z-score normalized to ensure they followed a distribution with a mean of zero and a variance of one, thereby enhancing training stability. Additionally, for some overexposed images, brightness was adaptively adjusted by calculating the average pixel value of adjacent lung regions to enhance soft tissue contrast. To further focus on the lung parenchyma, the K-means algorithm is used to classify the image center into foreground (tissue) and background (air). A global threshold is set based on the average values of the two cluster centers for binarization. Morphological erosion and dilation operations are then applied to remove isolated noise and fill internal voids, ultimately generating a complete lung mask used to crop out background interference. Concurrently, the original nodule annotations are converted into normalized bounding box coordinates in YOLO format (center x, y, and width/height) to ensure a one-to-one correspondence between the labels and the preprocessed images. Finally, the entire dataset is randomly split into training, validation, and test sets in a 7:2:1 ratio for model training, hyperparameter tuning, and generalization performance evaluation. This entire process achieves the conversion from raw medical images to standardized input samples, providing a high-quality data foundation for detection and prediction models.

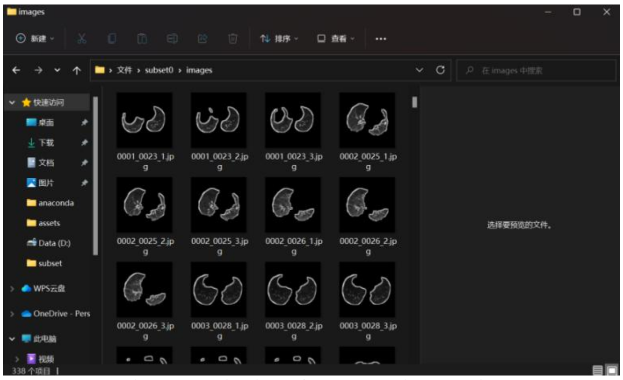

Fig. 4. A Selection of Preprocessing Results.

Growth prediction models also use these datasets to train their ability to segment and extract nodules.

C. Construction of a Model for Detecting Lung Nodules

This project aims to incorporate the MHSA attention mechanism into the original YOLOv8 detection model to help it focus more on key information in images and improve the model's accuracy. It also includes a visualization tool that allows users to draw annotation boxes simply by uploading an image.

Multi-Head Self-Attention (MHSA): A deep learning model based on the self-attention mechanism, first proposed by Vaswani et al. in 2017. It can automatically capture long-range dependencies between different positions within an input sequence, thereby enabling the model to better understand contextual information and improve its performance. The core idea of MHSA is to divide a linear transformation into multiple heads, with each head computing an attention matrix, and concatenating the outputs of all heads to form the final representation. In feature extraction networks, a stacked-layer approach is generally used to enhance feature extraction capabilities. However, lung nodules have complex spatial distributions and blurred edges, making it difficult to accurately define their boundaries; they can be located anywhere in the lung and may be obscured by pulmonary vessels or pulmonary texture. These characteristics of pulmonary nodules result in poor feature extraction capabilities in general feature extraction networks. This can lead to missed or false-positive detections of pulmonary nodules. Therefore, this paper incorporates an MHSA module into the backbone network. Multi-head attention enables the model to simultaneously focus on both low-quality images—where nodule boundaries are difficult to define precisely—and high-quality images with accurately annotated nodule

International Journal of Scientific Engineering and Science
Volume 10, Issue 8, pp. 22-27, August 2026. ISSN (Online): 2456-7361

boundaries. This not only enhances the model's ability to understand and represent the input data but also allows it to concentrate on the aspects most relevant to the current task, thereby improving the model's nodule recognition and detection capabilities.

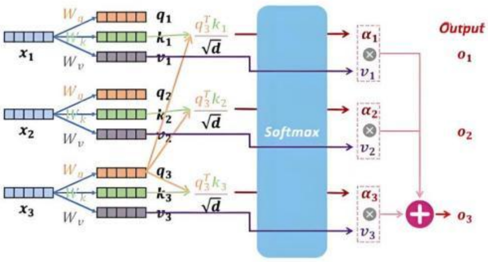

Fig. 5. Multi-Head Attention Mechanism Architecture.

