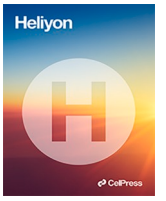
Heliyon 10 (2024) e23508

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Attention-guided deep neural network with a multichannel architecture for lung nodule classification

Rong Zheng [a], Hongqiao Wen [b], Feng Zhu [c,d], Weishun Lan [e,*]

[a] Department of Gynecology, Maternal and Child Health Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430070, China
[b] School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China
[c] Department of Cardiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[d] Clinic Center of Human Gene Research, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[e] Department of Medical Imaging, Maternal and Child Health Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430070, China

ARTICLE INFO

Keywords:
Attention mechanism
Deep learning
lung nodule classification

ABSTRACT

Detecting and accurately identifying malignant lung nodules in chest CT scans in a timely manner is crucial for effective lung cancer treatment. This study introduces a deep learning model featuring a multi-channel attention mechanism, specifically designed for the precise diagnosis of malignant lung nodules. To start, we standardized the voxel size of CT images and generated three RGB images of varying scales for each lung nodule, viewed from three different angles. Subsequently, we applied three attention submodels to extract class-specific characteristics from these RGB images. Finally, the nodule features were consolidated in the model's final layer to make the ultimate predictions. Through the utilization of an attention mechanism, we could dynamically pinpoint the exact location of lung nodules in the images without the need for prior segmentation. This proposed approach enhances the accuracy and efficiency of lung nodule classification. We evaluated and tested our model using a dataset of 1018 CT scans sourced from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI). The experimental results demonstrate that our model achieved a lung nodule classification accuracy of 90.11 %, with an area under the receiver operator curve (AUC) score of 95.66 %. Impressively, our method achieved this high level of performance while utilizing only 29.09 % of the time needed by the mainstream model.

1. Introduction

Lung cancer continues to hold the title as the most frequently identified cancer and the primary contributor to cancer-related fatalities on a global scale. The timely detection of lung cancer through chest computed tomography (CT) scans can significantly enhance the chances of survival for individuals diagnosed with this disease [1]. A "lung spot" observed on a chest CT scan is referred to as a lung nodule, which can either be benign or malignant [2]. Malignant nodules are the primary culprits behind lung cancer and closely resemble benign nodules during the initial stages. Consequently, it is both essential and demanding to precisely distinguish between benign and malignant lung nodules.

* Corresponding author.
E-mail addresses: oscinho@yahoo.co.uk, lws7375@126.com (W. Lan).

https://doi.org/10.1016/j.heliyon.2023.e23508
Received 7 June 2023; Received in revised form 15 November 2023; Accepted 5 December 2023
Available online 10 December 2023

Lately, the dominant approach has been to employ computer-aided diagnosis systems (CADs) for the automated classification of lung nodules. Typically, image-based lung CADs follow a sequence of steps, which include nodule delineation, extracting relevant features, and the actual classification process [3,4]. The proposed method first performs nodule segmentation from the original chest CT images, followed by feature extraction from each segmented nodule. Subsequently, a classifier is trained to classify the nodules as benign or potentially malignant based on the extracted features. As a prerequisite for nodule feature extraction, nodule segmentation is a time-consuming process and can influence the performance of lung nodule classifiers. Incorrect nodule delineation can result in the extraction of unreliable features, which in turn can produce inaccurate results.

At present, the adoption of deep learning techniques [5] in the field of biomedicine, particularly for the purposes of detection and classification, has garnered significant interest among researchers and radiologists. Building on the remarkable achievements of deep learning in various domains like image classification and speech recognition, a growing number of advanced models and theories have been put forth for computer-aided diagnosis systems focused on lung-related tasks. Lately, attention mechanisms have demonstrated significant success in natural language processing (NLP) and detailed image recognition [6]. In image recognition, the attention model not only aids in pinpointing a specific area but also amplifies diverse representations of objects within that area. This highlights the importance of models incorporating attention mechanisms [7]. perform better in complex scenery such as lung CADs. Models using attention mechanisms notably improve the analysis performed in comparison to counterparts without attention regulation. It can be very helpful to biomedical image analysis by regulating the attention of the deep models properly [8,9].

In the rest of the paper, the application of deep learning and attention mechanism models in lung nodule classification, along with the models proposed by us, is discussed in Section 2. Section 3 provides details regarding the dataset, the specific architecture of the models, and the implementation methods. Section 4 presents an analysis of relevant metrics and comparative experiments. Some conclusions are listed in Section 5.

