

RESEARCH

Open Access



A novel fusion algorithm for benign-malignant lung nodule classification on CT images

Ling Ma¹, Chuangye Wan¹, Kexin Hao¹, Annan Cai¹ and Lizhi Liu^{2*}

Abstract

The accurate recognition of malignant lung nodules on CT images is critical in lung cancer screening, which can offer patients the best chance of cure and significant reductions in mortality from lung cancer. Convolutional Neural Network (CNN) has been proven as a powerful method in medical image analysis. Radiomics which is believed to be of interest based on expert opinion can describe high-throughput extraction from CT images. Graph Convolutional Network explores the global context and makes the inference on both graph node features and relational structures. In this paper, we propose a novel fusion algorithm, RGD, for benign-malignant lung nodule classification by incorporating Radiomics study and Graph learning into the multiple Deep CNNs to form a more complete and distinctive feature representation, and ensemble the predictions for robust decision-making. The proposed method was conducted on the publicly available LIDC-IDRI dataset in a 10-fold cross-validation experiment and it obtained an average accuracy of 93.25%, a sensitivity of 89.22%, a specificity of 95.82%, precision of 92.46%, F1 Score of 0.9114 and AUC of 0.9629. Experimental results illustrate that the RGD model achieves superior performance compared with the state-of-the-art methods. Moreover, the effectiveness of the fusion strategy has been confirmed by extensive ablation studies. In the future, the proposed model which performs well on the pulmonary nodule classification on CT images will be applied to increase confidence in the clinical diagnosis of lung cancer.

Keywords Lung cancer, Lung nodule classification, Information fusion, Deep convolutional neural network, Graph convolutional network, Computed tomography (CT)

Introduction

Cancer is one of the most deadly diseases and lung cancer is the most common and among the deadliest cancers in the world [1]. Recent studies show that the 10-year relative survival rate can increase by 20% through annual screening with low-dose computed tomography (CT) in early stages to capture the cancer heterogeneity in a

non-invasive way [2–4] and most lung cancers arise from small malignant nodules [3, 5]. However, malignant nodule recognition by radiologists reading every CT scan is extremely time-consuming and subjective. Therefore, many computer-aided diagnosis (CAD) systems being powerful tools to aid radiologists in identifying malignant lung nodules have been proposed in recent years [6]. Most current CAD systems for lung nodule classification can be divided into two respects, one is handcraft-based CAD systems, and the other is deep learning-based CAD systems.

The handcraft-based CAD systems used the intensity, texture, and shape features, like local binary patterns [7], wavelet features [8], histogram of oriented gradient

*Correspondence:

Lizhi Liu
liulzh@sysucc.org.cn

¹ College of Software, Nankai University, Tianjin 300350, China

² Department of Radiology, Sun Yat-Sen University Cancer Center, Guangzhou 510060, Guangdong, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(HOG) [9], gray level co-occurrence matrix (GLCM) [10], 3D surface feature [11], selected the distinguishing features [12–14] and then trained a designed classifier, such as the support vector machine (SVM) [15], Bayesian [16], and random forest [17], or combined classifier [18], for pulmonary nodule classification.

The deep learning-based CAD systems used convolutional neural networks (CNNs) for lung nodule classification from CT images [19, 20]. Many deep learning-based CAD systems have achieved good performance for lung nodule classification. However, benign and malignant nodules sometimes look similar and hard to distinguish from one another, and most deep learning-based CAD systems use the CNN models which include a set of convolutional and pooling layers that can be operated on the data with a Euclidean structure. Hence, exploring and fusing the non-Euclidean characteristic becomes an attractive and reasonable choice to improve the classification results. Graph convolutional networks (GCNs) have a great expressive power to learn non-Euclidean representations and have demonstrated superior performance in the classification task [21].

In this paper, we propose a fusion algorithm, RGD, combining the Radiomics, and GCN features into the Deep CNNs to form a more complete feature representation for benign-malignant lung nodule classification on CT images. First, we calculate the radiomics features to describe high-throughput extraction from the lung nodules from CT images, because pathological studies have demonstrated that there is increased heterogeneity within malignant lung nodules, which is not appreciable in radiological studies by the naked eye but can be quantified with radiomics [22]. Second, we adopt several different CNN models to extract the self-learned features. Next, these deep features are fed into a graph convolutional network to model the relations among features and furnish more promising insights underlying the features in non-Euclidean space. Finally, ensemble learning fuses the informative knowledge to achieve knowledge discovery and better predictive performance via voting schemes in an adaptive way. To the best of our knowledge, this is the first work fusing the radiomics, deep CNN features, and graph features representation for lung nodule classification on CT images. Compared with the existing simple concatenation or fully connected fusion methods, our proposed RGD algorithm can effectively separate malignant from benign nodules by fusing several CNN features in graph structure to mine the discriminating features and eliminate the redundant representations.

Related work

In recent years, deep learning methods have gradually been applied and designed for lung nodule classification.

Some methods designed a special CNN architecture for the lung nodule classification by involving the 2D and 3D space information [23], multi-scale [24, 25], multi-view [26, 27], multi-crop [28], local-global information [29], dilated convolutions [30], attention mechanism [31], manifold representation [32], lung nodule shape and margin feature analysis [33].

Some methods fused multiple models to improve the classification ability. Dey et al. proposed a fused model for the problem of diagnostic classification between benign and malignant lung nodules in CT images, including a basic 3D CNN, a novel multi-output network, a 3D DenseNet, and an augmented 3D DenseNet with multi-outputs [34]. Zhao et al. constructed a hybrid CNN of LeNet and AlexNet by combining the layer settings of LeNet and the parameter settings of AlexNet for pulmonary nodule classification [35]. Zhao et al. integrated different CNNs and adopted transfer learning to utilize deep convolutional neural networks for lung nodule classification [36]. Wang et al. proposed an adaptive-boost deep learning method by composing a series of multi-layer perceptions (MLPs) to obtain a strong classifier [37]. Onishi et al. used the Wasserstein GAN to generate the multiplanar images of the pulmonary nodule and performed three DCNNs on these generated nodule images for pulmonary nodule classification in CT images [38]. Xu et al. proposed a method called MSCS-DeepLN to evaluate lung nodule malignancy. They trained and combined three CNN models with multi-scale input cropped from CT images for learning the multi-level contextual features and preserving diversity [39]. Jiang et al. used attentive and ensemble 3D dual path networks (DPNs) to improve the representation ability for pulmonary nodules classification [40]. Jiang et al. used neural architecture search to automatically search 3D network architectures and fused the outputs of n different neural architectures to get the final prediction of pulmonary nodules [41]. Liu et al. used the combinations of multiple CNNs and multiple traditional machine learning models, like k -nearest-neighbor, logistic regression, and SVM, based on handcrafted features in terms of morphology, density, curvature, and margin gradient, for predicting the likelihood of malignancy of pulmonary nodules from the CT scans [42].

