Original Article

Multiclassifier fusion based on radiomics features for the prediction of benign and malignant primary pulmonary solid nodules

Yao Shen[1,2], Fangyi Xu[1], Wenchao Zhu[1], Hongjie Hu[1], Ting Chen[2], Qiang Li[2]

[1]Department of Radiology, Sir Run Run Shaw Hospital Affiliated with the School of Medicine of Zhejiang University, Hangzhou 310016, China;
[2]Department of Radiology, Yinzhou Hospital Affiliated with the School of Medicine of Ningbo University, Ningbo 315040, China
Contributions: (I) Conception and design: Y Shen; (II) Administrative support: H Hu; Q Li; (III) Provision of study materials or patients: F Xu; T Chen; (IV) Collection and assembly of data: Y Shen; (V) Data analysis and interpretation: W Zhu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.
Correspondence to: Hongjie Hu, Director. Department of Radiology, Sir Run Run Shaw Hospital Affiliated with the School of Medicine of Zhejiang University. Email: hongjiehu@zju.edu.cn.

Background: To test the ability of a multiclassifier model based on radiomics features to predict benign and malignant primary pulmonary solid nodules.
Methods: Computed tomography (CT) images of 342 patients with primary pulmonary solid nodules confirmed by histopathology or follow-up were retrospectively analyzed. The region of interest (ROI) of the images was delineated, and the radiomics features of the lesions were extracted. The feature weight was calculated using the relief feature selection algorithm. Based on the selected features, five classifier models were constructed: support vector machine (SVM), random forest (RF), logistic regression (LR), extreme learning machine (ELM), and K-nearest neighbor (KNN). The precision, recall rate, and area under the receiver operating characteristic curve (AUC) were used to evaluate the prediction performance of each classifier. The prediction result of each classifier was first weighted, and then all the prediction results were fused to predict the nodule type of unknown images. The prediction precision, recall rate, and AUC of the fusion classifier and single classifier were compared. Cross-validation was used to evaluate the generalization of the fusion classifier, and *t*- and F-tests were performed on the five classifiers and fusion classifier.
Results: For each ROI, 450 features in four major categories were extracted and were analyzed using the relief feature selection algorithm. According to the weights, 25 highly repetitive and nonredundant stable features that played a major role in pulmonary nodule classification were selected. The fusion classifier's prediction performance (prediction precision =92.0%, AUC =0.915) was superior to those of SVM (prediction precision =75.3%, AUC =0.740), RF (prediction precision =89.1%, AUC =0.855), LR (prediction precision =68.4%, AUC =0.681), ELM (prediction precision =87.0%, AUC =0.830), and KNN (prediction precision =77.1%, AUC =0.702). The fusion classifier showed the best null hypothesis performance in the *t*-test (P=0.035) and F-test (P=0.036).
Conclusions: The multiclassifier fusion model based on radiomics features had high prediction value for benign and malignant primary pulmonary solid nodules.

Keywords: Pulmonary nodule; radiomics; classifier; cross-validation

Introduction

Pulmonary nodules usually refer to round or irregular lesions in the lung that are no more than 3 cm in diameter. With the advancement of spiral computed tomography (CT) scanning, reconstruction technology, and low-dose chest CT screening, the detection rate of pulmonary nodules is increasing. However, since the same kind of shadow can be cast by different diseases and different shadows can be cast by the same disease, benign and malignant nodules can be confused. In 2012, Dutch scholar Lambin *et al.* (1) first proposed the concept of radiomics. Kumar *et al.* (2) defined it as "high-throughput extraction and analysis of a large number of advanced quantitative imaging features from CT, magnetic resonance imaging (MRI), and positron emission tomography (PET)". In recent years, radiomics has played an important role in the identification of benign and malignant lesions, judgment of the malignancy of tumors, selection of treatment methods, and monitoring of therapeutic effect, and it has guiding significance for the development of personalized treatment plans (3-6). In this study, based on traditional imaging differential diagnosis, a prediction model of benign and malignant pulmonary solid nodules based on radiomics was constructed using a feature selection algorithm combined with multiple classifiers, and the prediction performance of the classifiers was quantitatively evaluated.

Methods

Clinical information

The clinical, pathological and imaging information of 342 patients with pulmonary solid nodules confirmed by histopathology or follow-up in the Sir Run Run Shaw Hospital affiliated with the School of Medicine of Zhejiang University and Yinzhou Hospital affiliated with the School of Medicine of Ningbo University from January 2015 to December 2018 were retrospectively analyzed. The inclusion criteria were (I) isolated solid nodules in the lung were identified on chest CT examination, and each nodule had a clear and complete thin layer to reconstruct the data in Digital Imaging and Communications in Medicine (DICOM) format; (II) diagnosis was confirmed by histopathology or follow-up after clinical treatment; (III) the reconstructed thin-layer image had no obvious calcification and/or fat content; (IV) there was no history of extrapulmonary malignancies; (V) it was untreated before CT examination.

