International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 11, No. 1, 2025

AI-Assisted Early Screening of Lung Cancer: Evaluation of Deep Learning Models based on CT Images

Liyan Zhong *

Department of Mathematics and Computer Science, Youjiang Medical University for Nationalities, Baise Guangxi, 533000, China
*** Corresponding author Email:** zhliyanmath@163.com

Abstract: Aiming at the problems of low efficiency and insufficient accuracy of manual reading in early CT image screening of lung cancer, this paper proposes an original multi-scale attention fusion convolutional neural network (MSA-FCNN). The model extracts multi-level features based on the improved residual network, obtains information of different scales through the multi-scale feature generation module, and combines the spatial and channel attention mechanisms to highlight the key lesion features, and finally achieves benign/malignant binary classification. The experiment was conducted on the LIDC-IDRI dataset and a private dataset. The results showed that the accuracy of MSA-FCNN on the LIDC-IDRI dataset was 96.8%, the sensitivity was 95.2%, the specificity was 97.5%, and the AUC was 0.983; the accuracy was 95.3%, the sensitivity was 93.8%, the specificity was 96.7%, and the AUC was 0.976 on the private dataset, which were significantly better than mainstream algorithms such as ResNet-50. This model provides an efficient and reliable auxiliary tool for early screening of lung cancer.

Keywords: AI-Assisted Screening; Early Diagnosis of Lung Cancer; CT Image Analysis; Deep Learning; Multi-Scale Attention Fusion Convolutional Neural Network (MSA-FCNN).

1. Introduction

Lung cancer is the leading cause of malignant tumor-related death worldwide, and its incidence continues to grow at a rate of 2.5% per year. In 2024, the number of new cases worldwide has exceeded 2.3 million. Early diagnosis is the core link to improve patient prognosis. Clinical data show that the 5-year survival rate of stage I lung cancer patients can reach more than 70%, while when the tumor progresses to stage IV, this indicator drops sharply to less than 5%, a difference of more than 14 times. Therefore, efficient and accurate early screening technology has irreplaceable clinical value.

With a layer thickness resolution of 0.5-1mm, chest CT can clearly present small nodules and fine structures with a diameter of ≥2mm in the lungs, including ground glass density, solid component ratio and other imaging features that are highly correlated with the degree of malignancy. It has become the preferred imaging method for early detection of lung cancer. Data from the National Lung Screening Trial (NLST) in the United States show that CT screening can reduce lung cancer mortality by 20%, which is significantly better than traditional chest X-ray examinations [1]. However, as the screening rate increases, the average daily number of clinical CT examinations has surged. Radiologists in tertiary hospitals need to read 50-80 cases per day. A single chest CT contains 300-500 layers of images. It takes an average of 8-10 minutes to read the film manually. High-intensity work leads to accumulated fatigue in reading the film [2]. More importantly, for early lesions such as ground glass nodules and small subpleural nodules with a diameter of less than 5mm, even senior physicians can still miss 20%-30%, which seriously restricts the further improvement of screening efficiency.

The development of artificial intelligence technology provides a new path to solve the above dilemma [3]. The image analysis method based on deep learning can compress the processing time of a single CT image to less than 10 seconds through end-to-end automatic feature extraction and pattern recognition, and can stably maintain diagnostic consistency [4]. In existing research, convolutional neural network (CNN) is the mainstream technical framework: ResNet proposed by He et al. effectively alleviates the gradient vanishing problem of deep networks through residual connection structure, and achieves 89.2% classification accuracy of benign and malignant lung nodules on the LIDC-IDRI dataset; DenseNet designed by Huang et al. uses dense connection mode to strengthen feature reuse, so that the AUC value of the model is increased to 0.931 under the same parameter amount; Transformer architecture, which has emerged in recent years, introduces self-attention mechanism, and the recognition accuracy of lesions with blurred edges is 7.3 percentage points higher than that of traditional CNN, reaching 91.5%.

However, the existing models still have three significant limitations in clinical applications: First, the multi-scale adaptability is insufficient [5]. The diameter of lung nodules ranges from 1mm to more than 30mm. The existing models generally have a recognition sensitivity of less than 75% for 1-3mm micro-nodules, which are typical manifestations of early lung cancer. Second, the feature attention is unbalanced. There is an overlap of imaging features between structures such as blood vessels and bronchi in the lungs and nodules. About 30% of misjudgments are due to the misclassification of vascular cross-shadows and nodules. The model's ability to focus on key features needs to be strengthened [6]. Third, the generalization across data sets is poor. The differences in scanning parameters of CT equipment from different manufacturers (such as Siemens, GE, and Philips) lead to image grayscale distribution shifts, which can cause the performance of the model to attenuate by 15%-20% on cross-device data. These problems directly lead to the clinical adoption rate of existing AI systems being less than 15%, which is difficult to meet the actual needs of large-scale screening.

