BMC Pulmonary Medicine

RESEARCH

Open Access

GC-WIR : 3D global coordinate attention wide inverted ResNet network for pulmonary nodules classification

Wenju Wang[1†], Shuya Yin[1*†], Fang Ye[1], Yinan Chen[2], Lin Zhu[2] and Hong Yu[2]

Abstract

Purpose Currently, deep learning methods for the classification of benign and malignant lung nodules encounter challenges encompassing intricate and unstable algorithmic models, limited data adaptability, and an abundance of model parameters.To tackle these concerns, this investigation introduces a novel approach: the 3D Global Coordinated Attention Wide Inverted ResNet Network (GC-WIR). This network aims to achieve precise classification of benign and malignant pulmonary nodules, leveraging its merits of heightened efficiency, parsimonious parameterization, and robust stability.

Methods Within this framework, a 3D Global Coordinate Attention Mechanism (3D GCA) is designed to compute the features of the input images by converting 3D channel information and multi-dimensional positional cues. By encompassing both global channel details and spatial positional cues, this approach maintains a judicious balance between flexibility and computational efficiency. Furthermore, the GC-WIR architecture incorporates a 3D Wide Inverted Residual Network (3D WIRN), which augments feature computation by expanding input channels. This augmentation mitigates information loss during feature extraction, expedites model convergence, and concurrently enhances performance. The utilization of the inverted residual structure imbues the model with heightened stability.

Results Empirical validation of the GC-WIR method is performed on the LUNA 16 dataset, yielding predictions that surpass those generated by previous models. This novel approach achieves an impressive accuracy rate of 94.32%, coupled with a specificity of 93.69%. Notably, the model's parameter count remains modest at 5.76M, affording optimal classification accuracy.

Conclusion Furthermore, experimental results unequivocally demonstrate that, even under stringent computational constraints, GC-WIR outperforms alternative deep learning methodologies, establishing a new benchmark in performance.

Keywords Classification of pulmonary nodules, 3D wide inverted residual network, 3D global coordinate attention mechanism

Introduction

Among all cancer types, lung cancer stands out with the highest incidence rate [1]. The death rate of cancer is closely related to the time when the cancer is first detected. The earlier it is detected, the lower the incidence. Notably, pulmonary nodules constitute the primary early indicators of lung cancer [2]. In the contemporary context, Computed Tomography (CT)

†Wenju Wang and Shuya Yin contributed equally to this work.

*Correspondence:
Shuya Yin
223332934@st.usst.edu.cn
[1] University of Shanghai for Science and Technology, Jungong 516 Rd, Shanghai 200093, China
[2] Department of Radiology, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, Huaihai West Road NO.241, Shanghai 200030, China

emerges as the premier imaging modality for lung cancer detection. The precise interpretation of CT images and the timely differentiation between benign and malignant pulmonary nodules by medical practitioners stand as pivotal factors in elevating the survival probabilities of individuals afflicted with lung cancer [3]. Consequently, the development of automated and accurate classification and diagnostic technologies pertaining to malignant pulmonary nodules, harnessed through the analysis of CT images, assumes paramount significance within the realm of clinical medicine. As such, this domain has garnered substantial attention as a burgeoning research focus.

In recent times, the domain of deep learning has undergone rapid and extensive advancement, finding pervasive utility across domains such as computer vision [4] and medical image processing [5]. This paradigm shift has engendered remarkable strides. Presently, the frontiers of deep learning are progressively extending into the realm of intelligent diagnosis concerning benign and malignant pulmonary nodules. Tracing the evolutionary trajectory of deep learning's convolutional neural network technology, the methodologies for classifying and diagnosing pulmonary nodules based on CT images delineate three discernible categories: the Convolutional Neural Network (CNN) classification algorithm, the residual network classification algorithm, and the attention mechanism classification algorithm. These categories encapsulate the evolving landscape of techniques employed for accurate nodule assessment.

Convolutional Neural Network (CNN) The early stages of lung nodule detection predominantly relied on conventional machine learning techniques, often entailing manual feature extraction. Regrettably, these approaches often suffered from suboptimal generalization capabilities, failing to meet the sensitivity and other critical requirements of clinical medical applications [6]. In contrast, the convolutional neural network (CNN) represents a pioneering advancement in artificial neural networks, fashioned through emulating the structural intricacies of the human brain's cortical layers. It streamline the network's complexity by using three strategies: local receptive field, weight sharing and down-sampling. This architecture leads to more accurate identification of malignant pulmonary nodules within chest CT images, substantially enhancing diagnostic efficiency. Numerous scholars have devised diverse lung nodule detection algorithms grounded in CNN frameworks. Gao Dachuan et al. [7] proposed a hybrid approach that fuses CNN-learned features with traditional machine-extracted features to discern benign from malignant pulmonary nodules. The theory corroborates the CNN model's remarkable classification and recognition capabilities. However, the automatic detection of pulmonary nodules has not yet been achieved, and its accuracy still needs to be further improved. Tan Jiaxing et al. [8] developed a two-stage framework system entirely based on 2D CNN. This method significantly improved the accuracy of automatic detection of pulmonary nodules, but with a certain false positive rate. Eun Hyunjun et al. [9] introduced a 2D CNN model for automatic pulmonary nodule classification, rooted in single-view analysis. Their study curtailed false positives through parameter and network layer adjustments. Yet, single-view networks exhibit limitations in capturing feature information comprehensively. In this regard, Arnaud Arindra Adiyoso Setio et al. [10] addressed this limitation by proposing a multi-view lung nodule detection model founded on 2D CNNs. This algorithmic framework not only expands the learning capacity of non-representative nodule features but also achieves further reduction in false positives. Compared with the above 2D CNN model, which captures spatial information on the plane, 3D CNN model can capture spatial information on the three-dimensional. Since most CT medical images are three-dimensional data, the application of 3D CNN on CT images will be more conducive to the improvement of diagnostic accuracy. Patrice Monkam et al. [9, 11] advocated a multi-view 3D CNN approach to differentiate micro-nodules and non-nodules in CT images. Combining this model with the Extreme Learning Machine (ELM) and five 3D-CNNs yielded robust classification outcomes. Notably, 3D-CNN outperformed 2D CNN in volume image data analysis. Thiago Jose Barbosa Lima et al. [12] contributed to this landscape by designing an algorithm based on 3D CNN, utilizing three distinct 3D CNN architectures with varying input shapes and convolutional layers for discerning benign and malignant pulmonary nodules. This strategy bolstered model generalization. Gupta Anindya et al. [13] introduced a 3D convolutional neural network classification model, which effectively minimized false positives by autonomously identifying volume data for each candidate nodule. While both these algorithms markedly improved the performance of intelligent diagnosis models, they overlooked the detection of minute nodules. Kadhim Omar Raad et al. [14] used 3D-CNN and 3D-CT scanning images to classify benign and malignant tissues of micro-nodules, greatly improving the classification accuracy of micro-nodules. However, due to technical limitations, the above studies [11–14] were carried out at a single scale, and only samples of a single scale could be tested. Multi-scale technology proves advantageous, capturing a breadth of information including both global and local features of pulmonary nodules. Xie Dan et al. [15] pioneered a 3D CNN model founded on a multi-scale double-path network (DPN), leveraging 3D-CNN for spatial information extraction and a

multi-scale architecture for depth feature extraction. This approach enhanced micro-nodule feature extraction capabilities, albeit with a diminished performance for lower-resolution images. Zhao Dandan et al. [16] introduced a novel multi-scale CNN framework founded on distinct orthogonal 2D images, aimed at augmenting discriminant features of low-resolution images through frequency domain embedding. These 3D CNN methodologies, when contrasted with their 2D counterparts, amplify the capacity for lung nodule feature extraction, yet are hindered by bulky model parameters and suboptimal operational efficiency.

Residual Network (ResNet) Within the realm of convolutional neural networks, each layer's passage through convolutional kernels inevitably leads to some degree of information loss. With the increase of network depth, problems such as gradient disappearance will occur, and it is more difficult to train the model. To tackle these challenges, the concept of the residual network (ResNet) emerged, offering a solution to construct novel pulmonary nodule detection models founded on convolutional neural networks. Wenhao Deng et al. [17] amalgamated the residual network with the YOLO_v3 model for feature extraction and classification. This fusion effectively reduces false positives by enabling rapid pulmonary nodule localization in CT images. However, the model falls short in providing nuanced nodule classification. Zhang Yanan et al. [18] harnessed a multi-layer fusion of 3D ResNet to extract nodule features, thereby enabling nonlinear radial basis feature mapping and personalized intelligent decision-making for lung nodule diagnosis. However, the model needs to further explore other characteristics of nodule to improve its accuracy. Zhifeng Lin et al. [19] introduced a 3D residual network model grounded in VGG architecture, aiming to unearth vertical information from tumor CT images and enhance the classification efficacy of the model. This approach ameliorates the issue of imbalanced positive and negative medical data distribution, but it cannot judge the data with limited area and few features. Bharti Meenakshi et al. [20] developed a classification image preprocessing technique utilizing 3D ResNet in tandem with V-Net for segmentation. This approach leveraged the residual network for classifying lung nodules with scant features, thereby reducing false positives. But it can only be used to detect nodules with fewer false positives. Therefore, Tong Chao et al. [21] introduced a high-efficiency pulmonary nodule classification model underpinned by a combination of 3D Convolutional Neural Network (3D-CNN) and the Multi-Kernel Learning (MKL) algorithm. The model employed a 34-layer 3D ResNet to extract deep image features, proving effective not only in detecting nodules with reduced false positives but also those of greater complexity. On this basis, Han, Yu et al. [22] designed a hybrid model that integrated 3D CNN, 3D ResNet, and Fully Connected Neural Network (FCNN) components. By basing the network model on 3D ResNet, they reduced instances of missing nodule detection. However, challenges persisted in identifying early-stage lung nodules with smaller diameters. In pursuit of enhanced detection for smaller pulmonary nodulevs, Haiying Yuan et al. [23] proposed a 3D residual U-Net model. This approach effectively harnessed expansion convolutions with varied rates to extract both local and global nodule features, thereby proficiently discerning lung nodules of diverse sizes and shapes. Xianfang Hu et al. [24] built a fusion model of deep neural network (DNN) based on the 3D residual U-Net model. It adopts a series of image processing techniques such as feature extraction, feature selection and synthetic minority oversampling, which effectively improves the performance of the classification model of small pulmonary nodules. But with the deepening of network model, the residual network model in this category may result in small target detection part feature information is missing. This is also the reasons of the decrease of recognition.

