3D-DHA-MAFNet: A Multi-View Attention-Enhanced Deep Learning System for Lung Nodule Detection and Classification

Guigui Zhao[1], Qingchun Meng[2*], Jingliang Zhang[1]
[1]Medical College, Zhengzhou University of Industrial Technology, Zhengzhou 451150, China
[2]School of Information Engineering, Zhengzhou University of Industrial Technology, Zhengzhou 451150, China
[*]E-mail: yyshq8494@163.com
[*]Corresponding author

Keywords: low-dose computed tomography, three-dimensional convolutional neural network, intelligent diagnostic tool, lung cancer

Received: August 27, 2025

Lung cancer, one of the most lethal malignancies globally, has a high mortality rate, making early screening and accurate diagnosis crucial for improving patient survival rates. Low-dose computed tomography (LDCT) has significantly increased the detection rate of lung nodules, but the vast amount of imaging data poses a significant challenge for manual interpretation. This article proposed a lung nodule computer-aided quantitative diagnostic system based on a three-dimensional convolutional neural network (3D CNN), named 3D-DHA-MAFNet. By integrating deformable convolution network (DCN), hybrid attention mechanism (HAM), and multi-view adaptive fusion network (MAFN), this system significantly enhanced the detection and benign/malignant classification capabilities for small and irregular lung nodules. The model was trained and evaluated on the lung nodule analysis 2016 (LUNA16) and LIDC-IDRI datasets with a patient-level split in the ratio of 7:2:1. The number of training epochs was 200. Data augmentation methods included random rotation (±15°), translation (±5 pixels), and elastic deformation. The performance was compared with baseline models including 3D-UNet, AttentNet, 3DAGNet, 3D multi-scale capsule network (3D-MCN), and 3D multi-view SE CNN (3D MVSECNN). Experiments showed that on the LUNA16 dataset, the model achieved a detection sensitivity of 98.7% with a low false-positive rate of 1.2 per scan, which was a notable improvement over baseline models like 3D-UNet. On the LIDC-IDRI dataset, the classification accuracy for benign and malignant nodules reached 96.8%, with an AUC value of 0.983, and particularly, the sensitivity for malignant nodules was enhanced to 97.5%. Ablation studies verified the synergistic gains of each module, and cross-clinical dataset testing further demonstrated its strong generalization ability. This article presents a high-precision and interpretable intelligent diagnostic tool designed for clinical applications, which holds strong potential to promote the clinical translation of early lung cancer screening.

Povzetek: Članek predstavlja 3D-DHA-MAFNet, globoki učni sistem za zaznavanje in razvrščanje pljučnih nodulov (vozlišč) iz LDCT slik. Z uporabo deformabilnih konvolucij, hibridne pozornosti in večpogledne fuzije dosega visoko točnost, robustnost in dobro posplošljivost na kliničnih podatkih.

1 Introduction

Lung cancer is one of the world's deadliest malignancies. The IARC's *Global Cancer Statistics 2022* shows that there are about 2.5 million new cases and over 1.8 million deaths annually, accounting for 18.7% of all cancer deaths [1]. Early detection and benign/malignant judgment of pulmonary nodules are crucial for improving the five-year survival rate. Low-dose computed tomography (LDCT) screening can significantly increase early detection rates. For example, the NELSON study shows that it can reduce mortality by 24-33% [2]. However, as CT data increases, especially for small, atypical, or nodules with underlying diseases, the reading pressure increases, raising the risk of missed diagnoses and misdiagnoses.

Artificial intelligence (AI), especially computer-aided diagnosis (CAD) systems based on deep learning, has provided new solutions for intelligent pulmonary nodule detection. A three-dimensional convolutional neural network (3D CNN) can extract spatial contextual information from 3D volume images. Compared with 2D networks, it can better capture nodule morphology, density, and spatial distribution features [3]. Previous studies have shown that Li et al. [4] proposed a multi-scale 3D CNN combined with channel and spatial attention to improve small nodule detection rates. Hammad et al. [5] built a lightweight CNN combined with saliency maps, achieving a 99.89% classification accuracy on the LIDC-IDRI dataset.

Nevertheless, lung nodule AI still faces three major challenges. First, small nodules have weak features and blurred boundaries, which are easily confused with blood vessels, resulting in high missed detection rates [6]. Second, multi-view fusion methods are simple, mostly using static splicing or average weighting, lacking adaptive capabilities [7]. Third, the model's generalization is insufficient. It mostly relies on public datasets and lacks pathological gold standard support, showing unstable performance in real-world clinical settings. Class imbalance also affects model robustness, and Focal Loss or Dice Loss is used to alleviate this issue [8, 9].

To address the aforementioned issues, this study proposed the 3D-DHA-MAFNet system. The deformable convolution network (DCN) was introduced to enhance the adaptability to nodule shape variations, the hybrid attention mechanism (HAM) was used to focus on key areas and reduce false negatives, and the multi-view adaptive fusion network (MAFN) efficiently integrated multi-view information through dynamic weighting. Moreover, the system combined Focal Loss with MixUp/CutMix data augmentation to mitigate class imbalance and improve generalization in complex clinical scenarios. The systematic integration of DCN, HAM, and MAFN enabled the model to maintain high precision while possessing strong generalization ability, providing reliable support for clinical early lung cancer screening. In the future, with the development of multi-center data sharing and privacy protection technologies, the system is expected to achieve broader clinical applications, reduce the burden on physicians, and enhance diagnostic efficiency.

2 Related works

Early detection of lung nodules is critical for improving the survival rates of lung cancer patients. However, traditional 2D CNN suffer from information loss when processing 3D medical image data, limiting their effectiveness in lung nodule detection tasks. To address this, researchers have begun to explore the application of 3D CNN in lung nodule detection. Tan et al. [10] proposed a two-stage lung nodule detection system based on a 3D CNN for detecting candidate nodules and reducing false positives (FP), achieving a detection sensitivity of 97.5% with an average of only one FP per scan on the LIDC-IDRI dataset. Song et al. [11] introduced a 3D CNN and capsule network (CapsNet) model (3D CNN-CapsNet), which achieved a detection rate of 95.19%, sensitivity of 92.31%, specificity of 98.08%, and an F1 score of 0.95 on the early lung cancer program (ELCAP) dataset, notably outperforming other baseline methods. Almahasneh et al. [12] proposed a lung nodule detection model based on a 3D fully convolutional attention mechanism, AttentNet, which demonstrated strong performance in reducing FP on the lung nodule analysis 2016 (LUNA16) dataset, validating the effectiveness of the 3D attention mechanism in lung nodule detection.

Accurate segmentation of lung nodules is critical for subsequent feature extraction and classification tasks. 3D CNN can effectively leverage spatial information from volumetric data, thereby enhancing segmentation accuracy. Xiao et al. [13] proposed the 3D-Res2UNet model, which introduced a Res2Net module on the basis of 3D-UNet to achieve multi-scale feature extraction, thereby enhancing the ability to identify nodule boundaries, and achieved a Dice coefficient of 95.30% and a recall rate of 99.1% on the LUNA16 dataset. Xu et al. [14] designed a 2.5D V-Net combining separable 3D convolution (S3D) with a dual-branch feature fusion module (DFFM), which not only improved the accuracy of lung nodule segmentation but also demonstrated good noise resistance and adaptability to complex anatomical structures. Liu et al. [15] proposed the SCA-VNet model, which combines 3D coordinate attention and edge enhancement mechanisms, achieving a Dice coefficient of 87.50% on the LIDC-IDRI dataset, notably outperforming existing methods and validating the effectiveness of this strategy in lung nodule segmentation. In addition, 3DAGNet achieved efficient detection and localization of lung nodules by combining global search and attention mechanisms with multi-scale feature extraction modules [16].

