Journal of Universal Computer Science, vol. 31, no. 13 (2025), 1491-1512
submitted: 24/12/2024, accepted: 11/6/2025, appeared: 28/11/2025 *CC BY 4.0*

AMCF-CNN: Attention-Guided Multi-Scale Cross Fusion for Reducing False Positives in Lung Nodule Detection

Yongbin Li
(University Malaysia Sarawak, Sarawak, Malaysia
https://orcid.org/0000-0003-3163-8448, bynn456@126.com)

Xinyue Yang
(Zunyi Medical University, Zunyi, China
https://orcid.org/0009-0006-5414-6656, 2570431405@qq.com)

Linhu Hui
(Zunyi Medical University, Zunyi, China
https://orcid.org/0000-0001-9067-8197, 394015820@qq.com)

Enlin Fu
(Zunyi Medical University, Zunyi, China
https://orcid.org/0009-0007-7661-4948, fuenlin2000@126.com)

Stephanie Chua
(University Malaysia Sarawak, Sarawak, Malaysia
https://orcid.org/0000-0001-6550-6828, chlstephanie@unimas.my)

Abstract: False-positive reduction is a critical step in the automatic lung nodule detection system, playing a significant role in the early detection and diagnosis of lung cancer. However, accurately distinguishing between nodules and non-nodules remains a challenge due to their high morphological similarity. To effectively reduce the false positives (FPs) in automated nodule detection using computed tomography (CT), we propose an attention-guided multi-scale cross-fusion three-dimensional (3D) convolutional neural network (AMCF-CNN). The proposed network adopts an innovative multi-scale cross-fusion strategy to integrate nodule features across different scales and incorporates a SimAM-Res module to adaptively enhance key features, thereby improving overall feature representation. To further enhance global contextual awareness, a Global Modeling Module (GMM) based on self-attention mechanism is introduced, enabling complementary fusion of local structural details and global semantic context, thereby enhancing structural discrimination in complex imaging backgrounds. We conducted experiments on the LUNA16 dataset consisting of 888 chest CT cases, and evaluated model performance using 10-fold cross-validation. Experimental results demonstrate that the proposed model achieves sensitivities of 0.983 and 0.984 at 4 and 8 FPs/scan, respectively, with a balanced accuracy of 0.983 and a competitive performance metric (CPM) score of 0.936. Compared to existing methods, AMCF-CNN achieves competitive performance in false-positive reduction, offering a more accurate and robust solution for lung nodule detection and demonstrating strong potential for real-world clinical applications.

Keywords: False-positive reduction; Multi-scale feature fusion; Attention mechanism; SimAM-Res; Global modeling; 3D convolutional neural network
Categories: J.3
DOI: 10.3897/jucs.145223

1 Introduction

Lung cancer is one of the cancers with the highest incidence rate and mortality in the world [Siegel et al. 2020]. In its early stages, lung cancer often presents no obvious symptoms, causing many patients to miss the optimal window for treatment. Therefore, early diagnosis and treatment are crucial for improving patient survival rates [Camarlinghi et al. 2013]. A typical feature of early-stage lung cancer is the presence of lung nodules, defined as round or irregularly shaped lesions in the lung with a diameter of 3mm or larger [Gould et al. 2007]. With the advancement of medical imaging technology, computed tomography (CT) scanning has become the primary method for detecting lung nodules. However, CT scans generate a large volume of image data, and lung nodules often occupy only a few pixels, making manual detection challenging. To address this issue, researchers had developed computer-aided techniques to assist radiologists in the automatic detection of lung nodules, which not only improved the accuracy of early lung cancer diagnosis but also significantly reduced the rates of missed and incorrect diagnoses [Firmino et al. 2014].

Currently, computer-aided detection (CAD) systems for lung nodule detection generally adopted a two-stage approach: (1) candidate nodule detection, and (2) false-positive (FP) reduction for the detected candidates [El-Bana et al. 2020] [Ding et al. 2017]. In the candidate detection stage, the system aimed to identify as many candidate nodules as possible to ensure high sensitivity, without considering specificity, which allowed for a high false-positive rate [Setio et al. 2016]. In the false-positive reduction stage, the system classified the detected candidates into true nodules and false positives, thereby reducing the false-positive rate. This stage was the most critical part of CAD systems for lung nodule detection [Sluimer et al. 2006], but it also presented significant challenges. The main challenges arised from the variability in size, shape, and structure of nodules, as well as frequent adhesion to adjacent organs like the trachea and blood vessels. This resulted in nodule morphological features that were highly similar to other structures in the thoracic cavity, such as lymph nodes, thus increasing the difficulty of false-positive reduction. Consequently, many researchers had focused on developing robust and efficient algorithms to differentiate nodules from non-nodules. Figure 1 shows examples of nodules and non-nodules.

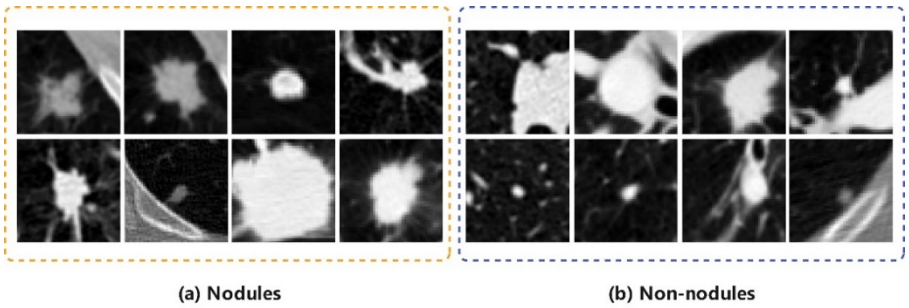

Figure 1: Examples of nodules and non-nodules in the LUNA16 dataset

In recent years, convolutional neural networks (CNNs) have been widely applied in reducing false positives in lung nodule detection [Setio et al. 2016] [Zhao et al. 2022] [Cao et al. 2019] [Dou et al. 2016]. For example, Ding et al. [Ding et al. 2017] used 3D nodule patches as inputs for a 3D CNN, significantly reducing the false-positive rate. Vipparla et al. [Vipparla et al. 2021] proposed an attention-based multi-patch 3D CNN model with a hybrid fusion architecture, achieving a CPM of 0.931 on the LUNA16 dataset. Although existing CNN-based models have made substantial progress in false-positive reduction, they still faced several challenging limitations. First, conventional CNNs have limitations in modeling long-range dependencies. Due to their local receptive field limitations [Vaswani et al. 2017], CNNs inherently operate as local window-based modeling mechanisms, primarily focusing on fine-grained features such as edges, textures, and intensity variations. However, in complex imaging backgrounds, nodules often coexist with anatomically similar structures such as blood vessels and airways. Relying solely on these local features makes it difficult to accurately differentiate between tissue types, which can easily lead to misclassification. Second, lung nodules exhibit diverse sizes and shapes, ranging from tiny to large nodules. Existing multi-scale feature fusion methods often rely on direct concatenation strategies, which may neglect fine-grained interactions between features and potentially introduce information redundancy or conflicts [Kim et al. 2019]. These issues constrain the model's ability to discriminate complex candidate regions from multiple perspectives and have become one of the key challenges in current false-positive reduction tasks.