Attention Mechanism (Attention) The concept of attention mechanism finds its roots in the study of human vision. Confronted with the limitations of information processing, humans naturally concentrate on select information while ignoring the rest. The attention mechanism mimics this property and can be used to focus on important feature information. Consequently, researchers have integrated attention mechanisms with existing neural network models such as convolution and residual networks to enhance the precision of pulmonary nodule detection. Alejandra Moreno et al. [25] introduced the self-focused (MSA) module within multi-scale networks to accentuate feature maps associated with abnormal nodules. They coupled multi-scale representations acquired through receptive field blocks (RBF) to more accurately pinpoint potential malignancies. However, the computational weight of the self-attention mechanism remains a challenge. Wang Ruinan et al. [26] developed the DPCA-Net, a streamlined and efficient 3D lung nodule detection model. Central to this model is a novel two-path channel attention block (DPCA), which enhances the efficacy of context information propagation. Qi Yongjun et al. [27] advanced a multi-scale depth residual channel attention network model, incorporating a channel attention module to capture image features across various scales and channel relationships. This technique preserves spatial structures and high-frequency details within low-resolution images, enhancing visual quality. While both approaches leverage channel attention mechanisms to discern the significance of individual

channels, they fall short in capturing important regional features when spatially transforming data through clipping, translation, or rotation. Addressing this, Mai Juanyun et al. [28] proposed a multi-head detection algorithm based on the 3D Squeezed Spatial Attention Network (MHSnet). This method orchestrates the model's focus on pivotal positions by adjusting pixel interdependencies. This adaptive strategy promotes nodule recognition by paying attention to surrounding information and adapting to different regions, thereby suppressing false positive rates. Han liangJiang et al. [29] adopted a context attention mechanism to simulate context correlations between adjacent locations, merging it with spatial attention to autonomously locate regions pertinent to tuberculosis classification. This approach effectively locates potential nodule features. In fact, channel and spatial attention can be synergistically combined to leverage their technical strengths. To propagate spatial information from coding layers to decoding layers, Dong Ting et al. [30] introduced a Residual Network algorithm (ResAANet) employing both channel and spatial attention modules for sensitive pulmonary nodule feature extraction. This technique mitigates the loss of nodule characteristics during forward transmission. While the amalgamation of channel-space attention mechanisms enhances model representational capacity and reduces irrelevant target interference, it does augment computational load. The Convolutional Block Attention Module (CBAM) streamlines convolutional structures and feature fusion operations, effectively curtailing module complexity and computational demands. Zhang Guanglu et al. [31] cited this CBAM to extract representative multi-scale nodal features for a 3D convolutional neural network for pulmonary nodule detection. Zhang Weiguo et al. [32] proposed a U-net structure detection algorithm based on attention module (CBAM) to realize adaptive feature learning and feature weights. This model does not require a large amount of data, but has low efficiency and needs to be enhanced in optimizing the network model. Hang liangJiang et al. [33] devised a robust deep lung nodule classifier by integrating CBAM with 3D CNN and ResNet architectures, effectively enhancing the model's feature extraction capabilities, robustness, and efficiency. Nevertheless, room for further improvement remains.

However, despite the advances highlighted, these models often manifest complexity and limited adaptability, rendering them less stable when confronted with diverse data. Furthermore, the incorporation of attention mechanisms, while demonstrably advantageous, can exacerbate model intricacy. This is predominantly due to the resource-intensive pooling and convolution operations commonly associated with attention mechanisms, which significantly augment the model's parameter count and computational demands. To address these challenges, we propose a GC-WIR deep neural network model for the automatic and accurate classification diagnosis of benign and malignant pulmonary nodules. The GC-WIR model represents a fusion of a 3D Broadened Inverted Residual Network and a 3D Global Coordinate Attention Mechanism. This synthesis yields a model characterized by remarkable accuracy, robust stability, potent convergence properties, and parsimonious parameterization. The principal contributions outlined in this paper are succinctly encapsulated as follows:

(1) In this paper, we employ the neural network model GC-WIR for the classification task concerning benign and malignant pulmonary nodules. The architecture of this model unfolds across four distinct stages. In the first three stages, the computational features of the 3D Widened Inverted Residual Network (3D WIRN) and the 3D Global Coordinate Attention Mechanism (3D GCA) synergistically operate. In the last stage, the benign and malignant pulmonary nodules were predicted and exported. Specifically, the 3D WIRN augments input channels to bolster feature extraction potency, thereby enhancing the model's capability. Meanwhile, the 3D GCA computes input features by leveraging both three-dimensional information transformation and the coordinates of three-dimensional space. Experimental findings amply corroborate the efficacy of GC-WIR, showcasing its ability to achieve optimal classification accuracy with the least number of parameters. The model not only boasts strong convergence but also exhibits remarkable stability, transcending the benchmarks set by existing advanced classification methodologies.
(2) GC-WIR constructs a 3D Wide Inverted Residual Network(3D WIRN), which not only widens the channel while reducing the depth of the residual network, but also uses the inverted residual structure to thin the network. This strategic combination achieves a twofold outcome-ameliorating the computational burden associated with the model's complexity and enhancing the model's stability. Importantly, while preserving the intrinsic performance of the original model, this adaptation accelerates the model's convergence rate. Empirical evidence gleaned from convergence experiments underscores the effectiveness of this approach. Remarkably, merely 300 iterations of training yield results comparable to those requiring 600 iterations in standard models. This expeditious convergence significantly contributes to the model's overall effi-

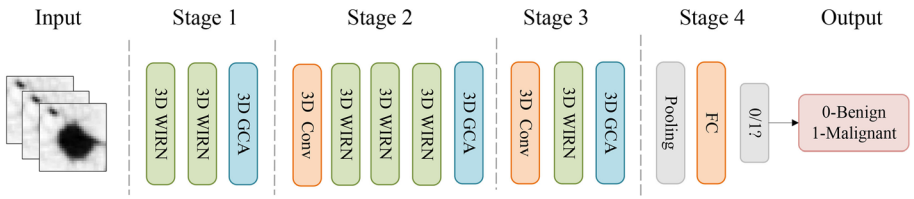

Fig. 1 3D GC-WIR model algorithm framework

ciency and reinforces its viability as a robust classification algorithm for pulmonary nodules.

(3) GC-WIR integrates a meticulously devised 3D Global Attention Mechanism (3D GCA), which admirably retains spatial position information spanning both horizontal and vertical axes. The incorporation of this attention mechanism heralds multifaceted advantages. Firstly, it amplifies global interactive representations, thereby enhancing the performance of deep neural networks. Secondly, it effectively captures diverse spatial information from multiple dimensions, thereby conferring a substantial upsurge in the classification accuracy of pulmonary nodules. One remarkable divergence from prior attention mechanisms is GC-WIR's approach to computation. Unlike preceding methods that heavily rely on convolution and pooling operations, this attention mechanism harnesses the power of three-dimensional information transformation coupled with the coordinates of three-dimensional space to calculate input features. This innovative strategy not only yields enhanced performance but also substantially curtails the parameter count. This parsimonious approach bolsters computational efficiency, rendering the model well-suited for real-world applications.

Methods

The proposed GC-WIR network, as outlined in this paper, synergistically leverages two fundamental 3D neural network techniques for the classification of benign and malignant pulmonary nodules. These techniques encompass the Wide Inverted ResNet (WIRN) and Global Coordinate Attention (GCA), elaborated upon in "3D Wide inverted resNet (3D WIRN)" and "3D Global coordinate attention mechanism" sections respectively. The network architecture, meticulously elaborated in Fig. 1, encompasses a four-stage classification and prediction processing paradigm for input CT images undergoing classification and detection. In specific terms: Stage 1: Comprises two layers of 3D WIRN in tandem with a 3D GCA module. Stage 2: Encompasses a solitary 3D convolutional layer, followed by three WIRN layers and a 3D GCA module. Stage 3: Parallels stage 2, integrating a 3D convolutional layer, WIRN layers, and a 3D GCA module, with a reduced number of WIRN layers. Stage 4: Culminates in the generation of predictive values via global averaging pooling and the fully connected layer (FC). This final stage facilitates the determination of binary labels, differentiating benign from malignant pulmonary nodules. Notably, an output of 0 indicates benign, while an output of 1 signifies malignancy. This holistic approach accomplishes the crucial task of classifying and adjudicating input pulmonary nodule images, providing a reliable avenue for effective classification and diagnosis.

3D Wide Inverted ResNet (3D WIRN)

The proposed residual network addresses the issues of vanishing and exploding gradients encountered during neural network training. Nonetheless, with the increase in network depth, training a residual network becomes progressively slow. To mitigate this concern, the concept of the wide residual network [34] introduces a new parameter-termed the "broadening factor"-building upon the original residual module. This operation widens the number of convolution cores and increases the width while reducing the depth of the residual network. This strategic design, however, leads to a proliferation of parameters. In the pursuit of model compactness, speed enhancement, and sustained performance, the approach of inverted residuals is incorporated within the residual block to streamline the network. This entails the introduction of a 3D Wide Inverted Residual network (WIRN), a core component applied to stages 1 to 3 of the network framework. The 3D WIRN configuration encompasses an initial convolutional layer, a 3D inverted residual block, and a normalized activation function, as visually presented in Fig. 2. Within this structure, the 3D inverted residual block predominantly employs widening operations to accentuate valuable features within the data. The methodology consists of four distinct steps, each executed as follows:

(1) Input convolution

The image data, denoted as the input set X, is subjected to a 3×3 3D initial convolution $Conv3D_{3\times3}$ operation to yield the resultant X_1, as depicted in Eq. 1.