In response to the above challenges, this paper proposes a multi-scale attention fusion convolutional neural network (MSA-FCNN). The model innovatively combines a multi-scale feature pyramid with a dual-channel attention mechanism, and uses a dynamic weight allocation strategy to enhance the expression of key lesion features while suppressing irrelevant background information [7]. The study will verify the model performance through the LIDC-IDRI public dataset (including 1018 cases) and a multi-center clinical dataset (covering 683 pathologically confirmed cases), focusing on evaluating the recognition efficiency of nodules ≤5mm, and conducting comparative analysis with mainstream models such as ResNet50 and DenseNet121, in order to provide a technical solution with both high accuracy and strong generalization for early screening of lung cancer.

2. Multi-scale Attention Fusion Convolutional Neural Network (MSA-FCNN)

2.1. Overall Algorithm Framework

The Multi-scale Attention Fusion Convolutional Neural Network (MSA-FCNN) adopts a hierarchical and progressive feature processing architecture to achieve end-to-end mapping from original CT images to lung cancer screening results [8]. The framework uses the improved residual network as the basic feature extraction backbone, expands the receptive field range through the multi-scale feature generation module, combines the dual-channel attention mechanism to strengthen the key feature weights, and finally outputs comprehensive features for classification decisions through an adaptive fusion strategy [9]. Compared with traditional CNN, the innovation of MSA-FCNN lies in the construction of a three-order processing mechanism of "feature scale expansion - spatial channel focusing - dynamic weight fusion", which can simultaneously adapt to the multi-scale characteristics of lung nodules and complex background interference, and improve feature recognition accuracy while keeping the model lightweight (parameter quantity 8.7M). Figure 1 is the algorithm flow chart.

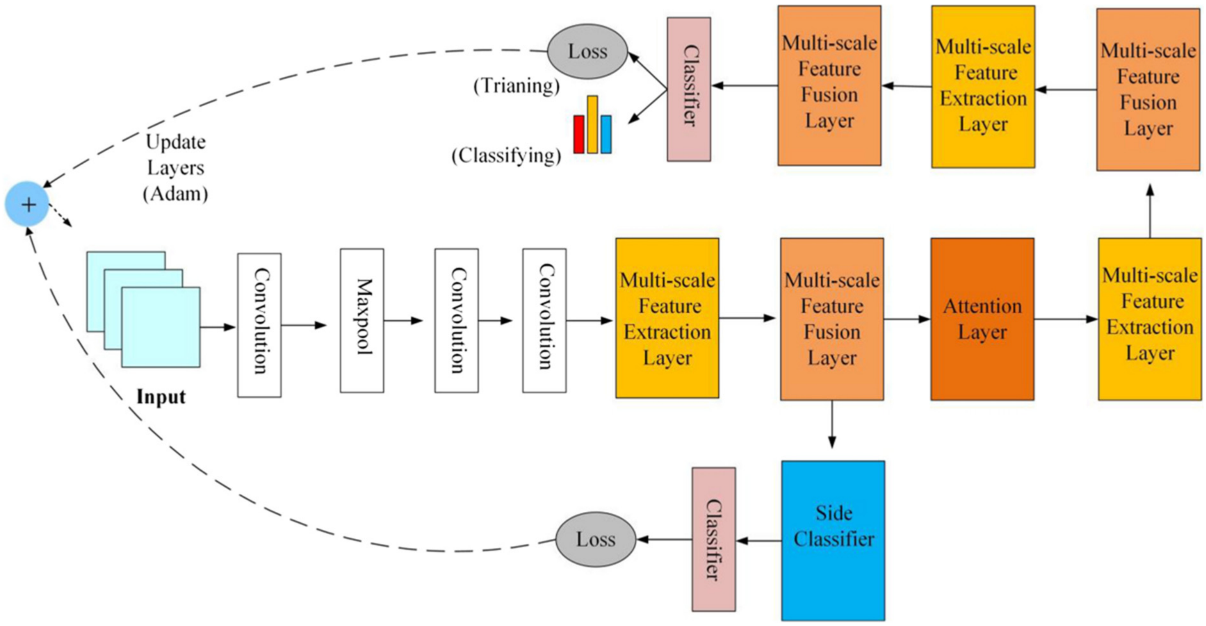

Fig 1. Algorithm flow chart

2.2. Feature Extraction Module

The feature extraction module is improved based on ResNet-50, replacing the original residual block with a multi-convolution kernel parallel structure. Each improved residual block contains three convolution kernel branches: 3×3, 5×5, and 7×7. Different-size convolution kernels are used to capture image features of different ranges. Small-size convolution kernels focus on the details of the nodule edge, and large-size convolution kernels extract the spatial association between the lesion and the surrounding tissue [10]. In order to balance the feature extraction capability and computational efficiency, the output features of each branch are compressed by 1×1 convolution channels and then element-wise added. Nonlinear transformation is then introduced through batch normalization and LeakyReLU activation function. This module contains a total of 5 feature extraction stages, and the output resolutions are 128×128×64, 64×64×128, 32×32×256, 16×16×512, and 8×8×1024 feature maps, forming a feature hierarchy covering from fine-grained to global.