To overcome these limitations, we propose an attention-guided multi-scale cross-fusion 3D convolutional neural network (AMCF-CNN), designed from the perspectives of global semantic modeling and cross-scale feature integration. To tackle the issue of insufficient modeling of long-range dependencies, we design a Global Modeling Module (GMM), which employs a parallel structure combining convolution and self-attention mechanisms [Vaswani et al. 2017], enabling the establishment of global spatial relationships while retaining local texture features. At the same time, to overcome the limitations in multi-scale feature interaction, we adopt a multi-scale cross-fusion (MCF) strategy to progressively integrate global and local information from different scales. This strategy is inspired by the SparseInst architecture [Cheng et al. 2022], ensuring sufficient information flow and interaction between different paths during fusion, avoiding information redundancy and conflicts often caused by traditional direct concatenation. Finally, to further enhance the fine-grained representation capability in multi-scale fusion, a SimAM residual (SimAM-Res) module is introduced at various scales. The SimAM-Res module adaptively adjusts feature maps using the simple attention module (SimAM) [Yang et al. 2021], dynamically computing attention weights based on local variations to highlight key features while suppressing irrelevant ones.

The main contributions are as follows:

(1) A staged MCF strategy that performed a bidirectional cross fusion of multi-scale features was proposed to enhance the model's adaptability to lung nodules of varying sizes. Additionally, the SimAM-Res module was designed to adaptively adjust the local responses in feature maps, improving the efficiency of cross-scale feature fusion.

(2) The GMM module was designed to enhance global modelling capabilities. By employing a parallel structure of convolution and self-attention mechanisms, the GMM

effectively captured long-distance dependencies in the feature maps, modelling the relationships between nodules and the background from a global perspective.

(3) The proposed model was evaluated on the LUNA16 dataset, demonstrating its high sensitivity and accuracy, outperforming most existing methods and validating the effectiveness in false-positive reduction.

The structure of the paper is organized as follows: Section 2 reviews related work on false-positive reduction; Section 3 describes the proposed method in detail; Section 4 presents the dataset, preprocessing steps, training details, and evaluation metrics; Section 5 analyzes and discusses the experimental results; and Section 6 summarizes this paper and proposes future research directions.

2 Related work

In the early studies on false-positive reduction in lung nodule detection, researchers primarily relied on handcrafted features combined with traditional machine learning algorithms for classification. For example, Jacobs et al. [Jacobs et al. 2014] extracted 128 descriptive features related to shape, brightness, texture, and contextual information, and applied Linear Discriminant Analysis (LDA) for classification. Setio et al. [Setio et al. 2015] computed a set of 24 intensity-, shape-, and spatial-based features and used a Support Vector Machine (SVM) classifier, while Liu et al. [Liu et al. 2016] combined 22 texture and shape features with a Random Forest (RF) model, achieving a sensitivity of 0.924 at 4.5 FPs/scan. Although these methods laid the foundation for early research, they heavily depended on expert-crafted features, which limited their generalizability—especially when dealing with complex nodule morphologies and heterogeneous backgrounds.

With the advancement of deep learning, CNNs have gradually become the mainstream approach for false-positive reduction in lung nodule detection tasks [Setio et al. 2016] [Zhao et al. 2022] [Haiying et al. 2021]. Representative works include Setio et al. [Setio et al. 2016], who constructed a CNN model using multi-view image patches for candidate screening, and Zhao et al. [Zhao et al. 2022], who integrated spatial-temporal and frequency-domain information from three orthogonal planes to enhance fine-grained representation, achieving a sensitivity of 0.952 and specificity of 0.981. While these methods improved the model's ability to recognize structural patterns, they did not fully exploit the synergy between features at different scales, thereby limiting their effectiveness in comprehensively discriminating complex structures.

To address this limitation, researchers have introduced multi-scale or multi-path architectures to enhance the model's ability to capture complex morphologies and diverse receptive fields [Cao et al. 2019] [Dou et al. 2016] [Yuan et al. 2021] [Kim et al. 2019] [Zuo et al. 2020] [Xiao et al. 2019]. For example, Cao et al. [Cao et al. 2019] designed a multi-branch 3D CNN structure to extract volumetric features, while Dou et al. [Dou et al. 2016] constructed parallel pathways based on multi-scale blocks to obtain multi-level contextual information. Yuan et al. [Yuan et al. 2021] employed a multi-path architecture with varying receptive fields to integrate spatial, morphological, and contextual features, achieving a CPM score of 0.881 and a sensitivity of 0.952 at 4 FPs/scan on the LUNA16 dataset. However, most of these methods rely on direct concatenation for feature fusion, which fails to effectively capture fine-grained semantic dependencies across scales and may introduce redundancy or conflict.

Building on this, Kim et al. [Kim et al. 2019] proposed a progressive integration strategy that performs stage-wise interactive fusion of multi-scale features, effectively mitigating the semantic fragmentation caused by naive concatenation and achieving a CPM score of 0.929, indicating more stable feature integration. This trend highlights that achieving semantic consistency and deep interaction across multi-scale features has become a critical direction for improving false-positive discrimination performance.

To further enhance the model's discriminative capability in complex structures, recent studies have introduced attention mechanisms to guide the network in focusing on critical regions [Vipparla et al. 2021] [Lai et al. 2021] [Gu et al. 2022] [Sun et al. 2021] [Hao et al. 2023]. Attention mechanisms enable the model to concentrate on salient areas while suppressing redundant features, thereby improving the discriminability of selected representations. For example, Lai et al. [Lai et al. 2021] incorporated the Squeeze-and-Excitation (SE) attention block [Hu et al. 2018] into CNNs to enhance the expression of prominent feature maps through channel-wise attention. Gu et al. [Gu et al. 2022] employed a combination of cross-attention and weighted fusion strategies to improve the precision of multi-scale feature interaction. Sun et al. [Sun et al. 2021] proposed an attention-embedded complementary-stream CNN that integrated the CBAM attention module [Woo et al. 2018] to jointly model spatial and channel attention, achieving a sensitivity of 0.920 at 4 FPs/scan on the LUNA16 dataset. Although these methods enhanced feature selection, most of them focused primarily on local feature optimization. In recent years, self-attention mechanisms [Vaswani et al. 2017] have drawn increasing attention due to their strong global modeling capabilities. By establishing direct connections between any spatial positions, self-attention captures long-range semantic dependencies and enables deep fusion across scales and regions, providing a new path for improving structural discrimination in complex backgrounds.

3 Methods

This section introduces the proposed AMCF-CNN framework, with its network architecture shown in Figure 2. The framework consisted of four main components: multi-scale feature extraction, multi-scale cross-fusion, a global modelling module, and a classification module.

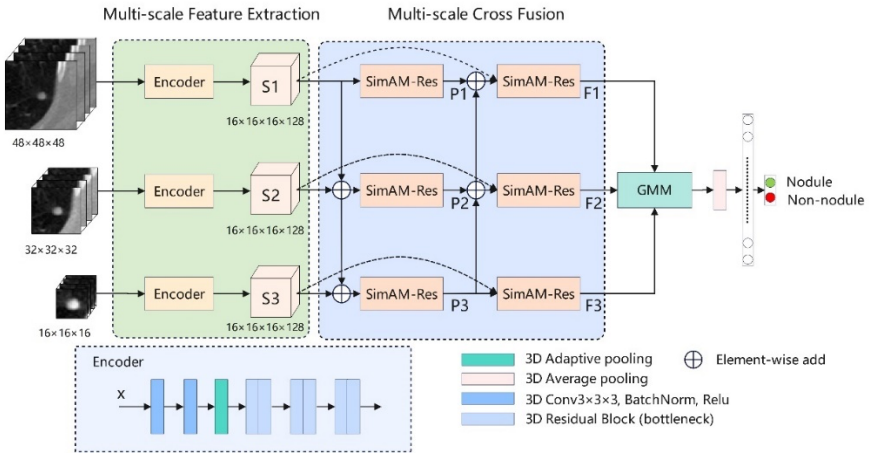

Figure 2: The AMCF-CNN framework for reducing false positives

First, the model extracted multi-scale features from CT images of different resolutions through a multi-scale feature extraction module to capture features at various levels of detail and global information. In the multi-scale cross-fusion stage, a phased fusion strategy was adopted to progressively integrate features from different scales, achieving sufficient interaction and information sharing. In each fusion step, the SimAM-Res module was introduced to enhance the representation of key features.