2. Related work

As mentioned in the introduction, deep learning methods and attention mechanisms have shown great potential in accurately detecting and distinguishing lung nodules. For example, Xie et al. [10] used the GLCM-based texture descriptors and Fourier shape descriptor to explore the nodule's heterogeneity in voxel values and heterogeneity in shapes, respectively, and combined both descriptors with the information learned by a nine-layer Convolutional Neural Networks (CNN) for lung nodule classification at the decision level. Setio et al. [11] developed a false positive reduction method constituting a combination of committee-fusion, late-fusion and mixed-fusion of nine 2D CNNs acting over 2D patches that were extracted using nine views of a volumetric object. However, the prior studies typically focused on obtaining higher scores of classification results by extracting richer visual features or using deeper and wider neural networks. All these methods require more expertise and more time for computation. Shen et al. [12] proposed a multi-scale CNN that captures lung nodule heterogeneity via extracting discriminative features from alternatingly stacked layers. Nobrega et al. [13] used transfer-learning method to evaluate the effects of the combination of different CNN models and classifiers on lung nodule classification. Despite the advantages of deep learning techniques in simplifying the procedure of lung nodule analysis, the accuracy of lung nodule classification still needs improvement. Takuma et al. [14] evaluated the effectiveness of the exponentiation method for improving the performance of a CNN in the task of classifying lung nodules on CT images according to malignancy level. Experimental results improved accuracy of the CNN and the ability to adjust sensitivity and specificity by selecting the exponent value. Yan et al. [15] improved the accuracy and efficiency of CT scanning of lung nodules by combining convolutional neural networks with the Snake optimization algorithm, and demonstrated its effectiveness by comparing it with other methods in the IQ-QTH/NCCD-Lung Cancer Dataset. Fang et al. [16] proposed an automatic detection framework for pulmonary nodules based on two-dimensional convolutional neural networks. To overcome the limitations of three-dimensional neural networks that take a long time and reconstruct images that may cause information leakage, the framework was fine-tuned using the LUNA16 dataset to obtain the best mAP50 value and the second-best recall value. Fu et al. [17] proposed an attention-enhanced multi-task network model, which introduces an attention mechanism and multi-task learning. The multi-task learning module can make the network benefit from multiple tasks by sharing features and attention-enhancing modules to improve the overall performance, and demonstrated significant performance on the well-known dataset LIDC-IDR. Zhao et al. [18] propose adaptive and attentive 3D Convolutional Neural Network for automatic detection of pulmonary nodules. The model is divided into two stages. In the first stage, the proposed 3D CNN approach fuses high resolution features to extract the local detail shape information of pulmonary nodules. The Res-class 3D CNN exploits the multi-scale 3D convolution kernel to extract the multilevel contextual information of nodules. The effectiveness of the model is demonstrated on the LUNA16 dataset. Without using any extra localization modules or processing steps, both global and local representations of input images can be noticed simultaneously by deep models with an attention mechanism. Such ability is equivalent to stimulating deep models to observe the input image more "carefully," which is much more valuable for biomedical image analysis in small sample cases.

The above provided an overview of research on the automatic detection of pulmonary nodules using image processing and computer vision techniques. In the current study, we present a novel model aimed at enhancing the accuracy of nodule recognition. The primary contribution of this research lies in its potential to enhance early cancer detection, expedite timely medical interventions, and contribute to the advancement of medical image analysis.

This study introduces a deep neural network model featuring a multi-channel attention mechanism designed for the classification of pulmonary nodules in chest CT images. When applied to the task of lung nodule classification, the attention mechanism allows for a specific focus on the nodule of interest while disregarding other lung structures that are unrelated to the classification objective. By focusing on key areas of the image, such as lung nodules, and ignoring unimportant areas like blood vessels and lung walls, the attention model achieves reliable classification results. Unlike nodule segmentation, the attention mechanism can be trained and

optimized during the training process, which reduces the time consumption. Additionally, the use of a multiview and multiscale strategy sufficiently quantifies nodule characteristics, leading to accurate classification results. Extensive experiments were conducted on the publicly available dataset Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [19–21]. The experimental findings illustrate that our suggested approach surpasses the performance of other cutting-edge deep learning models. The significance and contributions of this study are as follows.

- Improved Pulmonary Nodule Detection: By proposing a new approach that combines attention mechanisms with CNN models, our aim is to enhance the accuracy and efficiency of CT image detection.
- Automation and Efficiency: Manual identification of medical images is time-consuming and prone to human errors. Our model, combining attention mechanisms with CNN models, not only achieves reliable classification results by focusing on crucial regions of the image but also reduces time consumption compared to nodule segmentation methods.
- Comparative Research Analysis: To demonstrate the effectiveness of our approach, we conducted comparisons and evaluations of different models using the same dataset.
- Potential Clinical Significance: The CNN model combined with attention mechanisms could have significant clinical implications. Early detection of malignant nodules in CT images can aid doctors in formulating timely treatment strategies for patients.

3. Materials and methods

3.1. Dataset

Within the Cancer Imaging Archive (TCIA), the LIDC-IDRI database comprises 1018 chest CT scans, each with varying numbers of slices, all sized at 512x512. For every scan, four seasoned thoracic radiologists annotated the nodule locations and associated characteristics, including attributes like calcification and malignancy, in an XML file linked to that scan. The malignancy of each nodule was assessed using a 5-point scale, ranging from benign to malignant. To decrease the influence of inconsistent evaluations of nodule malignancy, we took all radiologists' annotations into consideration by calculating the mean malignancy level for each labeled nodule, and marked the nodule as benign (mean malignancy level lower than 2.75), uncertain (the mean malignancy level from 2.75 to 3.25), or malignant (mean malignancy level higher than 3.25). A total of 1184 benign, 1040 uncertain, and 427 malignant nodules were obtained. Based on previous experiences [22,23], uncertain nodules are excluded here. The resulting dataset is shown in Table 1.

3.2. Data pre-processing

Distinguishing between benign and malignant lung nodules is a primary area of focus for us once pulmonary nodules have been identified. When compared to the entire chest, lung nodules occupy a relatively small volume, which implies that CT scans contain a significant amount of extraneous background information. Based on the central coordinates of the pulmonary nodules provided by the radiologist in the annotations, a simple center-crop method was used to place the lung nodules in the center of the pre-processed CT image without additional parameters or calculations, while effectively filtering useless background information. For each CT scan, a pre-processing procedure was performed for further computation. The data pre-processing procedure is shown in Fig. 1.