In the task of classifying the benignity or malignancy of lung nodules, 3D CNN can extract spatial features of nodules to assist doctors in diagnosis. In recent years, a variety of enhanced 3D CNN architectures have been proposed to improve the accuracy and robustness of classification performance. Yang et al. [17] proposed a 3D multi-view SE CNN (3D MVSECNN), which enhanced the modeling ability of features from different views by introducing SE modules. Experimental results on the LIDC-IDRI dataset showed that the model achieved an accuracy of 96.04% and a sensitivity of 98.59% in the binary classification task, and an accuracy of 87.76% in the ternary classification task, all of which were superior to other advanced methods. In addition, Afshar et al. [18] developed a 3D multi-scale capsule network (3D-MCN), which combined the advantages of 3D input, multi-scale feature extraction, and capsule networks. Experiments on the LIDC-IDRI dataset showed that the model achieved an accuracy of 93.12%, a sensitivity of 94.94%, and an AUC of 0.9641, showing strong performance. The improved 3D attention CNN proposed by Huang et al. [19] combines feature fusion and hybrid loss functions, enhancing the interaction of different channel and spatial features in the task of lung nodule classification, thereby improving the classification ability of small and ill-defined nodules. On the NLST dataset, the model achieved an accuracy of 85.3% and an AUC of 0.9042, validating the effectiveness of attention mechanisms and hybrid loss functions in improving the model's discriminative ability. This echoes the idea adopted in this article's method of using the HAM module, both emphasizing the enhancement of key feature areas through attention to improve the recognition performance of indistinct nodules. The V-3DResNets model proposed by Prithvika and Anbarasi [20], based on residual network variants and slice grouping design, improves the accuracy of nodule localization and detection in lung nodule detection through multi-scale feature extraction and slice-aware mechanisms. This method emphasizes the full utilization of inter-slice information in volumetric data, which helps to improve the detection rate of small nodules. This strategy is similar to the goal of this article's MAFN module in multi-view feature fusion, both aiming to dynamically integrate spatial information to enhance the perception of nodule heterogeneity and small-sized nodules.

In summary, in recent years, based on 3D CNN, researchers have significantly improved the performance of lung nodule CAD systems by introducing methods such as multi-view, attention mechanisms, multi-scale feature fusion, improved loss functions, and false-positive reduction strategies (Table 1). However, further research is required to enhance the generalization capability of these models, reduce their computational complexity, and improve their interpretability and reliability in clinical settings.

Table 1: Summary of relevant work on lung nodule detection and classification using 3D deep learning methods

References	Datasets	Key Metrics (Accuracy/Sensitivity/AUC)	Main Structural Innovations	Limitations
Tan et al. [10]	LIDC-IDRI	Sensitivity: 97.5%, FP: 1 per scan	Two-stage 3D CNN for detection and FP reduction	Limited generalization ability for small nodules
Song et al. [11]	ELCAP	Accuracy: 95.19%, Sensitivity: 92.31%	3D CNN + CapsNet	High computational cost
Almahasneh et al. [12]	LUNA16	–	AttentNet: 3D fully convolutional network with attention	Classification results not reported
Xiao et al. [13]	LUNA16	Dice coefficient: 95.30%, Recall: 99.1%	3D-Res2UNet with multi-scale feature extraction	Focused solely on segmentation tasks
Xu et al. [14]	LIDC-IDRI	–	2.5D V-Net with S3D and DFFM	Limited to segmentation tasks
Liu et al. [15]	LIDC-IDRI	Dice coefficient: 87.50%	SCA-VNet with 3D coordinate attention and edge enhancement mechanism	Low Dice coefficient for complex nodules
Yang et al. [17]	LIDC-IDRI	Accuracy: 96.04%, Sensitivity: 98.59%	3D MVSECNN with SE module	Limited multi-view adaptive fusion capability
Afshar et al. [18]	LIDC-IDRI	Accuracy: 93.12%, AUC: 0.9641	3D-MCN	High model complexity
Huang et al. [19]	NLST	Accuracy: 85.3%, AUC: 0.9042	An improved 3D attention CNN that combines feature fusion and hybrid loss functions	Lower performance on public datasets

3 System design and methods

To enhance the detection capability of small, edge-blurred, or irregularly shaped lung nodules in LDCT images, this article proposed a 3D CNN system integrating DCN, HAM, and MAFN (3D-DHA-MAFNet) (Figure 1). The system comprised three core modules: Feature extraction module: Combining DCN with HAM to improve the representation ability of small nodules with complex shapes; MAFN: Dynamically weighting multi-directional features through an SE module to achieve information integration; Optimization module: Introducing Focal Loss to address class imbalance and supplements with MixUp/CutMix data augmentation to alleviate the scarcity of 3D imaging samples.

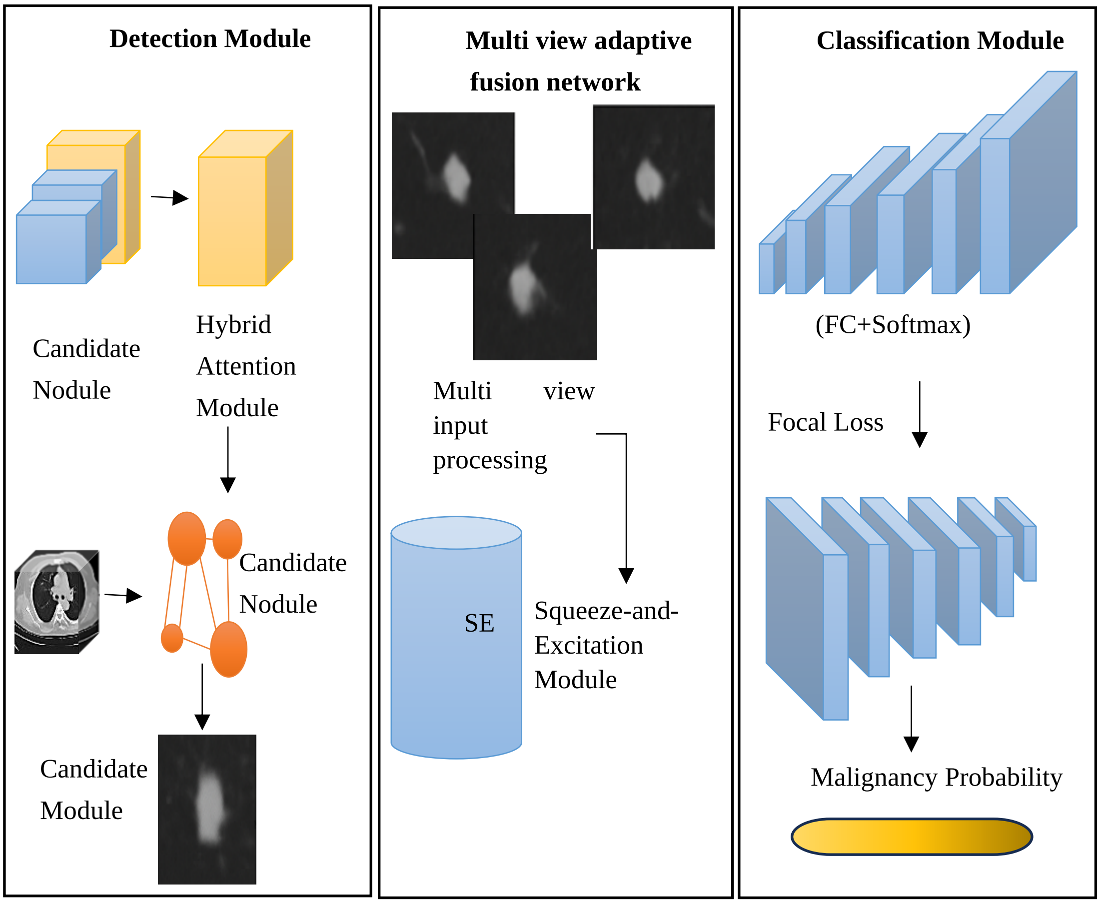

Figure 1: Schematic diagram of the three major modules of 3D-DHA-MAFNet

3.1 DCN and HAM feature extraction module

To enhance the detection ability of small-sized and irregular-shaped nodules (such as ground-glass nodules), DCN and HAM were introduced into the Backbone in this study. The Backbone is composed of four layers of 3D convolution, with parameters for each layer as follows: Conv1 (3×3×3, stride 1, channels 32, ReLU + BatchNorm3D), Conv2 (3×3×3, stride 2, channels 64), Conv3 (3×3×3, stride 2, channels 128), Conv4 (3×3×3, stride 2, channels 256). DCN is embedded between Conv2 and Conv3 to dynamically capture nodule edges in mid-level features; HAM is embedded behind Conv4 to enhance attention to key areas in high-level semantic features.