To further enhance the ability for global context modelling, the GMM module was integrated, employing a parallel structure of convolution and self-attention mechanisms to capture long-distance dependencies within the feature maps. The fused feature maps underwent adaptive pooling for dimensionality reduction, followed by a fully connected layer to classify the input as either a nodule or non-nodule.

3.1 Multi-scale Feature Extraction

In the task of false positive reduction, a single-scale input often struggled to effectively capture the varying characteristics of nodules across different sizes and shapes. To address this, a multi-scale input module that extracts multi-level features from input blocks of different sizes was designed. Larger blocks were used to capture broader contextual information, such as the interactions between nodules and complex structures like blood vessels and tracheas, while smaller blocks focus on detail-rich local information, such as texture and edge features, thus preventing the loss of important feature details.

Specifically, this study proposed the design of three different feature extraction paths, using image blocks of 48×48×48, 32×32×32, and 16×16×16 as inputs to the model. In each path, two consecutive 3×3×3 convolutional layers were used to perform preliminary feature extraction on the input image blocks, capturing low-level spatial features and enhancing local information representation. Subsequently, adaptive pooling layers were applied to normalize the feature maps to a consistent size of 16×16×16, ensuring uniformity across different scales and preparing the features for subsequent fusion. After the feature maps were unified, six 3D bottleneck residual

blocks [Tajbakhsh et al. 2016] were employed to further extract high-level features. The bottleneck residual blocks alleviated the problem of gradient vanishing through residual connections, making the training of deep networks more stable. Finally, three feature maps with a unified size of 16×16×16 were obtained through the feature encoder, denoted as S1, S2, and S3, respectively.

3.2 SimAM-Res Mudule

The structure of the SimAM-Res module is shown in Figure 3. It combined the design of residual blocks with the SimAM mechanism, similar to SE-ResNet [Hu et al. 2018]. The residual block used a combination of 1×1×1, 3×3×3, and 1×1×1 convolutions to compress and expand the channel dimensions while effectively extracting multi-scale features. This structure not only reduced computational complexity but also retained rich features. Each convolutional layer was followed by Batch Normalization and a ReLU activation function to enhance the model's nonlinear representation capability and improve training stability.

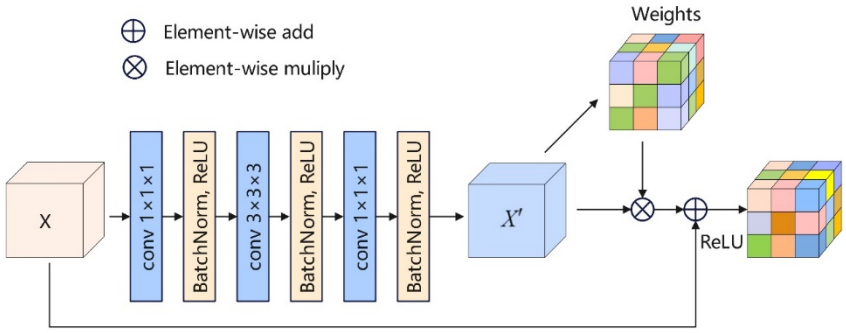

Figure 3: The structure of SimAM-Res Module

After the convolution operation, SimAM adaptively adjusted the feature map. SimAM [Yang et al. 2021] was a lightweight attention mechanism inspired by neuroscience, designed to capture salient information in features without requiring additional convolutions or parameterized operations. Its purpose was to perceive local variations in the feature map and assign adaptive attention weights to each position, thereby emphasizing important features and suppressing irrelevant information. This approach achieved a balance between fine details and global information, enhancing the effectiveness of feature fusion.

The specific calculation process is as follows: First, the local variance of the input feature X' is calculated, with a smoothing term λ introduced to prevent division by zero errors. The resulting local variation is shown in Eq. (1):

$$e = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 + \lambda \tag{1}$$

Where N is the total number of elements in the feature map and μ is the feature mean. Next, the attention weight y_i of each position is calculated by measuring the deviation of each feature value from the local mean, as defined in Eq. (2):

$$y_i = \frac{(x_i-\mu)^2}{4e} + 0.5 \tag{2}$$

Then, the attention weight is mapped to [0, 1] using the Sigmoid function, as shown in Eq. (3):

$$weight = \sigma(y) \tag{3}$$

Finally, the attention weights are element-wise multiplied with the feature map to obtain the weighted output features, as shown in Eq. (4):

$$X'_{out} = X' \cdot weight \tag{4}$$

This adaptive feature enhancement approach in SimAM could automatically highlight features that were significantly relevant to the classification task while reducing the interference from redundant information. In the last step of the SimAM-Res module, a residual connection was performed by element-wise adding the original input features to the SimAM-adjusted features, followed by a ReLU activation function to further enhance the non-linear representation, ultimately generating the output feature map.

3.3 Multi-scale Cross Fusion

In the feature fusion stage, the MCF strategy to perform cross-fusion of feature maps from different resolutions was designed, aiming to effectively integrate local and global information. Drawing inspiration from the multi-level feature transmission and fusion concept in the SparseInst architecture [Cheng et al. 2022], MCF adopted a bidirectional fusion strategy to facilitate the flow of multi-scale information across different paths, ensuring thorough transmission and preservation of feature information.

First, the feature maps S1, S2, and S3 from three different scales were fused in a top-down manner, step by step. The larger-scale feature maps mainly provided global background information, helping to capture the macroscopic structure of the lungs, while the smaller-scale feature maps focused on local details, allowing a more refined depiction of the diversity and intricacies of lung nodules. This fusion approach enabled the model to capture both local details and global information simultaneously, enhancing its adaptability to lung nodules of various sizes and shapes. The fused features were then processed by the SimAM-Res module to obtain feature maps P1, P2, and P3. In each fusion step, the SimAM-Res module adaptively adjusted the features, allowing multi-scale information to flow and interact efficiently across different levels, thereby achieving more effective cross-scale fusion. Subsequently, the MCF module performed a bottom-up fusion, sequentially merging P3 with P2, and then fusing with P1, to further enhance the transmission of information.

To optimize the fused features, the merged feature stream underwent further processing through residual connections and the SimAM-Res module. The residual connections were inspired by the design principles of bi-directional feature pyramid network (BiFPN) [Liu et al. 2018], enabling cross-layer information transmission and preventing degradation of information during deep-layer propagation. Specifically, the initially fused features S1 were concatenated with the processed features from P1, forming a new feature representation that was then fed into a second SimAM-Res module. This design retained the original feature information while adaptively adjusting the fused features through the SimAM-Res module, thereby further enhancing the model's ability to represent multi-scale features.

This multi-level, bidirectional cross-fusion strategy not only established a closer connection between high-resolution and low-resolution features but also ensured the diversity and expressive capability of the fused features through the combination of

residual connections and the SimAM-Res module. This approach provided a more robust feature representation for lung nodule detection.

3.4 Global Modeling Module

In the task of false-positive reduction, relying solely on local features was often insufficient for accurately distinguishing lung nodules from other similar structures. To enhance global context awareness, a GMM module that used convolution and self-attention mechanisms [Vaswani et al. 2017] in parallel was designed to capture long-distance dependencies between features, thereby improving the feature expression capability and classification performance. The structure of GMM is shown in Figure 4.