YOLOv8 Object Detection Model: As a highly efficient single-stage object detection algorithm, the YOLOv8 object detection model is structured into three main components: the backbone network, the neck network, and the detection head. In the YOLOv8n network architecture, the backbone network employs data augmentation techniques to enhance the detection performance for small objects. As the core network structure for feature extraction, the backbone's C2f module combines object information with contextual information, thereby enhancing the expressive power of the features. Additionally, the SPPF module unifies the processing of images of different sizes and concatenates the output results, enabling the effective extraction and integration of feature information at different levels within the image. The neck network consists primarily of two parts and is used for feature fusion and processing. Among these, the Feature Pyramid Network (FPN) uses a top-down approach to construct multi-scale feature maps ranging from low-level details to high-level semantic information, further enhancing the model's ability to detect objects of different sizes; the Path Aggregation Network (PAN) employs a bottom-up approach to enrich the fusion of features at different levels, improving the network's ability to represent features in complex backgrounds. The Head network is the output component of YOLOv8n, responsible for decoding the feature maps processed by the Neck layer. By utilizing input information at different scales, the Head network can effectively capture the feature information of objects of corresponding sizes, overcoming the limitations of traditional methods in object detection.

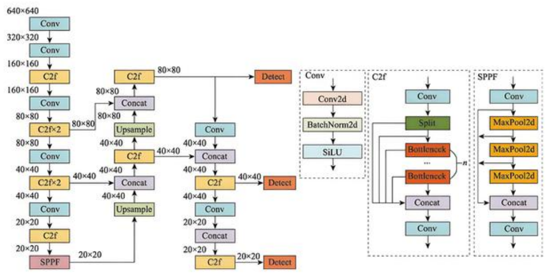

Fig. 6. YOLOv8 Network Architecture Diagram.

This project aims to introduce the MHSA attention mechanism after the SPPF layer in the YOLOv8 model. Without increasing the model's computational load or the number of parameters, this approach significantly enhances the feature extraction capabilities of the backbone network and improves the model's ability to detect small lung nodules.

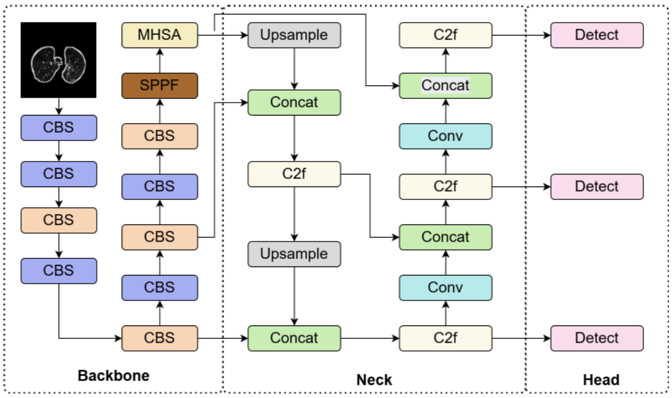

Fig. 7. YOLOv8-attention Network Architecture Diagram.

D. Construction of a Model for Predicting the Growth of Lung Nodules

Growth Prediction Model (PredNet): Certain patterns in nodule growth can be identified by analyzing their growth over a previous period. A nodule grows along its upper edge during the first time interval and continues to grow in a similar direction during the second time interval. Based on this premise, a nodule growth prediction network is established that treats the continuous growth process of a nodule as a spatial transformation problem. It predicts the growth pattern for the next time interval based on the growth pattern observed in the previous time interval, incorporating the lengths of both time intervals as a key reference factor.

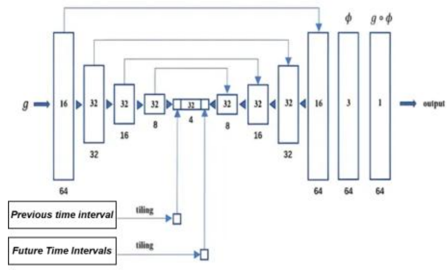

Fig. 8. PredNet Network Architecture Diagram.

The predicted node growth is given by g ∘ φ, where g represents the nodule growth image from the previous time step, with dimensions 64×64×64×1, and φ represents a registration field with dimensions 64×64×64×3. The numbers below and within each rectangle indicate the size of the 4-dimensional matrix, while the long and short arrows in the figure represent the skip connections and the upsampling and downsampling processes, respectively.

The PredNet network adopts an encoder-decoder architecture similar to that of the U-Net network, which has achieved great success in the field of medical deep learning. The PredNet network takes as input the growth history of a

International Journal of Scientific Engineering and Science
Volume 10, Issue 8, pp. 22-27, August 2026. ISSN (Online): 2456-7361

nodule over a past period (i.e., the nodule's growth images and the duration of that period, measured in months) and a future period (measured in months), and outputs the registration field from the previous growth image to the next growth image. The spatial transformation function used is as follows:

$$g \circ \phi\,(\mathrm{p}) = \sum_{q \in Z(p')} g(q) \prod_{d \in (x,y,z)} (1 - |p'_d - q_d|) \quad (1)$$

Here, p is any point in the original image, p ́ is the position of the transformed point p calculated based on the registration field, Z(p ́) denotes the integer voxel point closest to p ́, and d represents the dimension of the image, which in this case refers to the three dimensions: the x-axis, y-axis, and z-axis.