DCN: The standard 3D convolution operates on the input feature map x at the fixed sampling points $\{d_l\}_{l=1}^{L}$, and its output is as follows.

$$y(d_0) = \sum_{l=1}^{L} w_l \cdot (xd_0 + xd_l) \tag{1}$$

d_0 denotes the position coordinates on the output feature map; $d_l \in Z^3$ represents the convolution kernel sampling offsets; w_l represents the convolutional kernel parameters. The irregular structure of nodule edges means that fixed sampling limits the ability to model non-Euclidean structures. DCN achieves dynamic sampling by learning offsets $\Delta d_l \in R^3$.

$$y(d_0) = \sum_{l=1}^{L} w_l \cdot (xd_0 + xd_l + x\Delta d_l) \tag{2}$$

This mechanism endows the receptive field with the ability of nonlinear deformation, enhancing the ability to capture morphological features such as nodule edges, protrusions, or cavities.

HAM: The HAM module enhances features by combining channel attention and spatial attention. The channel attention (SE) globally weights each channel.

$$Z_C = GAP(x_c) = \frac{1}{V}\sum_{v=1}^{V} x_c\,(v) \tag{3}$$

$$S_C = \sigma(W_2 \cdot ReLU(W_1 \cdot Z_c)) \tag{4}$$

$$X_c^{att} = S_C \odot X_c \tag{5}$$

Spatial attention further highlights significant regions by performing max pooling and average pooling across the channel dimension, concatenating the results, and then generating the attention map MSM_SMS through a 3×3×3 convolution.

$$F_{avg} = \frac{1}{C}\sum_{c=1}^{C} F_c, F_{max} = max_c F_c \tag{6}$$

$$M_S = \sigma(Conv_{3\times3\times3}([F_{avg} \oplus F_{max}])) \tag{7}$$

$$F^{out} = F^{in} \otimes M_s \tag{8}$$

F_{avg} denotes the result of average pooling on the channel dimension, and F_{max} denotes the result of maximum pooling on the channel dimension, with dimensions $[D,H,W]$; C represents the number of channels in the feature map; Conv3×3×3 indicates a 3D convolution operation with a kernel size of 3×3×3 and an output channel count of 1, used to generate the spatial attention map M_s; F^{in} represents the feature map weighted by channel attention (output of the SE module); F^{out} represents the output feature map weighted by both spatial and channel attention, which is the final output of this module. The HAM helps enhance the model's response to small nodules and blurry areas, improving localization accuracy and edge discrimination.

3.2 MAFN

Lung nodules display considerable morphological variations across different anatomical planes (axial, coronal, and sagittal). Relying exclusively on a single viewing plane may lead to the omission of critical structural information. To fully exploit the redundant and complementary features in multi-view images, this article designed a MAFN based on three-view input and SE-guided fusion mechanism. This structure comprised three key components: view slicing generation, multi-view feature extraction, and attention-guided fusion.

3.2.1 Multi-view input processing

It is assumed that the original lung nodule cubic input be a tensor.

$$x \in \mathbb{R}^{D\times H\times W} \quad (9)$$

D, *H*, *W* represent the depth (number of axial slices), height, and width of the cube, respectively. To achieve multi-view modeling, 2D slice sequences in three anatomical directions are extracted from this tensor.

Axial slices are directly extracted along the axial direction to form the original input sequence.

$$x_A = x \in \mathbb{R}^{D\times H\times W} \quad (10)$$

The dimensions of the original tensor are rearranged to obtain sagittal view slices.

$$x_s = permute(x, [2,0,1]) \in \mathbb{R}^{D\times H\times W} \quad (11)$$

Similarly, the coronal view is obtained through transposition.

$$x_C = permute(x, [1,2,0]) \in \mathbb{R}^{D\times H\times W} \quad (12)$$

The inputs from the three directions are fed into a structurally shared CNN for feature extraction, resulting in feature vectors for each view.

$$F_A = F_{CNN}(x_A), F_S = F_{CNN}(x_S), F_C = F_{CNN}(x_C) \quad (13)$$

$F_{\mathrm{CNN}}(\cdot)$ represents the feature extraction network with shared parameters, which is used to maintain structural consistency and reduce model redundancy.

All views are cropped to a 32×32×32 voxel region centered on the nodule and undergo rotation normalization (alignment of the axial, coronal, and sagittal planes with the scanning coordinate system) to ensure multi-view spatial consistency.

3.2.2 SE-guided fusion mechanism

Direct concatenation of multi-view features may introduce a large amount of redundant or irrelevant information. To enhance the discriminative ability of feature fusion, this article introduced the SE module, which assigned importance weights to each view through a dynamic weighting mechanism.

It is assumed that any view feature is $f_i \in \mathbb{R}^c$, where $i \in \{\mathrm{A,S,C}\}$ represents the view index, the α_i for this view is first calculated through an SE attention branch.

$$\alpha_i = \sigma(W_2 \cdot \delta(W_1 \cdot f_i) \quad (14)$$

After calculating the attention weights α_A, α_S, α_C for the three views, the features are fused in a weighted manner to obtain a unified multi-view representation.

$$\mathrm{f}_{fusion} = \alpha_A \cdot \mathrm{f}_A + \alpha_S \cdot \mathrm{f}_S + \alpha_C \cdot \mathrm{f}_C \quad (15)$$

In multi-view feature fusion, although direct concatenation followed by convolution or fully connected layers is commonly used, it increases the feature dimension, leading to redundancy and computational overhead. The SE-based weighted sum method is more parameter-efficient, allowing dynamic adjustment of the importance of each view, reducing redundancy, improving training stability, and intuitively reflecting the contribution of each view, thereby enhancing interpretability. In multi-view or multi-modal image analysis, this method achieves better inference efficiency while maintaining performance [21, 22]. Therefore, this study adopted this strategy to balance performance, efficiency, and interpretability.

3.3 Classification optimization and training strategies

To effectively enhance the robustness and generalization ability of the model in classifying the benignity or malignancy of lung nodules, this article designed two key training optimization strategies: First, Focal Loss was introduced to address the class imbalance issue caused by the relatively small proportion of malignant nodule samples; second, MixUp and CutMix data augmentation techniques were jointly used to effectively increase the diversity of training samples in 3D medical imaging.

3.3.1 Focal loss

Because benign nodule samples far outnumber malignant ones, traditional cross-entropy loss is easily dominated by the majority class, resulting in insufficient model recognition ability for malignant nodules. To focus on hard-to-classify samples and reduce the risk of false negatives, this study combined Focal Loss with cross-entropy loss at a weight ratio of 0.75:0.25, where Focal Loss targeted difficult samples and cross-entropy ensured overall training stability.

The expression for Focal Loss is as follows.

$$\zeta_{focal} = -\alpha_t(1-\hat{y}_t)^2 \log(\hat{y}_t) \quad (16)$$

$\hat{y}_t \in [0,1]$ means the model's predicted probability for the true class. The class balance factor, $\alpha_t \in [0,1]$, is used to increase the weight of minority classes (such as malignant nodules). The focusing parameter, 2, is used to adjust the penalty for easily classified samples and is set to 2 in this article. When $\hat{y}_t$ approaches 1, that is, when the sample is easily correctly classified, the loss value will be substantially compressed. Conversely, when $\hat{y}_t$ is small (indicating that the model is not confident in its prediction for the sample), the loss function will assign a higher penalty, causing the model to focus more on difficult samples. Through this mechanism, Focal Loss substantially improves the model's sensitivity to minority classes such as malignant nodules and effectively reduces the FN rate (FNR).

3.3.2 Mixed augmentation strategy (MixUp + CutMix)

Medical images, especially LDCT lung nodule data, face the problem of limited sample size, and the training of 3D models requires an even larger sample scale. To address this, MixUp and CutMix, two mixed augmentation strategies, were incorporated to improve the diversity of sample distributions and enhance boundary robustness from the perspectives of global feature synthesis and local region perturbation, respectively.

MixUp is a global linear interpolation augmentation method. It generates new training samples by linearly mixing two samples and their corresponding labels.

$$\widetilde{\boldsymbol{X}} = \lambda \boldsymbol{X}_i + (1-\lambda)\boldsymbol{X}_j, \widetilde{\boldsymbol{Y}} = \lambda \boldsymbol{Y}_i + (1-\lambda)\boldsymbol{Y}_j \quad (17)$$

X_i, X_j mean the volume images of two input samples. Y_i and Y_j represent their corresponding labels (encoded using one-hot encoding). λ~Beta(α,α) represents the mixing weight, with α typically ranging from 0.2 to 0.4.