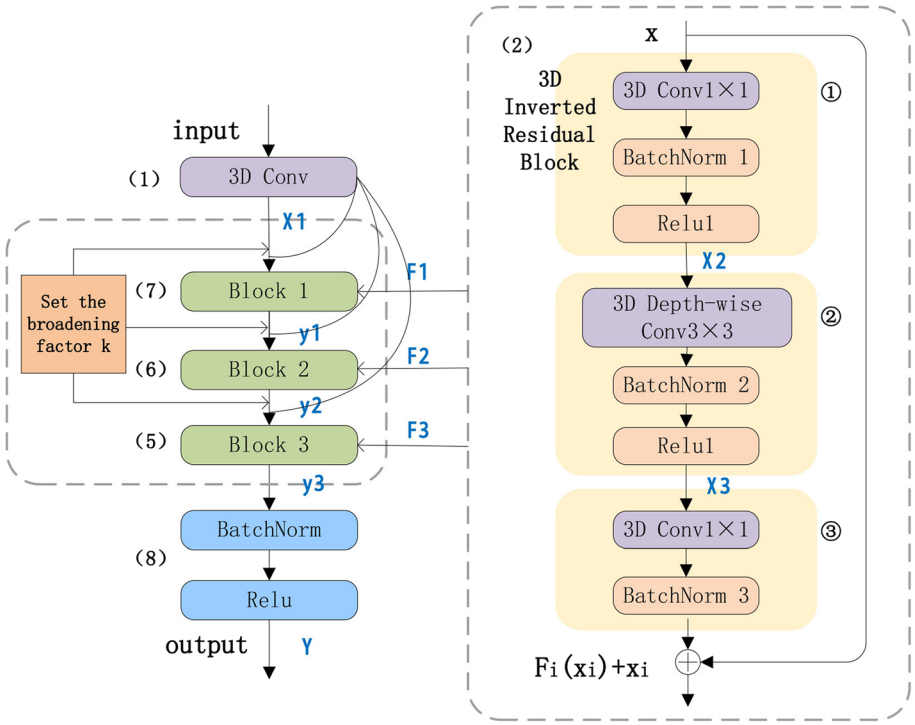

Fig. 2 Schematic diagram of the 3D widen inverted residual network

$$X_1 = Conv3D_{3\times3}(X) \tag{1}$$

(2) 3D inverted residuals $Block_i$

The network comprises three instances of residual blocks denoted as $Block_i (i \in [1,3])$. Within each *Block*, the same inverted residual structure is applied, consisting of three convolutional layer structures: $Conv_{1\times1}$, Depth wise $Conv_{3\times3}$ and $Conv_{1\times1}$. The subsequent steps illustrate the working principle of a single $Block_i$:

① 3D $Conv_{1\times1}$: In the initial layer of the 3D network, the 3D convolutional layer $Conv3D_{1\times1}$ is employed to facilitate the mapping of the input from a lower-dimensional space to a higher-dimensional one. Notably, the count of convolutional kernels within $Conv3D_{1\times1}$ corresponds to n times the quantity of input channels represented by X_1. The output of X_1 after normalization and activation function calculations is denoted as X_2, which is represented by Formula 2 as the input for the subsequent module.

$$X_2 = ReLU(BatchNorm(Conv3D_{1\times1}(n \times X_1))) \tag{2}$$

② 3D Depth-wise $Conv_{3\times3}$: In order to enhance the efficacy of data processing pertaining to X_2, the subsequent layer within the 3D network employs Depth-wise $Conv3D_{3\times3}$ for the purpose of feature extraction, as elaborated in Eq. 3. Here,X_3 signifies the output, and s denotes the expansion ratio.

$$X_3 = ReLU(BatchNorm(Conv3D_{3\times3}(\frac{X_2}{s}))) \tag{3}$$

③ 3D $Conv_{1\times1}$: The convolution operation employed within this module utilizes a 1×1 convolutional kernel with the purpose of increasing dimensionality. This mapping aids in the transformation of high-dimensional features into a lower-dimensional space, thereby facilitating data compression. Consequently, this process contributes to the reduction in network size while enhancing the model's expressiveness. The output X_3 processed by a convolutional layer and BatchNorm is combined with the initial output X_1 in step (1), yielding a residual architecture as depicted in Eq. 4. Here, the resultant output F_i is characterized as a residual function of $Block_i (i \in [1,3])$.

$$F_i = BatchNorm(Conv3D_{1\times1}(X_3)) + X_1 \tag{4}$$

(3) Channel widening

The primary objective of the widening operation is to incorporate a widening factor k (where k is a multiple of the number of output feature maps of the convolutional layer) into the output channel of the 3D inverted residuals block $Block_i$. This serves to broaden the model by augmenting the quantity of channels. The introduced widening factor k contributes to the expansion of the width across three distinct blocks. Assuming

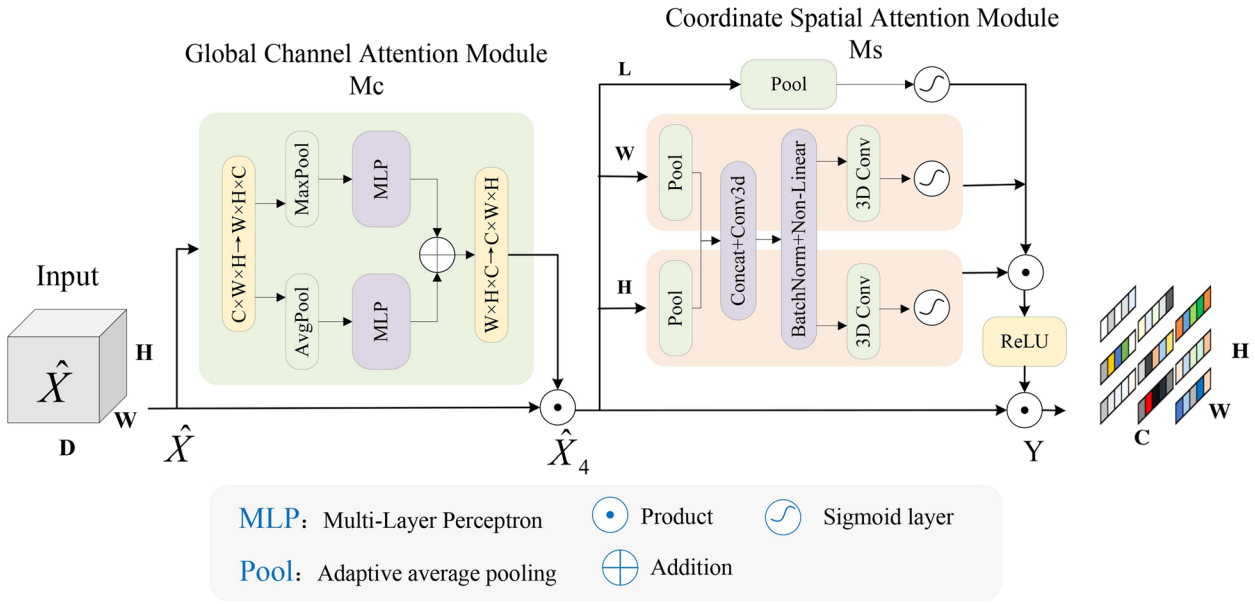

Fig. 3 The 3D global coordinate attention mechanism

that y_i represents the output result of the residual block ($i \in [1, 3]$) in the network, which concurrently functions as the subsequent layer's input denoted as $Block_{i+1}$, and F_i signifies the residual function of $Block_i$. After three layers of channel widening, the final output is y_3. This augmentation can be mathematically represented by Eq. 5.

$$y_3 = k \otimes F_3(X_1, y_2) \tag{5}$$

Equation 5, the solution for y_2 is illustrated by the expression presented in formula 6.

$$y_2 = k \otimes F_2(X_1, y_1) \tag{6}$$

Referring to Eq. 6, the solution for y_1 can be found in formula 7.

$$y_1 = k \otimes F_1(X_1) \tag{7}$$

(4) Normalization and activation functions

To mitigate the risk of overfitting, normalized BatchNorm and activation function are added after $Block_3$. This sequential process yields the resultant output Y from the complete 3D broadened inverted residual network. This outcome is demonstrated in formula 8.

$$Y = ReLU\left(BatchNorm(y_3)\right) \tag{8}$$

3D Global coordinate attention mechanism

In recent years, the integration of attention mechanisms has yielded enhancements in model performance. However, due to information reduction and dimensional segregation, these mechanisms are constrained to utilizing visual representations from limited receptive fields. Conventional attention mechanisms are commonly globally encoded using Global Average Pooling (GAP); nevertheless, the features derived from GAP exhibit limited diversity. When these techniques compress global spatial information into channel descriptors, preserving the spatial positional information essential for capturing spatial structures in visual tasks becomes challenging. In addition, prior attention mechanisms often only capture a single spatial information from one dimension, such as CBAM [35] and DPCA [26] mentioned above. To address these limitations, this paper proposes a 3D Global Coordinate Attention Mechanism, grounded in the Coordinate Attention (CA) [36] approach. This innovative mechanism upholds spatial positional information along both horizontal and vertical coordinates. The GCA attention mechanism encompasses two principal modules: the global channel attention module and the spatial attention module, the architectural depiction of which is illustrated in Fig. 3 Within these modules, the global channel attention module gleans global information via dimension interchange and computes the weight attributed to each channel within the input image. Meanwhile, the spatial coordinate attention module primarily computes spatial positional information from three dimensions: length, width, and height.