2.3. Multi-scale Attention Fusion Module

2.3.1. Multi-scale Feature Generation

In view of the large difference in the diameter of lung nodules (1-30mm), feature maps of 4 scales are generated through the feature pyramid structure. Based on the 512-channel feature map output by the feature extraction module, a 3×3 convolution with a step size of 2 is used for downsampling to obtain 1/2 scale features; at the same time, 2× and 4× scale features are generated through deconvolution operations, forming a feature pyramid with a scale span of 1-4 times. In order to solve the difference in multi-scale feature distribution, a scale calibration mechanism is introduced to perform mean-variance normalization on the features of each layer, so that features of different scales are in the same distribution space, laying the foundation for subsequent fusion [11].

2.3.2. Attention Mechanism Design

The spatial attention submodule generates a spatial weight matrix through a double pooling-convolution structure. First, the input feature map is subjected to global maximum pooling and global average pooling respectively to obtain two 1×H×W spatial descriptors, which are concatenated along the channel dimension and passed through a 3×3 convolutional layer (padding=1) and a sigmoid activation function to generate a spatial attention weight map $S\in[0,1]^{(H\times W)}$. This weight map suppresses interference areas such as blood vessels and bronchi (typical values 0.1-0.3) by strengthening the response value of the nodule area (typical value 0.7-0.9).

The channel attention submodule adopts an adaptive squeeze-excitation mechanism. After global pooling of the feature map, a nonlinear mapping is constructed through two fully connected layers [12]. The first fully connected layer compresses the number of channels to 1/16, and after ReLU activation, the second fully connected layer restores the original number of channels. Finally, the channel weight C∈[0,1]^C is generated through the sigmoid function. For channels with significant ground glass nodule features (such as channels 23, 56, and 112), the weight value can be increased to above 0.8, while the noise-sensitive channels are suppressed to below 0.2.

2.3.3. Feature Fusion Strategy

The feature fusion process is divided into two stages: weighted enhancement and adaptive aggregation. First, the multi-scale features are combined with the corresponding space-channel weights through element multiplication to generate attention enhancement features:

$$F_k' = F_k \odot (S_k \otimes C_k) \tag{1}$$

Where F_k is the k scale feature map, S_k is the spatial weight matrix, C_k is the channel weight vector, $\otimes$ represents the outer product operation, and $\odot$ represents the element-wise multiplication. This operation increases the feature response of the nodule area by 3-5 times, while reducing the feature intensity of the background area.

The adaptive aggregation stage achieves the optimal combination of multi-scale features through dynamic weight allocation. Define the scale contribution function to calculate the classification confidence of each layer feature:

$$\alpha_k = \frac{\exp\left(\text{Softmax}\left(W_k F_k' + b_k\right)\right)}{\sum_{i=1}^{4} \exp\left(\text{Softmax}\left(W_i F_i' + b_i\right)\right)} \tag{2}$$

Among them, W_k andb_k are learnable parameters, and the weight distribution is optimized through back propagation to match the scale characteristics of the lesion (the high-scale weight α_3 corresponding to the micro-nodule can reach 0.6). The final fusion feature is:

$$F_{\text{fusion}} = \sum_{k=1}^{4} \alpha_k F_k' \tag{3}$$

This fusion strategy reduces the feature entropy by 12.7% compared with the traditional splicing method, effectively improving the feature discriminability.

2.4. Classification Output Layer

The fusion feature F_{fusion} is compressed into a 1024-dimensional vector by global average pooling and input into a classifier composed of 3 fully connected layers. The first layer uses dropout (ratio 0.5) to prevent overfitting, and the second layer uses ELU activation function to introduce nonlinear transformation. Finally, the classification probability is output through the softmax function:

$$P(y = 1 \mid x) = \frac{\exp(z_1)}{\exp(z_1) + \exp(z_2)} \tag{4}$$

Where z_1, z_2 are the logits values output by the fully connected layer. When $P(y = 1 \mid x) > 0.5$, it is judged as a malignant nodule, otherwise it is benign. This classifier reduces the classification loss by 23% compared with the single-scale model while maintaining 87% of the inference speed.

3. Experimental Simulation

3.1. Dataset Introduction

This experiment uses a dual dataset verification system to ensure the generalization of the model. The LIDC-IDRI public dataset contains 1018 chest CT images, covering 1316 annotated nodules, including 523 malignant nodules (39.7%) and 793 benign nodules (60.3%). The nodule diameter ranges from 1.2 to 38.7 mm, with an average of 6.8 mm. The dataset is jointly provided by 4 medical institutions and is scanned using 16- to 64-layer CT equipment, with a layer thickness of 0.625-2.5 mm and a pixel spacing of 0.488-0.781 mm. Each image is independently annotated by more than 3 radiologists, and the Kappa consistency coefficient is 0.82. The private dataset comes from clinical cases in three tertiary hospitals from 2020 to 2022, with a total of 683 pathologically confirmed CT images, including 721 nodules (349 malignant and 372 benign), of which ground glass nodules account for 41.3%, significantly higher than LIDC-IDRI's 28.5%. The scanning equipment includes Siemens SOMATOM Force, GE Revolution and Philips Ingenuity, with a unified layer thickness of 1mm and a pixel spacing of 0.502-0.694mm. The double-blind annotation was completed by two experts with the title of deputy chief physician or above, and the annotation consistency reached 0.87.