Some methods fused multiple tasks to improve the performance of pulmonary nodule classification. Hussein et al. proposed a 3D CNN-based multi-task learning for nodule classification. They acquired another task-dependent feature representation for six high-level nodule attributes for improving the benign-malignant nodule classification accuracy [43]. Dai et al. proposed a new method named attribute-lung-nodule classification to combine the two classification tasks, pulmonary nodule

benign-malignant classification, and pulmonary nodule image attributes classification, into a deep learning network to improve the accuracy of pulmonary nodule classification [44]. Khosravan et al. proposed a 3D deep multi-task CNN to tackle these two problems, false positive nodule reduction, and nodule segmentation jointly to improve the computer-aided diagnosis systems' performance [45].

Some methods fused multiple features to improve the benign-malignant nodule classification accuracy. Chen et al. combined the general low-level Haarlike and HOG features, the description of nine semantic features, and the heterogeneous computational features derived from the deep learning models to provide richer quantitative assessments of nodules for better support of the diagnostic decision and management [46]. Buty et al. used both appearance distinctions described by deep convolutional neural networks and 3D surface variations modeled by spherical harmonics to characterize the lung nodules [47]. Wang et al. fused the hand-crafted and deep features to train the cost-sensitive random forest for classifying the lung nodules in chest X-ray images [48]. Zhu et al. constructed features by concatenating the learned deep 3D DPN features, nodule size, and raw 3D cropped nodule pixels and employed a gradient boosting machine for pulmonary nodule classification [49]. Xie et al. proposed an algorithm that fused the texture features by employing a GLCM-based texture descriptor and shape features by using a Fourier shape descriptor to characterize the heterogeneity of nodules and deep convolutional neural network model-learned information at the decision level for lung nodule classification [50]. Li et al. proposed a fusion algorithm that combined the twenty-nine handcrafted features including nine intensity features, eight geometric features, and twelve texture features based on GLCM, and the CNN features from AlexNet, VGG-16, and Multi-crop Net. Then the combined features were used as the input for the SVM coupled with the sequential forward feature selection method to select the optimal feature subset and construct the classifiers [51].

Although these fused models can capture semantic and syntactic information in local descriptors well, they combine the multiple information in a simple and direct strategy and might ignore underlying complex relationships among these different types of information. In contrast, our proposed method fusing several CNN features in graph structure can capture the non-Euclidean representation and relational features for distinguishing malignant and benign lung nodules.

In this work, we propose a fusion algorithm, named RGD, for benign-malignant lung nodule classification on CT images by fusing the Radiomics, GCN features, and Deep CNNs features. Radiomics is the use of quantitative

imaging features extracted from medical images to characterize tumor pathology or heterogeneity and radiomics features have successfully elucidated subtle relationships between image characteristics and disease status [52]. The CNN models can highly exploit the stationarity and compositionality properties and have demonstrated their power in identifying malignant lung nodules from CT data [53]. In addition, the combination of the radiomics and CNN features has been proven to be an efficient route [54]. Graph neural networks have rich relational structures and can preserve global structure information [55], and it can enable our model to explore the existing and potential relationships between neighboring features to learn discriminative features. Therefore, the proposed RGD, which fuses the Radiomic, GCN, and multiple Deep CNN features, is suitable for the feature representation. The RGD will be a powerful artificial intelligence aid for identifying malignant lung nodules in clinical practice and effectively detecting early lung cancer and then reducing mortality in lung cancer patients in the future.

Method

We propose a fusion algorithm for benign-malignant lung nodule classification on CT images as depicted in Fig. 1. As shown in Fig. 1, the proposed model (RGD) fuses the Radiomics, and GCN features into the Deep CNNs to form a more complete feature representation. Specifically, radiomics feature extraction is performed on CT images to describe the texture appearance. At the same time, the five different CNN models are separately trained to capture the complementary features from the CT images, then the GCN takes the CNN features as the input features and outputs the aggregated optimal features. Finally, the radiomics features and the GCN features are combined with the highest level CNN representation learned at the output layer of five 3D CNNs respectively to generate decision scores. Those decision scores from the five CNN models are ensemble for enhancing the classification performance.

Radiomics features extraction

Perinodular and intranodular radiomics features corresponding to texture information that are believed to be of interest based on expert opinion are extracted from CT images. We involve local binary pattern (LBP), the histogram of oriented gradient (HOG), and gray level co-occurrence matrix (GLCM) for radiomics feature extraction, because of their advantages of the acquirement of more detailed information and the invariances to monotonic gray-level changes and geometric orientations.

LBP is an effective descriptor that codes the gray levels of an image by comparing the central pixel with its neighbors, and the result is counted as a binary number

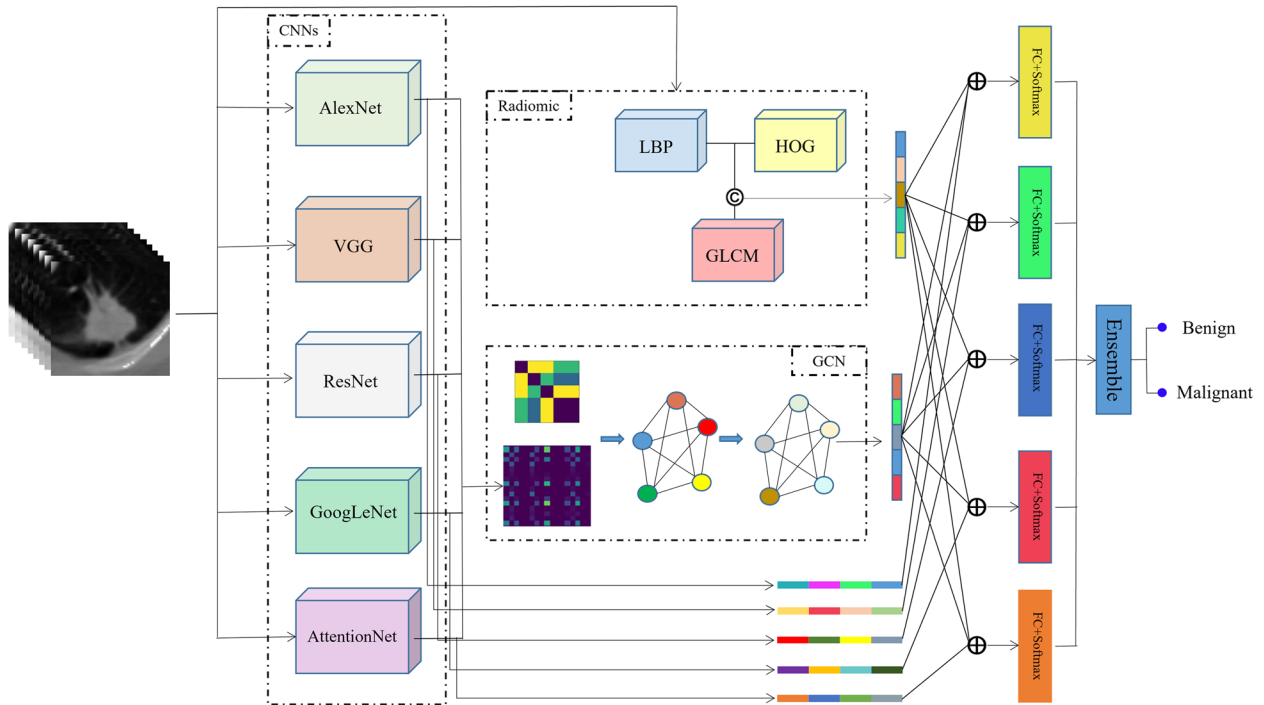


Fig. 1 The flowchart of our RGD model

converted to a decimal number that substitutes the central pixel value. In our work, we use the volume LBP for modeling the texture of lung nodules in 3D CT scans. Given a central pixel in the slice, a pattern number is computed by comparing its value with those of its neighborhoods