This strategy achieves continuous modeling in both the image space and the label space, enhancing the model's ability to represent the transition regions of class boundaries and reducing the risk of overfitting.

CutMix is a local masking augmentation method. It forms new training samples by randomly cutting out a cuboid region from one image and replacing it with the same location region from another image.

$$\widetilde{\boldsymbol{X}} = \boldsymbol{M} \circledast \boldsymbol{X}_i + (\boldsymbol{1} - \boldsymbol{M}) \circledast \boldsymbol{X}_j, \widetilde{\boldsymbol{Y}} = \lambda \boldsymbol{Y}_i + (1-\lambda)\boldsymbol{Y}_j \quad (18)$$

$M \in \{0,1\}^{D \times H \times W}$ means a randomly generated binary cuboid mask, indicating the region to be replaced. $\odot$ denotes element-wise multiplication. $\lambda = \frac{\sum \mathrm{M}}{\mathrm{D \cdot H \cdot W}}$ represents the proportion of the retained region and is also used to calculate the weight of the new label. By locally replacing regions, CutMix disturbs the model's regional focus, effectively enhancing the model's robustness to image structural uncertainty and preventing it from "memorizing" the training set based on local features.

During the training process, the switching between MixUp and CutMix was done randomly on a batch-wise basis, meaning that each mini-batch randomly decided whether to use MixUp or CutMix with a fixed probability of 0.5 to achieve the complementarity of the two strategies. MixUp provides overall smooth feature transitions, while CutMix introduces local perturbations to challenge the model's discriminative boundary capabilities. Together, they effectively improve the model's generalization ability and alleviate overfitting caused by limited data.

3.4 Model training and inference process

In Figure 2, the training process first cropped and normalized the lung nodule region from LDCT images into a fixed-size 3D cube, then generated three sets of 2D slices along the three orthogonal axes as input to a 3D CNN with the same structure. The network introduced DCN to enhance spatial modeling ability, embeded HAM to capture local and global features, and integrated multi-view features through the SE mechanism during the fusion stage, finally outputting the probability of nodule malignancy through a fully connected layer. During training, Focal Loss addressed class imbalance, combined with MixUp and CutMix to enhance sample diversity, achieving end-to-end optimization. The parameter settings were as follows: Focal Loss α_t=0.75, γ=2; Cosine Annealing learning rate with an initial value of 0.001 and T_max=200; MixUp and CutMix were randomly switched with a probability of 0.5, with α values of 0.3 and 1.0, respectively.

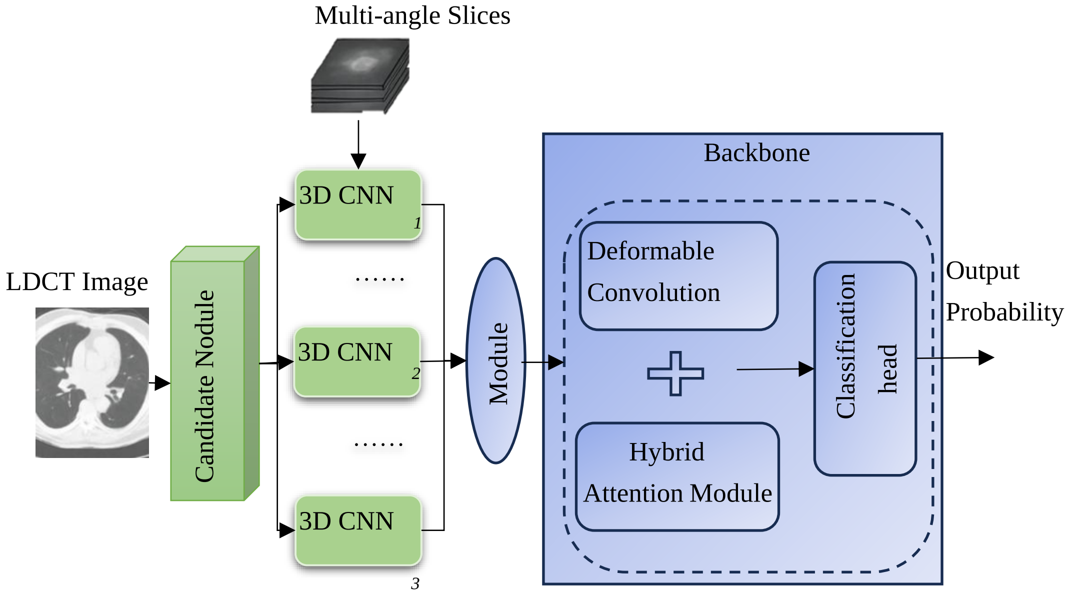

Figure 2: Model training and inference flowchart

4 Experiments and results analysis

To verify the effectiveness of 3D-DHA-MAFNet in lung nodule detection and classification, experiments were conducted on the LIDC-IDRI and LUNA16 datasets and comparisons were performed with 3D-UNet, AttentNet, 3DAGNet, 3D-MCN, 3D-MVSECNN, 3D-ResAttNet, and 3D ViT-CNN. The experiments were carried out on an NVIDIA A100 GPU, using PyTorch and the Adam optimizer, with an initial learning rate of 0.001, a batch size of 8 and 200 training epochs. Data augmentation included random rotation (±15°), translation (±5 pixels), and elastic deformation. The network's Backbone was a 3D-ResNet18 with four residual blocks, using 3×3×3 convolutional kernels with 64/128/256/512 channels, followed by BatchNorm and ReLU after convolution.

4.1 Experimental datasets and evaluation metrics

The effectiveness and robustness of 3D-DHA-MAFNet were verified in this study using two authoritative public lung nodule datasets. The LIDC-IDRI dataset, released by the NLST, contains 1,018 LDCT scans with a total of 2,636 nodules, of which 1,423 are benign and 1,213 are malignant. Each nodule is provided with its center location, boundary contour, and a malignancy likelihood score based on a 5-point scale. In this article, nodules with an average score ≤3 are defined as benign, and those with a score ≥4 are defined as malignant. Nodules with a standard deviation of the score greater than 1.2 (about 8%) are excluded. The LUNA16 dataset is derived from LIDC-IDRI. After removing nodules without annotations or with a diameter of less than 3 mm, it included 888 scans and 1,186 valid nodules, which were used for standard evaluation of automatic detection tasks. The data were divided into training, validation, and test sets at a ratio of 7:2:1 at the patient level to ensure no overlap, thereby guaranteeing the independence and generalizability of the evaluation. All results were based on stratified 5-fold cross-validation and reported as mean ± standard deviation. Hyperparameters were only optimized on the validation set without using nested cross-validation.

The evaluation metrics for the detection task included sensitivity, FP per scan (FP/Scan), and Dice coefficient.

Sensitivity is used to measure the model's ability to identify all true nodules.

$$Sensitivity = \frac{TP}{TP+FN} \tag{19}$$

True positives (TP) refer to the number of true nodules that are correctly detected. FN: the number of true nodules that are missed by the detection model.

FP/Scan indicates the number of false detections per CT scan, reflecting the model's false alarm rate. The calculation method is as follows.

$$FP/Scan = \frac{False\ Positives}{Number\ of\ Scans} \tag{20}$$

The Dice coefficient is a measure of the overlap between the predicted nodule region by the model and the actual nodule annotation.

$$Dice = \frac{2\times|P\cap G|}{|P|+|G|} \tag{21}$$

P and G are the voxel sets of the predicted and true nodules, respectively.

For the classification task, the model is evaluated using accuracy, area under the ROC curve (AUC), sensitivity, specificity, and F1-score.

Accuracy reflects the proportion of overall correct classifications made by the model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{22}$$

TN represents the number of truly benign samples that are correctly identified as benign, while FP represents the number of benign samples that are falsely identified as malignant.

AUC is used to measure the model's ability to distinguish between benign and malignant nodules. The closer the AUC is to 1, the better the classification performance.