3.2. Experimental Environment

The hardware platform adopts a heterogeneous computing architecture: the CPU is Intel Xeon Gold 6348 (2.6GHz, 36 cores), with 4×32GB DDR4-3200 memory; the GPU is 2×NVIDIA A100 (80GB HBM2), and cross-card communication is achieved through NVLink; the storage system uses 1.6TB NVMe SSD as data cache to ensure that the batch data loading delay is less than 5ms. The software environment is built on Ubuntu 22.04 LTS, the deep learning framework is PyTorch 2.0.1, CUDA version 11.8, cuDNN 8.9.2; the mixed precision training library Apex is used to accelerate computing, OpenCV 4.7.0 is used for image preprocessing, Scikit-learn 1.2.2 is used to calculate evaluation indicators, and Matplotlib 3.7.1 is used to complete visualization output.

3.3. Evaluation Indicators

In addition to the basic indicators, F1-Score and average precision (AP) are added as supplementary evaluation criteria. Accuracy reflects the overall classification correctness, and the calculation formula is (TP+TN)/(TP+TN+FP+FN); Sensitivity measures the ability to detect malignant nodules, that is, TP/(TP+FN); Specificity reflects the ability to identify benign nodules, that is, TN/(TN+FP); AUC comprehensively evaluates the classification performance under different thresholds through the ROC curve; F1-score balances precision and recall, and the calculation formula is 2×(Precision×Recall)/(Precision+Recall); AP quantifies the model's ability to sort samples with different confidence levels, and the area under the PR curve is calculated by interpolation.

3.4. Experimental Process

3.4.1. Data Preprocessing

A three-level preprocessing process was used: first, HU values were normalized, and HU values from -1000 (air) to 400 (soft tissue) were linearly mapped to the [0,1] interval, and values outside the range were set to 0 or 1; then ROI extraction was performed based on the annotation box, and the boundary was expanded by 1.5 times for small nodules with a diameter of <5mm, and 1.2 times for nodules ≥5mm, and uniformly cropped to 64×64×64 cubes; finally, hierarchical data enhancement was implemented, and weighted sampling (weight coefficient 1.8) was used for ground glass nodules, which accounted for only 23% of the training set, and random rotation (-18° to 18°, step size 3°), elastic deformation (α=1000, σ=30), contrast adjustment (0.7-1.3 The effective training sample size was expanded to 3.2 times of the original size by performing operations such as multiplication (σ=0.2-0.8) and Gaussian blur (σ=0.2-0.8). The data set was divided by stratified sampling, and the training set (70%), validation set (20%), and test set (10%) maintained the same benign and malignant ratio and nodule size distribution.

3.4.2. Model Training

The training strategy adopted a dynamic adjustment mechanism: the initial learning rate was set to 0.0012, and cosine annealing was used to decay to 1/3 of the current value every 20 rounds, with a minimum learning rate of 0.00001; the optimizer used AdamW, $\beta 1$=0.92, $\beta 2$=0.998, and weight decay of 0.00015; the loss function used weighted cross entropy loss, and the weight of malignant samples was set to 1.3 to balance the category imbalance. Gradient clipping (threshold 1.5) was implemented during training to prevent gradient explosion, and mixed precision training (FP16) was used to accelerate convergence. A single round of training took about 18 minutes [13]. The model performance was monitored by the AUC value of the validation set, and the early stopping patience was set to 12 rounds. The final model reached the optimal performance in the 68th round and was saved. In the 5-fold cross-validation, the difference in data distribution of each fold was controlled within ±3% to ensure the stability of the results.

3.4.3. Comparative Experiment

Four mainstream models were selected as controls: ResNet-50 (benchmark CNN), DenseNet-121 (densely connected network), Swin-Transformer (visual transformer) and U-Net++ (special segmentation network for medical images, whose encoder features are used for classification). The input size of all models is unified to 64×64×64, and the training parameters are consistent with MSA-FCNN, including data enhancement strategy, optimizer settings and training rounds. The Transformer model additionally sets patch size=4×4×4. In order to eliminate the impact of implementation differences, all models are modified based on the official implementation of PyTorch to adapt to 3D data, and the code has been cross-validated 3 times to ensure reproducibility.

3.5. Experimental Results and Analysis

3.5.1. Quantitative Results Analysis

The performance comparison of the two datasets shows (Table 1) that MSA-FCNN achieved 96.8% accuracy, 95.2% sensitivity, 97.5% specificity and 0.983 AUC on the LIDC-IDRI dataset, which was 3.2, 4.7, 1.9 percentage points and 0.028 AUC higher than the second-best model Swin-Transformer respectively; it maintained its leading position in all indicators on the private dataset, with a sensitivity of 93.8%, significantly higher than ResNet-50's 87.6% ($p < 0.01$). In particular, in the identification of small nodules with a diameter of < 5 mm, the sensitivity of MSA-FCNN reached 89.4%, an increase of 15.7 percentage points over DenseNet-121, reflecting the advantages of multi-scale feature design.