CT examination

The CT scan was performed using Siemens Definition AS 64 and Philips Brilliance 64-row multilayer spiral CT scanner. The scanning range was from the thorax entrance to the adrenal level. The scanning parameters were as follows: tube voltage of 120 kVp, tube current automatic adjustment technique, pitch of 1.2, collimation of 64×0.625 mm, reconstruction layer thickness/layer spacing of 1.25 mm/0.625 mm, matrix of 512×512, reconstruction convolution function of B70f, window width of 1,200 HU, window level of –600 HU, and all images exported in DICOM format.

Image analysis

Segmentation of region of interest (ROI) of nodules

Using ITK-SNAP software (Version 3.4.0, http://www.itksnap.org/), two radiologists with 5 years of experience manually and semiautomatically delineated the maximum boundary of the nodules layer by layer. Blood vessels and bronchus were avoided. The longest diameter of the nodule and the presence of the spicule sign, lobulation sign, vacuole sign, and vessel convergence sign were extracted and recorded (*Figure 1A,B*). The ROI segmentation was checked by one senior radiologist.

Image and data preprocessing

The images and data were preprocessed using image binarization and data normalization. Image binarization (*Figure 1C*) was done to set the gray value of each pixel on the images to 0 or 1. Binary images are conducive to further processing of the images so that image information other than the ROI can be eliminated to avoid the introduction of noise. In addition, in order to eliminate the influence of feature vectors of different dimensions on the analysis results, the extracted original radiomics feature data were standardized. All the feature vectors after processing were on the same order of magnitude and conformed to the standard normal distribution, that is, the mean was 0 and the standard deviation was 1. The conversion function was $x^* = (x–\mu)/\sigma$, where x, μ, and σ are the actual value, statistical average, and standard deviation of all feature vectors, respectively.

Radiomics feature extraction

The radiomics features of all segmented nodules were extracted using the Matlab2018b software (http://www.
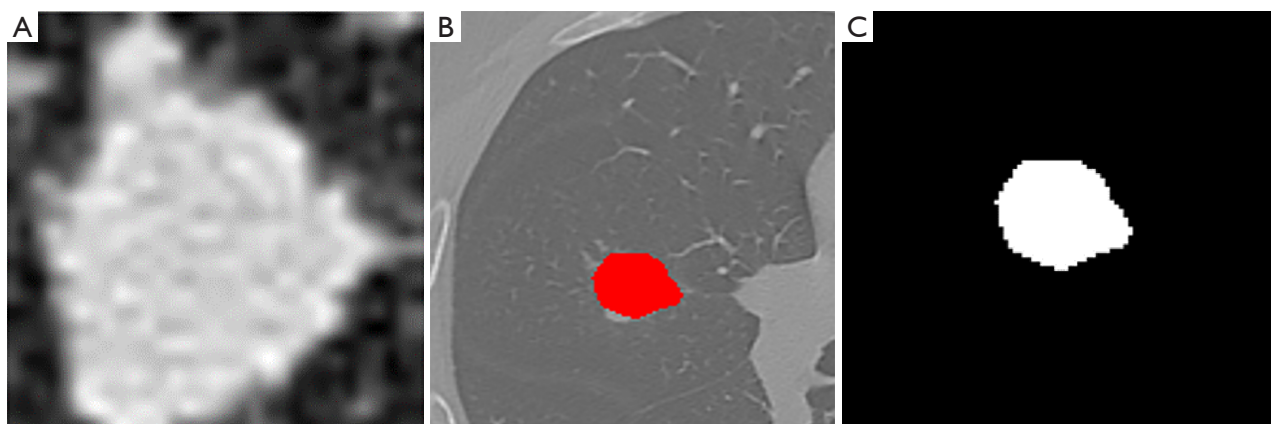
Figure 1 Pulmonary CT images with pulmonary nodules (red represents the boundary of the nodule). (A) Original CT image; (B) ROI of the nodule; (C) binarized image of the nodule. CT, computed tomography; ROI, region of interest.

mathworks.com/) (MathWorks Co., USA). A total of 450 features in four major categories were extracted from all nodules, including geometric features, texture features, gray-level features, and wavelet features. Texture features included the common gray-level cooccurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), neighborhood gray tone difference matrix (NGTDM), and neighborhood gray-level difference matrix (NGLDM). The local binary pattern (LBP) was also used to describe local texture features of the images.