Since the points in the image contain only integer coordinates, interpolation is performed here across the three dimensions for the eight adjacent points. This spatial transformation function is differentiable, so the relevant parameters in the PredNet network can be iteratively updated via backpropagation. Subsequently, based on the registration field and the aforementioned spatial transformation function, the next nodule growth image can be obtained.

At the same time, the future risk level of the nodule can also be predicted based on its growth rate and the risk levels from the previous two iterations. The complete flowchart is as follows:

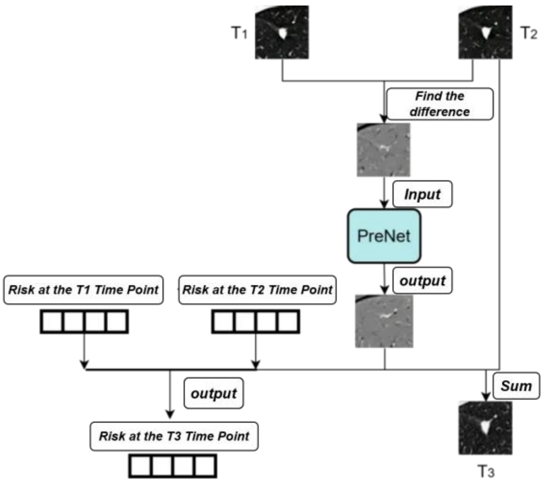

Fig. 9. Complete Flowchart for Growth Forecasting.

E. Model Training

a. Experimental Environment

- PyTorch
- Python 3.9
- Anaconda 4.10.3
- Numpy,Sklearn,Visdom
- Visualization Tool Visdom
- Windows11
- 12th Gen Intel(R) Core(TM) i5-1235U 1.30 GHz

b. YOLOv8 Detection Model Training Parameters

- Task Type: Object detection
- Model: YOLOv8n (with MHSA attention)
- Input Dimensions: 640×640×1 (grayscale CT slice)
- Epochs: 300
- Pretrained: True (COCO pretrained weights)
- Loss function: CIoU Loss + BCE Loss (YOLOv8 native)
- Performance evaluation: Precision, Recall, mAP50, mAP50-95, etc.

c. PredNet Growth Forecasting Model Training Parameters

- Task Type: Deformation field regression (image-to-image registration)
- Model: PredNet (encoder-decoder with skip connections)
- Input Dimensions: 128×128×1 (grayscale CT slice, two-channel after concatenation)
- Epochs: 300
- Loss function: MSE Loss + Smoothness Regularization (λ=0.1)
- Performance evaluation: Reconstruction MSE, Smoothness Loss, Visual inspection of warped images

IV. ANALYSIS AND OPTIMIZATION OF EXPERIMENTAL

A. Training Results and Application Performance of the YOLOv8 Model

This paper evaluates the lung nodule detection model using precision (P), recall (R), mAP50 (mean average precision at 50% IoU), as well as the number of model parameters, computational complexity, and detection speed as evaluation metrics, which assess the model ' s performance from different perspectives. Precision reflects the model's ability to predict specific categories accurately; recall indicates how often the model detects a specific category; and AP represents the harmonic mean of precision and recall, providing an estimate of both values. If we denote TP as true positive samples, FP as false positive samples, TN as true negative samples, and FN as false negative samples, the metric equations can be defined as:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \quad (2)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \quad (3)$$

$$AP = \int_0^1 P(R)\mathrm{d}r, mAP = \frac{1}{N}\sum_{i=0}^{n} AP\mathrm{i} \quad (4)$$

The training results are shown in the figure below

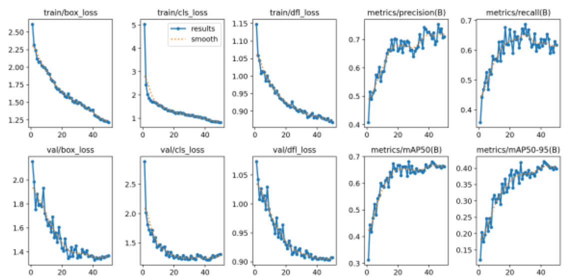

Fig. 10. Summary of Changes to Various Parameters.

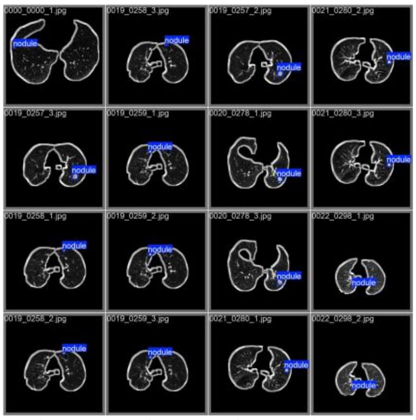

Fig. 11. Sample Annotation Results.