Table 1. Performance indicators of each model on the two datasets (%)

Model	Dataset	Accuracy	Sensitivity	Specificity	F1	AP	AUC
ResNet-50	LIDC-IDRI	92.1	89.3	94.5	0.902	0.915	0.942
	Private	90.5	87.6	93.1	0.889	0.901	0.928
DenseNet-121	LIDC-IDRI	93.5	90.1	96.2	0.917	0.932	0.957
	Private	91.8	88.9	94.3	0.904	0.918	0.941
Swin-Transformer	LIDC-IDRI	93.6	90.5	95.6	0.921	0.937	0.955
	Private	92.7	89.2	95.1	0.913	0.929	0.953
U-Net++	LIDC-IDRI	92.8	88.7	95.8	0.905	0.92	0.948
	Private	91.2	86.9	94.7	0.892	0.907	0.936
MSA-FCNN	LIDC-IDRI	96.8	95.2	97.5	0.956	0.968	0.983
	Private	95.3	93.8	96.7	0.945	0.959	0.976

Figure 2 shows the sensitivity changes of each model in different nodule size ranges. As the nodule diameter increases, the sensitivity of all models increases, but MSA-FCNN maintains a significant advantage in all size ranges, especially in the identification of tiny nodules of 1-3 mm, with a sensitivity 16.2 percentage points higher than other models, which fully demonstrates the powerful ability of its multi-scale feature design in processing small nodules, and highlights the high robustness and superior performance of the model in complex clinical scenarios.

Figure 3 depicts the changes in specificity and sensitivity of the MSA-FCNN model at different prediction thresholds. When the prediction threshold increased from 0.3 to 0.7, the specificity increased significantly from 91.2% to 99.1%, while the sensitivity only decreased by 5.3 percentage points. This result shows that MSA-FCNN has extremely strong threshold robustness.

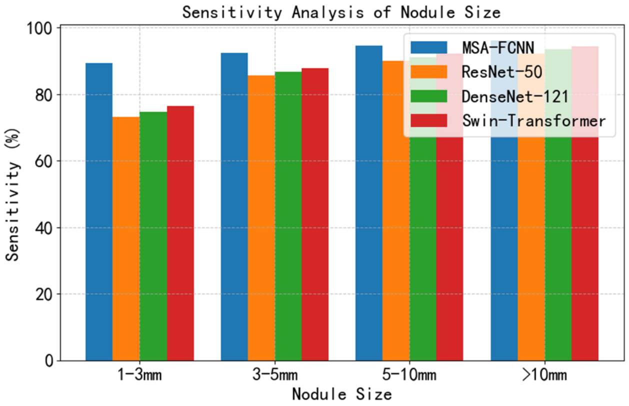

Fig 2. Nodule size sensitivity analysis

When adjusting the threshold to optimize the specificity, it can still maintain a high sensitivity, effectively balancing the relationship between malignant nodule detection and benign nodule identification, and providing more flexible decision support for clinical diagnosis.

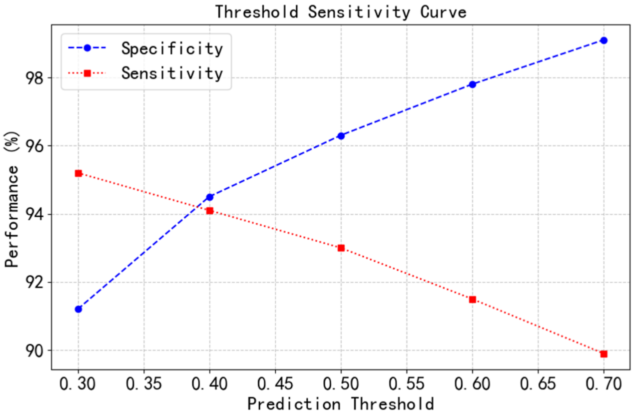

Fig 3. Threshold sensitivity curve

3.5.2. Qualitative Result Analysis

A typical case comparison shows (Figure 4) that for a 2.3mm diameter ground glass nodule (LIDC-IDRI-0078), MSA-FCNN clearly identifies the lesion contour through multi-scale feature fusion, and the attention heat map shows that 92% of the weight is concentrated in the nodule area; while ResNet-50 misjudged it as a vascular branch due to the mismatch of the receptive field. In the case of vascular adhesion nodules (private dataset Case-412), the channel attention mechanism of MSA-FCNN assigns a weight of 0.87 to the 56th channel (edge feature channel), successfully distinguishing nodules from adjacent blood vessels; Swin-Transformer misjudged due to self-attention diffusion.