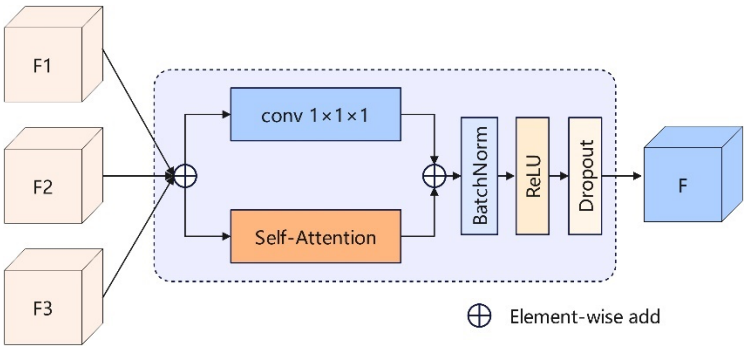

Figure 4: The structure of Global Modeling Module

The GMM module received feature inputs from multi-scale cross-fusion and passed these features to two parallel paths: one path used a 1x1x1 convolution operation to extract local features without changing the spatial dimensions, thereby enhancing the expression of local information; The other path used a self-attention mechanism to model the global information associations. The self-attention mechanism generated attention weights by computing the similarity of each position in the feature map, thereby strengthening the feature response of important positions. Specifically, each position of the feature map was considered as a query, while the other positions were treated as keys and values. The similarity between the Query and Key is computed via dot product, followed by Softmax normalization to generate the attention weight matrix. This matrix is then multiplied by the Value to obtain the globally modeled feature representation, as shown in Eq. (5):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

Here, Q, K, and V represent the Query, Key, and Value matrices, respectively, and d_k is the dimension of the Key matrix.

The GMM module added the output features from both the convolution and self-attention paths and then performed Batch Normalization, ReLU activation. Additionally, a Dropout layer was incorporated to prevent overfitting. This design enabled the GMM module to maintain efficient local feature extraction while significantly enhancing the capture of global information, thereby improving the performance of lung nodule classification in complex backgrounds.

4 Experiment

4.1 Dataset

The public dataset used in this study was the LUng Nodule Analysis 2016 challenge (LUNA16) [Setio et al. 2017]. The LUNA16 dataset was built by Setio et al. based on the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [Armato et al. 2011], and it served as a benchmark for the objective evaluation of automatic lung nodule detection algorithms. The LIDC-IDRI database contained 1,018 CT scans, each with a transverse plane size of 512×512, annotated by four radiologists in a two-phase process. In the LUNA16 dataset, the original LIDC-IDRI data underwent strict selection, excluding images with slice thickness greater than 2.5 mm or missing slices, ultimately retaining 888 CT scans. Additionally, LUNA16 filtered out nodules with a diameter less than 3 mm, keeping only 1,186 nodules annotated by 3 to 4 experts as the ground truth, with an average diameter of 8.3 mm.

In the false positive reduction challenge, LUNA16 provided two candidate datasets: V1 and V2. Each dataset contained the patient identifier, the central coordinates and the corresponding category of the candidate nodules. Since V1 had been deprecated, training and validation was conducted using V2. The V2 candidate dataset included 754,975 candidate nodules detected using five different nodule detection methods [Jacobs et al. 2014] [Setio et al. 2015] [Murphy et al. 2009] [Tan et al. 2011] [Traverso et al. 2017], among which 1,557 were annotated as true nodules, and 1,166 matched the 1,186 ground truth nodules.

4.2 Preprocessing

The number of nodules and non-nodules in the LUNA16 dataset was extremely imbalanced, approximately 1:484. This severe imbalance can lead to class bias during training and reduce the model's sensitivity to true nodules. To address this issue, a series of data augmentation strategies were applied to balance the class distribution to enhance the model's robustness and adaptability. Specifically, geometric transformations were performed on the nodule samples: rotating the images by 90°, 180°, and 270° in the transverse plane; translating 1 to 2 pixels along each axis to simulate displacement variations; performing random mirror flipping along the X-axis; and adding Gaussian noise with a mean of 0 and variance of 0.1. Figure 5 shows representative examples of the augmentation strategies. These augmentation operations resulted in 149,472 augmented true nodule samples, significantly alleviating the class imbalance problem. Additionally, to ensure standardized processing of the CT images, their intensity values were limited to the range of -1000 to 400 Hounsfield Units (HU) and these values were normalized to a range of 0 to 1.

Figure 5: Representative examples of image augmentation strategies. (a) original image; (b) 180° rotation; (c) left translation; (d) horizontal flipping; (e) gaussian noise addition (mean=0, variance=0.1).

For performance evaluation, 888 chest CT scans were randomly divided into 10 subsets, and 10-fold cross-validation was used to systematically evaluate the model's performance. The detailed data distribution is presented in Table 1. In each fold, eight subsets were used for training, one subset was designated as a validation set to select the optimal model, and the remaining one subset served as the test set. The overall performance was reported as the average of evaluation metrics across the 10 test sets. Notably, data augmentation was performed exclusively on true nodules within the training set to alleviate class imbalance.

Folds	Scans	Nodules			Non-nodules		
		Training	Validation	Testing	Training	Validation	Testing
1	710/89/89	119,904	170	138	603,579	70,842	78,997
2	710/89/89	115,776	181	170	608,349	74,277	70,842
3	710/89/89	116,928	158	181	603,349	75,792	74,277
4	710/89/89	117,984	170	158	601,213	76,413	75,792
5	710/89/89	120,960	127	170	601,441	75,564	76,413
6	710/89/89	122,496	154	127	677,854	76,517	75,564
7	710/89/89	123,168	120	154	601,337	74,943	76,517
8	711/88/89	119,232	195	120	604,182	74,293	74,943
9	712/88/88	116,928	144	195	603,345	75,780	74,293
10	711/89/88	122,400	138	144	598,641	78,997	75,780

Table 1: Distribution of lung nodule candidates in training, validation, and testing sets for 10-fold cross-validation

To ensure consistent pixel spacing across different CT scans, nearest-neighbor interpolation was used to spatially normalize all images, setting the pixel spacing to 1 mm. This preprocessing step eliminated variations due to different equipment and scan parameters, ensuring uniformity in the input data. According to the statistics of the LUNA16 dataset, as shown in Figure 6, there was a significant variation in nodule sizes. To capture features at different scales, three input block sizes (16×16×16, 32×32×32 and 48×48×48) were designed, covering 90.56%, 99.92%, and 100% of the nodules, respectively. This multi-scale design enabled the model to not only capture the fine details of tiny nodules but also analyze the relationship between the nodules and surrounding anatomical structures in a broader context.

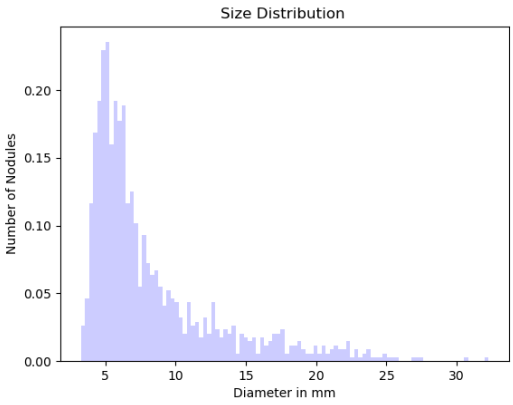

Figure 6: Distribution of nodule sizes in the LUNA16 dataset

4.3 Experimental Setup

The experimental setup was based on the Windows 11 operating system, and hardware consisting of an Intel Core i7-12700K CPU with a main frequency of 3.6 GHz and 32 GB of RAM. All network models were implemented using Python 3.8.8 and trained with the deep learning framework PyTorch 1.8.1 on an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory. The training process was set for 100 epochs, with a batch size of 8, using the stochastic gradient descent (SGD) optimization algorithm. The initial learning rate was 0.01, momentum was set to 0.9, and the weight decay coefficient was 0.0001. To prevent overfitting, an early stopping strategy was applied, dynamically halting the training based on the performance of the validation set.