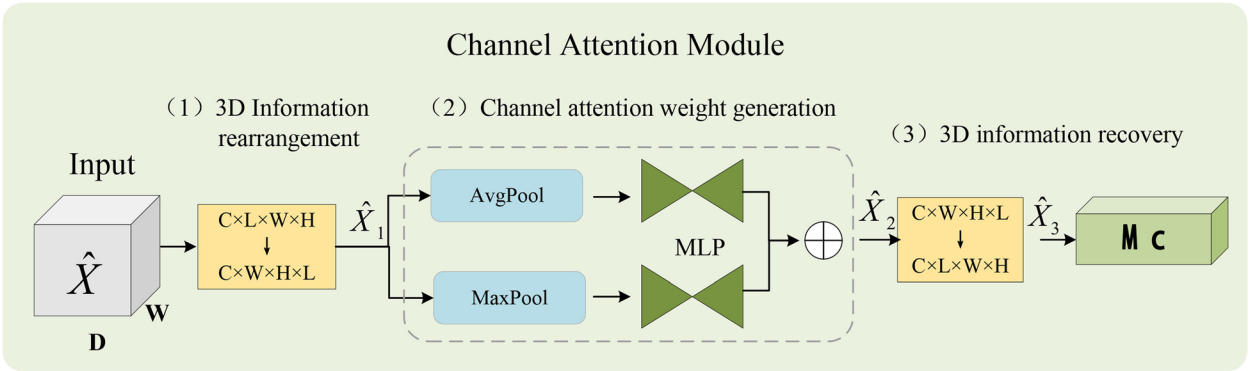

Fig. 4 Global channel attention module

3D Global channel attention module

The implementation of the 3D Global Channel Attention can be primarily divided into three sequential steps. Its module structure is shown in the Fig. 4

(1) Rearrangement of three-dimensional information

$\hat{X} \in R^{C\times L\times H\times W}$ represents the feature map of the input, and L, H, and W stand for the length, height, and width of the input, respectively. C denotes the number of channels within$\hat{X}$. In the channel attention submodule, a three-dimensional arrangement becomes essential to conserve information across the three dimensions. To enhance the global interaction representation and improve the performance of the deep neural network, a reorganization of information across the three dimensions is rearranged. The method of information reorganization is depicted in formula 9.

$$\hat{X}(C\times L\times W\times H)\rightarrow \hat{X}_1(C\times W\times H\times L) \quad (9)$$

(2) Generation of channel attention weights

Following the reorganization of the three-dimensional feature information, a combination of maximum pooling and average pooling techniques is employed to proficiently aggregate significant cues regarding distinctive object features. This amalgamation aims to deduce more refined channel attention. The procedural steps of this process can be succinctly expressed through formula 10.

① $\hat{X}_1$ is employed to consolidate spatial information from feature maps via both average pooling and maximum pooling, yielding two distinct spatial context descriptions.

② MLP is used to learn the characteristics of channel dimensions after pooling and the importance of each channel to amplify the spatial dependence of cross-dimensional channels. This MLP is constructed as a multi-layer perceptron comprising two fully connected layers and a ReLU activation layer.

③ The results obtained from the MLP learning are incorporated and added to produce the feature vector $\hat{X}_2$.

$$\hat{X}_2 = MLP\left(AvgPool\left(\hat{X}_1\right)\right) + MLP\left(MaxPool\left(\hat{X}_1\right)\right) \quad (10)$$

(3) 3D information recovery

The three-dimensional information within $\hat{X}_2$ is restructured to yield $\hat{X}_3$. This process can be represented by the formula 11.

$$\hat{X}_2(C\times W\times H\times L)\rightarrow \hat{X}_3(C\times L\times W\times H) \quad (11)$$

Where $\hat{X}_3$ can be denoted as M_c, signifying the computed outcome of the complete global channel attention module.

Following the execution of the global channel attention procedure, the ultimate output outcome $\hat{X}_4$ from the global channel attention module is derived by element-wise multiplication between the input $\hat{X}$ and M_C of the feature graph. This resultant output then serves as the input for the subsequent coordinate attention phase. This process is succinctly depicted using formula 12.

$$\hat{X}_4 = M_C(\hat{X})\otimes\hat{X} \quad (12)$$

3D spatial coordinate attention module

The implementation of the spatial coordinate attention module can be primarily divided into three distinct steps: direction pooling, sequential convolution, and fusion of features across the three-dimensional space. Its module structure is shown in the Fig. 5.

(1) Three-dimensional space direction pooling

The spatial details encompassing length, width, and height within the output result $\hat{X}_4$ obtained from the channel attention mechanism are independently

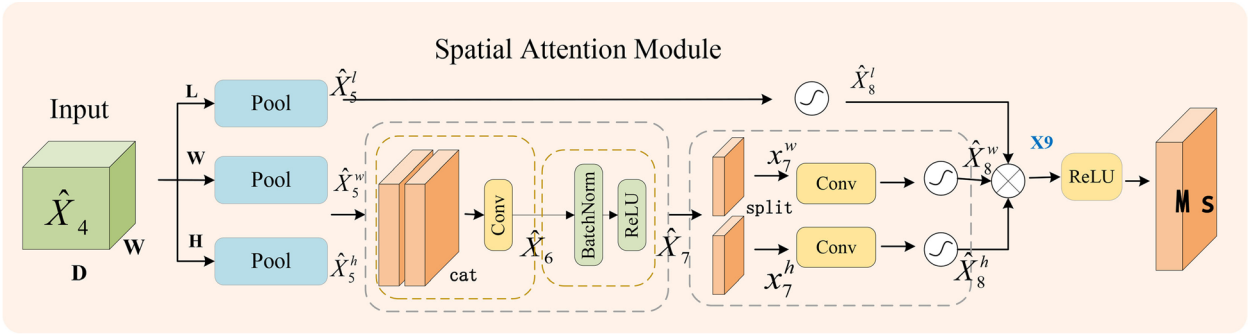

Fig. 5 Space coordinates attention module

subjected to pooling. This strategic approach facilitates the encapsulation of information from all three dimensions within the 3D input image, thereby encoding the spatial characteristics. Subsequently, this process transforms the information into encoded representations of three distinct one-dimensional features. Refer to formulas 13 through 15 for the specific formulations. In the formula, $\hat{X}_5^l$, $\hat{X}_5^w$ and $\hat{X}_5^h$ respectively represent the results after pooling in the three directions of length, width and height.

$$\hat{X}_5^l = AvgPool\left(\hat{X}_4, L\right) \tag{13}$$

$$\hat{X}_5^w = AvgPool\left(\hat{X}_4, W\right) \tag{14}$$

$$\hat{X}_5^h = AvgPool\left(\hat{X}_4, H\right) \tag{15}$$

(2) Series convolution

To comprehensively capture positional information, a concatenation of both broad coordinates and elevated coordinates is employed, merging them along the spatial dimension. This concatenation, denoted as $\hat{X}_6$, is then passed through a shared convolutional transformation function $Conv3D_{1\times1}$ for calculation. Refer to Eq. 16 for a concise representation of this process.

$$\hat{X}_6 = Conv3D_{1\times1}\left(\left[\hat{X}_5^w, \hat{X}_5^h\right]\right) \tag{16}$$

The resultant $\hat{X}_6$ is achieved subsequent to BatchNorm and ReLU processing, as illustrated in Eq. 17.

$$\hat{X}_7 = ReLU\left(BatchNorm\left(\hat{X}_6\right)\right) \tag{17}$$

(3) Fusion of features across three-dimensional space

The spatial information within the three dimensions necessitates re-fusion to extract features that have been individually processed. The specific procedural steps encompass:

① The concatenated spatial dimension information $\hat{X}_7$ is divided into two independent tensors $x_7^h \in R^{C/R\times H}$ and $x_7^w \in R^{C/R\times W}$.

② x_7^h and x_7^w are individually subjected to transformation through three-dimensional convolution denoted as $Conv_{1\times1}$, resulting in tensors with the same count of channels as the input $\hat{X}_4$. This convolutional transformation maps the three dimensions independently to yield $\hat{X}_8^l$, $\hat{X}_8^h$ and $\hat{X}_8^w$. Refer to formulas 18 through 20 for the precise mathematical representations of these transformations.

$$\hat{X}_8^l = Sigmoid\left(\hat{X}_5^l\right) \tag{18}$$

$$\hat{X}_8^h = Sigmoid\left(Conv_{1\times1}\left(x_7^h\right)\right) \tag{19}$$

$$\hat{X}_8^w = Sigmoid\left(Conv_{1\times1}\left(x_7^w\right)\right) \tag{20}$$

③ The three dimensional information $\hat{X}_8^l$, $\hat{X}_8^h$ and $\hat{X}_8^w$ within $\hat{X}_8$ are multiplied to obtain the output feature $\hat{X}_9$ of the three dimensional fusion process. This outcome is illustrated in formula 21.

$$\hat{X}_9 = \hat{X}_8^l \otimes \hat{X}_8^w \otimes \hat{X}_8^h \tag{21}$$

④ Following the application of the ReLU activation function to $\hat{X}_9$, the resultant calculation outcome M_s for the entire spatial coordinate attention module can be derived. This relationship is outlined in Eq. 22.

$$M_S(\hat{X}_4) = ReLU(\hat{X}_9) \tag{22}$$

To amalgamate both the channel and spatial attention submodules, while simultaneously retaining the original input information, the outcome $M_s(\hat{X}_4)$ from the spatial attention phase is multiplied with the input $\hat{X}_4$ from the channel attention phase. This resultant product

is denoted as Y. This process is succinctly captured by formula 23.

$$Y = M_s(\hat{X}_4) \otimes \hat{X}_4 \tag{23}$$

Experiment

Experimental environment and configuration

The hardware employed for all experiments in this study comprises an Nvidia 3080 graphics card, 64GB of random access memory (RAM), and a 2TB solid-state drive (SSD). The operating system utilized is Ubuntu 20.04. The development software stack encompasses PyCharm 2022, PyTorch 1.13.0, and CUDA 11.7. The initial learning rate is configured at 0.0002.