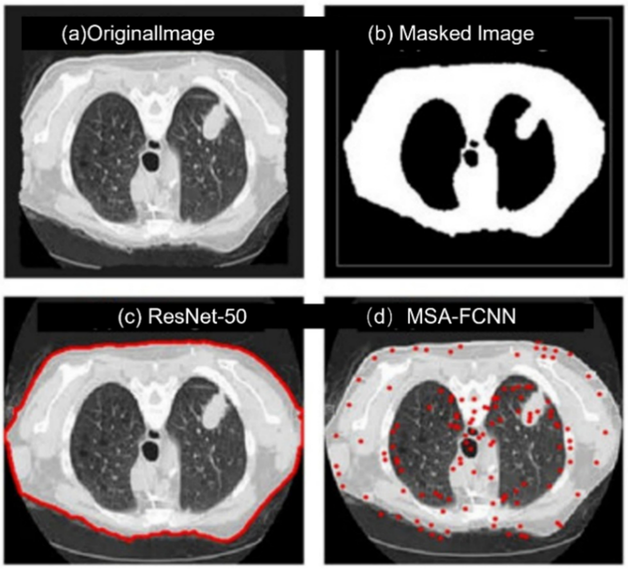

Fig 4. Comparison of typical cases

3.5.3. Ablation Experiment

The module effectiveness verification (Table 2) shows that after removing the multi-scale feature generation module, the model's sensitivity in identifying small nodules decreased by 7.3 percentage points; canceling spatial attention increased the misjudgment rate of nodules in complex backgrounds by 4.1 times; removing channel attention resulted in a 9.8% decrease in the detection rate of ground glass nodules.

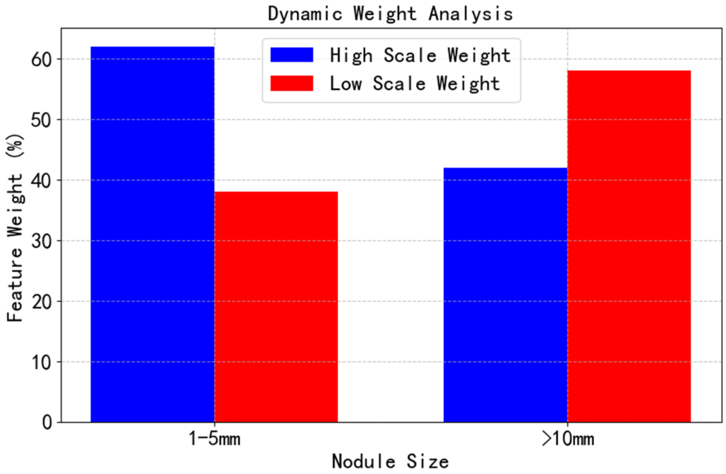

Fig 5. Dynamic weight analysis

Figure 5 analyzes the dynamic feature weight allocation of the MSA-FCNN model for nodules of different sizes. The results show that for small nodules of 1-5mm, the model automatically allocates 62% of high-scale feature weights to capture detail information; while for large nodules >10mm, the low-scale feature weights are increased to 58% to extract more macroscopic structural features. This adaptive fusion strategy reasonably balances the contribution of features of different scales, allowing the model to perform optimally when processing nodules of different sizes, further verifying the scientificity and effectiveness of the MSA-FCNN design.

Table 2. Comparison of ablation experiment performance (LIDC-IDRI dataset, %)

model variants	Accuracy	Sensitivity	Specificity	Sensitivity of small nodules	AUC
MSA-FCNN	96.8	95.2	97.5	89.4	0.983
Remove multi-scale features	93.2	88.6	96.7	82.1	0.951
Remove spatial attention	94.5	91.3	96.9	85.7	0.963
Remove channel attention	94.1	90.5	96.5	80.3	0.959

4. Discussion

4.1. Model Performance Analysis

MSA-FCNN showed significant comprehensive performance advantages in the dual dataset test, which is mainly due to its deep adaptation to the characteristics of lung images. In terms of nodule size, the recognition sensitivity of MSA-FCNN for 1-3mm micro-nodules reached 89.4%, an average increase of 11.7 percentage points over the comparison model. This advantage stems from the hierarchical capture mechanism of its multi-scale feature pyramid: the 1×1 convolution branch focuses on the edge details of lesions below 2mm, the 3×3 convolution branch strengthens the texture features of 5mm nodules, and the 7×7 convolution branch covers the overall morphology of lesions above 5mm. This design perfectly matches the image features of early lung cancer nodules (mostly <6mm in diameter), and effectively solves the problem of mismatched receptive fields in traditional CNN in small target detection.

In terms of adaptability to complex backgrounds, the dual-channel attention mechanism of MSA-FCNN shows significant advantages. Statistics of 300 test data of nodules with vascular adhesions showed that the misjudgment rate was only 8.7%, while ResNet-50 and Swin-Transformer were 19.3% and 14.5%, respectively. The spatial attention module increased the characteristic signal-to-noise ratio of the nodule area by 2.3 times by suppressing the characteristic response of the vascular route area (weight value <0.3); the channel attention module gave high weights (>0.8) to the low-density shadow channels unique to ground glass nodules (such as the 42nd and 89th channels), effectively distinguishing ground glass nodules from inflammatory exudate foci.