Different chest CT scans have different numbers of slices in the LIDC-IRDI dataset. To obtain the same spatial resolution for each scan, spline interpolation was used to resample the slices to a unified voxel size of $1.0mm \times 1.0mm \times 1.0mm$. In the next step, a center-crop approach was applied to obtain a cube containing a labeled nodule from the resampled CT scans. Subsequently, three 2D slices were extracted from the cube using center-crop, from three different perspectives: transverse, sagittal, and coronal. To facilitate the calculation of the network, these 2D slices were normalized (Equation (1)). The normalized formula is as follows:

$$X_{norm} = \frac{X\text{-}X_{min}}{X_{max}\text{-}X_{min}} \tag{1}$$

Where X_{norm} is the normalized data, X is the original data, X_{max} and X_{min} are maximum and minimum values in the raw data, respectively.

Ultimately, the 2D slices obtained from three distinct viewpoints were merged into a unified RGB image, where each slice corresponds to a distinct color channel. Conventional practice involves training a separate model for each view plane using its corresponding grayscale image, resulting in a substantial increase in training time as the number of models expands. Furthermore, relying solely on grayscale images fails to harness the model's complete capabilities. Therefore, by consolidating the three grayscale images into a single RGB image, the proposed approach diminishes the need for multiple models, accelerates the training process, and retains nodule characteristics to the fullest extent possible.

Table 1
Experimental dataset.

Label	Benign	Uncertain	Malignant
Mean malignancy level	0–2.75	2.75–3.25	3.25–5
Number	1184	1040	427

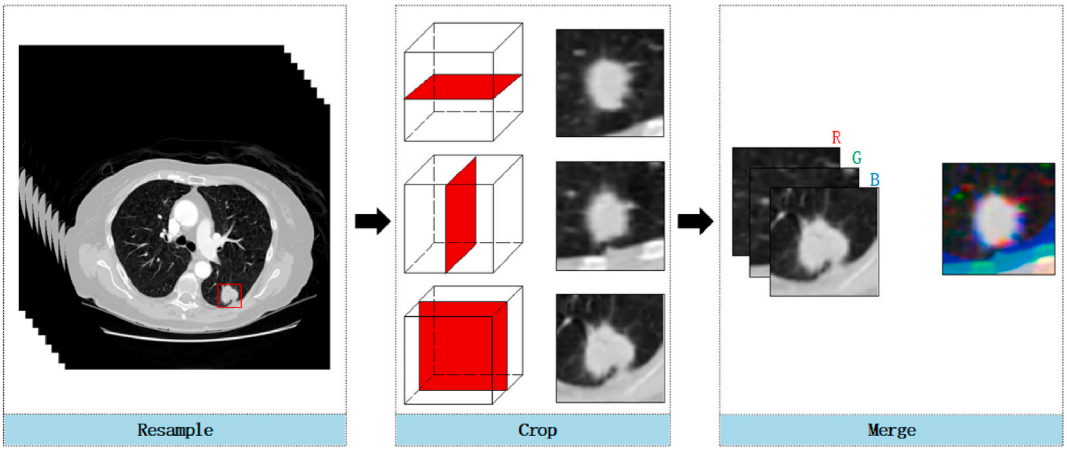


Fig. 1. Data pre-processing procedure. Firstly, we used spline interpolation to resample the original CT scans to a unified voxel size. Then, a cube was cropped from resampled CT scans and contains a complete nodule. On each cube, we cropped three 2D slices from three different perspectives. Finally, we merged three grayscale images to form an RGB image.

In addition, lung nodules have a large variation in both size and morphological characteristics, which leads to the fact that multiscale features play a crucial role in network performance. Thus, for each nodule, we cropped three cubes with different scales of $60 \times 60 \times 60$, $50 \times 50 \times 50$ and $40 \times 40 \times 40$. The three cubes were processed with the same pre-processing procedure as shown in Fig. 1. Finally, we obtained three RGB images of different scales for each nodule, which will be input as data sources into the deep neural network model.

Compared to conventional pre-processing methods, our proposed approach eliminates the need for complicated segmentation of the nodule's edges, reducing both the time and computational resources required. Furthermore, the use of a single RGB image instead of multiple grayscale images reduces the number of models needed, further decreasing the time cost of the training process.

3.3. Attention model

The idea of attention in deep neural networks is inspired by the human visual attention system. Spatial attention allows humans to selectively process visual information through prioritization of an area within the visual field [24] and significantly improve both recognition and detection performance, especially in images with cluttered backgrounds [25]. Following the same principle, neural networks can be trained to focus on specific parts of an input signal that appear to be more strongly related to the task, which means deep learning models with attention mechanism can concentrate on subtle differences between similar pictures.

The main differences between malignant nodules and benign nodules are shape and size. Traditional models for lung nodule classification often require a complicated pre-processing procedure to accurately quantify nodule characteristics. However, a deep learning model with attention mechanism can focus on the main differences, and ignore the background information such as blood

Table 2
Structure of attention model.

Layer	Output Size	Parameters
Conv1	114×114	7×7, 64, stride 2
Max pooling	56×56	3×3, stride 2
Residual Unit	56×56	$\begin{pmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{pmatrix} \times 1$
Attention Module	56×56	Attention
Residual Unit	28×28	$\begin{pmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{pmatrix} \times 1$
Attention Module	28×28	Attention
Residual Unit	14×14	$\begin{pmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{pmatrix} \times 1$
Attention Module	14×14	Attention
Residual Unit	7×7	$\begin{pmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{pmatrix} \times 3$
Average Pooling	1×1	7×7, stride 1
FC(Softmax)	2	None

vessels and lung walls. Thus, some pre-processing steps such as nodule segmentation can be replaced by attention module to some extent. Besides, as a part of the network, attention mechanism can be regulated during training time conveniently. For all these advantages, we built a deep neural network with an attention model. The attention model is based on a residual attention network [26] structure, which is shown in Table 2.