$$3DLBP_{L,P,R} = \sum_{q=0}^{3P+1} v_q 2^q, \tag{1}$$

where $L, P,$ and R mean the interval between the previous and posterior neighboring scans, the number and radius of neighbors around the central pixel in one scan, respectively. v is joint distribution in a local neighborhood of a series of CT scans and can be represented by

$$V = v \begin{pmatrix} B(g_{s_c-L,c} - g_{s_c,c}), & B(g_{s_c-L,0} - g_{s_c,c}), & \dots, \\ B(g_{s_c-L,P-1} - g_{s_c,c}), & B(g_{s_c,0} - g_{s_c,c}), & \dots, \\ B(g_{s_c,P-1} - g_{s_c,c}), & B(g_{s_c+L,0} - g_{s_c,c}), & \dots, \\ B(g_{s_c+L,P-1} - g_{s_c,c}), & B(g_{s_c+L,c} - g_{s_c,c}) \end{pmatrix} \tag{2}$$

$$= v(v_0, v_1, v_2, \dots, v_{3P+1}), B(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where $g_{s_c,c}$ is the gray value of the central pixel in the slice s and $g_{s,p}$ is a circularly symmetric neighbor set, corresponding to the gray value of the p pixel on a circle of radius R from the previous and posterior neighboring slices. In our method, we set the $L, P,$ and R as one, four, and one, respectively. According to Eq. 1, we sample and

threshold neighboring points to get a binary code, and then calculate its distribution. Finally, the histograms are normalized for forming the 3D LBP features.

HOG feature vector which has a good description of the lung nodules' appearance and involves the spatial information for the voxel's neighbors in the features, is calculated for distinguishing between benign and malignant nodules. Each voxel within a cell is applied the filter mask $[-1, 0, 1]$ to its neighboring voxels in all three dimensions for calculating the gradient vector G , whose magnitude is obtained using azimuth θ .

$$(\theta(x, y), \theta(z, xy)) = \left(\tan^{-1} \frac{G_y}{G_x}, \tan^{-1} \frac{G_z}{\sqrt{G_x^2 + G_y^2}} \right), \tag{3}$$

where $\theta(x, y)$ and $\theta(z, xy)$ are the angles of the 3D gradient when projected to XY -plane and the angle between the gradient and z -axis, $G_x, G_y,$ and G_z are the gradients along the X, Y and Z directions, respectively. Then these histograms are normalized to obtain the 3D HOG features.

GLCM can reflect the second-order conditional probability distribution by counting the frequency of the pair of voxels with specific directions and distances. To generate the 3D GLCM features, we sample along the direction linking the center voxel with its nearest 26 neighbors and consider 13 (half of 26) directions because of the redundant information from the opposite directions. The distances between pixel pairs along the 13 directions in the

z-axis are chosen as 1, 2, 3, and 4, and the distances along four directions on the XY-plane are 1, 2, 4, and 8. The co-occurrence matrix can be defined by

$$M_{ij}(\theta, d) = \sum_{p \in D} \begin{cases} 1 & \text{if } V(p) = i \text{ \& } V(p + d(\cos \theta, \sin \theta)) = j \\ 0 & \text{otherwise} \end{cases}, \quad (4)$$

where $V(p)$ is the pixel value of the point p within the domain D in the 3D volume V , i and j are the pair of pixel values, and d is the distance between the pair of points along the direction θ . Then we can obtain fifty matrixes and use their energies, entropies, contrasts, and homogeneities as the 3D GLCM features to enrich the information of the lung nodules.

Finally, the radiomics feature, FR , is represented by linking the LBP, HOG, and GLCM features.

GCN features extraction

CNN can extract effective and robust features adaptively and automatically without dependence on human subjectivity. We utilize the five different architectures of CNN models, AlexNet, GoogLeNet, VGG, ResNet, and AttentionNet for providing a rich and complementary description of lung nodules.

AlexNet is a simple convolutional neural network architecture that consists of five consecutively stacked convolutional layers with a combination of max pooling followed by three fully connected layers. GoogLeNet is a 22-layer deep convolutional neural network that can choose between multiple convolutional filter sizes in each block. VGG has several convolutional layers of 3×3 or 1×1 filters and three fully connected layers. ResNet is a short name for the residual network which can learn the residual function by using the skip connection connecting activations of a layer to further layers. AttentionNet is short for residual attention network which is constructed by stacking multiple attention modules based on the ResNet.

Different models reflect different perspectives on the feature description of nodules. AlexNet tends to capture the important information covering a large area. GoogLeNet holds attention to extract the features at different scales, VGG can generate arbitrarily complex decision features with the help of the non-linear activation functions, ResNet is inclined to solve the problem of vanishing gradient and learn the identity features, and AttentionNet adaptively generates attention-aware features.

Aiming to mine the relationship among these deep features, the GCN is used for learning a function of features in a graph structure, instead of linking these deep features from the different CNN models directly.

A graph consists of a set of nodes and edges, $G = (V, A)$ where node $V_i \in V$ represents the

features of the i -th node and edge $A_{ij} \in A$ represents the similarity between the features of the i -th and j -th node. In our work, the nodes are defined by $V = \{F_{AlexNet}, F_{GoogLeNet}, F_{VGG}, F_{ResNet}, F_{AttentionNet}\}$ where the features are extracted by the five different CNN models. The edges are the adjacency matrix A , representing the divergence of different models, which can be defined as

$$A_{ij} = 1 - similarity(Net_i, Net_j), \quad (5)$$

where $similarity(Net_i, Net_j)$ measures the correlation between the ability of i -th and j -th Net. It is calculated by

$$similarity(Net_i, Net_j) = \sum_{k \in S} \frac{Bf(k)}{S} \quad (6)$$

$$Bf(k) = \begin{cases} 1, & C_{Net_i} = C_{Net_j} \\ 0, & C_{Net_i} \neq C_{Net_j} \end{cases},$$

where S is the number of training samples. $Bf(.)$ is the binary function producing 1 when the model Net_i and Net_j make the same prediction on the same sample in the training data or producing 0 when they make a different prediction.