The loss function used in this study is binary cross-entropy (BCE) loss, which quantifies the discrepancy between the predicted probability and the ground truth label, and is computed according to Eq. (6):

$$BCE(p, y) = -(ylog(p) + (1 - y)\log(1 - p)) \tag{6}$$

where y represents the ground true label (1 for nodule, 0 for non-nodule), and p denotes the probability that the sample is predicted as a true nodule. To address the issue of data imbalance, weighted adjustments were applied to the positive and negative samples. Additionally, during data loading, the order of the training data was shuffled randomly to reduce the model's dependence on the sample sequence.

4.4 Evaluation Metrics

The main task was to classify lung nodule candidates and reduce false positives. The primary evaluation metrics were sensitivity and specificity, which are defined in Eq. (7) and (8), respectively:

$$Sensitivity = TP/(TP + FN) \tag{7}$$

$$Specificity = TN/(TN + FP) \tag{8}$$

where TP (true positive) is the number of correctly predicted nodules, FN (false negative) is the number of nodules that were not detected, TN (true negative) is the number of correctly predicted non-nodules, and FP (false positive) is the number of

non-nodules incorrectly predicted as nodules. To provide a more comprehensive evaluation of the model's performance under class imbalance, balanced accuracy (BAcc) [Thölke et al. 2023] was also included as a supplementary metric. BAcc is the average of sensitivity and specificity, offering a more accurate assessment of the model's performance on imbalanced datasets compared to individual metrics.

Additionally, the competition performance metric (CPM) [Niemeijer et al. 2010], which was widely used in the LUNA16 challenge, was adopted as the primary evaluation metric. CPM assesses the model's performance across different false-positive rates, defined as the average sensitivity at specific values of false positives per scan (FPs/scan), namely 0.125, 0.25, 0.5, 1, 2, 4, and 8, as shown in Eq. (9):

$$CPM = \frac{1}{7}\sum_{i=\{0.125,0.25,0.5,1,2,4,8\}} Sensitivity_{fpr=i} \quad (9)$$

where fpr represents the FPs/scan, and $Sensitivity_{fpr=i}$ denotes the sensitivity when $fpr = i$.

5 Results and Discussion

5.1 Experimental Results

To validate the effectiveness of the proposed AMCF-CNN model, experiments on the LUNA16 dataset was conducted and the results were compared with several existing methods. Table 2 presents the sensitivity of different models at various FPs/scan thresholds, along with the CPM scores for performance comparison. Figure 7 shows the free receiver operating characteristic (FROC) curves for the different methods.

Methods	**Sensitivity of different FPs/scan**							**CPM**
	0.125	**0.25**	**0.5**	**1**	**2**	**4**	**8**	
[Setio et al. 2016]	0.669	0.760	0.831	0.892	0.923	0.944	0.960	0.854
[Zhao et al. 2022]	0.719	0.798	0.882	0.946	0.952	0.952	0.952	0.896
[Kim et al. 2019]	**0.904**	**0.931**	**0.943**	0.947	0.952	0.956	0.962	**0.942**
[Gu et al. 2022]	0.631	0.743	0.813	0.889	0.927	0.956	0.979	0.848
[Yuan et al. 2021]	0.748	0.815	0.861	0.903	0.929	0.952	0.962	0.881
[Dou et al. 2016]	0.677	0.834	0.927	**0.972**	**0.981**	**0.983**	0.983	0.908
[Xiao et al. 2019]	0.833	0.867	0.932	0.953	0.960	0.977	**0.984**	0.929
[Vipparla et al. 2021]	0.821	0.869	0.935	0.968	0.971	0.976	0.982	0.931
AMCF-CNN (Proposed)	0.819	0.884	0.940	0.966	0.978	**0.983**	**0.984**	0.936

Table 2: Performance comparison of different methods

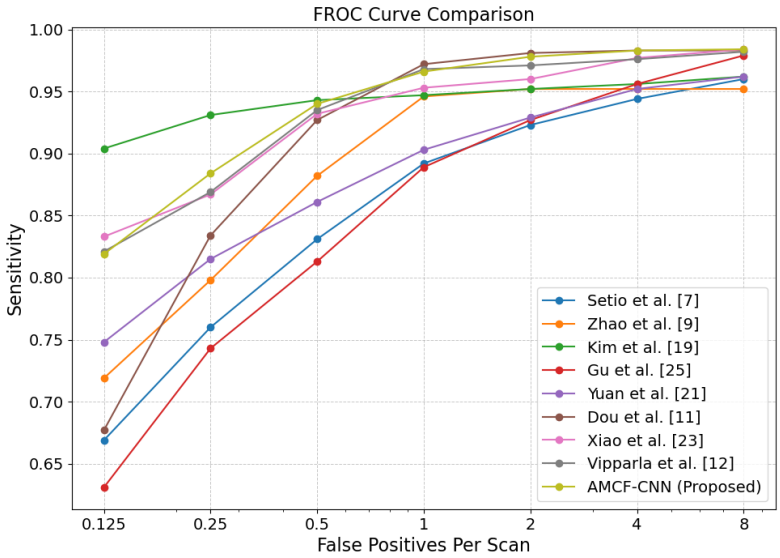

Figure 7: The FROC curves of different methods

The experimental results demonstrate that the proposed model achieved a CPM score of 0.936, outperforming most existing methods. Notably, it showed excellent performance at higher FPs/scan (4 and 8), with sensitivities reaching 0.983 and 0.984, respectively.

Compared to the 2D CNN methods of Setio et al. [Setio et al. 2016] and Zhao et al. [Zhao et al. 2022], the proposed model achieved an overall CPM improvement of 0.082 and 0.04, respectively. This indicated that the proposed model not only enhanced multi-scale feature interactions but also overcame the limitations of 2D CNNs in capturing global information. The stepwise multi-scale integration approach by Kim et al. [Kim et al. 2019] showed excellent performance at low false-positive rates, achieving a sensitivity of 0.904 at 0.125 FPs/scan and a CPM of 0.942. However, its sensitivity at 4 and 8 FPs/scan was lower than that of the proposed model, demonstrating that AMCF-CNN exhibited stronger adaptability and robustness in complex backgrounds and high false-positive rates. In contrast, the multi-path and multi-receptive field fusion approaches used by Gu et al. [Gu et al. 2022] and Yuan et al. [Yuan et al. 2021] improved feature extraction but performed inadequately at low false-positive rates due to the lack of efficient global information modelling. Dou et al. [Dou et al. 2016] improved CPM using a multi-model ensemble strategy, but their sensitivity at 0.125 FPs/scan was only 0.677, revealing limitations in low false positive detection conditions. The models of Xiao et al. [Xiao et al. 2019] and Vipparla et al. [Vipparla et al. 2021] achieved high CPM scores of 0.929 and 0.931 respectively, but their sensitivities at 4 and 8 FPs/scan were slightly lower than our model, further confirming that AMCF-CNN has a distinct advantage in multi-scale feature interaction and handling complex backgrounds.