The attention model primarily consists of one convolutional layer, two pooling layers, one fully connected layer, three attention modules, and multiple residual units. The convolutional kernel size is 7 × 7 with a stride of 2. The max-pooling layer uses a pooling kernel size of 3 × 3 and a stride of 2, while the average pooling layer employs a pooling kernel size of 7 × 7 with a stride of 1. Specific parameters are detailed in Table 2. The residual unit serves as the fundamental learning component within the ResNet-50 model [27], and the attention module structure is shown in Fig. 2. In our proposed model, each attention module is composed of two branches. The first branch, known as the trunk branch, processes the input features, while the second branch, known as the soft mask branch, generates an attention mask to control the output of the trunk branch. The attention mask acts as a feature selector during inference and a gradient filter during backpropagation. This attention mechanism generates a feature map with attention weights that highlight the most relevant areas of the input image. This approach allows our model to dynamically identify the most salient regions of the CT scan for pulmonary nodule analysis, which is an essential step for accurate classification. The process can be expressed as provided in (Equation (2)).

$$H(x) = (1 + M(x)), T(x) \tag{2}$$

Where $M(x)$ is the attention mask generated by the Soft Mask Branch range in [0,1], $T(x)$ is the feature map generated by Trunk Branch, and $H(x)$ is the output of attention module.

The number of output layer neurons was set to 2 to accommodate the needs of binary results of lung nodule classification. Then we pre-trained three attention models at three scales separately.

3.4. Multichannel attention model

Scale is an important factor to consider in automatic nodule-type classification [8,28]. Integration of features in different scales can sufficiently quantify nodule characteristics. The multi-scale sampling strategy is motivated by the clinical fact that nodule sizes vary remarkably, ranging from less than 3 mm to more than 30 mm in the LIDC-IDRI datasets. Patches of different scales will provide different information. Patches with small scales can provide the details of the nodules, while patches with large scales can provide information surrounding the tumor tissue.

Although 3D CNN [29–31] can extract richer features, the structure takes more computational resources, which usually needs a very long training time. Extending the use of 2D CNN to the analysis of volumetric medical images on a slice-by-slice basis, together with data augmentation, enables us to have more training samples [7,32]. Besides, Volumetric data can be decomposed into fixed tri-planar views (sagittal, coronal, and axial planes), which can preserve the features of 3D nodules as much as possible. Different from the methods provided by Setio et al. [7], we fed multi-view patches from the same nodule to one model instead of three models to save computational resources. In general, neural networks for analyzing CT have only one input channel, because a piece of a CT scan is a grayscale image with only one color channel. As illustrated in Fig. 1, multiview patches from the same nodule can be combined into an RGB image which will be fed to the model with all features from multiview patches.

We propose a multichannel attention model shown in Fig. 3. The attention mechanism allows the model to dynamically capture the location of lung nodules in the images without segmenting the lung nodules in advance. It combines multi-scale and multi-view strategies to exploit the information in the original CT images. The model consists of three attention submodels, and the outputs of all submodels are combined using a post-fusion strategy, i.e., fusion at a richer feature level. Two neurons in the output layer of each submodel are connected to the classification layer using a Softmax activation function. This approach allows the model to achieve high classification accuracy with as few preprocessing steps as possible. The output of the classification layer is the prediction results of the multichannel attention model (Equation (3)), and the result can be expressed as:

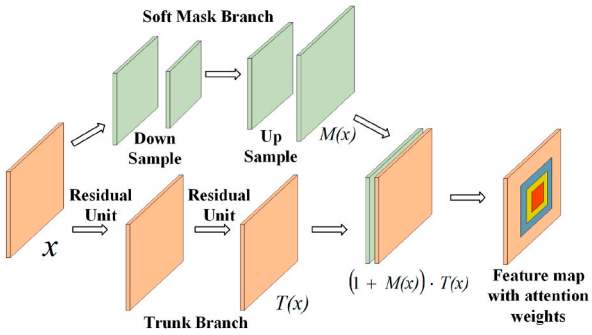


Fig. 2. Structure of attention module. The trunk branch performs feature processing and the soft mask branch generates a mask of attention as a control gate for neurons of the trunk branch. The output of attention module is a feature map with attention weights.

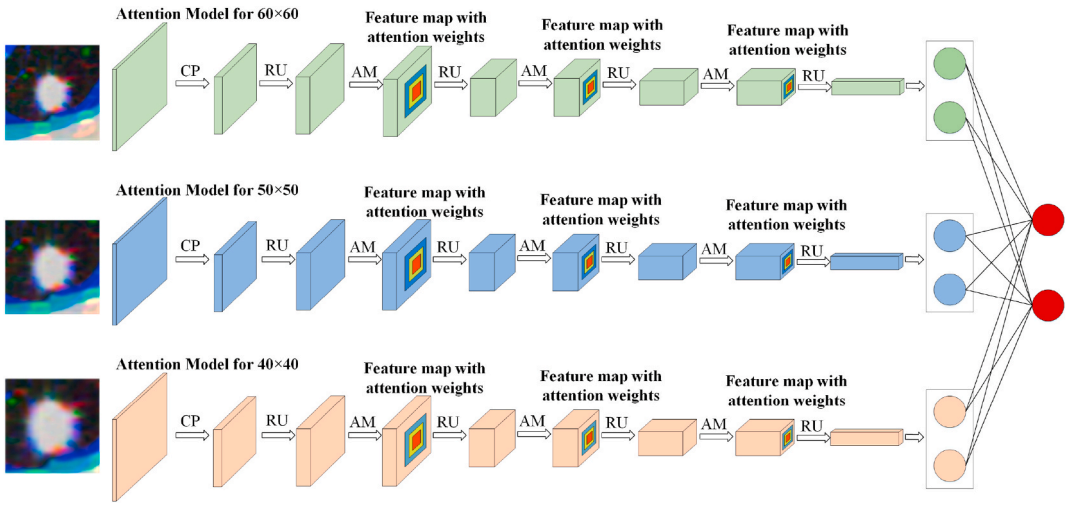


Fig. 3. Multichannel attention model. It consists of three attention submodels. The two neurons of the output layer of each submodel are connected to a classification layer using the Softmax activation function. For each channel, we used inputs of different scales to train, and integrated features under different scales. CP means convolution and pooling, RU means residual unit, and AM means attention module.