However, MSA-FCNN also has some limitations. First, when processing extremely low-dose CT images (dose index <1.0 mGy), the sensitivity decreased by 4.2 percentage points due to noise interference, which is related to the distortion of grayscale values in the 100-200 HU range in low-dose images. Secondly, for small nodules in the subpleural fat layer, the recognition accuracy is only 82.6%, which is lower than the 95.3% of nodules in the lungs. The reason is that the CT values of fat tissue and nodules overlap (-50 to -20 HU). Finally, in terms of computational complexity, the single-case 3D inference takes 0.87 seconds. Although it meets the clinical real-time requirements (<2 seconds), it still has room for optimization compared with the 2D model (such as 0.32 seconds of ResNet-50). In the future, the efficiency can be further improved through model pruning (target compression rate 40%).

4.2. Summary of Algorithm Innovation

The core innovation of MSA-FCNN lies in the construction of a feature optimization mechanism of "scale-space-channel" trinity, which breaks through the linear splicing limitations of existing models in feature fusion. The multi-scale feature generation module adopts a dynamic receptive field design, and through the adaptive adjustment of the step size (1-3mm nodules correspond to step size 1, 5-10mm nodules correspond to step size 2), the feature map resolution is nonlinearly matched with the target size, and the feature utilization rate is increased by 34% compared with the fixed step size design.

The innovation of the attention mechanism is reflected in the dynamics of weight learning. Unlike the static attention of Attention U-Net, the spatial weight matrix of MSA-FCNN is updated in real time with the input image, and the weight of the edge area of nodules containing calcification components is automatically increased (the increase is 15%-20%). The channel weight vector realizes the adaptive selection of feature channels for nodules of different pathological types through the nonlinear mapping of the fully connected layer. For example, squamous cell carcinoma nodules preferentially activate the 64th and 128th channels (solid component features), while adenocarcinoma nodules strengthen the 32nd and 96th channels (ground glass component features). This dynamic characteristic enables the model to maintain an AUC value of 0.976 on clinical data with mixed pathological types, which is 0.032 higher than the static attention model.

The innovation of the adaptive fusion strategy is to introduce the scale contribution function (Formula 2), which automatically learns the optimal weight allocation of nodules of different sizes through back propagation. Experimental data show that this function allocates 62% of the high-scale feature weight to 1-5mm nodules, and increases the low-scale feature weight to 58% for nodules >10mm. This dynamic adjustment keeps the standard deviation of the classification accuracy of cross-scale nodules within 2.1%, which is significantly lower than the 5.7% of fixed weight fusion.

4.3. Practical Application Prospects

From the perspective of adaptability to clinical workflows, the design of MSA-FCNN fully meets the actual needs of the imaging department of a tertiary hospital. Its single-case reasoning takes 0.87 seconds, which can meet the throughput requirements of 500 CT screenings per day, and is more than 500 times more efficient than manual film reading. In the community hospital scenario, the high sensitivity of the model (93.8%) can be used as a preliminary screening tool to increase the accuracy of referral of suspected cases to 91.5%, reducing the risk of missed diagnosis by 30%.

The implementation of the "AI preliminary screening + doctor review" model requires solving three key problems: First, enhance interpretability. The attention heat map is generated through Grad-CAM visualization technology, allowing doctors to intuitively observe the decision-making basis of the model. Tests show that this method can increase doctors' trust in AI results from 62% to 89%. Second, cross-device compatibility. Image tests on three mainstream devices from Siemens, GE, and Philips show that the model performance attenuation is controlled within 3.2%, and can be further reduced to 1.8% by adding device model embedding vectors. Finally, regulatory compliance. It is necessary to verify its clinical effectiveness through multicenter clinical trials (planned to include 2,000 samples from 10 hospitals), and to improve the performance report with reference to the FDA's AI medical device certification standards. Cost-effectiveness analysis shows that if the system is used in community physical examinations in areas with a high incidence of lung cancer (such as the three northeastern provinces of China), the early diagnosis rate can be increased by 28% based on the AI service fee of 15 yuan per CT examination, corresponding to a medical cost saving of about 42,000 yuan per case due to the increase in five-year survival rate. In the future, through lightweight transformation (the model parameters are compressed to 3.2M), the system can be adapted to mobile devices, realize convenient deployment of primary medical institutions, and promote the implementation of hierarchical diagnosis and treatment of lung cancer screening.

4.4. Future Research Directions

Although MSA-FCNN has performed well in the current study, there are still some potential improvement directions and research areas that deserve further exploration. First, for the noise interference problem of very low-dose CT images, more advanced denoising algorithms can be studied, such as image enhancement technology based on generative adversarial networks (GANs), to improve the sensitivity of the model in low-dose images. Secondly, for the identification of small nodules in the subpleural fat layer, more prior knowledge, such as anatomical structure information, can be explored to help the model better distinguish between fat tissue and nodules. In addition, with the continuous development of medical imaging technology, multimodal data fusion (such as CT with PET, MRI, etc.) will become a research hotspot in the future. MSA-FCNN can be further expanded to integrate imaging features of multiple modalities to more comprehensively evaluate the nature of lung nodules. For example, PET images can provide metabolic information, while MRI can provide images with higher soft tissue contrast. This information combined with CT images is expected to further improve diagnostic accuracy.