Experimental dataset

The LUNA16 dataset was employed for our experiments. The LUNA16 dataset [37] is a subset of the larger LIDC-IDRI dataset [38], specifically focused on lung nodules. The LIDC-IDRI dataset comprises 1018 low-dose CT images from 1010 patients across different regions. It adheres to internationally recognized standards for benign and malignant lung nodule classification. Following anonymization, experienced radiologists annotated the lesions and sizes of lung nodules, resulting in a standardized dataset. These images were acquired using various imaging protocols, including different kilovoltages (120 kV-140 kV) and slice thicknesses (0.6 mm - 5.0 mm). The LUNA16 dataset consists of 888 images (in mhd format) remaining after excluding CT images from LIDC-IDRI with slice thickness greater than 3mm. Each image in LUNA16 contains a series of axial slices of the chest, which vary depending on the scanning machine and patient characteristics.

Performance evaluation criteria

To evaluate the classification performance of lung nodule, we employed three widely utilized metrics: Accuracy, Sensitivity and Specificity [33]. Accuracy(Acc) represents the probability of being correctly judged for actual positive and actual negative, and is defined by formula (24). Sensitivity(Sens) refers to the proportion of the samples that are actually positive that are correctly judged to be positive. In other words, it quantifies the probability of correctly identifying a patient with a positive condition. Sensitivity is calculated using formula (25) . Specificity(Spec) represents the ratio of correctly classified negative samples to the total number of actual negative samples. It quantifies the probability of accurately diagnosing a patient as negative. The formula for Specificity is given by (26). In these Eqs. 24 to 26, TP, FN, FP, and TN stand for true positive, false negative, false positive, and true negative respectively. Higher values for these evaluation metrics indicate better classification performance.Moreover, ROC curves and AUC values [39] are common metrics used to assess classifier performance, effectively balancing sensitivity and specificity across different thresholds. They serve as widely adopted quantitative tools in research domains.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{24}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{25}$$

$$Specificity = \frac{TN}{FP + TN} \tag{26}$$

Table 1 Comparison of pulmonary nodule classification algorithms

Model	Acc ↑	Sensv ↑	Spec ↑	Para(M)↓
Multi-scale CNN [40]	86.84	-	-	-
Nodule-level 2D CNN [41]	87.3	88.5	86	-
Vanilla 3D CNN [41]	87.4	89.4	85.2	-
Multi-crop CNN [42]	87.14	-	-	-
NAS-Lung [33]	89.56	76.19	<u>89.19</u>	<u>7.84</u>
DeepLung [43]	90.44	81.42	-	141.57
AE-DPN [29]	90.24	**92.04**	88.94	678.69
Fast CapsNet [44]	<u>91.84</u>	89.11	-	52.2
GC-WIR(ours)	**94.32**	<u>91.49</u>	**93.69**	**5.76**

The best and second-best results in each column are shown in bold and underlined, respectively

Experimental results and analysis

Performance comparison with advanced algorithms

(1) Comparison of basic performance indicators

In order to demonstrate the superiority of GC-WIR, we conducted a comparative analysis against other state-of-the-art methods using metrics such as Accuracy, Sensitivity, Specificity, and Parameter Count. The compared methods encompass Multi-scale CNN [38], Vanilla 3D CNN [39], Multi-crop CNN [40], NAS-Lung [33], DeepLung [41], AE-DPN [29], and Fast CapsNet [42]. The comparison results are shown in Table 1.

The experimental outcomes unequivocally demonstrate the notable superiority of our proposed algorithm framework in comparison to other methodologies. Our approach achieves the highest Accuracy and Specificity metrics at 94.32% and 93.69%, respectively. Furthermore, it attains the second-best Sensitivity. Remarkably, GC-WIR surpasses Fast CapsNet [42] by 2.48% in Accuracy and 2.38% in Sensitivity, highlighting its broad applicability and proficiency in advanced pulmonary nodule classification. This remarkable performance is

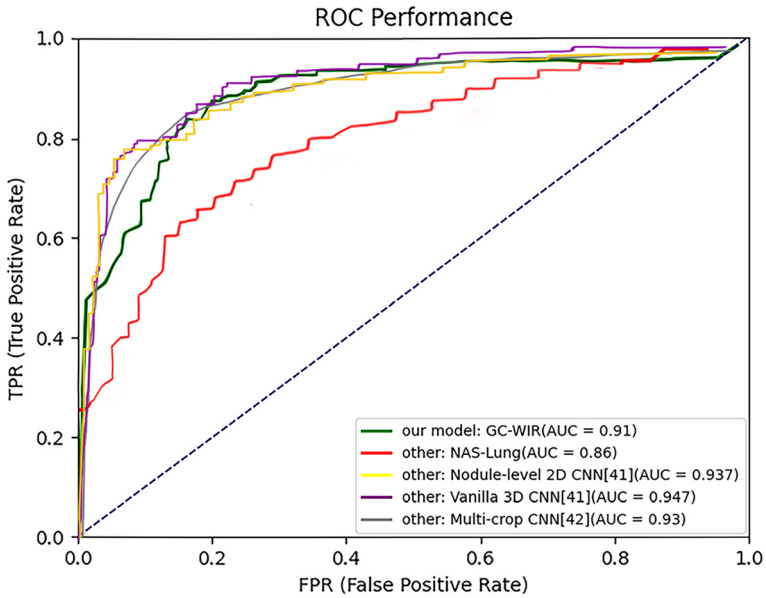

Fig. 6 Comparison of ROC curves of five different models

primarily attributed to the design of a 3D Broadened Inverted Residual Network (3D WIRN) within the GC-WIR framework. This network design augments feature extraction efficacy by increasing the number of input channels. This not only mitigates information loss during feature extraction but also substantially enhances model accuracy.

Compared with the traditional method, this method can enlarge the feature calculation and enhance the convergence without substantially inflating parameter count. It is worth noting that AE-DPN [29] attains optimal Sensitivity due to the implementation of two attention mechanisms to elevate pulmonary nodule classification performance. However, this comes at the cost of a significantly larger model size compared to GC-WIR. In contrast, GC-WIR only necessitates 5.76M parameters to achieve the highest classification accuracy. This efficiency is attributed to the proposed Global Coordinate Attention (GCA) mechanism, which primarily computes input features through 3D information transformation and spatial coordinates. Unlike previous attention mechanisms that extensively use convolution and pooling, our method accommodates both channel and orientation-related positional information in a flexible and lightweight manner. Collectively, these results substantiate that our proposed method attains superior classification performance and holds an advanced standing in the domain of pulmonary nodule classification.

(2) Comparison of the ROC curve

The ROC curve is a tool used to evaluate the performance of binary classifiers by illustrating the relationship between true positive rate (TPR) and false positive rate (FPR) at different thresholds. To evaluate the performance of the proposed algorithm in this context, ROC curves for the GC-WIR model, the NAS-Lung model [33], as well as the Nodule-level 2D CNN [41], Vanilla 3D CNN [41], and Multi-crop CNN [42] models, were plotted for comparison (see Fig. 6). As shown in the figure, the curves for these models all start from the lower-left corner and extend towards the upper-right corner. This indicates that as the FPR increases, the TPR also rises correspondingly for all models, demonstrating their effectiveness in both identifying positive cases and reducing misclassification of negative cases. Within the threshold range of [0, 0.2] for the FPR, the Nodule-level 2D CNN [41], Vanilla 3D CNN [41], and Multi-crop CNN [42] models achieve a TPR above 0.85 more rapidly compared to the GC-WIR model (proposed in this study) and the NAS-Lung model. However, in the threshold range of [0.2, 0.6], the GC-WIR model stabilizes more quickly than other methods, with the exception of the model Vanilla 3D CNN model [41]. The NAS-Lung model [33] shows slightly inferior performance in this range. In the threshold range of [0.6, 1.0], the ROC curves for these models exhibit similar performance, with all models demonstrating excellent classification of both positive and negative samples. Nevertheless, in terms of AUC values, the Nodule-level 2D CNN [41], Vanilla 3D CNN [41], and Multi-crop CNN [42] models achieve AUCs of 93.7%, 94.7%, and 93%, respectively, whereas the GC-WIR model achieves an AUC of 91%. Although the AUC value for the GC-WIR model is not the highest, it achieves the greatest

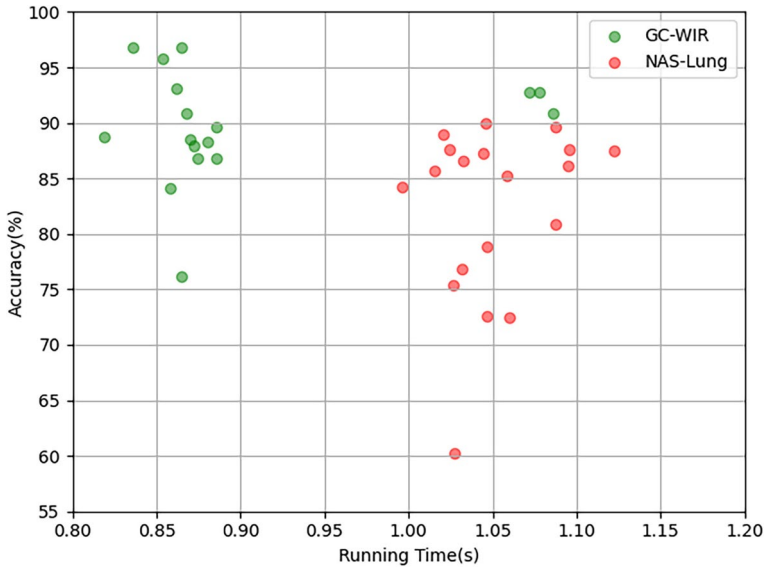

Fig. 7 The comparison of inference speed and corresponding accuracy

classification accuracy of 94.32% (see Table 1). Under similar dataset conditions, the NAS-Lung model [33] also achieves a lower AUC of 86%. This discrepancy may be attributed to differences in the datasets used. The Nodule-level 2D CNN [41], Vanilla 3D CNN [41], and Multi-crop CNN [42] models utilize the LIDC-IDRI dataset, whereas the GC-WIR and NAS-Lung models [33] use the Luna16 lung nodule dataset. The Luna16 dataset includes CT images with multiple benign or malignant nodules, which may lead to an imbalance in the distribution of positive and negative samples. This imbalance could cause performance fluctuations or instability at certain thresholds, thereby affecting the AUC value. Additionally, the use of ten-fold cross-validation in this study may result in variations in AUC values depending on different training and testing splits. Nevertheless, the GC-WIR model demonstrates commendable performance in the ROC curve comparison.