To evaluate the statistical significance of the performance differences between the proposed method and other reported approaches, a one-sample t-test [Gerald 2018] was conducted based on the CPM scores obtained from 10-fold cross-validation, using the

CPM value of each compared method as the hypothetical population mean. The results showed that, for most methods, the performance differences were statistically significant ($p < 0.05$). However, no statistically significant difference was observed when compared to [Kim et al. 2019] ($p = 0.081$) and [Vipparla et al. 2021] ($p = 0.136$). These findings confirm the effectiveness and robustness of the proposed AMCF-CNN method.

In summary, the proposed AMCF-CNN achieved a good balance between performance and robustness. Especially at high false-positive rates, the proposed model demonstrated exceptional detection capability, effectively addressing the similarities among complex structures.

5.2 Ablation Experiment

To evaluate the contribution of each module and strategy, a series of ablation experiments was designed. By removing or replacing key modules, multiple variant models were constructed to analyze the specific impact of each design on the final performance. The experimental design is as follows:

(1) AMCF-CNN-RE: The SimAM mechanism was removed from the SimAM-Res module, leaving only the standard ResNet block [Tajbakhsh et al. 2016] to assess the role of SimAM in adaptive feature adjustment.

(2) AMCF-CNN-SE: The SimAM-Res module was replaced with the SE-ResNet block [Hu et al. 2018] to explore the relative effectiveness of the SE and SimAM mechanisms in feature enhancement.

(3) AMCF-CNN-NC: Cross-fusion paths of top to bottom and bottom to top were omitted, while residual connections in the MCF strategy were retained to evaluate the effect of the cross-fusion paths.

(4) AMCF-CNN-SF: Only a top-down fusion path was used without a reverse path to assess the improvement in model performance resulting from the bidirectional cross-fusion.

(5) AMCF-CNN-NR: The residual connections in the MCF strategy were removed, and the cross-fusion paths were employed along with SimAM-Res module-based feature fusion to evaluate the impact of residual connections.

(6) AMCF-CNN-GC: The parallel structure of convolution and self-attention in the GMM module was replaced with a simple 1×1×1 convolution to analyze the effect of the GMM module on global information modelling.

Methods	**Sensitivity of different FPs/scan**							**CPM**
	0.125	**0.25**	**0.5**	**1**	**2**	**4**	**8**	
AMCF-CNN-RE	0.774	0.868	0.931	0.956	0.959	0.961	0.962	0.916
AMCF-CNN-SE	0.793	0.880	0.943	0.961	0.966	0.966	0.966	0.926
AMCF-CNN-NC	0.749	0.877	0.934	0.944	0.954	0.957	0.958	0.910
AMCF-CNN-SF	0.784	0.880	**0.944**	0.963	0.966	0.968	0.968	0.925
AMCF-CNN-NR	0.780	0.872	0.937	0.958	0.970	0.973	0.974	0.923
AMCF-CNN-GC	0.776	0.856	0.904	0.944	0.960	0.968	0.971	0.911
AMCF-CNN (Proposed)	**0.819**	**0.884**	0.940	**0.966**	**0.978**	**0.983**	**0.984**	**0.936**

Table 3: Performance comparison of different AMCF-CNN variants under various experimental settings

Table 3 summarizes the performance of each model variant under different experimental settings, illustrating the impact of individual modules and strategies on the overall performance, thereby validating the importance of each design. To further assess the robustness of each module's contribution to the overall performance improvement of AMCF-CNN, paired t-tests [Xu et al. 2017] were conducted based on the CPM scores obtained from 10-fold cross-validation. The results of the paired t-tests indicate that, compared to each ablated variant, AMCF-CNN demonstrates statistically significant performance improvements ($p < 0.05$).

5.2.1 Effectiveness of SimAM-Res Module

The comparison between AMCF-CNN-RE and AMCF-CNN-SE in the ablation experiments demonstrated the advantages of the SimAM-Res module in adaptive feature adjustment. Specifically, when the SimAM mechanism was removed and the standard ResNet structure was used (AMCF-CNN-RE), the CPM decreased from 0.936 to 0.916. This indicates that the SimAM mechanism significantly improved the effectiveness of feature selection, with its ability to adaptively adjust feature weights, enhancing the expression of important features while suppressing irrelevant information. After replacing the SimAM-Res module with the SE-ResNet block (AMCF-CNN-SE), the CPM is 0.926, which was an improvement over the standard ResNet block but still did not reach the original SimAM-Res module's performance. This suggested that SimAM outperformed the SE attention mechanism in capturing local variations and fine-grained features.

To present these results more intuitively, FROC curves were plotted as shown in Figure 8, comparing the sensitivity of different models under various FPs/scan conditions. It showed that AMCF-CNN achieved higher sensitivity across most FPs/scan values compared to AMCF-CNN-RE and AMCF-CNN-SE, further validating the effectiveness of the SimAM mechanism.

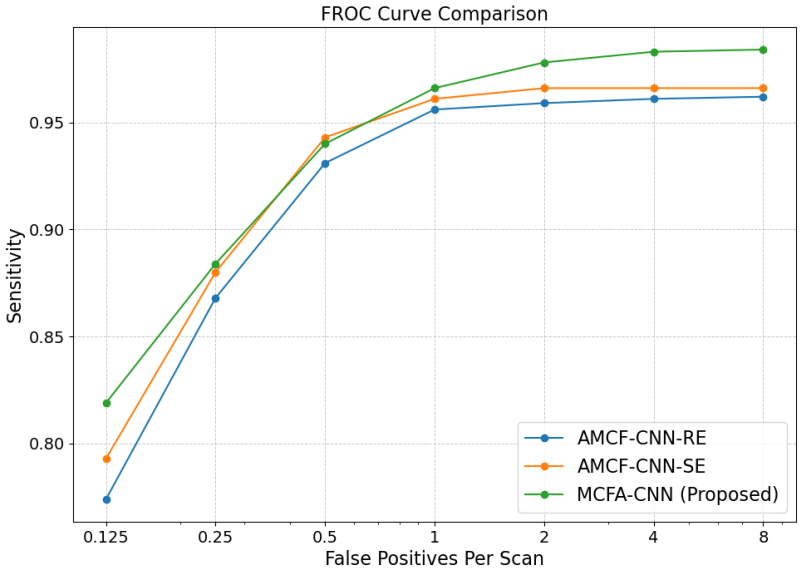

Figure 8: FROC curves for evaluating the effect of the SimAM-Res module

5.2.2 Effectiveness of MCF Module

The evaluation of the AMCF-CNN-NC and AMCF-CNN-SF models in the ablation experiments allowed for assessing the impact of the bidirectional cross-fusion paths within the MCF. When the bidirectional cross-fusion paths were removed (AMCF-CNN-NC), the CPM dropped to 0.910, indicating that cross-fusion was crucial for the effective integration of multi-scale information. When only a single top-down path was applied (AMCF-CNN-SF), the CPM improved to 0.925 but still fell short of the bidirectional fusion result of 0.936. This suggested that bidirectional cross-fusion paths could more comprehensively integrate multi-scale features, allowing global information and local details to complement each other across different levels. Additionally, when residual connections in the MCF were removed (AMCF-CNN-NR), the CPM decreased to 0.923, which was lower than AMCF-CNN's 0.936. This demonstrated that residual connections helped retain the original feature information and enhance the depth of information flow, further improving the effectiveness of feature fusion.

As shown in Figure 9, AMCF-CNN with the complete configuration of MCF strategy has higher sensitivity than other ablation variants, especially under low false-positive rates. This further confirmed the critical role of bidirectional cross-fusion and residual connections in multi-scale feature integration and enhancing model performance.