$$p = f\left(\sum_{k=1}^{3}\sum_{j=1}^{2} w_{kj} s k_{kj}\right) \tag{3}$$

Where $\{w_{kj}; k = 1, 2, 3; j = 1, 2\}$ is the assemble of weights between the output layer of each submodel and the classification layer, $\{s_{kj}; k = 1, 2, 3; j = 1, 2\}$ is the output of each submodel. The summation over j means the weighted sum of the output in each attention submodel. The summation over k means the weighted sum of the output of three submodels. The function $f(.)$ is the Softmax activation.

Finally, we trained the model to obtain the final classification results.

4. Results and discussion

We utilized the multi-channel attention model on the LIDC-IDRI dataset, implementing data augmentation alongside a 5-fold cross-validation approach. Data augmentation helps mitigate the overfitting challenges often encountered in deep learning models by introducing variations to the dataset. To achieve this, we created augmented data for each training patch through random image transformations, including translation, rotation, and horizontal or vertical flips. The objective was to ensure that the model was exposed to a diverse range of data perspectives during training, preventing it from encountering the same image twice. This approach promotes improved generalization of the model. The translation step was selected from [0, 20] voxels, and the rotation angle was randomly selected from [0°, 180°]. Then, all nodule patches were resized to 224 × 224. The training dataset was randomly partitioned into five segments. In each training cycle, one of these segments served as the test set, while the remaining segments constituted the training set, creating a 5-fold cross-validation. The ultimate score was computed as the average of these five scores, and we could also calculate standard deviation metrics for these five scores.

Throughout the training procedure, specific parameters were set as follows: the learning rate was established at 0.001, the batch size was set to 16, the maximum number of training epochs was limited to 200, and the chosen optimizer was Rmsprop. [33], respectively. Moreover, we randomly chose 20 % of the training patches to form a validation set and terminated the training process even before reaching the maximum epoch number if the error on the other 80 % of the training patches continues to decline but the error on the validation set stops decreasing. The program was written in Python® under a Windows® 10 operating system on a workstation with Nvidia Geforce RTX 2080Ti. The difficulty of benign-malignant classification is reflected in the gradual improvement of its technique by previous models. Our model differs from the comparison method as it employs a multi-channel model with an attention mechanism that enables the network to accurately locate nodule locations during the recognition process, and fusing the

Table 3
Comparison experimental results.

Methods	Accuracy(%)	AUC(%)
Shen et al. [35] (Multicrop CNN)	87.14	93.0
Nobrega et al. [9] (ResNet50+SVM-RBF)	88.41	93.19
Liu et al. [36] (DenseNet)	88.31	93.35
Xie et al. [6] (Deep + visual features)	89.53	96.65
Our work	90.11 ± 0.24	95.66 ± 0.17

outputs of each single-channel model improves the recognition accuracy. As can be seen from Table 3, our model performs competitively and improves overall accuracy when using the same dataset. Table 3 shows that the performance was assessed by the mean of the obtained accuracy and area under the receiver operator curve (AUC) score. Besides, the confusion matrix [34] is shown in Fig. 4. Accuracy shows the performance of our model in classifying nodules as malignant or benign. The AUC is sensitive to an imbalance among the classes. Interpreting the metrics of a confusion matrix in the context of the attention-guided deep neural network with a multichannel architecture for lung nodule classification is challenging due to the complexity of the model, potential class imbalances, the presence of attention mechanisms, and the need for external validation. It's essential to consider these limitations and address them in a way that aligns with the clinical and diagnostic needs of the application. Additionally, efforts should be made to make the model's outputs more interpretable and trustworthy for healthcare professionals.

The Grad-Cam algorithm [37] was applied to calculate the heat map to evaluate whether the results are reliable and show how the attention mechanism works. The heat map is shown in Fig. 5. The area's color, from red to blue, means the effects on classification results from important to unimportant. We can observe that nodules are surrounded by attention areas, which proves that the classification results are predicted based on lung nodules rather than other incorrect information exactly. The heat map also shows that the complicated nodule segmentation procedure can be replaced by attention mechanism due to the reason that the purpose of nodule segmentation, which is similar to attention mechanism, is to enable the model to focus on nodules. For malignant nodules, the model can identify the size and shape information of the nodules. For benign nodules, since the size of benign nodules is generally small, the prediction results are more affected by the scale. One possible improvement is to use U-Net [38] to segment nodule images with fewer backgrounds or use smaller scales. But the nodules that fit on the lung wall, are difficult to be accurately located by the model.

Lung nodules have a large variation in both size and morphological characteristics, which leads to the fact that it is important to make full use of the information from original CT images. The results show that the multichannel and multiview attention model performs well in background suppression and foreground enhancement in feedforward operation. The attention mechanism is useful to refine the deep feature representations of lung nodules. By regulating the attention of deep models, deep models can observe input images more "carefully", which is expected to improve the performance of models in lung nodule classification.