(3) Comparison of inference time and accuracy

To demonstrate the proposed model's lightweight nature alongside its rapid inference capabilities, comparative experiments on inference speed and corresponding accuracy were conducted. For fair comparison, considering the open-source availability of code and datasets based on LUNA16, we performed comparative experiments between the GC-WIR model and the NAS-Lung algorithm [33] over approximately 20 epochs. As depicted in Fig. 7. During each epoch of training, two key metrics were recorded: inference time and model accuracy. In the figure, the horizontal axis represents inference time, indicating the average time the model requires from input data to output prediction, measured in seconds (s); while the vertical axis denotes model performance as Accuracy metric. From the experimental results, it is evident that the GC-WIR model (highlighted in green) averages 0.85 seconds per epoch and achieves an accuracy exceeding 90%. In contrast, under similar hardware conditions, the NAS-Lung model requires an average inference time of 1.05 seconds to achieve 90% training accuracy. This study demonstrates that the GC-WIR model maintains high classification accuracy within shorter inference times. This further underscores the effectiveness of the attention module in GC-WIR, which not only lightweightens the model sufficiently but also effectively reduces inference time during model training.

All these results collectively demonstrate that the proposed method in this study achieves superior classification performance and demonstrates advancement.

Convergence test

To prove the robust convergence of GC-WIR, we conducted tests on both the training and test sets of the LUNA16 dataset. Figure 8 illustrates the classification accuracy trends over 600 iterations for both sets, while Fig. 9 showcases the accuracy outcomes over 300 iterations. As evident from two figures, our network model with only 300 iterations is capable of achieving experimental results comparable to those after 600 iterations. This substantial reduction in computation and experimental costs is attributed to the inherent design of the 3D WIRN within GC-WIR. The incorporation of residual structures allows for direct connections across layers, thus facilitating smoother model convergence. The inverted residual structure performs identity mapping

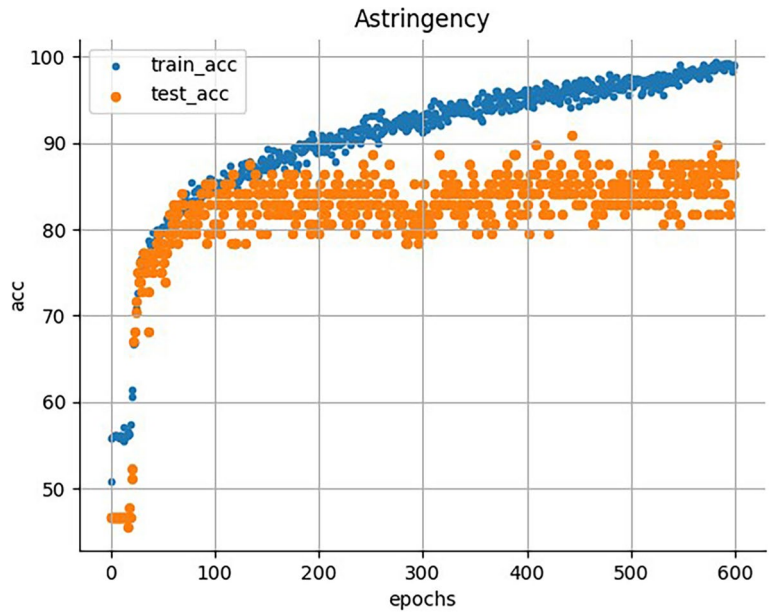

Fig. 8 Classification accuracy graph for 600 iterations of the GC-WIR model

and spatial transformation in higher dimensions, thus effectively reducing information loss and gradient confusion. This enhancement significantly augments model efficiency. In essence, the convergence analysis reaffirms that GC-WIR not only excels in achieving high classification accuracy but also demonstrates efficient and expedited convergence behaviors.

Stability test

To establish the stability of the proposed algorithm GC-WIR algorithm, we conducted a 10-fold cross-validation experiment on the LUNA16 dataset [37]. In this experiment, the entire dataset was partitioned into 10 subsets, with each subset containing 89 CT case images. One subset was randomly reserved as the validation dataset for the validation procedure, while the remaining nine

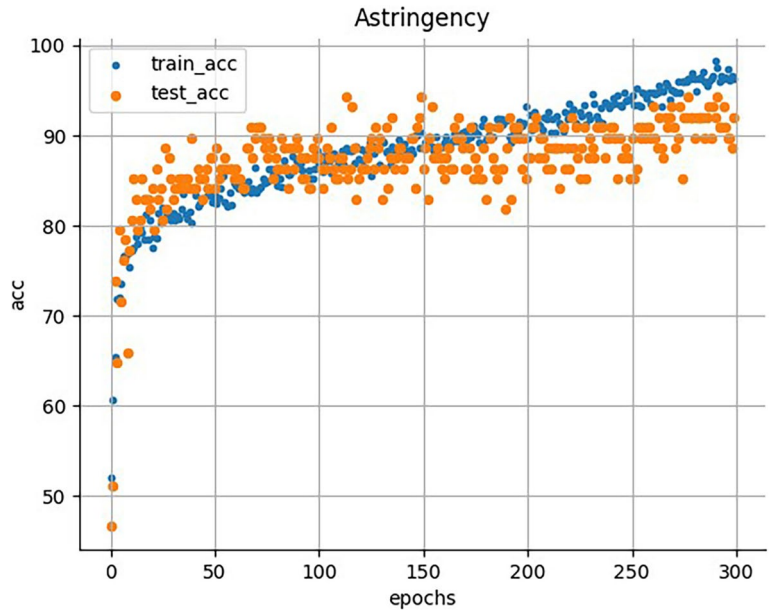

Fig. 9 Classification accuracy graph for 300 iterations of the GC-WIR model

subsets were used for training. The ten-fold cross-validation was executed in the following four steps:

(1) One category was randomly selected for validation, while the remaining categories were used for training. The samples in the training set and the validation set were non-overlapping.

(2) The model described in the article is trained using the data from the training set. After each epoch of training, both training accuracy and validation accuracy are computed for comparison. The purpose is to evaluate the model's generalization ability and performance during training, in order to select the optimal model configuration.

(3) Subsequently, the trained model is evaluated using the test set, and this evaluation generates the resultant output file.

(4) This cross-validation process is repeated 10 times, with each subset validated once. After completing ten-fold cross-validation, the aggregated results from these ten iterations yield a single estimation.

To ensure fairness, the open-source NAS-Lung [33] algorithm model was reproduced under equivalent hardware conditions in this paper. Both the NAS-Lung and GC-WIR models underwent training and comparison across these ten distinct categories. The resulting changes in accuracy were charted and visualized in Fig. 10. It's important to note that the data used for each model was randomly extracted from the LUNA16 dataset. As apparent from the chart, these models exhibit divergent behaviors. With the exception of classification 1, the accuracy of GC-WIR exceeded that of NAS-Lung, which remained stable at about 90%. Moreover, the break line trend of GC-WIR is relatively smooth, without a lot of fluctuations. On the contrary, NAS-Lung demonstrated a more varied accuracy pattern, with the lowest accuracy dropping below 80%. This experiment underscores that the proposed GC-WIR model demonstrates robust stability and broad applicability across different data classifications. Nonetheless, there is still potential for further improvement in performance.

Table 2 Acc, Sens, Spec and Para(M) in ablation experiments

Model	Acc ↑	Sensv ↑	Spec ↑	Para(M)↓
Base(Res+CBAM)	87.78	71.43	89.19	7.84
Res+GAM	90.38	87.75	80.63	11.12
Res+CA	89.42	85.71	80.81	8.85
Res+GCA	**90.38**	75.3	87.36	**2.7**
GCA+WRN	91.34	85.72	82.92	4.35
GCA+WIRN(GC-WIR)	**94.32**	**91.49**	**93.69**	5.76

The best results in each column are shown in bold

Ablation study

The algorithm GC-WIR described in this paper consists of four stages, where the first three stages leverage the joint application of two technologies: 3D Global Coordinate Attention (3D GCA) and 3D Wide Inverted Residual Network (3D WIRN). To showcase that the optimal results are achieved only when both of these methods are combined, a set of ablation experiments was conducted to explore the individual contributions of these techniques. The results of the ablation experiments are presented in Table 2. Within the same algorithm framework, the model formed by the combination of residual network (ResNet) and attention module CBAM serves as the base case, denoted as "Base (Res+CBAM)". It is observed that the accuracy of this base model only reaches 87.78%, which falls short of the accuracy requirements for precise pulmonary nodule classification.