Based on the above definition, the GCN can fuse the features by following the layer-wise propagation rule

$$H^{l+1} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right), \quad (7)$$

where \tilde{A} is the adjacency matrix with added self-connections, which is defined by $\tilde{A} = A + I$, I is the identity matrix, \tilde{D} is the degree matrix of \tilde{A} . $H^{(l)}$ and $W^{(l)}$ are the output and trainable weight matrix at the l -th layer, respectively, and $H^{(0)}$ is the input of the graph convolutional network. In our work, the GCN consists of two layers. $\sigma(.)$ is the activation function, ReLU.

Finally, the fused features by GCN, FG , are obtained by concatenating the features of five nodes at the last layer.

Classification decision

Each CNN model can predict the probability of each nodule belonging to the malignant or benign nodules according to its training schedule. To improve the classification performance, we fine-tune each CNN model by mixing the radiomics features (FR), GCN features (FG), and its original CNN features on its trained model for obtaining complementary and useful information from the other models. To address class imbalance during training and focus learning on hard misclassified nodules, the Focal Loss is used for training each model, which is formulated as follows

$$L = \begin{cases} -\alpha(1 - y')^\gamma \log(y'), & y = 1 \\ -(1 - \alpha)y'^\gamma \log(1 - y'), & y = 0 \end{cases} \quad (8)$$

where y and y' are the true labels and predicted probabilities of nodules, γ and α are two hyperparameters, adjusting the rate of hard misclassified nodules and balancing the importance of malignant and benign nodules. α is set to 0.25 and γ is set to 2 in our experiments.

The ensemble learning integrates the multiple results from the above models into a consistency function for the final classification decision with weighted averaging. Let N be the number of models (five in this work), parameter w_i is the weight parameter, and it can be defined by the proportion of accuracy of all models, like

$$W_i = \frac{Acc_i}{\sum_{j=1}^N Acc_j} \quad (9)$$

Since the weight of the single model can evaluate its confidence levels of classification ability, the predictions from the models, P_{Net_i} , are fused to get the final probability P belonging to the malignant or benign nodule using a weighting scheme, by

$$P = \sum_{i=1}^N W_i \times P_{Net_i} \quad (10)$$

Experiments

Dataset and implementation details

The proposed method was evaluated on the LIDC-IDRI dataset [56], which contained 1018 clinical chest CT scans with lung nodules annotated by four experienced radiologists. The malignancy of each nodule was evaluated using a 5-point scale from benign to malignant. In this study, only the nodules whose diameters are bigger than 3 mm and which were annotated by at least three radiologists are included. We calculated the mean malignancy rating scale of each lung nodule and annotated a nodule whose malignancy rating scale ≤ 2.75 as benign, a nodule whose malignancy rating in $[2.75, 3.125]$ as uncertain, and a nodule whose malignancy rating scale ≥ 3.125 as malignant. We obtained a final total of 798 lung nodules, including 313 malignant nodules and 485 benign nodules. In addition, we split these nodules into ten disjoint subsets nearly evenly to conduct 10-fold cross-validation experiments and guarantee that the nodules in different subsets came from different patients to avoid bias in measuring performance.

We normalized the slice thickness of these CT scans to 1 mm, and cut the center of the nodules into the cubes with $56 \times 56 \times 8$, which were reshaped into $128 \times 128 \times 8$. The proposed method was mainly

implemented using Python 3.7 in the PyTorch framework. The experimental platform was equipped with an NVIDIA GeForce RTX 3080 Ti with 12 GB of memory. The proposed model was trained using the Adam optimizer and the maximum iteration number of 100. The initial learning rate was set as 0.0001 and decreased by a weight decay of e^{-4} .

Evaluation metrics

To evaluate the proposed method quantitatively, six metrics were used. They are accuracy, sensitivity, specificity, precision, F1 Score, and area under the receiver operator curve (AUC). Let TP, TN, FP, and FN be the number of true positives, true negatives, false positives, and false negatives, respectively. The true positive means that a malignant nodule is correctly classified as malignant. The false negative means that a malignant nodule is incorrectly classified as benign. In the same way, true negative means that a benign nodule is correctly classified as benign, whereas false positive means that a benign nodule is incorrectly classified as malignant. The six metrics are calculated as follows:

1. Accuracy measures the ratio of the number of correctly classified nodules to the number of all nodules, like

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

2. Sensitivity measures the proportion of malignant nodules that are identified correctly, and it is also called recall. It is calculated as follows

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \quad (12)$$

3. Specificity measures the proportion of benign nodules that are identified correctly and it is calculated as follows

$$Specificity = \frac{TN}{TN + FP} \quad (13)$$

4. Precision is the fraction of retrieved true positive instances among the retrieved positive instances. It is defined by

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

5. The F1 Score is a statistical analysis of binary classification and it is calculated from the precision and recall as

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{15}$$

6. The AUC is sensitive to an imbalance among the classes. The AUC is the area under the ROC curve which is a graphical plot that illustrates the diagnostic ability of a binary classification method as its discrimination threshold is varied. The AUC is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. For a predictor f , an unbiased estimator of its AUC can be expressed by

$$AUC(f) = \frac{\sum_{t_0 \in D^0} \sum_{t_1 \in D^1} B(f(t_0) < f(t_1))}{|D^0| \times |D^1|}, \tag{16}$$

where $B(f(t_0) < f(t_1))$ denotes an indicator function which returns 1 iff $f(t_0) < f(t_1)$, otherwise return 0. D^0 is the set of negative samples and D^1 is the set of positive samples.

Classification results

The classification results of the 10-fold cross-validation evaluating the performance of the proposed RGD model are shown in Table 1. In Table 1, we can see that we can achieve a mean accuracy of $93.25\% \pm 0.021$, a sensitivity of $89.22\% \pm 0.045$, a specificity of $95.82\% \pm 0.032$, a precision of $92.46\% \pm 0.058$, F1-score of 0.9114 ± 0.029 , and AUC of 0.9629 ± 0.018 . Those good and reliable results with small standard deviations show the effectiveness and robustness of the proposed method on lung nodule classification.

Ablation study

In this section, we perform ablation studies to further analyze the effectiveness of the fusion of the radiomics, GCN, and CNN features.

We conduct the classification performance on the single CNN model, the fusion of the radiomics and CNN, the fusion of the GCN and CNN, the fusion of the radiomics, GCN, and single CNN, the fusion of the five CNN models, and list their classification accuracies along with our RGD model in the 10-fold cross-validation experiments in Table 2, where the $CNN_1, CNN_2, CNN_3, CNN_4,$ and CNN_5 represent the AlexNet, VGG13, Resnet34, Attention56, and GoogLeNet, respectively, CNN_s represents the fused DNN model which merges the above five CNN features together, R and G are short for radiomics and GCN features, respectively, and ‘+’ means the fusion. In Table 2, we can see that the performance can be improved by involving the radiomics or GCN features for most CNN models. Especially, embedding GCN features into the CNN models could greatly improve classification performance. The fusion of the radiomics, GCN, and CNN has a significantly increased rate of 3.24%, 2.37%, 2.06%, 3.80%, and 2.53% compared with the original CNN model, respectively. The CNN_s model has higher average accuracy than the single CNN model, which proves that the multiple and different CNN models can extract discriminant features. The $CNN_s + R$ model has a reduction rate of 1.3% compared with our RGD, which demonstrates again our GCN merging in RGD is a better way of fusing features compared with the fully connected DNN merging. In addition, our RGD achieves the highest accuracy by integrating the multiple decisions from the fused CNN models into a consistent decision with ensemble averaging.