Although accurate processing time comparison of different methods is difficult to obtain, we list the time spent by two other methods and our multichannel model in Table 4. In Table 4, we present the experimental results obtained using replication techniques. All comparisons were conducted on the same dataset and computer configuration to ensure the comparability of the results. The time required for manual segmentation in the Massive-feat method was not accounted for, as the segmentation was pre-existing in the dataset. However, it was observed that the process of hand-crafted feature extraction in the Massive-feat method proved to be highly time-consuming, exceeding 10 h. Furthermore, the testing time for a single nodule using this method was considerably longer compared to other approaches. Compared with mainstream models, the proposed model performs better in time consumption. And the time consumption is 29.09 % of ResNet-50.

It has been proved that the multichannel architecture does have a positive effect on the lung nodule classification task [8]. Comparing networks of the same width, i.e. the same number of input channels, we can see that the combination of multichannel architecture and attention model achieves higher scores in both accuracy and AUC as shown in Table 5.

Training and validation/testing plots were used to demonstrate the loss curve and accuracy curve of the training process (Fig. 6). Within the plots, the solid line represents the training progress and the dashed line represents the verification progress. Based on the trend within the plots, when the training ended, the training dataset converged to 89.7 % accuracy and 1.65 loss.

5. Conclusions

This paper introduces an innovative deep learning approach for the classification of lung nodules in chest CT scans, integrating an attention mechanism [41]. In order to make the most of the valuable information in the original CT images, we employed a multi-faceted approach for feature extraction, incorporating multiple views and channels to preprocess the input images. By employing the attention mechanism to quantify nodule characteristics, our model achieved a high classification accuracy with minimal

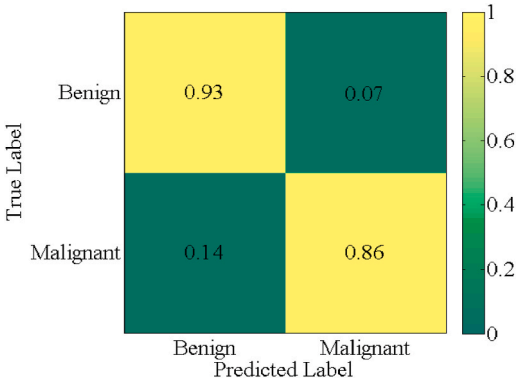


Fig. 4. Confusion matrix.

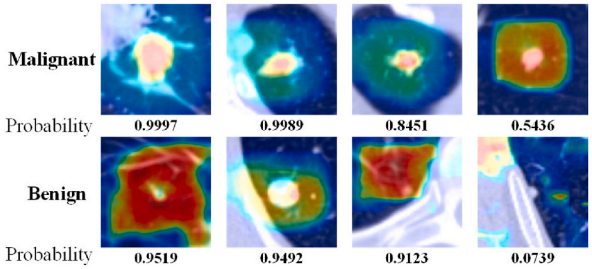


Fig. 5. Heat map and probability predicted by the model.

Table 4
Time consumption comparison.

Methods	Nodule Segmentation	Training Time	Test Time pre Nodule
Massive-feat [39]	Manual	10h15min	32.76s
ResNet-50 [27]	None	1h50min	0.33s
Multicrop CNN [35]	None	47min01s	0.23s
Our work	None	32min06s	0.07s

Table 5
Performance of the same width networks.

Methods	Accuracy (%)	AUC (%)
Shen et al. [8] (Multiscale CNN)	86.84	Not given
Xie et al. [40] (MVKBC, using view1∼3)	90.1	94.17
Our work	90.11 ± 0.24	95.66 ± 0.17

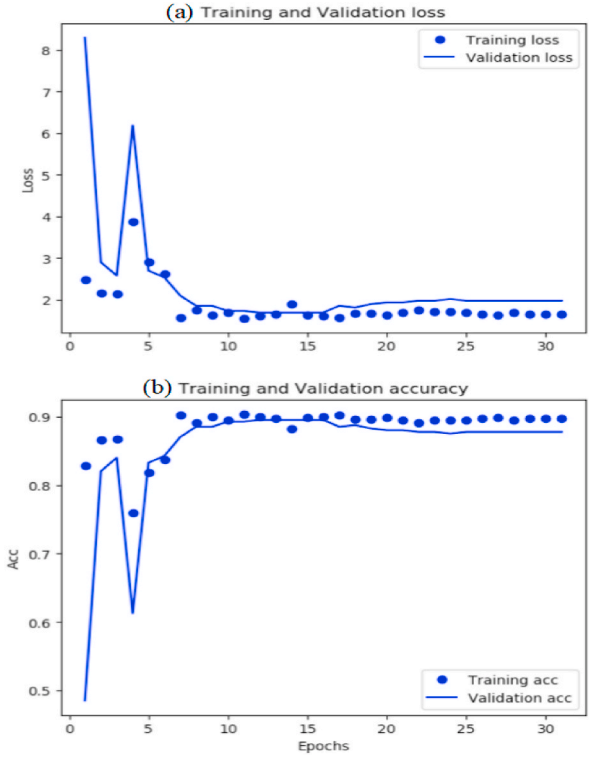


Fig. 6. Training and validation/testing plots.

preprocessing. Unlike traditional slice segmentation methods, we quantified nodule features by employing attention mechanisms. The use of attention mechanisms allowed us to optimize nodule features further during the training process, enabling our model to achieve high classification accuracy with minimal preprocessing. We conducted training and evaluation using a dataset comprising 1018 CT scan images from the LIDC-IDRI database. Additionally, we compared our model with several other mainstream models under the same conditions. The results demonstrated a nodule classification accuracy of 90.11 % and an AUC score of 95.66 %. Notably, our approach achieved this level of performance while consuming only 29.09 % of the time needed by the mainstream ResNet-50 model.