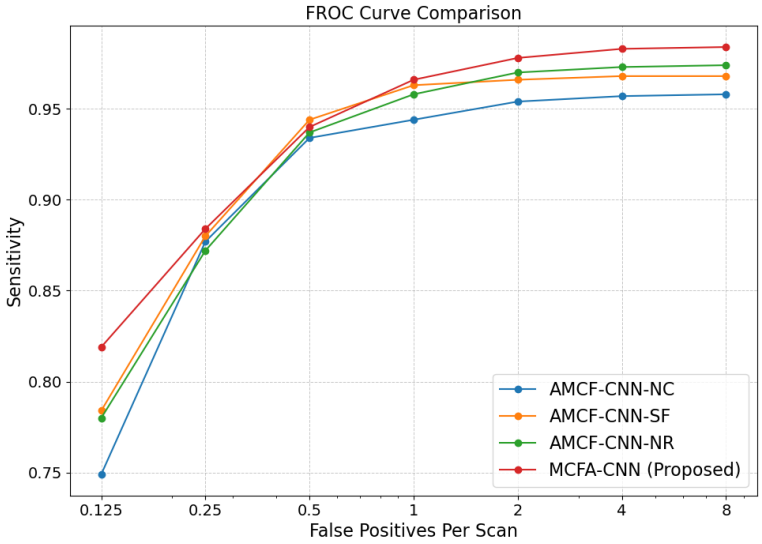

Figure 9: FROC curves for evaluating the effect of the MCF module

5.2.3 Effectiveness of GMM Module

To verify the role of the GMM module in global modelling, the parallel structure of convolution and self-attention in the GMM was replaced with a standard 1×1×1 convolution (AMCF-CNN-GC). The experimental results showed that the CPM for AMCF-CNN-GC was 0.911, which was lower than the 0.936 achieved by AMCF-CNN. This indicated that the self-attention mechanism had a significant advantage in

capturing long-distance dependencies between features and enhancing global information representation. Compared to using only local convolution operations, self-attention was better at distinguishing lung nodules from similar structures in complex backgrounds.

As seen in the FROC curves in Figure 10, the sensitivity of AMCF-CNN outperformed AMCF-CNN-GC across all false-positive rates, with the gap being particularly pronounced at lower false-positive rates. This further demonstrated that the GMM module, by combining convolution and self-attention mechanisms, not only improved the completeness of feature representation but also significantly enhanced the model's detection sensitivity.

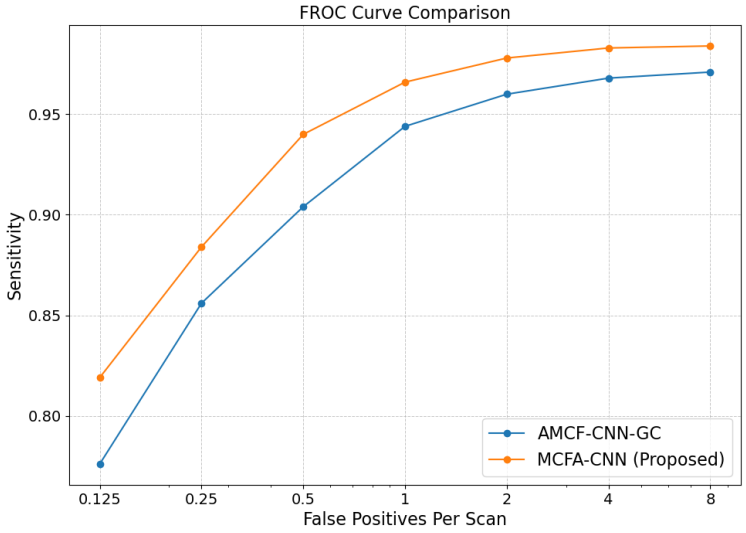

Figure 10: FROC curves for evaluating the effect of the GMM module

5.3 Analysis

The proposed AMCF-CNN, by combining the MCF strategy, SimAM-Res module, and GMM module, significantly enhanced the performance in false-positive reduction. To further explore the advantages of this approach, the performance of AMCF-CNN was compared with several existing methods in terms of sensitivity, specificity, accuracy and BAcc, as shown in Table 4.

Methods	Sensitivity	Specificity	Accuracy	BAcc
[Zuo et al. 2020]	0.877	0.993	0.978	0.935
[Hao et al. 2023]	0.926	0.987	**0.987**	0.957
[Zhao et al. 2022]	0.952	0.981	0.967	0.967
[El-Bana et al. 2020]	0.964	**0.994**	0.970	0.979
AMCF-CNN (Proposed)	**0.984**	0.983	0.983	**0.983**

Table 4: Performance comparisons of Sensitivity, Specificity, Accuracy, and BAcc among the proposed method and other methods

In terms of sensitivity, AMCF-CNN achieved the highest value of 0.984, surpassing all other compared methods. This indicates that the proposed model demonstrates superior capability in detecting pulmonary nodules. High sensitivity is particularly crucial for early lung cancer screening, as it enhances the detection rate of potential patients and supports earlier intervention and treatment. Regarding specificity, AMCF-CNN also exhibited excellent performance, reaching 0.983, which outperformed the methods of Zuo et al. [Zuo et al. 2020], Hao et al. [Hao et al. 2023], and Zhao et al. [Zhao et al. 2022], and was only slightly lower than that of El-Bana et al. [El-Bana et al. 2020] (0.994). This result reflects that AMCF-CNN maintains near-optimal accuracy in correctly identifying negative cases (non-nodule samples).

Regarding the overall classification performance, AMCF-CNN achieved a BAcc of 0.983, surpassing all compared methods and achieving the highest classification accuracy. As a comprehensive metric for evaluating sensitivity and specificity, the high BAcc value indicates that the model possesses good adaptability and robustness when dealing with class-imbalanced and complex background data. In clinical practice, lung nodule screening is often affected by data imbalance, imaging noise, and complex tissue variations, which can easily lead to missed or false diagnoses. Effective early lung cancer screening requires a proper balance between sensitivity and specificity. Insufficient sensitivity may result in missed diagnoses and delayed treatment, while inadequate specificity can lead to a large number of unnecessary follow-up examinations and invasive procedures, increasing patient burden and wasting healthcare resources. The balanced performance of AMCF-CNN across multiple metrics contributes to improving the overall efficiency and accuracy of the screening system, reducing unnecessary clinical interventions, alleviating the diagnostic workload of radiologists, and providing more reliable technical support for early lung cancer screening and follow-up management.

6 Conclusions

The proposed AMCF-CNN framework achieved significant improvements in reducing false-positive rates for lung nodule detection. By introducing the MCF strategy, the model effectively alleviates the semantic fragmentation problem caused by direct feature concatenation in traditional multi-scale methods. The GMM, combining 3D convolutions and self-attention mechanisms, overcomes the limitations of local receptive fields and substantially enhances discrimination performance in complex backgrounds. Additionally, the SimAM-Res module employs a parameter-free adaptive weighting mechanism to suppress redundant information while enhancing key feature representations. On the LUNA16 dataset, AMCF-CNN achieved a competitive CPM score of 0.936, with sensitivities of 0.983 and 0.984 at 4 and 8 FPs/scan, respectively. Notably, the model demonstrated an excellent balance between sensitivity and specificity (BAcc = 0.983), ensuring both high detection capability and clinical practicality. While the effectiveness of the method has been validated on the LUNA16 benchmark, clinical validation remains essential for real-world deployment. Future work will focus on external validation across diverse devices and patient populations, conducting prospective studies to quantify the model's impact on diagnostic accuracy and workflow efficiency, and further optimizing inference speed to meet real-time clinical demands. Overall, AMCF-CNN provides a robust and practical solution for

early lung cancer detection and serves as a valuable reference for advancing future research and applications in medical image analysis.