Recall rate and specificity are employed to evaluate the model's capability in detecting malignant and benign nodules, respectively. A high recall rate reflects the model's strong performance in identifying malignant nodules and a correspondingly low risk of FN. A high specificity indicates that the model can effectively avoid misdiagnosing benign nodules as malignant, reducing the burden of FP. Recall rate is calculated in the same way as sensitivity, and the equation for calculating specificity is as follows.

$$Specificity = \frac{TN}{TN+FP} \tag{23}$$

The F1-score is a harmonic mean metric used in the evaluation of classification model performance.

$$F1 - score = 2\times\frac{Precision\times Recall}{Precision+Recall} \tag{24}$$

To quantify the statistical significance of performance differences among different models, this study employed the two-tailed paired t-test for hypothesis testing. Specifically, the performance metrics (such as accuracy, sensitivity, AUC) obtained from each fold of the five-fold cross-validation were regarded as paired sample points. The mean differences between 3D-DHA-MAFNet and each baseline model at these sample points were compared. The null hypothesis (H_0) was that there was no significant difference in performance between the two models. If the calculated P-value was less than the significance level α (set at 0.05), the null hypothesis was rejected, and the performance difference was considered statistically significant.

4.2 Performance comparison of the detection module

To verify the effectiveness of the proposed 3D-DHA-MAFNet model in lung nodule detection, comparative experiments were conducted with the classic 3D-UNet, AttentNet, and 3DAGNet, and the results are shown in Figure 3. On the public LUNA16 dataset, the 3D-DHA-MAFNet achieved notably better performance in the lung nodule detection task than existing models.

Compared with advanced models such as 3D-UNet, AttentNet, and 3DAGNet, its overall sensitivity was increased to 98.7%, and the FP rate was controlled at 1.2 per scan, with improvements of 5.6% and 2.3 per scan, respectively, demonstrating higher detection ability and stronger false alarm suppression effect. In the detection of small nodules with a diameter of less than 5 mm, the sensitivity was increased to 95.2%, which was 8.1% higher than that of 3D-UNet (*P*<0.001), indicating the model's advantage in high-difficulty recognition tasks. The performance improvement can be primarily attributed to the incorporation of the DCN and HAM modules, which strengthened the adaptive receptive field construction for capturing nodule shape edges and enhanced the focus on low-contrast regions, respectively. Finally, the Dice coefficient was increased to 94.5%, reflecting the model's strong performance in nodule localization and segmentation accuracy.

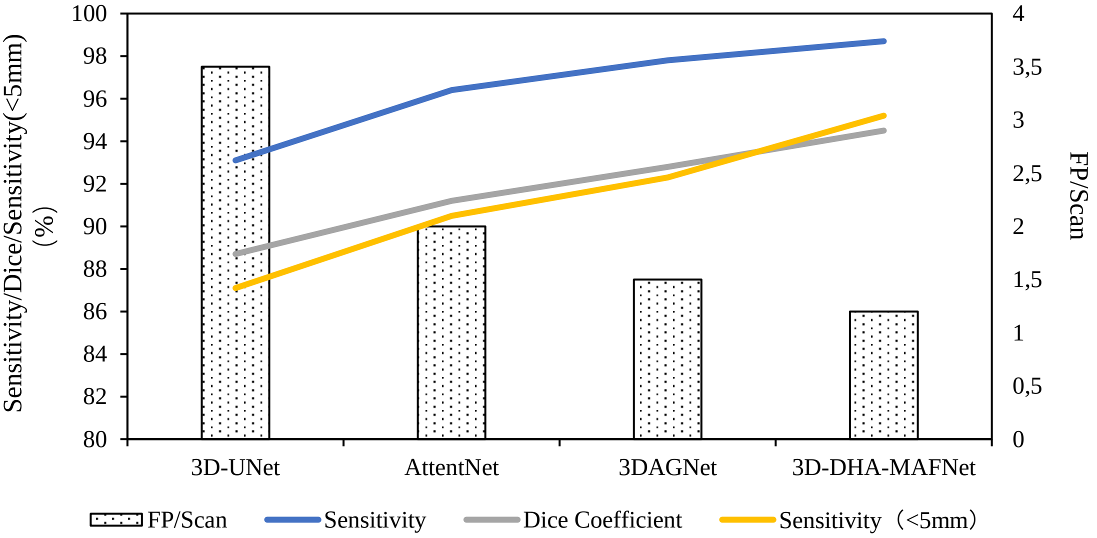

Figure 3: Performance comparison of lung nodule detection (LUNA16 dataset)

4.3 Performance analysis of benign and malignant nodule classification

In the task of lung nodule malignancy classification, 3D-DHA-MAFNet outperformed 3D-MCN and 3D-MVSECNN on the LIDC-IDRI dataset (Figure 4). Under five-fold cross-validation, the average classification accuracy reached 96.8%, with an AUC of 0.983, higher than that of 3D-MVSECNN (accuracy 96.0%, AUC 0.975). The F1 score for malignant nodule identification was 0.971, with sensitivity and specificity of 97.5% and 96.1% respectively, showing significant differences compared with baseline models (*P*<0.01). The performance improvement was mainly due to Focal Loss mitigating class imbalance, which increased the sensitivity for malignant nodules from 93.8% to 97.5%, and the MAFN module dynamically fusing multi-view information to enhance the recognition ability of heterogeneous structural features, thereby increasing the classification accuracy by 3.6% compared with single-view models.

To evaluate the robustness of the model in real-world clinical settings, a multi-type interference test set was constructed, including motion artifacts (σ=1.5), Gaussian noise (variance 0.01), low-dose artifacts, and tiny nodules (diameter <3 mm). On this test set, the classification accuracy of 3D-DHA-MAFNet still reached 95.4%, with a sensitivity of 93.7% for tiny nodules, significantly higher than that of 3D-MVSECNN (92.1%) and 3D-MCN (91.8%), and with an AUC>0.94 in images containing noise and motion artifacts, demonstrating strong robustness, which was attributed to the MixUp and CutMix augmentation strategies.

In summary, 3D-DHA-MAFNet still performed stably in heterogeneous data environments, with strong clinical practicability and potential for promotion.

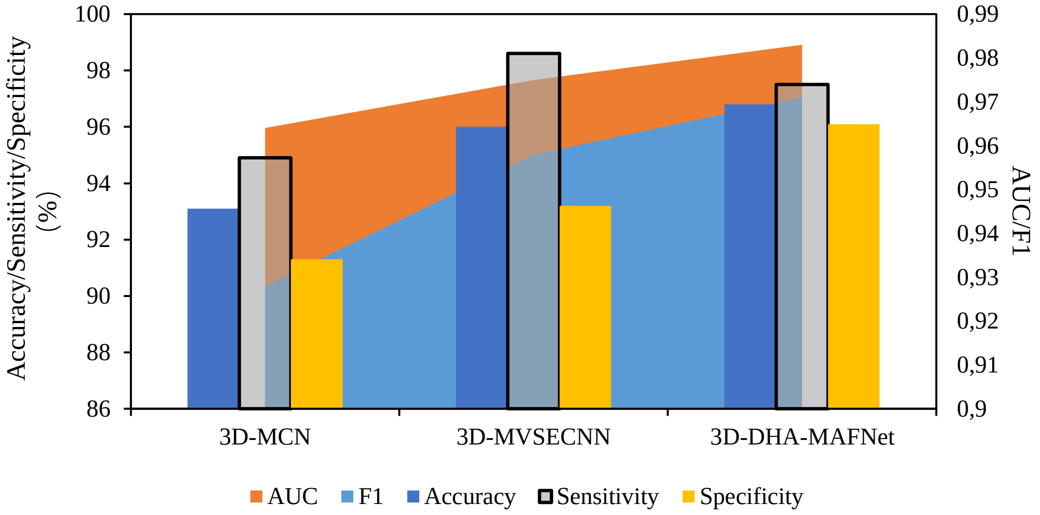

Figure 4: Performance comparison of benign and malignant classification

4.4 Comparison of efficiency and interpretability with the latest 3D CNN/transformer models

To evaluate the practical application value of 3D-DHA-MAFNet, this study compared it with 3D-ResAttNet and 3D ViT-CNN in terms of computational efficiency, model complexity, and interpretability (Figure 5, Figure 6). Experiments on an NVIDIA A100 GPU showed that 3D-DHA-MAFNet had 28.5M parameters, lower than ViT-CNN (45.2M) and slightly higher than 3D-ResAttNet (25.8M); each training epoch took about 12 minutes, and inference was 0.28 seconds per scan, meeting the near real-time clinical needs. In terms of interpretability, the saliency maps generated by the HAM module combined with the multi-view fusion of MAFN can intuitively display the nodule regions the model focused on, and assist doctors in understanding the decision-making process with Grad-CAM. In comparison, 3D-ResAttNet provided local attention but lacked multi-view integration; ViT-CNN captured global information but had difficulty in precisely locating small nodules. In summary, 3D-DHA-MAFNet achieved high performance while maintaining low computational cost and strong clinical interpretability.