Funding

This research was supported by the National Natural Science Foundation of China under Grant 62075172.

Ethical considerations

This article does not contain any studies with human or animal subjects performed by the any of the authors.
Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.

Data availability statement

Data associated with this study has not been deposited into a publicly available repository. Data associated with his study may be made available on reasonable request through the corresponding author.

CRediT authorship contribution statement

Rong Zheng: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Formal analysis, Data curation, Conceptualization. **Hongqiao Wen:** Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Feng Zhu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis. **Weishun Lan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are very grateful for the CT images provided by the LIDC-IDRI lung nodule database, which provided a great help for our study.

References

[1] R.L. S, K. D, M, A, J. Cancer Statistics, 2017. CA, Cancer J. Clin 67 (2017) 7–30, https://doi.org/10.3322/caac.21387.
[2] C.G. Slatore, R.S. Wiener, A.D. Laing, What is a lung nodule? Am. J. Respir. Crit. Care Med. 193 (2016) P11–P12, https://doi.org/10.1164/rccm.1937P11.
[3] T.W. Way, L.M. Hadjiiski, B. Sahiner, H.-P. Chan, P.N. Cascade, E.A. Kazerooni, N. Bogot, C. Zhou, Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours, Med. Phys. 33 (2006) 2323–2337, https://doi.org/10.1118/1.2207129.
[4] F. Han, G. Zhang, H. Wang, B. Song, H. Lu, D. Zhao, H. Zhao, Z. Liang, A Texture Feature Analysis for Diagnosis of Pulmonary Nodules Using LIDC-IDRI Database, Northeastern University, Shenyang, Liaoning 110819, China, 2013.
[5] Y. LeCun, Y. LeCun, Y. Bengio, G. Hinton, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.
[6] J. Fu, H. Zheng, T. Mei, Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition, Univ Sci & Technol China, Hefei, Anhui, Peoples R China, 2017, pp. 4476–4484, https://doi.org/10.1109/cvpr.2017.476 [1] Microsoft Res, Beijing, Peoples R China [2].
[7] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, Recurrent models of visual attention, Advances in neural information processing systems 27 (2014) 2204–2212. https://doi.org/10.48550/arXiv.1406.6247.
[8] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, G. Yang, Is attention all you need in medical image analysis? A review, arXiv preprint arXiv (2023), 12775, https://doi.org/10.48550/arXiv.2307.12775, 2307.
[9] Y. Xie, B. Yang, Q. Guan, J. Zhang, Q. Wu, Y. Xia, Attention mechanisms in medical image segmentation: a Survey, arXiv preprint arXiv (2023), 17937, https://doi.org/10.48550/arXiv.2305.17937, 2305.
[10] Y. Xie, J. Zhang, Y. Xia, M. Fulham, Y. Zhang, Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT, Shaanxi Key Lab of Speech & Image Information Processing (SAIIP), School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China Centre for Multidisciplinary Convergence Computing (CMCC), School of Computer Sci 42 (2018) 102–110, https://doi.org/10.1016/j.inffus.2017.10.005.
[11] F.C. Arnaud Arindra Adiyoso Setio, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sanchez, Bram van Ginneken. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks, IEEE Trans. Med. Imaging 35 (2016) 1160–1169, https://doi.org/10.1109/tmi.2016.2536809.