In this study, the Residual Network (ResNet) is integrated with global attention [45], coordinate attention [36], and 3D Global Coordinate Attention (3D GCA). These composite network models are designated as "Res+GAM", "Res+CA, and "Res+GCA" respectively. Upon conducting a thorough comparison among these models, it has been deduced that 3D GCA attains the highest classification

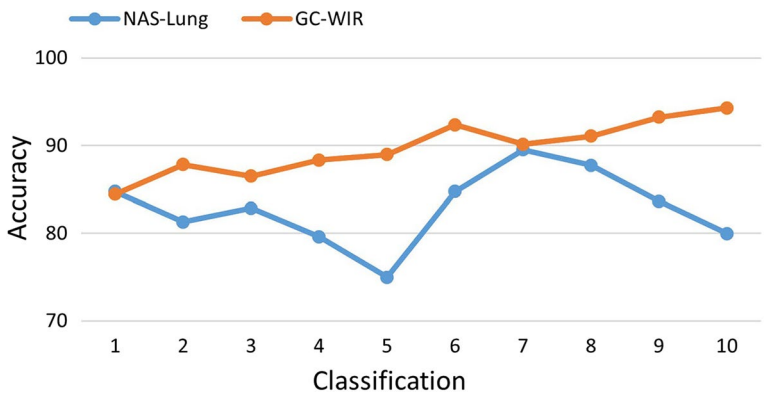

Fig. 10 Stability comparison of GC-WIR and NAS-Lung [33] on the LUNA16 datasetl

accuracy while utilizing the least number of parameters. Its peak accuracy can reach 90.38%, with a mere 2.7 M parameters, rendering it both adaptable and lightweight. The models "GCA+WRN" and "GCA+WIRN" denote the amalgamation of 3D GCA with widened residual networks [34] and widened inverted residual networks respectively. It's evident that the introduction of 3D WIRN yields significant enhancements in accuracy, sensitivity, and specificity through the utilization of the inverted residual structure. However, this improvement comes with a parameter increment of 3.06 M as compared to the residual network, which is unavoidable. The quest to further reduce model parameters and computational demands remains a focus for future work. In summation, the optimum classification outcomes are only attainable through the combined utilization of 3D GCA and 3D WIRN, effectively avoiding the challenge of excessive parameter growth.

Discussion

This paper investigates a 3D GC-WIR model built upon deep learning technology to discern between benign and malignant small pulmonary nodules. The model's design is primarily rooted in the connectivity patterns and attention mechanisms of neural networks, enabling it to achieve precise pulmonary nodule classification with flexibility, efficiency, and rapid convergence.

In previous algorithms, researchers typically employed Convolutional Neural Network (CNN) models for the classification and processing of lung nodule images. CNNs primarily adopt a local connection and weight-sharing design, enhancing the network's feature extraction capabilities. However, CNNs focus on a neural network's characteristics as a group of neurons, with each input corresponding to a single output. This tendency can render the fully connected mode of CNNs redundant and inefficient, especially when dealing with networks with a high number of layers, making training challenging and susceptible to issues like gradient explosion and gradient vanishing. To address these issues, the Residual Network (ResNet) architecture introduced skip connections, which effectively circumvent problems related to deep networks. However, as network models have continued to evolve, architectures like ResNet may inadvertently result in the loss of certain feature information for small target detection, consequently diminishing recognition performance. In order to tackle this challenge, we have designed the 3D Inverted Residual Network (WIRN). This novel architecture primarily incorporates three widened residual blocks to enlarge the channel connections, thereby reducing network depth while simultaneously increasing network width. Additionally, within each widened residual block, we apply the inverted connection structure, which engages in identity mapping and spatial transformation in higher dimensions. This strategy effectively mitigates information loss and gradient obfuscation, thereby enhancing the model's learning capability.

In many images, the feature information often presents a complex and intricate pattern. When utilizing a Convolutional Neural Network (CNN) to process images, our objective is to have it selectively extract the crucial input features, rather than all of them. Consequently, the task of ensuring that CNNs focus their attention on the key features becomes of paramount importance. The incorporation of an attention mechanism provides a means to achieve adaptive attention within the network, and its application serves to eliminate extraneous information in the input image. Currently, the channel attention mechanism enables the calculation of the significance of each channel in the input image, thereby enhancing the network's feature representation capabilities. Meanwhile, spatial attention seeks to elevate the representation of critical regions, transforming the spatial information within the input image while preserving essential details. In this paper, we introduce the 3D Global Coordinate Attention mechanism (GCA), which not only combines the strengths of both channel attention and spatial attention but also deviates from prior attention mechanisms employing convolution and pooling. The GCA operates by exchanging 3D information to acquire global channel features. It captures spatial coordinate information across three dimensions, generating weighted values that modulate the output. As a result, the specific region of interest is enhanced, while the irrelevant background information is attenuated.

The combination of 3D WIRN and 3D GCA synergistically harnesses their respective strengths, resulting in enhanced recognition capabilities for pulmonary nodule images and achieving optimal classification performance. This assertion is further substantiated by ablation experiments. To evaluate the effectiveness of this approach, we compared it with five other deep learning methods, namely Vanilla 3D CNN [39], NAS-Lung [33], DeepLung [41], AE-DPN [29], and Fast CapsNet [42]. While a method based on two-dimensional Convolutional Neural Networks (CNN) has been proposed and applied to extract stratification and discriminant features for pulmonary nodule classification, it's important to note that most CT medical images are three-dimensional data. Consequently, the utilization of 3D CNNs on CT images is expected to improve diagnostic accuracy significantly. Indeed, the introduction of Vanilla 3D CNN [39] successfully validated this point. When compared to 2D methods at the slice level, 3D CNNs at the nodule level can effectively integrate nodule-level features and context features across three dimensions in 3D patches to some extent, resulting in improved performance. NAS-Lung [33] combines residual networks with Convolutional

Block Attention Module (CBAM) to achieve suboptimal specificity, demonstrating the efficacy of the attention mechanism in image processing. However, it does not notably enhance sensitivity. DeepLung [41] introduces two deep 3D Dual Path Networks (DPN) for nodule detection and classification. While it offers high performance and low resource utilization, its network structure is complex, and computational demands are substantial. AE-DPN [29] employs two sensitivity-optimal attention mechanisms, but it involves a significantly larger number of parameters compared to all other models included in the comparison.

Compared to the aforementioned models, GC-WIR achieves the highest classification accuracy with only 5.76 M parameters, positioning it at an advanced level in the realm of pulmonary small nodule classification. This remarkable feat can be attributed to GC-WIR's ingenious design, specifically the utilization of 3D WIRN, which enhances feature calculations by widening channels while simultaneously reducing the depth of the residual network. Consequently, this approach mitigates information loss during feature extraction, leading to improved classification accuracy, ultimately reaching a peak accuracy of 94.32%. Additionally, it accelerates model convergence and enhances overall stability. Furthermore, GC-WIR incorporates a flexible and lightweight attention mechanism known as 3D GCA. By converting 3D information and calculating input image features across multiple dimensions, this technique avoids the excessive use of convolution and pooling layers, resulting in a reduction in the number of parameters and computational workload.

This study has several limitations that warrant consideration. Firstly, the experimental data used in this study were obtained from the LUNA16 dataset, meaning that all the images used for training and testing originated from the same geographical region. However, due to the limited availability of publicly accessible datasets pertaining to pulmonary nodules on the Internet, it is challenging to ensure the robustness and practicality of the model. In order to enhance the model's reliability, a larger and more diverse set of sample examples for training and testing would be highly beneficial. Furthermore, since we utilized the publicly available LUNA16 dataset to validate the superiority of our model, the imaging protocols in this dataset vary due to differences in scanners and patients. Therefore, we did not account for parameters such as kV and collimation that could affect accuracy. These tasks also require the collection, scanning, and comparative analysis of real-world data. In the future, we aim to gather and curate authentic lung nodule imaging data to construct a dedicated medical CT image dataset for KV, thickness, and collimation. This effort will facilitate further analysis of the influence of various data acquisition factors on classification accuracy. Secondly, expanding the sample size by incorporating CT images of small pulmonary nodules with distinctive characteristics could potentially lead to an increase in the sensitivity index of the model, further enhancing its performance. In addition, we will verify the quantity and characteristics of positive and negative samples in the dataset, and enhance feature informativeness and classifier discriminability through feature engineering methods such as feature scaling, dimensionality reduction, or feature combination. In the future, we may also explore ensemble techniques to combine predictions from multiple classifiers and apply them across various datasets. This approach typically stabilizes the ROC curves of models and further improves the AUC value. Particularly, by selecting appropriate features and models based on dataset characteristics, continuous improvement of evaluation metrics will be pursued through experimental validation and results analysis. Lastly, it is important to note that the model presented in this paper is primarily focused on CT images. Future research endeavors may explore the integration of other multimodal images to broaden the scope and capabilities of the model.

Conclusion

In this paper, we introduce GC-WIR, a model designed for the precise classification of pulmonary nodules, characterized by high precision, stability, rapid convergence, and parameter efficiency. GC-WIR addresses the issues associated with complex and unstable traditional network models, limited data adaptability, and excessive parameter counts through the incorporation of 3D WIRN to enhance feature calculation. Additionally, GC-WIR integrates a versatile and lightweight attention mechanism known as 3D Global Coordinate Attention (3D GCA). A comprehensive set of experiments was conducted using the LUNA16 dataset. The results demonstrate that, in comparison to advanced deep learning methods, GC-WIR achieves the highest classification accuracy while utilizing the fewest parameters, ultimately yielding superior classification prediction results.

Acknowledgements

The authors would like to acknowledge the LUNA16 dataset for providing valuable data used in this study. Also, The authors are grateful to the editor and reviewers for their constructive comments, which have significantly improved this work.

Authors' contributions

W.W designed the initial research methodology, constructed the experimental design, and participated in the discussion and revision of the paper. S.Y wrote the first draft of the paper and built the algorithm model, and participated in the discussion and revision of the paper. She is responsible for the planning, design and implementation of the entire study. F.Y was involved in the design of the study and the collection of the dataset. Y.C and L.Z is responsible for

data collection, processing and analysis. H.Y vouched for the integrity and accuracy of the research.

Funding
This work was supported in part by the National Natural Science Foundation of China (8207070786), Young Scientists Fund of the National Natural Science Foundation of China (82302188), Shanghai Pujiang Program (22PJD069), Shanghai Health Research Foundation for Talents (2022YQ060), State Commission of Science Technology of Shanghai (22Y11911100), Shanghai Municipal Health Commission (20204Y0201), and Medical engineering cross project from University of Shanghai for Science and Technology(1021309706). The funders had no involvement in the collection, analysis or interpretation of data, nor in the writing of the report and in the decision to submit the article for publication.