Table 1 Performance of proposed RGD model in 10-fold cross-validation experiments

Folds	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
1	93.83%	84.38%	100%	100%	0.9153	0.9566
2	95.00%	90.63%	97.92%	96.67%	0.9355	0.9785
3	95.00%	96.77%	93.88%	90.91%	0.9375	0.9921
4	88.61%	87.10%	89.58%	84.38%	0.8571	0.9516
5	90.91%	82.76%	95.83%	82.31%	0.8727	0.9231
6	93.59%	90.00%	95.83%	93.10%	0.9153	0.9701
7	96.20%	90.32%	100%	100%	0.9492	0.9758
8	94.87%	93.55%	95.74%	93.55%	0.9355	0.9629
9	92.31%	83.33%	97.92%	96.15%	0.8929	0.9688
10	92.21%	93.33%	91.49%	87.50%	0.9032	0.9496
Mean	$93.25\% \pm 0.021$	$89.22\% \pm 0.045$	$95.82\% \pm 0.032$	$92.46\% \pm 0.058$	0.9114 ± 0.029	0.9629 ± 0.018

Table 2 Accuracy of ablation study in 10-fold cross-validation experiments

Model	1	2	3	4	5	6	7	8	9	10	Mean
<i>CNN</i> ₁	89.71%	92.71%	88.54%	87.36%	87.50%	88.24%	95.83%	85.12%	91.37%	89.67%	89.61%
<i>CNN</i> ₁ + <i>R</i>	84.50%	91.66%	89.58%	87.36%	88.54%	83.48%	90.63%	83.78%	88.99%	89.66%	87.82%
<i>CNN</i> ₁ + <i>G</i>	92.71%	92.71%	93.75%	89.44%	92.71%	91.07%	95.83%	93.45%	94.79%	89.18%	92.56%
<i>CNN</i> ₁ + <i>R</i> + <i>G</i>	92.71%	91.67%	93.75%	89.45%	92.71%	91.07%	95.83%	93.45%	94.79%	89.66%	92.51%
<i>CNN</i> ₂	91.67%	91.25%	91.25%	88.58%	90.00%	89.46%	91.17%	91.25%	93.75%	89.42%	90.78%
<i>CNN</i> ₂ + <i>R</i>	88.54%	92.50%	91.25%	88.58%	91.25%	92.14%	95.00%	91.25%	91.07%	90.67%	91.23%
<i>CNN</i> ₂ + <i>G</i>	92.71%	92.50%	95.00%	87.42%	90.00%	94.64%	95.00%	92.32%	92.50%	91.63%	92.37%
<i>CNN</i> ₂ + <i>R</i> + <i>G</i>	91.67%	93.75%	96.25%	88.67%	91.25%	92.32%	97.50%	93.75%	92.50%	91.63%	92.93%
<i>CNN</i> ₃	91.67%	86.25%	91.25%	88.58%	90.00%	89.46%	91.17%	91.25%	93.75%	89.42%	90.28%
<i>CNN</i> ₃ + <i>R</i>	91.67%	90.00%	90.00%	92.33%	90.00%	85.89%	92.50%	88.57%	90.00%	91.92%	90.29%
<i>CNN</i> ₃ + <i>G</i>	91.67%	92.50%	93.75%	86.17%	88.75%	92.14%	95.00%	93.75%	92.50%	90.38%	91.66%
<i>CNN</i> ₃ + <i>R</i> + <i>G</i>	92.71%	92.50%	95.00%	88.67%	90.00%	90.89%	95.00%	92.50%	92.50%	91.63%	92.14%
<i>CNN</i> ₄	92.86%	89.29%	94.05%	89.29%	85.90%	88.33%	86.25%	87.50%	90.83%	91.25%	89.56%
<i>CNN</i> ₄ + <i>R</i>	89.77%	92.50%	91.25%	88.57%	87.50%	87.50%	90.00%	88.75%	93.75%	92.50%	90.21%
<i>CNN</i> ₄ + <i>G</i>	93.19%	93.75%	95.00%	87.14%	88.75%	93.33%	96.25%	93.75%	93.75%	91.25%	92.62%
<i>CNN</i> ₄ + <i>R</i> + <i>G</i>	93.19%	93.75%	95.00%	87.14%	90.00%	94.58%	97.50%	93.75%	92.50%	92.50%	92.99%
<i>CNN</i> ₅	89.77%	90.00%	92.50%	91.07%	86.25%	89.58%	92.32%	93.75%	92.08%	90.00%	90.73%
<i>CNN</i> ₅ + <i>R</i>	89.77%	90.00%	95.00%	88.57%	88.57%	91.25%	93.75%	95.00%	91.25%	90.00%	91.32%
<i>CNN</i> ₅ + <i>G</i>	92.05%	92.50%	96.25%	88.57%	88.75%	92.08%	96.25%	95.00%	92.50%	91.25%	92.52%
<i>CNN</i> ₅ + <i>R</i> + <i>G</i>	93.19%	96.25%	95.00%	88.39%	88.75%	93.75%	96.25%	95.00%	92.50%	91.25%	93.03%
<i>CNN</i> ₅	93.19%	92.50%	90.00%	87.14%	88.75%	92.50%	93.75%	93.75%	90.00%	95.00%	91.66%
<i>CNN</i> ₅ + <i>R</i>	93.19%	92.5%	90.00%	87.14%	90.00%	93.75%	95.00%	93.75%	91.25%	93.75%	92.03%
<i>RGD</i>	93.83%	95.00%	95.00%	88.61%	90.91%	93.59%	96.20%	94.87%	92.31%	92.21%	93.25%

Then, we test the effect of different adjacency matrices in GCN. We record the effect of different weight functions for the edges in the graph and show the accuracy in Fig. 2. The adjacency matrix *A* defines the relationship of nodes in the graph. “*A*-0” is a (0,1)-matrix in which the entries outside the main diagonal are all zero. “*A*-1” represents a complete graph whose adjacency matrix is an all-ones matrix. “*A*-*S*” represents a weight matrix, measuring the similarity of nodes, while “*A*-1-*S*” means our

proposed adjacency matrix representing the divergence of nodes. In Fig. 2, we can see that any GCN with different adjacency matrices can achieve the above accuracy of 91.9%, which outperforms the single CNN model shown in Table 2. Those results prove the effectiveness of GCN again. In addition, our proposed adjacency matrix measuring the divergence of different models can obtain the highest accuracy, which demonstrates that our method can make full use of important information from the

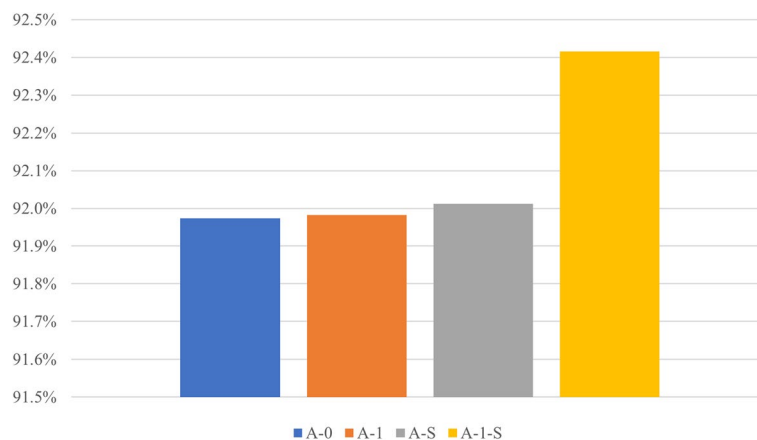


Fig. 2 Accuracy of ablation study with the different adjacency matrices in GCN

different models and fuse them better for obtaining the discriminative features.