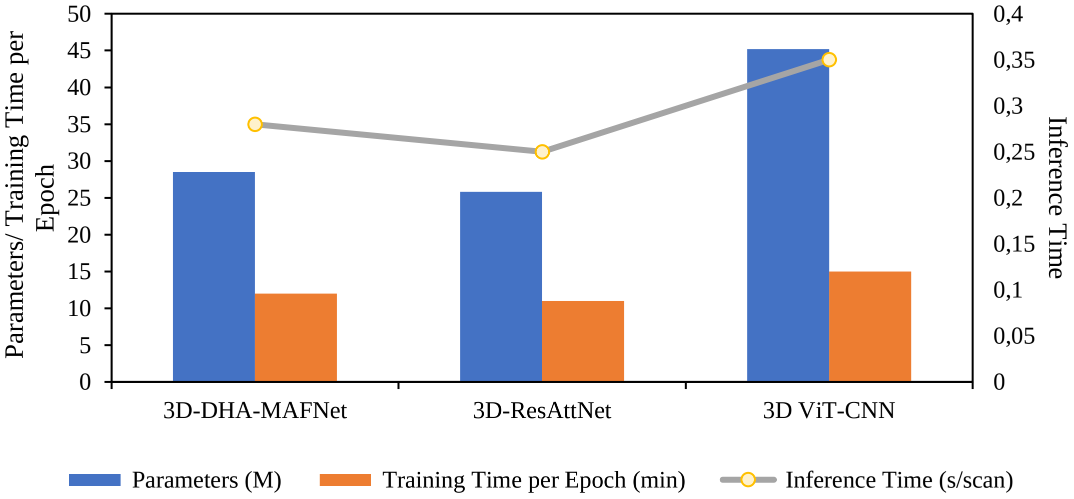

Figure 5: Comparison of computational efficiency and model complexity with the latest 3D CNN/Transformer models

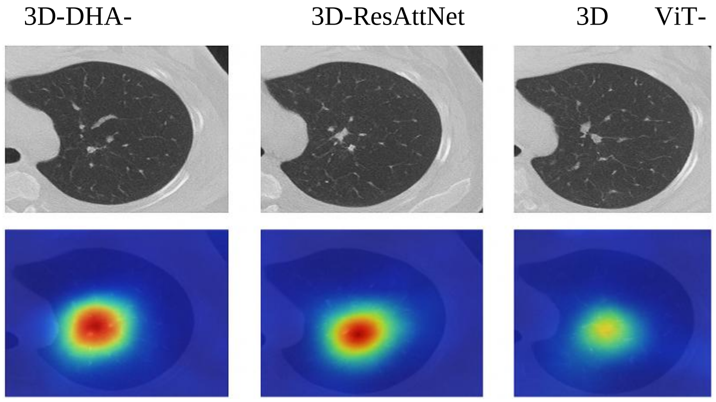

Figure 6: Comparison of Grad-CAM visualization for 3D-DHA-MAFNet, 3D-ResAttNet, and ViT-CNN (Red hotspots: Areas of high attention)

4.5 Ablation study

To verify the contributions of each module to the model's performance, systematic ablation experiments were designed (Figure 7). The results showed that after the DCN module was removed, the sensitivity of small nodule detection decreased by 4.2%, and the Dice coefficient decreased by 3.2%, indicating that DCN is crucial for capturing nodule shape variations. Removing the HAM module resulted in a 3.1% decrease in the Dice coefficient (*P*<0.05) and a significant decline in overall classification performance, indicating its central role in enhancing the discriminability of local features. Moreover, disabling the MAFN reduced the classification accuracy from 96.8% to 94.1%, highlighting the importance of multi-view synergy in representation learning. Even in terms of data augmentation strategies, jointly canceling MixUp and CutMix would cause the F1 value to drop to 0.949, emphasizing the positive impact of multi-strategy integration on model robustness. Overall, there was a clear synergistic gain among the modules, and the combination of the three can substantially enhance the model's perception and classification capabilities for complex and irregular nodules.

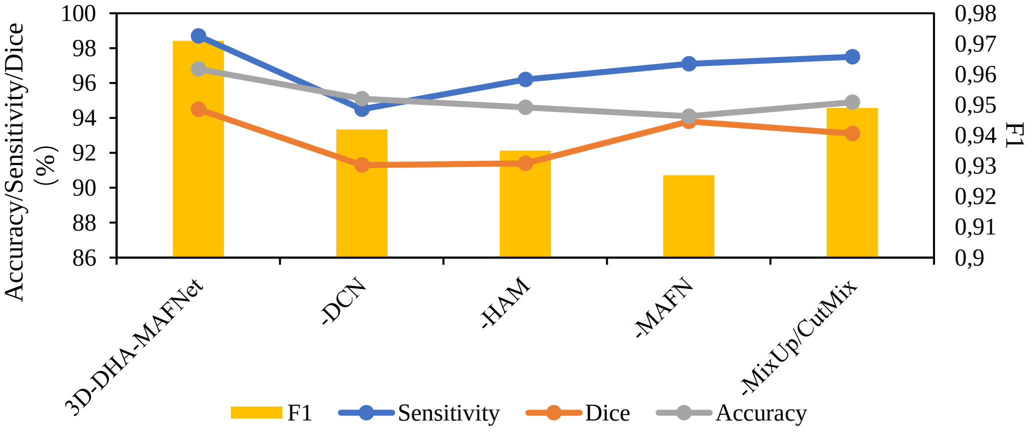

Figure 7: Comparison of module ablation experiment results

4.6 Cross-dataset generalization ability test (internal clinical data)

To verify the generalization ability of 3D-DHA-MAFNet, this study selected a clinical CT dataset of 200 cases confirmed by pathology (92 malignant and 108 benign nodules, with diameters ranging from 3 to 28 mm, scanned by Philips with slice thickness of 1.0-2.5 mm, 120 kVp, 100-250 mAs), which had undergone anonymization. All nodules were annotated by experienced radiologists, with pathological biopsy results serving as the gold standard. The model performed excellently on this dataset (Figure 8, Figure 9), with a sensitivity of 96.2%, accuracy of 94.7%, and AUC of 0.962, significantly outperforming 3D-MVSECNN, 3DAGNet, AttentNet, 3D-UNet, D-ResAttNet, and ViT-CNN (*P*<0.05). Although the performance was slightly lower than that on the LIDC-IDRI test set (due to the influence of scanning standards and artifacts), 3D-DHA-MAFNet had the smallest drop, demonstrating strong cross-source transferability and clinical applicability.