Availability of data and materials
The LUNA16 dataset used in this study was available online at https://www.sciencedirect.com/science/article/pii/S1361841517301020.

Data availability
Code related to the GC-WIR model can be available online at https://github.com/shuyaYin2018/open-code-of-GC-WIR.git.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]University of Shanghai for Science and Technology, Jungong 516 Rd, Shanghai 200093, China. [2]Department of Radiology, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, Huaihai West Road NO.241, Shanghai 200030, China.

Received: 29 February 2024 Accepted: 4 September 2024
Published online: 20 September 2024

References

1. Zhao J, Zhang C, Li D, Niu J. Combining multi-scale feature fusion with multi-attribute grading, a CNN model for benign and malignant classification of pulmonary nodules. J Digit Imaging. 2020;33:869–78.
2. Zhu L, Zhu H, Yang S, Wang P, Yu Y. HR-MPF: high-resolution representation network with multi-scale progressive fusion for pulmonary nodule segmentation and classification. EURASIP J Image Video Process. 2021;2021(1):1–26.
3. Zhang X, Wang K, Zhang X, Huang S. Pulmonary Nodule Classification of CT Images with Attribute Self-guided Graph Convolutional V-Shape Networks. In: PRICAI 2021: Trends in Artificial Intelligence: 18th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2021, Hanoi, Vietnam, November 8–12, 2021, Proceedings, Part I 18. Springer; 2021. pp. 280–292.
4. Qiu J, Liu J, Shen Y. Computer Vision Technology Based on Deep Learning. In: 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), vol. 2. IEEE; 2021. pp. 1126–1130.
5. Tripathi A, Goel A. A Survey on Exploring Deep Learning in Medical Image Processing. In: 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). IEEE; 2021. pp. 412–418.
6. Zhang F, Song Y, Cai W, Lee MZ, Zhou Y, Huang H, et al. Lung nodule classification with multilevel patch-based context analysis. IEEE Trans Biomed Eng. 2013;61(4):1155–66.
7. Gao D, Nie S. Method for identifying benign and malignant pulmonary nodules combing deep convolutional neural network and hand-crafted features. Acta Opt Sin. 2020;40(24):2410002.
8. Tan J, Huo Y, Liang Z, Li L. A comparison study on the effect of false positive reduction in deep learning based detection for juxtapleural lung nodules: CNN VS DNN. In: Proceedings of the Symposium on Modeling and Simulation in Medicine,The Society for Modeling and Simulation International, 11315 Rancho Bernardo Road, Suite 139 San Diego, California 92127. 2017. p. 1–8.
9. Eun H, Kim D, Jung C, Kim C. Single-view 2D CNNs with fully automatic non-nodule categorization for false positive reduction in pulmonary nodule detection. Comput Methods Prog Biomed. 2018;165:215–24.
10. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, Van Riel SJ, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging. 2016;35(5):1160–9.
11. Monkam P, Qi S, Xu M, Li H, Han F, Teng Y, et al. Ensemble learning of multiple-view 3D-CNNs model for micro-nodules identification in CT images. IEEE Access. 2018;7:5564–76.
12. Lima TJB, de Araiújo FHD, de Carvalho Filho AO, Rabêlo RdAL, Veras RdMS, Mathew MJ, Evaluation of data balancing techniques in 3D CNNs for the classification of pulmonary nodules in CT images. In: 2020 IEEE Symposium on Computers and Communications (ISCC). IEEE; 2020. pp. 1–6.
13. Gupta A, Saar T, Martens O, Le Moullec Y, Sintorn IM. Detection of pulmonary micronodules in computed tomography images and false positive reduction using 3D convolutional neural networks. Int J Imaging Syst Technol. 2020;30(2):327–39.
14. Kadhim OR, Motlak HJ, Abdalla KK, Developing a CAD System to Detect Pulmonary Nodules from CT-Scan Images via Employing 3D-CNN. In: 2021 2nd Information Technology To Enhance e-learning and Other Application (IT-ELA). IEEE; 2021. pp. 136–41.
15. Xie D, Tang C, Li Y, Liu X, Zhuang M. Pulmonary nodules detection via 3D multi-scale dual path network. In: 2021 7th International Conference on Computer and Communications (ICCC). IEEE; 2021. pp. 980–984.
16. Zhao D, Liu Y, Yin H, Wang Z. A novel multi-scale CNNs for false positive reduction in pulmonary nodule detection. Expert Syst Appl. 2022;207:117652.
17. Deng W, Wang Z, Ren X, Zhang X, Wang B, Yang T. YOLO_v3-Based Pulmonary Nodules Recognition System. In: The 10th International Conference on Computer Engineering and Networks. Springer; 2021. pp. 11–19.
18. Zhang Y, Zhao J, Wu W, Qiang Y, Jia L. Multi-level learning based on 3D CT image integrated medical clinic information for accurate diagnosis of pulmonary nodules. Concurr Comput Pract Experience. 2022;34(17):e6998.
19. Lin Z, Zheng J, Hu W. Using 3D convolutional networks with shortcut connections for improved lung nodules classification. In: Proceedings of the 2020 2nd International Conference on Big Data Engineering. 1601 Broadway, 10th Floor New York, NY 10019-7434: Association for Computing Machinery; 2020. p. 42–49.
20. Bharti M, Choudhary J, Singh DP. Detection and Classification of Pulmonary Lung Nodules in CT Images Using 3D Convolutional Neural Networks. In: 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1. IEEE; 2022. pp. 1319–1324.
21. Tong C, Liang B, Su Q, Yu M, Hu J, Bashir AK, et al. Pulmonary nodule classification based on heterogeneous features learning. IEEE J Sel Areas Commun. 2020;39(2):574–81.
22. Han Y, Qi H, Wang L, Chen C, Miao J, Xu H, et al. Pulmonary nodules detection assistant platform: an effective computer aided system for early pulmonary nodules detection in physical examination. Comput Methods Prog Biomed. 2022;217:106680.
23. Yuan H, Wu Y, Cheng J, Fan Z, Zeng Z. Pulmonary nodule detection using 3-D residual U-Net oriented context-guided attention and multi-branch classification network. IEEE Access. 2021;10:82–98.
24. Hu X, Gong J, Zhou W, Li H, Wang S, Wei M, et al. Computer-aided diagnosis of ground glass pulmonary nodule by fusing deep learning and radiomics features. Phys Med Biol. 2021;66(6):065015.
25. Moreno A, Rueda A, Multi-Scale Martinez F. A, Network Self-Attention, to Discriminate Pulmonary Nodules. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). IEEE; 2022. pp. 1–4.
26. Ruinan W, Muqing W. DPCA-Net: dual path with 3D channel attention for pulmonary nodule detection. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC). IEEE; 2020. pp. 1186–1190.

27. Qi Y, Gu J, Li W, Tian Z, Zhang Y, Geng J. Pulmonary nodule image super-resolution using multi-scale deep residual channel attention network with joint optimization. J Supercomput. 2020;76:1005–19.
28. Mai J, Wang M, Zheng J, Shao Y, Diao Z, Fu X, et al. MHSnet: Multi-head and Spatial Attention Network with False-Positive Reduction for Pulmonary Nodules Detection. 2022. arXiv preprint arXiv:2201.13392.
29. Jiang H, Gao F, Xu X, Huang F, Zhu S. Attentive and ensemble 3D dual path networks for pulmonary nodules classification. Neurocomputing. 2020;398:422–30.
30. Dong T, Wei L, Ye X, Chen Y, Hou X, Nie S. Segmentation of ground glass pulmonary nodules using full convolution residual network based on atrous spatial pyramid pooling structure and attention mechanism. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi. 2022;39(3):441–51.
31. Zhang G, Zhang H, Yao Y, Shen Q. Attention-Guided Feature Extraction and Multiscale Feature Fusion 3D ResNet for Automated Pulmonary Nodule Detection. IEEE Access. 2022;10:61530–43.
32. Zhang W, Cui L. Detection algorithm of pulmonary nodules based on deep learning. In: 2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). IEEE; 2021. pp. 185–188.
33. Jiang H, Shen F, Gao F, Han W. Learning efficient, explainable and discriminative representations for pulmonary nodules classification. Pattern Recog. 2021;113:107825.
34. Zagoruyko S, Komodakis N. Wide residual networks. 2016. arXiv preprint arXiv:1605.07146.
35. Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). arXiv; 2018. p. 3–19.
36. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2001 L Street N.W., Suite 700, Washington, DC 20036-4928: IEEE Computer Society; 2021. p. 13713–13722.
37. Setio AAA, Traverso A, De Bel T, Berens MS, Van Den Bogaard C, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal. 2017;42:1–13.
38. Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys. 2011;38(2):915–31.
39. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
40. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. In: Information Processing in Medical Imaging: 24th International Conference, IPMI 2015, Sabhal Mor Ostaig, Isle of Skye, UK, June 28-July 3, 2015, Proceedings 24. Springer; 2015. pp. 588–599.
41. Yan X, Pang J, Qi H, Zhu Y, Bai C, Geng X, et al. Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2d and 3d strategies. In: Asian Conference on Computer Vision. Springer; 2016. pp. 91–101.
42. Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, et al. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recogn. 2017;61:663–73.
43. Zhu W, Liu C, Fan W, Xie X, Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In: 2018 IEEE winter conference on applications of computer vision (WACV). IEEE; 2018. pp. 673–81.
44. Mobiny A, Van Nguyen H. Fast capsnet for lung cancer screening. In: International conference on medical image computing and computer-assisted intervention. Springer; 2018. pp. 741–749.
45. Liu Y, Shao Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions. 2021. arXiv preprint arXiv:2112.05561.

Publisher's Note