Acknowledgements

This work was supported by Universiti Malaysia Sarawak (UNIMAS); the Youth Science and Technology Talent Growth Project of Ordinary Higher Education Institutions in Guizhou Province (黔教合 KY 字[2022] No. 281); Zunyi Medical University Undergraduate Innovation and Entrepreneurship Training Program (ZYDC202402321).

References

[Armato et al. 2011] Armato III S G, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans[J]. Medical physics, 2011, 38(2): 915-931.

[Camarlinghi et al. 2013] Camarlinghi N. Automatic detection of lung nodules in computed tomography images: training and validation of algorithms using public research databases[J]. The European Physical Journal Plus 2013, 128(9): 110.

[Cao et al. 2019] Cao H, Liu H, Song E, et al. Multi-branch ensemble learning architecture based on 3D CNN for false positive reduction in lung nodule detection[J]. IEEE access 2019, 7: 67380-67391.

[Cheng et al. 2022] Cheng T, Wang X, Chen S, et al. Sparse instance activation for real-time instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022: 4433-4442.

[Ding et al. 2017] Ding J., Li A., Hu Z., Wang L. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks; Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention 2017: 559–567.

[Dou et al. 2016] Dou Q, Chen H, Yu L, et al. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection[J]. IEEE Transactions on Biomedical Engineering 2016, 64(7): 1558-1567.

[El-Bana et al. 2020] El-Bana S, Al-Kabbany A, Sharkas M. A two-stage framework for automated malignant pulmonary nodule detection in CT scans[J]. Diagnostics 2020, 10(3): 131.

[Firmino et al. 2014] Firmino M, Morais A H, Mendoça R M, et al. Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects[J]. Biomedical engineering online 2014, 13: 1-16.

[Gerald 2018] Gerald B. A brief review of independent, dependent and one sample t-test[J]. International journal of applied mathematics and theoretical physics, 2018, 9(2): 50-54.

[Gould et al. 2007] Gould M K, Fletcher J, Iannettoni M D, et al. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines[J]. Chest 2007, 132(3): 108S-130S.

[Gu et al. 2022] Gu Z, Li Y, Luo H, et al. Cross attention guided multi-scale feature fusion for false-positive reduction in pulmonary nodule detection[J]. Computers in Biology and Medicine 2022, 151: 106302.

[Haiying et al. 2021] Haiying Y, Zhongwei F, Ding D, et al. False-positive reduction of pulmonary nodule detection based on deformable convolutional neural networks[C]//2021 IEEE 9th International Conference on Bioinformatics and Computational Biology (ICBCB). IEEE, 2021: 130-134.

[Hao et al. 2023] Hao K, Cai A, Feng X, Ma L, Zhu J, Wang M, Zhang Y, Fei B. Lung nodule false positive reduction using a central attention convolutional neural network on imbalanced data[C]//Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling. SPIE 2023, 12466: 459-465.

[Hu et al. 2018] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition 2018: 7132-7141.

[Jacobs et al. 2014] Jacobs C, Van Rikxoort E M, Twellmann T, et al. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images[J]. Medical image analysis 2014, 18(2): 374-384.

[Kim et al. 2019] Kim B C, Yoon J S, Choi J S, et al. Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection[J]. Neural Networks, 2019, 115: 1-10.

[Lai et al. 2021] Lai K D, Nguyen T T, Le T H. Detection of lung nodules on ct images based on the convolutional neural network with attention mechanism[J]. Annals of Emerging Technologies in Computing (AETiC) 2021, 5(2): 78-89.

[Liu et al. 2016] Liu J, Jiang H, Gao M, et al. An assisted diagnosis system for detection of early pulmonary nodule in computed tomography images[J]. Journal of medical systems, 2017, 41: 1-9.

[Liu et al. 2018] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition 2018: 8759-8768.

[Murphy et al. 2009] Murphy K, van Ginneken B, Schilham A M R, et al. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification[J]. Medical image analysis, 2009, 13(5): 757-770.

[Niemeijer et al. 2010] Niemeijer M, Loog M, Abramoff M D, et al. On combining computer-aided detection systems[J]. IEEE Transactions on Medical Imaging, 2010, 30(2): 215-223.

[Setio et al. 2015] Setio A A A, Jacobs C, Gelderblom J, et al. Automatic detection of large pulmonary solid nodules in thoracic CT images[J]. Medical physics, 2015, 42(10): 5642-5653.

[Setio et al. 2016] Setio A A A, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks[J]. IEEE transactions on medical imaging 2016, 35(5): 1160-1169.

[Setio et al. 2017] Setio A A A, Traverso A, De Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge[J]. Medical image analysis, 2017, 42: 1-13.

[Siegel et al. 2020] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2020[J]. CA: a cancer journal for clinicians 2020, 70(1): 7-30.

[Sluimer et al. 2006] Sluimer I, Schilham A, Prokop M, et al. Computer analysis of computed tomography scans of the lung: a survey[J]. IEEE transactions on medical imaging 2006, 25(4): 385-405.

[Sun et al. 2021] Sun L, Wang Z, Pu H, et al. Attention-embedded complementary-stream CNN for false positive reduction in pulmonary nodule detection[J]. Computers in Biology and Medicine, 2021, 133: 104357.

[Tajbakhsh et al. 2016] Tajbakhsh N, Shin J Y, Gurudu S R, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning?[J]. IEEE transactions on medical imaging 2016, 35(5): 1299-1312.

[Tan et al. 2011] Tan M, Deklerck R, Jansen B, et al. A novel computer-aided lung nodule detection system for CT images[J]. Medical physics, 2011, 38(10): 5630-5645.

[Thölke et al. 2023] Thölke P, Mantilla-Ramos Y J, Abdelhedi H, et al. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data[J]. NeuroImage, 2023, 277: 120253.

[Traverso et al. 2017] Traverso A, Torres E L, Fantacci M E, et al. Computer-aided detection systems to improve lung cancer early diagnosis: state-of-the-art and challenges[C]//Journal of Physics: Conference Series. IOP Publishing, 2017, 841(1): 012013.

[Vaswani et al. 2017] Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems 2017.

[Vipparla et al. 2021] Vipparla V K, Chilukuri P K, Kande G B. Attention based multi-patched 3D-CNNs with hybrid fusion architecture for reducing false positives during lung nodule detection[J]. Journal of Computer and Communications 2021, 9(04): 1-26.

[Woo et al. 2018] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV) 2018: 3-19.

[Xiao et al. 2019] Xiao Z, Du N, Geng L, et al. Multi-scale heterogeneous 3D CNN for false-positive reduction in pulmonary nodule detection, based on chest CT images[J]. Applied Sciences 2019, 9(16): 3261.

[Xu et al. 2017] Xu M., Fralick D., Zheng J.Z., Wang B., Changyong F. The differences and similarities between two-sample t-test and paired t-test. Shanghai Arch. Psychiatry. 2017;29:184.

[Yang et al. 2021] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning. PMLR 2021: 11863-11874.

[Yuan et al. 2021] Yuan H, Fan Z, Wu Y, et al. An efficient multi-path 3D convolutional neural network for false-positive reduction of pulmonary nodule detection[J]. International journal of computer assisted radiology and surgery, 2021, 16(12): 2269-2277.

[Zhao et al. 2022] Zhao D, Liu Y, Yin H, et al. A novel multi-scale CNNs for false positive reduction in pulmonary nodule detection[J]. Expert Systems with Applications 2022, 207: 117652.

[Zuo et al. 2020] Zuo W, Zhou F, He Y. An embedded multi-branch 3D convolution neural network for false positive reduction in lung nodule detection[J]. Journal of digital imaging 2020, 33(4): 846-857.