[12] W. Shen, M. Zhou, F. Yang, C. Yang, J. Tian, Multi-scale convolutional neural networks for lung nodule classification. Information processing in medical imaging : proceedings of the, conference 24 (2015) 588–599, https://doi.org/10.1007/978-3-319-19992-4_46.
[13] R.V.M. Da Nóbrega, S.A. Peixoto, S.P.P. Da Silva, P.P.R. Filho, Lung Nodule Classification via Deep Transfer Learning in CT Lung Images(Conference Paper). Programa de Pós-Graduao em Ciência da Computao (PPGCC), Instituto Federal de Educao, Ciência e Tecnologia Do Ceará (IFCE), Brazil (2018) 244–249, https://doi.org/10.1109/cbms.2018.00050.
[14] T. Usuzaki, K. Takahashi, H. Takagi, M. Ishikuro, T. Obara, T. Yamaura, M. Kamimoto, K. Majima, Efficacy of exponentiation method with a convolutional neural network for classifying lung nodules on CT images by malignancy level, Eur. Radiol. (2023) 1–11, https://doi.org/10.1007/s00330-023-09946-w.
[15] Yan, C.; Razmjooy, N. Optimal lung cancer detection based on CNN optimized and improved Snake optimization algorithm. Biomed. Signal Process Control. Vol.86, 105319, doi:10.1016/j.bspc.2023.105319.
[16] D. Fang, H. Jiang, W. Chen, Z. Qin, J. Shi, J. Zhang, Pulmonary nodule detection on lung parenchyma images using hyber-deep algorithm, Heliyon (2023), https://doi.org/10.1016/j.heliyon.2023.e17599.
[17] X. Fu, L. Bi, A. Kumar, M. Fulham, J. Kim, An attention-enhanced cross-task network to analyse lung nodule attributes in CT images, Pattern Recogn. 126 (2022), 108576, https://doi.org/10.1016/j.patcog.2022.108576.
[18] D. Zhao, Y. Liu, H. Yin, Z. Wang, An attentive and adaptive 3D CNN for automatic pulmonary nodule detection in CT image, Expert Syst. Appl. 211 (2023), 118672, https://doi.org/10.1016/j.eswa.2022.118672.
[19] S.G. Armato III, G. McLennan, L. Bidaut, The lung image database Consortium, (LIDC) and image database resource Initiative (IDRI): a completed reference database of lung nodules on CT scans, Med. Phys. 38 (2011) 915–931, https://doi.org/10.1118/1.3528204.
[20] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, et al., The cancer imaging archive (TCIA): Maintaining and operating a public information repository (Article), J. Digit. Imaging 26 (2013) 1045–1057, https://doi.org/10.1007/s10278-013-9622-7.
[21] S.G. Armato III, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, L.P. Clarke, Data from The Lung Image Database consortium (LIDC) and Image Database resource Initiative (IDRI): A completed reference database of lung nodules on CT scans (LIDC-IDRI), TCIA (2015) 9. https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254.
[22] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, Z. Liang, Texture feature analysis for computer-aided diagnosis on pulmonary nodules, J. Digit. Imaging 28 (2015) 99–115, https://doi.org/10.1007/s10278-014-9718-8.
[23] A.K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, N. Khandelwal, A combination of shape and texture features for classification of pulmonary nodules in lung CT images(article), J. Digit. Imaging 29 (2016) 466–475, https://doi.org/10.1007/s10278-015-9857-6.
[24] R.A.N.R. Rensink, I. Development, Cambridge Basic Research, Cambridge, MA, US. The Dynamic Representation of Scenes. Rensink, Ronald A vol. 7, Nissan Research & Development, Inc., Cambridge Basic Research, Cambridge, MA, US, 2000, pp. 17–42, https://doi.org/10.1080/135062800394667.
[25] R. Cichy, D. Pantazis, A. Oliva, Resolving human object recognition in space and time, Nat. Neurosci. 17 (2014) 455–462, https://doi.org/10.1038/nn.3635.
[26] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification; ; SenseTime Grp Ltd, Hong Kong, Hong Kong, Peoples R China, in: Tsinghua Univ, Beijing, Peoples R China ; Chinese Univ Hong Kong, Hong Kong, Hong Kong, Peoples R China, Beijing Univ Posts & Telecommun, Beijing, Peoples R China, 2017.
[27] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, Peoples R China, Beijing, 2016, pp. 770–778, https://doi.org/10.1109/cvpr.2016.90 [1] Microsoft Res.
[28] F. Ciompi, K. Chung, S.J.v. Riel, A.A.A. Setio, P.K. Gerke, C. Jacobs, E.T. Scholten, C. Schaefer-Prokop, M.M.W. Wille, A. Marchianò, et al., Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, The Netherlands; Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands; Department of Respiratory Medicine, Gentofte Hospital 7 (2017), https://doi.org/10.1038/srep46479.
[29] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. Mok, L. Shi, P. Heng, Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, IEEE Trans. Med. Imaging 35 (2016) 1182–1195, https://doi.org/10.1109/tmi.2016.2528129.
[30] Q. Dou, H. Chen, L. Yu, J. Qin, P. Heng, Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection, IEEE Trans. Biomed. Eng. 64 (2017) 1558–1567, https://doi.org/10.1109/tbme.2016.2613502.
[31] M.S. Sunarjo, H.S. Gan, High-performance convolutional neural network model to identify COVID-19 in medical images, Journal of Computing Theories and Applications 1 (1) (2023) 19–30, https://doi.org/10.33633/jcta.v1i1.8936.
[32] A P, P. K, I. C, F L, D. E, M N, Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network, in: Affiliations Department of Computer Science, University of Copenhagen, Denmark. Biomediq, Denma vol. 16, 2013, pp. 246–253, https://doi.org/10.1007/978-3-642-40763-5_31.
[33] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A. Mohamed, Deep Neural Networks for Acoustic Modeling in Speech Recognition, IEEE Signal Processing Magazine 29 (6) (2012) 82–97, https://doi.org/10.1109/MSP.2012.2205597.
[34] F. Mustofa, A.N. Safriandono, A.R. Muslikh, Dataset and feature analysis for Diabetes Mellitus classification using random forest, Journal of Computing Theories and Applications 1 (1) (2023) 41–48, https://doi.org/10.33633/jcta.v1i1.9190.
[35] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, J. Tian, Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification, Pattern Recogn. 61 (2017) 663–673, https://doi.org/10.1016/j.patcog.2016.05.029.
[36] Y. Liu, P. Hao, P. Zhang, X. Xu, J. Wu, W. Chen, Dense convolutional binary-Tree networks for lung nodule classification, Key Laboratory of Advanced Transducers and Intelligent Control Systems, Ministry of Education and Shanxi Province, College of Physics and Optoelectronics, Taiyuan University of Technology, Taiyuan, 030024, China; Science and Technology on Ne 6 (2018) 49080–49088, https://doi.org/10.1109/access.2018.2865544.
[37] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-Cam, Visual Explanations from deep networks via gradient-based localization (Conference paper), Georgia Institute of Technology, United States; Facebook AI Research, United States 2017 (2017) 618–626, https://doi.org/10.1109/iccv.2017.74.
[38] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: convolutional networks for biomedical image segmentation, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 9351 (2015) 234–241.
[39] H.J.W.L. Aerts, E.R. Velazquez, R.T.H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, et al., Correction: Corrigendum: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5 (2014), https://doi.org/10.1038/ncomms5006.
[40] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, W. Cai, Knowledge-based Collaborative deep learning for benign-malignant lung nodule classification on chest CT, IEEE Trans. Med. Imaging 38 (2019) 991–1004, https://doi.org/10.1109/tmi.2018.2876510.
[41] C. Xiaohan, W. Hongqiao, W. Shengzong, L. Weishun, Rong Zheng, J. Phys. Conf. Ser. (2021), 012185, 1754.