Next, we test the effect of different numbers of CNN models and record the classification performance in Table 3, where N is the number of involved CNN models. When N is four, we only involve the AlexNet, VGG, ResNet, and AttentionNet models. The ShuffleNet [57] is added to the five current models if N is six, while the Mobilenet [58] is added again if N is seven. In Table 3, we can see that the classification performance is getting better with the increase of N until N is six. The performance with N of six and seven is worse than that when N is four. It indicates that our model with $N = 5$ has reached the best result. In addition, the model has more parameters with the bigger N , but the difference is small and almost negligible. The reasoning time is almost the same, no matter what N is. Hence, the number of involved CNN models has a certain influence on the effectiveness and does not have much effect on the efficiency. The experiment shows that our proposed RGD model with $N = 5$ has the best classification performance.

Finally, we visualize the CNN features and GCN-embedded CNN features of test data using t-distributed stochastic neighborhood embedding (t-SNE) in Fig. 3, where the benign and malignant nodules are presented

by red pentagons and blue circles, respectively. In Fig. 3, we can see that the GCN-CNN features are more compact within classes and enforce a larger angular margin between classes compared with the CNN features. It can be concluded that the aggregation of the GCN feature can enhance feature discrimination and is beneficial to separating the benign and malignant nodules into two distinct groups.

Comparison with other methods

To further evaluate the performance of the proposed RGD, we compared our approach with the state-of-the-art methods of lung nodules in the public LIDC-IDRI database, and the compared results are shown in Fig. 4. Specifically, on the one hand, compared with these methods [23, 27–30, 32], which designed a novel single CNN model by involving the three-dimensional space, multi-scale, multi-view, local and global, and other information for the lung nodule classification, our method outperforms those single-model methods because our model is a mixture model combining the advantages of the different modeling strategies. On the other hand, compared with the multi-model fusion frameworks [34, 39–41], multi-task fusion frameworks [43, 44], and multi-features fusion frameworks [49–51], our model still achieves the

Table 3 Evaluation with different numbers of involved CNN models

N	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC	Size	Time
4	92.49%	87.94%	95.39%	92.70%	0.9016	0.9559	14.897743M	1.7ms
5	93.25%	89.22%	95.82%	92.46%	0.9114	0.9629	14.897747M	1.7ms
6	92.31%	87.10%	95.74%	93.10%	0.9000	0.9499	14.897751M	1.7ms
7	92.12%	87.95%	94.77%	91.85%	0.8971	0.9478	14.897755M	1.7ms

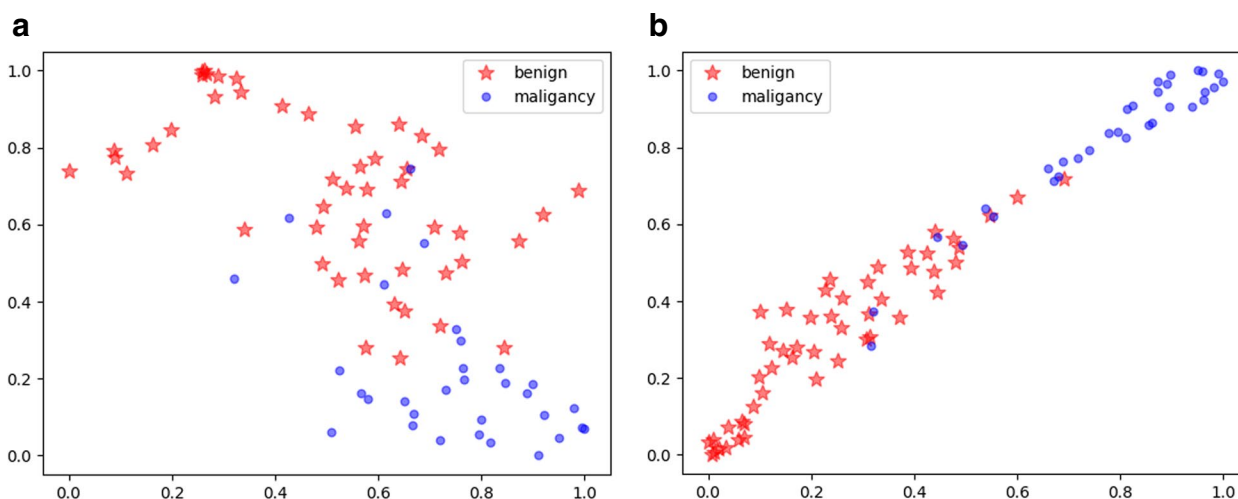


Fig. 3 t-SNE embedding using CNN features and the GCN-CNN features. The fused GCN-CNN features can present better differentiation of benign and malignant nodules compared with the CNN features. **a** t-SNE embedding of CNN feature. **b** t-SNE embedding of GCN-CNN feature

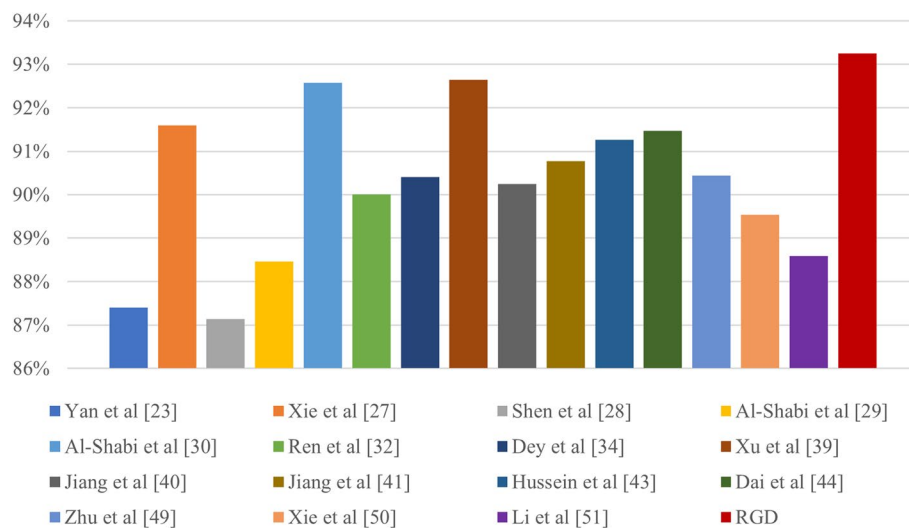


Fig. 4 The classification accuracy of the proposed RGD model and the state-of-the-art methods on the LIDC-IDRI dataset

highest classification accuracy because the radiomics can capture the relationships between CT image characteristics and nodule classes, graph learning can explore the potential relationships between different deep features to achieve discriminative features, and our model involves them into the five different CNN models training in ensemble schemes for robust performance. The comparison between our RGD model and state-of-the-art methods demonstrates the effectiveness of the proposed method in classifying malignant from benign nodules on this database.