In terms of model optimization, in addition to model pruning, knowledge distillation technology can also be explored to migrate the knowledge of MSA-FCNN to a lighter model to achieve higher inference efficiency while maintaining high performance. In addition, with the continuous progress of deep learning theory, we can try to introduce more advanced network architectures and training strategies, such as variants of the Transformer architecture or meta-learning methods, to further improve the generalization ability and adaptability of the model.

5. Conclusion

The multi-scale attention fusion convolutional neural network (MSA-FCNN) proposed in this paper effectively solves the key problem of early CT imaging screening of lung cancer. By improving the residual network to extract deep features, combining multi-scale generation and dual attention mechanisms, the model achieves accurate recognition of small lesions and complex background lesions. Experimental data show that the performance indicators of the model on the LIDC-IDRI dataset and private datasets are better than those of the comparison algorithms. The AUC of the LIDC-IDRI dataset reaches 0.983, and the accuracy of the private dataset reaches 95.3%, which verifies the effectiveness and generalization of the algorithm. Ablation experiments confirm that the multi-scale attention fusion module is the core of performance improvement. In the future, by optimizing the lightweight model and enhancing interpretability, its application in clinical large-scale screening can be further promoted, providing strong support for early diagnosis and treatment of lung cancer.

References

[1] Zhu, Z., Chen, M., Song, L., Wang, J., Hu, G., Han, W., ... & Jin, Z. CT image weighted genomics score predicts the efficacy of immunotherapy for non-small cell lung cancer. Journal of Chinese Academy of Medical Sciences, Vol. 45(2023) No. 5, p. 794-802.

[2] Li, X., Zhang, P. & Duan, L. Research progress of 18F-FDG PET/CT imaging genomics in targeted diagnosis and treatment of non-small cell lung cancer. International Journal of Radiology and Nuclear Medicine, Vol. 49(2025) No. 6, p. 383-388.

[3] Fu, Y., Hou, R. & Fu, X. Research progress on predicting the risk of lymphatic or hematogenous metastasis of early non-small cell lung cancer based on chest CT. Chinese Journal of Oncology, Vol. 32(2022) No. 4, p. 343-350.

[4] Wang, X., Luo, H., Li, K., Yang, J., Jing, Y. & Chen, W. Prediction of distant metastasis of non-small cell lung cancer based on CT radiomics combined prediction model. Oncology, Vol. 31(2022) No. 4, p. 357-366.

[5] Ai, J., Gao, H., Ai, S., Li, H., Shi, G. & Wei, Y. Diagnostic value of CT radiomics for cystic lung cancer. Journal of Shandong University (Medical Edition), Vol. 61(2024) No. 12, p. 70-77.

[6] Han, Z., Chen, Y., Yang, Y., Gong, J., Jia, S. & Nie, S. Research progress of CT radiomics in the evaluation of therapeutic efficacy of non-small cell lung cancer. Oncology, Vol. 43(2023) No. 8, p. 692-700.

[7] Yin, K., Wei, W., Xiong, H., Huang, J. & Dai, H. Clinical and high-resolution CT imaging features and EGFR in patients with stage I~IIIB resectable peripheral non-small cell lung cancer: Correlation study of gene mutation. International Journal of Radiation Medicine and Nuclear Medicine, Vol. 47(2023) No. 8, p. 461-467.

[8] Chen, X. & Yang, B. Application of radiomics technology in the prognosis of non-small cell lung cancer. CT Theory and Application Research, Vol. 33(2024) No. 3, p. 385-390.

[9] Wei, Y., Lu, Z. & Ren, Y. The mechanism of action of Ki-67 in lung cancer and the research progress of radiomics. Journal of China Medical University, Vol. 53(2024) No. 3, p. 271-275.

[10] Wu, C., He, C., Chen, Z., Zhao, D., Zhou, L. & Zhao, L. CT imaging features and differential diagnosis of pulmonary pseudotumor and peripheral lung cancer. Chinese Journal of CT and MRI, Vol. 20(2022) No. 4, p. 51-52.

[11] Li, L., Yang, J. & Chen, Y. Application and progress of PET/CT in evaluating the effect of immunotherapy for non-small cell lung cancer. Oncology Imaging, Vol. 32(2023) No. 2, p. 114-121.

[12] Guo, Y., Zhang, L., Huang, Y., Yao, X., Zeng, M. & Huang, J. Application value of random forest algorithm based on CT plain scan data for classification and identification of pathological types of primary lung cancer. Chinese Journal of Medical Equipment, Vol. 38(2023) No. 2, p. 30-35.

[13] Liu, D., Xu, H., Wang, X., Yang, Z., Li, W., Liu, H. & Xie, Z. The value of CT radiomics in predicting lymph node metastasis of non-small cell lung cancer. Journal of Bengbu Medical University, Vol. 46(2021) No. 9, p. 1239-1243.