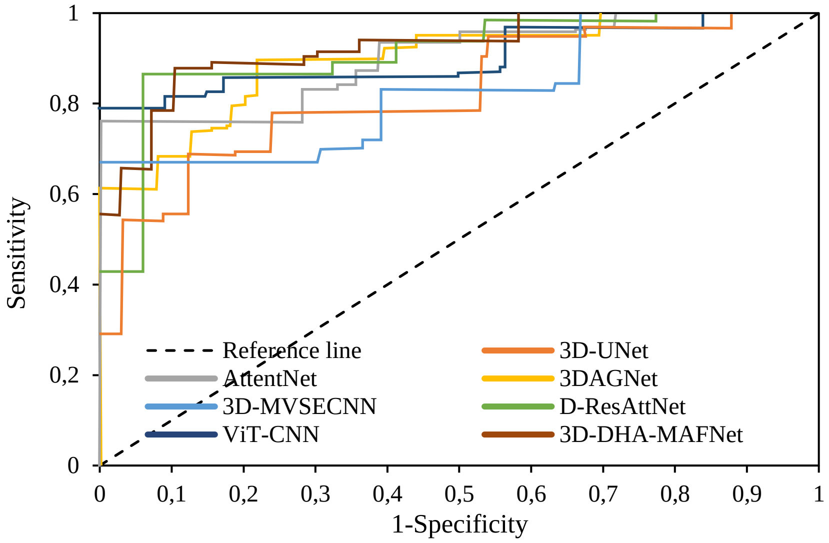

Figure 8: Classification ROC curves of different models

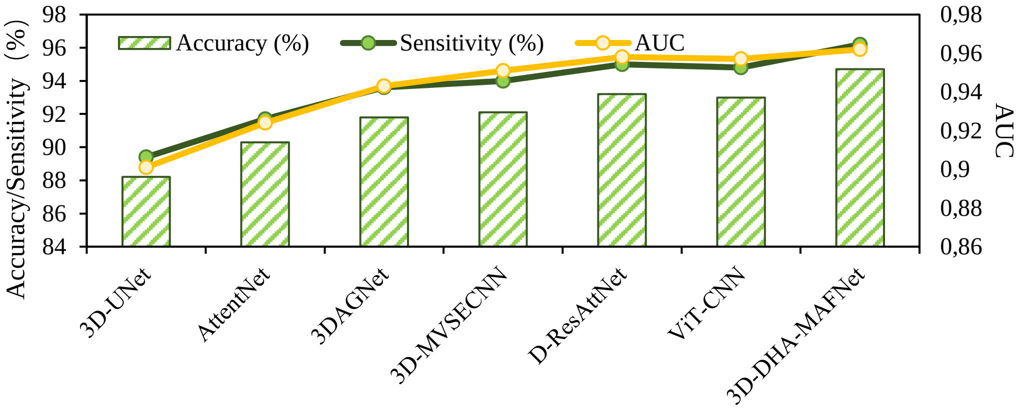

Figure 9: Performance comparison of different models across datasets (internal clinical dataset)

On the internal dataset, the sensitivity of 3D-UNet was only 89.4%, AttentNet and 3DAGNet decreased by about 3-5%, D-ResAttNet was 95.0%, and ViT-CNN had a sensitivity of 94.8% and an AUC of 0.957. In contrast, 3D-DHA-MAFNet, relying on the MAFN and HAM modules, maintained high robustness in the identification of small nodules and in heterogeneous lesion scenarios, showing the best performance and significant potential for clinical application.

4.7 Robustness analysis in heterogeneous data environments

To verify the stability of 3D-DHA-MAFNet in a multi-center clinical setting, this study collected CT data from 150 cases confirmed by pathology from three hospitals (GE, Siemens, Philips) (68 malignant and 82 benign cases, with slice thickness of 1.0-2.5 mm, 120 kVp, 100-250 mAs). On this heterogeneous dataset, the model achieved a classification accuracy of 94.2%, an AUC of 0.958, a sensitivity of 95.6%, and a specificity of 92.9%, all of which were significantly better than the comparison models. For different manufacturers' equipment and scanning protocols, the model's performance remained stable; on samples containing motion artifacts and noise, the accuracy was still >93%, demonstrating good device-independence, protocol-adaptability, and clinical robustness, providing reliable support for practical applications (Figure 10).

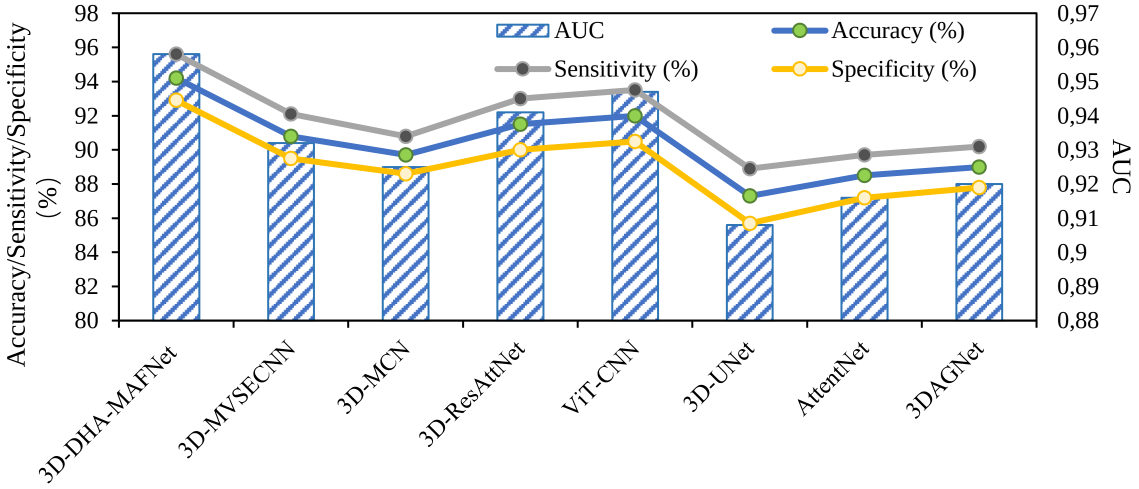

Figure 10: Comparison of benign-malignant classification performance of different models on multi-center heterogeneous clinical datasets

5 Discussion

The 3D-DHA-MAFNet system, integrating DCN, HAM, and MAFN, was proposed in this study. Experiments on public and internal clinical data have shown that the system performed excellently in the tasks of lung nodule detection and benign-malignant classification. Quantitative results indicated that in the LUNA16 detection task, the model had a detection sensitivity of 98.7%, with only 1.2 FP per scan, and a sensitivity of 95.2% for small nodules (<5 mm), significantly reducing the risk of missing early lesions. Its performance surpassed that of the classic 3D-UNet and recent advanced models such as AttentNet [12] and 3DAGNet [16]. In the LIDC-IDRI classification task, the accuracy was 96.8%, the AUC was 0.983, and the sensitivity for malignant nodules was as high as 97.5%, outperforming 3D-MVSECNN [17] and 3D-MCN [18]. Compared with the multi-scale feature pyramid of Li et al. [4] and the 3D multi-scale CNN of Peng et al. [6], this study enhanced feature representation through the synergy of DCN and HAM, and further improved the identification performance of malignant nodules by combining Focal Loss, MixUp, and CutMix strategies.

In terms of qualitative and interpretability aspects, the HAM module combined with Grad-CAM can generate clear attention heatmaps that highlight small and ill-defined nodules. The MAFN module demonstrates the contribution of different views to the decision-making through dynamic weights (αA, αS, αC) [17], enhancing model transparency and providing auxiliary references for clinical use. The performance improvement is mainly attributed to three aspects: the deep integration of DCN and HAM captures irregular shapes and key information; MAFN adaptively fuses multi-view features to enhance adaptability to nodule shape variations; optimized training strategies improve generalization and robustness. In addition, recent studies have shown that multi-layer optimization and multi-modal fusion can further enhance model robustness and cross-center adaptability. For example, RREASO algorithm by Li e al. [23] and the VEnDR-Net model [24] provide referable strategies for multi-model fusion and feature integration.

Despite the encouraging results, this study still has limitations. The scale of the training data is limited, with an insufficient number of pathologically confirmed malignant nodules (n=92), and the performance on non-public clinical data has declined (accuracy dropped from 96.8% to 94.7%), reflecting the challenge of generalization posed by different medical center equipment and scanning protocols. Although the introduction of DCN, HAM, and MAFN modules has improved performance, it has increased the number of parameters and computational complexity. Although the inference speed is 0.28 seconds per scan, optimization is still needed in resource-constrained environments. In the future, larger-scale, multi-institutional, and class-balanced datasets should be used to enhance generalizability. All clinical data were strictly anonymized, and the research protocol was approved by the IRB. The clinical application of AI-assisted diagnosis needs continuous attention to transparency, fairness, and validation in diverse populations to ensure safety and reliability.

From the perspective of translating research into clinical practice, 3D-DHA-MAFNet still needs to address challenges in many aspects, including multi-center validation to assess robustness, integration with existing workflows (such as PACS systems), regulatory approval, and standardized evaluation, enhancing clinically understandable interpretability, and data privacy and security protection. In the future, adaptive and robust control methods in complex dynamic systems, such as adaptive fuzzy control, output-feedback synchronization, and robust neural adaptive control [25-29], can be used for reference to dynamically adjust model parameters in the presence of noise interference, cross-center differences, or insufficient data, thereby enhancing the stability, generalizability, and clinical credibility of CAD systems.