Discussion

In this study, by fusing the Radiomics and GCN features into five different Deep CNN networks to improve the feature representation, a benign-malignant lung nodule on CT image classification algorithm, named RGD, is established. The extensive experimental results indicated that the proposed model demonstrated outstanding performance in differentiating malignant and benign pulmonary nodules on CT images.

Many automated lung nodule classification approaches have been proposed in the literature and proven to be effective. The deep convolutional neural networks are now increasingly being employed to learn highly representative, hierarchical image features. Multiple CNN models are combined at the feature or decision level to form a new feature representation or learn with optimized decision-making [34, 39–41]. Radiomics signatures have the advantage of differentiating benign and malignant nodules [59, 60]. The fusion of the radiomics and CNN models has been proven to be an efficient strategy [61–63].

In our lung nodule classification method, the radiomics, the graph convolutional network, and five different CNN networks are fused for the first time. Our method can learn the discriminative features between the different CNN features, based on the effective design of an adjacency matrix in a graph convolutional network. With the addition of radiomics, the performance is steadily and robustly improved. Hence, our fused method can effectively separate malignant from benign nodules when tested on the LIDC-IDRI dataset. Sufficient experimental results demonstrate that our method outperforms other fused methods [43, 44, 49–51].

There were some limitations in the current study. On the one hand, radiomics features are sensitive to nodule variability. As shown in Table 2, the model involving the radiomics cannot improve the performance in some folds for some CNN models. On the other hand, there may be overfitting problems for the GCN. In Table 3, we find that the number of nodes in GCN is larger and the performance is worse when N is bigger than five.

In the future, we will combine the radiomics in a more complex way, such as fusing the radiomics in front of the network or involving more radiomics features to extract more valuable and robust information about nodules. Moreover, we will integrate the attention module into the graph convolutional network to extract the significant features for reducing the overfitting problem.

Conclusion

In this study, we propose a fusion algorithm combining the Radiomics, and GCN features into the multiple Deep CNNs, named RGD, for benign-malignant lung nodule classification on CT images. We extract the

radiomics features of the pulmonary nodule to characterize the nodule phenotype, capture deep complementary features by using five different CNN models, and utilize the GCN to aggregate optimal features in non-Euclidean space. Finally, we fuse the radiomics features, and the GCN features into the 3D CNNs for generating multiple predictions and acquiring the classification result by ensemble decision. Experimental results demonstrate the effectiveness of the proposed method in extensive ablation studies and comparisons with the current state-of-the-art approaches on the LIDC-IDRI dataset. In the future, we will improve the proposed method and apply it to aid radiologists in the early lung cancer diagnosis process.

Acknowledgements

Not applicable.

Authors' contributions

Ling Ma and Chuangye Wan conceived of the presented idea and wrote the manuscript. Chuangye Wan performed most of the experiments. Xexin Hao and Annan Cai verified the analytical methods. Lizhi Liu supervised the project. All authors read and approved the manuscript.

Funding

This research was partly supported by the National Natural Science Foundation of China under Grant No. 61901234.

Availability of data and materials

The dataset analyzed during the current study is a public dataset named LIDC-IDRI and is available at <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254>.

Declarations

Ethics approval and consent to participate

This study proposed a new machine learning method for benign-malignant lung nodule classification on CT images. The CT images used in this study are from the public LIDC-IDRI dataset. We have gotten permission by accepting its license and accessed these images from the link, <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254>.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 7 March 2023 Accepted: 12 October 2023

Published online: 28 November 2023

References

- Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17–48.
- Pastorino U, Silva M, Sestini S, Sabia F, Boeri M, Cantarutti A, et al. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. *Ann Oncol.* 2019;30(7):1162–9.
- Davis AM, Cifu AS. Lung cancer screening. *JAMA.* 2014;312(12):1248–9.
- Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N Engl J Med.* 2017;377(9):849–61.
- Hawkins S, Wang H, Liu Y, Garcia A, Stringfield O, Krewer H, et al. Predicting malignant nodules from screening CT scans. *J Thorac Oncol.* 2016;11(12):2120–8.
- Thakur SK, Singh DP, Choudhary J. Lung cancer identification: a review on detection and classification. *Cancer Metastasis Rev.* 2020;39:989–98.
- Sorensen L, Shaker SB, De Bruijne M. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Trans Med Imaging.* 2010;29(2):559–69.
- Quellec G, Lamard M, Cazuguel G, Cochener B, Roux C. Wavelet optimization for content-based image retrieval in medical databases. *Med Image Anal.* 2010;14(2):227–41.
- 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.
- Lin PL, Huang PW, Lee CH, Wu MT. Automatic classification for solitary pulmonary nodule in CT image by fractal analysis based on fractional Brownian motion model. *Pattern Recognit.* 2013;46(12):3279–87.
- El-Baz A, Nitzken M, Khalifa F, Elnakib A, Gimel'farb G, Falk R, et al. 3D shape analysis for early diagnosis of malignant lung nodules. In: *Information Processing in Medical Imaging: 22nd International Conference, IPMI 2011, Kloster Irsee, Germany, July 3–8, 2011. Proceedings 22.* Springer; 2011. p. 772–783.
- Liu X, Ma L, Song L, Zhao Y, Zhao X, Zhou C. Recognizing common CT imaging signs of lung diseases through a new feature selection method based on Fisher criterion and genetic optimization. *IEEE J Biomed Health Inform.* 2014;19(2):635–47.
- Ma L, Liu X, Fei B. Learning with distribution of optimized features for recognizing common CT imaging signs of lung diseases. *Phys Med Biol.* 2016;62(2):612.
- Dy JG, Brodley CE, Kak A, Broderick LS, Aisen AM. Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE Trans Pattern Anal Mach Intell.* 2003;25(3):373–8.
- Mousa WA, Khan MA. Lung nodule classification utilizing support vector machines. In: *Proceedings. International Conference on Image Processing. vol. 3. IEEE; 2002. p. III.*
- Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. *Application Radiology.* 1993;186(2):415–22.
- Lee SLA, Kouzani AZ, Hu EJ. Random forest based lung nodule classification aided by clustering. *Comput Med Imaging Graph.* 2010;34(7):535–42.
- Ma L, Liu X, Song L, Zhou C, Zhao X, Zhao Y. A new classifier fusion method based on historical and on-line classification reliability for recognizing common CT imaging signs of lung diseases. *Comput Med Imaging Graph.* 2015;40:39–48.
- Mastouri R, Khelifa N, Neji H, Hantous-Zannad S. Deep learning-based CAD schemes for the detection and classification of lung nodules from CT images: A survey. *J X-ray Sci Technol.* 2020;28(4):591–617.
- Gu Y, Chi J, Liu J, Yang L, Zhang B, Yu D, et al. A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning. *Comput Biol Med.* 2021;137:104806.
- Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. *Comput Soc Networks.* 2019;6(1):1–23.
- Mitra S, Shankar BU. Integrating radio imaging with gene expressions toward a personalized management of cancer. *IEEE Trans Hum-Mach Syst.* 2014;44(5):664–77.
- 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: *Computer Vision—ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20–24, 2016, Revised Selected Papers, Part III.* Springer; 2017. p. 91–101.
- 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. p. 588–599.
- Liu X, Hou F, Qin H, Hao A. Multi-view multi-scale CNNs for lung nodule type classification from CT images. *Pattern Recognit.* 2018;77:262–75.
- Kang G, Liu K, Hou B, Zhang N. 3D multi-view convolutional neural networks for lung nodule classification. *PLoS ONE.* 2017;12(11):e0188290.

27. Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M, et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. *IEEE Trans Med Imaging*. 2018;38(4):991–1004.
28. 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 Recognit*. 2017;61:663–73.
29. Al-Shabi M, Lan BL, Chan WY, Ng KH, Tan M. Lung nodule classification using deep local-global networks. *Int J CARS*. 2019;14:1815–9.
30. Al-Shabi M, Lee HK, Tan M. Gated-dilated networks for lung nodule classification in CT scans. *IEEE Access*. 2019;7:178827–38.
31. Al-Shabi M, Shak K, Tan M. ProCAN: Progressive growing channel attentive non-local network for lung nodule classification. *Pattern Recognit*. 2022;122:108309.
32. Ren Y, Tsai MY, Chen L, Wang J, Li S, Liu Y, et al. A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification. *Int J CARS*. 2020;15:287–95.
33. Lei Y, Tian Y, Shan H, Zhang J, Wang G, Kalra MK. Shape and margin-aware lung nodule classification in low-dose CT images via soft activation mapping. *Med Image Anal*. 2020;60:101628.
34. Dey R, Lu Z, Hong Y. Diagnostic classification of lung nodules using 3D neural networks. In: IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE. 2018;2018:774–8.
35. Zhao X, Liu L, Qi S, Teng Y, Li J, Qian W. Agile convolutional neural network for pulmonary nodule classification using CT images. *Int J CARS*. 2018;13:585–95.
36. Zhao X, Qi S, Zhang B, Ma H, Qian W, Yao Y, et al. Deep CNN models for pulmonary nodule classification: model modification, model integration, and transfer learning. *J X-ray Sci Technol*. 2019;27(4):615–29.
37. Wang J, Chen X, Lu H, Zhang L, Pan J, Bao Y, et al. Feature-shared adaptive-boost deep learning for invasiveness classification of pulmonary subsolid nodules in CT images. *Med Phys*. 2020;47(4):1738–49.
38. Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, et al. Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. *Int J CARS*. 2020;15:173–8.
39. Xu X, Wang C, Guo J, Gan Y, Wang J, Bai H, et al. MSCS-DeepLN: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks. *Med Image Anal*. 2020;65:101772.
40. 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.
41. Jiang H, Shen F, Gao F, Han W. Learning efficient, explainable and discriminative representations for pulmonary nodules classification. *Pattern Recognit*. 2021;113:107825.
42. Liu S, Xie Y, Jirapatnakul A, Reeves AP. Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks. *J Med Imaging*. 2017;4(4):041308.
43. Hussein S, Cao K, Song Q, Bagci U. Risk stratification of lung nodules using 3D CNN-based multi-task learning. In: *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25–30, 2017, Proceedings 25*. Springer; 2017. p. 249–260.
44. Dai Y, Yan S, Zheng B, Song C. Incorporating automatically learned pulmonary nodule attributes into a convolutional neural network to improve accuracy of benign-malignant nodule classification. *Phys Med Biol*. 2018;63(24):245004.
45. Khosravan N, Bagci U. Semi-supervised multi-task learning for lung cancer diagnosis. In: 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE. 2018;2018:710–3.
46. Chen S, Qin J, Ji X, Lei B, Wang T, Ni D, et al. Automatic scoring of multiple semantic attributes with multi-task feature leverage: a study on pulmonary nodules in CT images. *IEEE Trans Med Imaging*. 2016;36(3):802–14.
47. Buty M, Xu Z, Gao M, Bagci U, Wu A, Mollura DJ. Characterization of lung nodule malignancy using hybrid shape and appearance features. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part 1*. Springer; 2016. p. 662–670.
48. Wang C, Elazab A, Wu J, Hu Q. Lung nodule classification using deep feature fusion in chest radiography. *Comput Med Imaging Graph*. 2017;57:10–8.
49. Zhu W, Liu C, Fan W, Xie X, Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In: IEEE winter conference on applications of computer vision (WACV). IEEE. 2018;2018:673–81.
50. Xie Y, Zhang J, Xia Y, Fulham M, Zhang Y. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Inf Fusion*. 2018;42:102–10.
51. Li S, Xu P, Li B, Chen L, Zhou Z, Hao H, et al. Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features. *Phys Med Biol*. 2019;64(17):175012.
52. Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Sci Rep*. 2017;7(1):588.
53. Tomassini S, Falconelli N, Sernani P, Burattini L, Dragoni AF. Lung nodule diagnosis and cancer histology classification from computed tomography data by convolutional neural networks: a survey. *Comput Biol Med*. 2022;146:105691.
54. Feng B, Chen X, Chen Y, Lu S, Liu K, Li K, et al. Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas. *Eur Radiol*. 2020;30:6497–507.
55. Chen M, Wei Z, Huang Z, Ding B, Li Y. Simple and deep graph convolutional networks. In: *International conference on machine learning*. PMLR; 2020. p. 1725–1735.
56. Armato SG III, McLennan G, 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.
57. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, IEEE; 2018. p. 6848–56.
58. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. 2017.
59. Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. *Clin Radiol*. 2020;75(1):13–9.
60. Zhang R, Shi J, Liu S, Chen B, Li W. Performance of radiomics models derived from different CT reconstruction parameters for lung cancer risk prediction. *BMC Pulm Med*. 2023;23(1):132.
61. Beig N, Khorrami M, Alilou M, Prasanna P, Braman N, Orooji M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. *Radiology*. 2019;290(3):783–92.
62. Lin X, Jiao H, Pang Z, Chen H, Wu W, Wang X, et al. Lung cancer and granuloma identification using a deep learning model to extract 3-dimensional radiomics features in CT imaging. *Clin Lung Cancer*. 2021;22(5):e756–66.
63. Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al. Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology*. 2019;292(2):365–73.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.