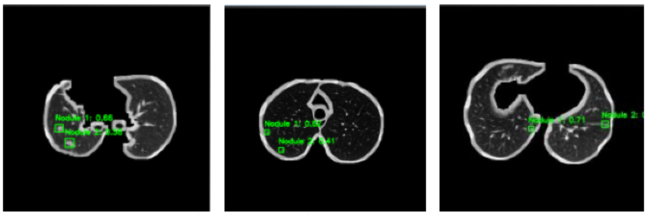

Fig. 12. Sample Annotation Results.

TABLE III. Data Table of Model Experiment Results.

Model	Precision	Recall	mAP50
YOLOv8n	0.855	0.840	0.805
YOLOv8n+MHSA	0.874	0.846	0.827

The experimental results show that the model performs well, accurately identifying lung nodules in CT images and estimating their size parameters. Furthermore, after incorporating the MHSA attention mechanism, all evaluation metrics improved significantly.

B. Performance of the PreNet Model

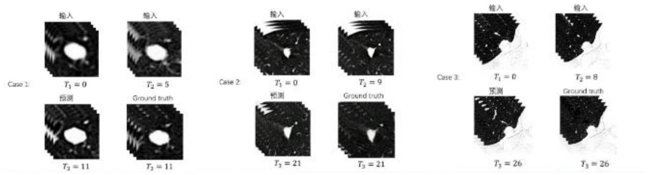

Fig. 13. Prediction of lung nodule growth; from left to right: stable nodule, enlarging nodule, shrinking nodule.

Figure 13 shows the visualization results of the lung nodule growth prediction model. As can be seen, the PredNet model not only performs well when nodules enlarge but also yields reliable predictions when nodules show little change or shrink. The model produces predictions that appear reasonably plausible and have some value for visual assessment, making them more meaningful for subsequent diagnostic procedures.

REFERENCES

[1] 叶良平,姚瑶,赵红,等.某单位 3 458 例体检人群肺结节检出情况分析[J].安徽医专学报,2025,24(01):5-7.
[2] 高瑞,刘苗艳,杨朝婷,等.深度学习在肺结节检测及评估中的研究进展[J].山西医药杂志,2025,54(03):201-205.
[3] 薛文赫,陈楠,刘伦旭.基于 CT 影像预测肺结节生长规律的研究进展[J/OL].中国胸心血管外科临床杂志,1-6[2025-05-05].
[4] 周孟然,徐君阳,卞凯,等.基于 YOLOv8-CGS 的肺结节 CT 图像检测[J/OL].重庆工商大学学报(自然科学版),1-13[2025-05-05].
[5] 郭相均,蒋朝根.基于改进 YOLOv8s 的肺结节检测算法[J].现代信息科技,2025,9(07):87-92.
[6] 孙雅静,郑爽,罗显志.基于改进 YOLOv8n 的手语识别算法[J/OL].湖北大学学报(自然科学版),1-12[2025-05-05].
[7] 韩芳芳.基于 CT 图像多维特征的肺结节检测和诊断方法研究[D].东北大学,2015.
[8] 包强强,唐思源,李擎乾,等.改进YOLOv8模型实现多类型肺结节检测[J].计算机科学与探索,2025,19(02):429-442.
[9] Zenger K S, Agarwal R, Auffermann F W. Using a limited field of view to improve training for pulmonary nodule detection on radiographs.[J].Journal of medical imaging (Bellingham, Wash.), 2025,12(5):051804.
[10] Zhou D, Xu H, Liu W, et al.LN-DETR: cross-scale feature fusion and re-weighting for lung nodule detection[J].Scientific Reports,2025,15(1):15543-15543.
[11] Tuminello S, Untalan M, Flores R, et al.A Response to the Letter to the Editor: "Clinical Value of Incidental Lung Nodule Detection".[J].Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer,2025,20(4):e58.
[12] Mark P, Papalia I, Lai K J, et al.Clinical application of convolutional neural network lung nodule detection software: An Australian quaternary hospital experience.[J].Journal of medical imaging and radiation oncology,2024,68(6):659-666
[13] J. L. Author, “Title of dissertation,” Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.
[14] Azama K, Tsuchiya N, Toyosato S, et al.Artificial intelligence-based lung nodule detection for pulmonary arteriovenous fistulas on chest computed tomography.[J].World journal of radiology,2026,18(4):119851.DOI:10.4329/WJR.V18.I4.119851.

This paper was supported by the 2025 South Anhui Medical University Undergraduate Innovation and Entrepreneurship Competition research project titled “Implementation of a Deep Learning-Based Intelligent Detection and Growth Prediction System for Lung Nodules” (Project No.: S202510368019).