6 Conclusion

To address the issues of insufficient feature representation of small nodules, limited multi-view fusion, and poor model generalization ability, this study proposed the 3D-DHA-MAFNet system, which integrated DCN, HAM, and MAFN modules to achieve efficient perception of irregular nodules, focusing on key areas, and dynamic fusion of multi-view information. On the LUNA16 and LIDC-IDRI datasets, the detection sensitivity reached 98.7%, with an FP rate of 1.2 per scan, showing outstanding performance for small nodules with a diameter of less than 5 mm. In benign-malignant classification, combined with Focal Loss and MixUp/CutMix, the model outperformed existing methods in terms of accuracy, AUC, and F1 value, and effectively alleviated the imbalance of malignant samples. Ablation experiments verified the synergistic gains of each module, and independent clinical data validation showed that the model still maintained high accuracy and stability across datasets. This system provides a high-precision, scalable, and clinically translatable CAD solution for lung nodules, which is conducive to the early screening and accurate diagnosis of lung cancer.

Funding statement

This study was supported by the influence mechanism of TRPM5 ion channel on bitter taste expression and screening of taste-masking drugs (242102310456).

Data availability statement

All data generated or analysed during this study are included in this article.

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 2024, 74(3):229-263. https://doi.org/10.3322/caac.21492

[2] Ramaswamy A. Lung cancer screening: Review and 2021 update. *Curr Pulmonol Rep*, 2022, 11(1):15-28. https://doi.org/10.1007/s13665-021-00283-1

[3] Lu Z, Long F, He X. Classification and segmentation algorithm in benign and malignant pulmonary nodules under different CT reconstruction. *Comput Math Methods Med*, 2022, 2022:3490463. https://doi.org/10.1155/2022/3490463

[4] Li Y, Hui L, Wang X, Zou L, Chua S. Lung nodule detection using a multi-scale convolutional neural network and global channel spatial attention mechanisms. *Sci Rep*, 2025, 15(1):12313. https://doi.org/10.1038/s41598-025-97187-w

[5] Hammad M, ElAffendi M, El-Latif AAA, Ateya AA, Ali G, Plawiak P. Explainable AI for lung cancer detection via a custom CNN on CT images. *Sci Rep*, 2025, 15(1):12707. https://doi.org/10.1038/s41598-025-97645-5

[6] Peng H, Sun H, Guo Y. 3D multi-scale deep convolutional neural networks for pulmonary nodule detection. *PLoS One*, 2021, 16(1):e0244406. https://doi.org/10.1371/journal.pone.0244406

[7] Pan J, He Z, Li Y, Zeng W, Guo Y, Jia L, Jiang H, Chen W, Lu Y. Atypical architectural distortion detection in digital breast tomosynthesis: a multi-view computer-aided detection model with ipsilateral learning. *Phys Med Biol*, 2023, 68(23). https://doi.org/10.1088/1361-6560/ad092b

[8] Tong C, Liang B, Zhang M, Chen R, Sangaiah AK, Zheng Z, Wan T, Yue C, Yang X. Pulmonary nodule detection based on ISODATA-improved faster RCNN and 3D-CNN with Focal loss. *Acm T Multim Comput (TOMM)*, 2020. https://doi.org/10.1145/3365445

[9] Li Y, Li P, Wang H, Gong X, Fang Z. CAML-PSPNet: A medical image segmentation network based on coordinate attention and a mixed loss function. *Sensors (Basel)*, 2025, 25(4):1117. https://doi.org/10.3390/s25041117

[10] Tan M, Wu F, Yang B, Ma J, Kong D, Chen Z, Long D. Pulmonary nodule detection using hybrid two-stage 3D CNNs. *Med Phys*, 2020, 47(8):3376-3388. https://doi.org/10.1002/mp.14161

[11] Song L, Zhang M, Wu L. Detection of low-dose computed tomography pulmonary nodules based on 3D CNN-CapsNet. *Electronics Letters*, 2023, 59(18). https://doi.org/10.1049/ell2.12952

[12] Almahasneh M, Xie X, Paiement A. AttentNet: Fully convolutional 3D attention for lung nodule detection. *Sn Comput. Sci.* 2025, 6, 292. https://doi.org/10.1007/s42979-025-03799-4

[13] Xiao Z, Liu B, Geng L, Zhang F, Liu Y. Segmentation of lung nodules using improved 3D-UNet neural network. *Symmetry*, 2020, 12(11):1787. https://doi.org/10.3390/sym12111787

[14] Xu X, Du L, Yin D. Dual-branch feature fusion S3D V-Net network for lung nodules segmentation. *J Appl Clin Med Phys*, 2024, 25(6):e14331. https://doi.org/10.1002/acm2.14331

[15] Liu JJ, Li YQ, Li WT, Li ZS, Lan YH. Multiscale lung nodule segmentation based on 3D coordinate attention and edge enhancement. *Electronic Research Archive*, 2024, 32(5): 3016-3037. https://doi.org/10.3934/era.2024138

[16] Jian M, Zhang L, Jin H, Li X. 3DAGNet: 3D deep attention and global search network for pulmonary nodule detection. *Electronics*, 2023, 12(10):14. https://doi.org/10.3390/electronics12102333

[17] Yang Y, Li X, Fu J, Han Z, Gao B. 3D multi-view squeeze-and-excitation convolutional neural network for lung nodule classification. *Med Phys*, 2023, 50(3):1905-1916. https://doi.org/10.1002/mp.16221

[18] Afshar P, Oikonomou A, Naderkhani F, Tyrrell PN, Plataniotis KN, Farahani K, Mohammadi A. 3D-MCN: A 3D multi-scale capsule network for lung nodule malignancy prediction. *Sci Rep*, 2020, 10(1):7948. https://doi.org/10.1038/s41598-020-

64824-5

[19] Huang YS, Wang TC, Huang SZ, Zhang J, Chen HM, Chang YC, Chang RF. An improved 3-D attention CNN with hybrid loss and feature fusion for pulmonary nodule classification. *Comput Methods Programs Biomed*, 2023, 229:107278. https://doi.org/10.1016/j.cmpb.2022.107278

[20] Prithvika P C S, Anbarasi L J. V-3DResNets: a 3D convolutional neural network based on residual network variants and slice grouping for pulmonary nodule detection. *Multimedia Tools & Applications*, 2024, 83(31). https://doi.org/10.1007/s11042-024-18485-5

[21] Hu J, Shen L, Sun G, Albanie S, Wu E. Squeeze-and-excitation networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, 32: 7132-7141. https://doi.org/10.1109/TPAMI.2019.2913372

[22] Zhu X, Wu Y, Hu H, Zhuang X, Yao J, Ou D, Li W, Song M, Feng N, Xu D. Medical lesion segmentation by combining multimodal images with modality weighted UNet. *Med Phys*, 2022, 49(6):3692-3704. https://doi.org/10.1002/mp.15610

[23] Li X, Zhang Y, Wang H. RREASO building structure physical parameter identification algorithm for structural damage identification. *Informatica*, 2024, 48(2):123-138. https://doi.org/10.31449/inf.v48i18.6156

[24] Mishra A, Gupta P, Singh R. VEnDR-Net: voting ensemble classifier for automated diabetic retinopathy detection. *Informatica*, 2024, 48(3):201-218. https://doi.org/10.31449/inf.v49i32.8899

[25] Hu K, Ling Y, Liu J. Image recognition technology for bituminous concrete reservoir panel cracks based on deep learning. *PLoS One*, 2025, 20(2):e0318550. https://doi.org/10.1371/journal.pone.0318550

[26] Bai Z, Li S, Liu H, Zhang X. Adaptive fuzzy backstepping control of fractional-order chaotic system synchronization using event-triggered mechanism and disturbance observer. *Fractal and Fractional*, 2022, 6(12):714. https://doi.org/10.3390/fractalfract6120714

[27] Sun Y, Liu Y, Liu L. Fixed-time adaptive synchronization of fractional-order memristive fuzzy neural networks with time-varying leakage and transmission delays. *Fractal and Fractional*. 2025, 9(4):241. https://doi.org/10.3390/fractalfract9040241

[28] Zouari F, Saad K B, Benrejeb M. Robust neural adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems. *International Review on Modelling & Simulations*, 2012, 5(5):2075-2103. https://doi.org/10.1016/j.neucom.2012.09.036

[29] Yang C, Zhang X. Adaptive fuzzy control for fractional-order networked control systems with input time delay and data loss. *Journal of Nonlinear Mathematical Physics*, 2024, 31(1). https://doi.org/10.1007/s44198-024-00201-4