Feature selection and model construction

A total of 342 samples were divided into training set, test set and verification set at a ratio of 7:2:1 using the random index method. The relief feature selection algorithm was used to screen the 450 radiomics features. After repetitive operations (an average of 10 repeats was used as the feature weight), 25 robust features with a major role in classification in the training set were screened. The classifiers were tested and verified based on the selected feature set, and a prediction model for distinguishing benign and malignant primary pulmonary solid nodules was constructed. The five classifiers were a support vector machine (SVM), random forest (RF), logistic regression (LR), extreme learning machine (ELM), and K-nearest neighbor (KNN). Finally, the weighted voting method (an algorithm that fuses classifiers (7) was used to fuse the prediction results of the above five classifiers, and each weight w_i of the weighted fusion was calculated using the Lagrangian and QR decomposition method. The Lagrangian was used to construct the objective function, and QR decomposition

was used to obtain the analytical solution. The classifier fusion method was as follows:

$$\varphi(x) = \frac{1}{2} \min \| y - h(y') \| \qquad [1]$$

$$h(y') = \sum_{i=1}^{5} w_i y'_i \qquad [2]$$

$$subject : \sum_{i=1}^{5} w_i = 1 \qquad [3]$$

The Lagrangian expression was as follows:

$$L(x, \lambda) = \phi(x) - \sum_{i=1}^{5} \lambda_i h(x) \qquad [4]$$

$$w_i = R_1^{-1} Q_1^T h(x_i) \qquad [5]$$

where w_i is the weight corresponding to the ith classifier, R_1^{-1}, Q_1^T are the QR decomposition of $h(x)$, and $h(x) = |Q_1 | Q_2| \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}$.

Cross-validation of classifiers

The method of simple cross-validation (8) was used to verify the robustness of the classifiers. The process was as follows: (I) random integers 1 to 342 (the sample size was 342) were randomly generated. (II) The data set was divided into a training set, a test set, and a verification set at a ratio of 7:2:1 using the random number index method. (III) The five classifiers were trained, tested, and verified separately, and the training, testing, and verification results of the fusion classifier were obtained based on the results of the five classifiers. (IV) The process of (I) to (III) was repeated 10 times.

Page 4 of 11

Shen et al. Multiclassifier fusion predicts primary pulmonary solid nodules

Statistics

SPSS 20.0 and Matlab2018b software were used for statistical analysis of the data in this study. Measurement data are expressed as mean ± standard deviation. Count data are expressed as a ratio or percentage. Differences in sex and imaging signs between the two groups were compared by the chi-squared test. The differences in age and the longest diameter of the lesions were compared using the two-independent-sample *T* test. P<0.05 was considered statistically significant. The diagnostic performance of the classifiers was described using precision, recall rate, and area under the receiver operating characteristic (ROC) curve (AUC).

$$\text{Precision} = TP/TP + FP \quad \text{Recall} = TP/TP + FN \quad [6]$$

Results

General information

This study enrolled 342 patients with solid pulmonary nodules. Among them, 171 patients (91 males and 80 females) had benign nodules, with an average age of 56.63±13.26 years, and the longest diameter of nodules was 1.61±0.60 cm. Of them, 120 cases were confirmed by histology, including 34 cases of inflammatory pseudotumor (inflammatory granuloma), 31 cases of tuberculosis, 25 cases of fungus infection, 17 cases of hamartoma, and 13 cases of sclerosing hemangioma. The remaining 51 cases were confirmed by follow-up after treatment to have nodules that shrank or disappeared. There were 171 patients (87 males and 84 females) with malignant nodules, with an average age of 60.92±10.28 years and a longest nodule diameter of 1.71±0.54 cm, all of which were confirmed by histology, including 51 cases of squamous cell carcinoma, 75 cases of adenocarcinoma, 9 cases of large cell carcinoma, 24 cases of small cell carcinoma, 9 cases of adenosquamous carcinoma, and 3 cases of sarcomatoid carcinoma. The general information of the patients is shown in *Table 1*.

Weight distribution of radiomics features

The entire experimental process is shown in *Figure 2*. The distribution of all 450 radiomics features between the two groups is shown in *Figure 3*. The features of benign *vs.* malignant nodules were significantly different, indicating that the extracted features objectively reflected the essential attributes of pulmonary nodules. The distribution of the top 25 features according to the weight after feature selection is shown in *Table 2* and *Figure 4*. The features with the greatest weights were mainly concentrated in texture features, wavelet features, and gray-level features.

The prediction performance of classifiers

This study employed a simple 10-fold cross-validation method to analyze the performance indicators of the classifiers, and we obtained the ROC curves (*Figures 5,6*), prediction precision, and recall curves (*Figures 7,8*) of each classifier. The fusion classifier demonstrated superior prediction performance in the test set (precision =92.0%±1.16%, recall rate =92.2%±1.22%, and AUC =0.915±0.019) and the verification set (precision =92.1%±1.25%, recall rate =92.3%±1.55%, and AUC =0.921±0.015) over any single classifier. After cross-validation, the performance indicators (precision, recall rate, and AUC) all fluctuated within a small range (*Table 3*), indicating that the fusion algorithm had strong robustness. The *t*-test and F-test were used to statistically analyze the mean and variance of the prediction performance indicators of the samples, and the results are expressed as P values. The smaller the P value, the greater the probability that the performance indicator of the samples can represent the entire population. The *t*-test (P=0.035) and F-test (P=0.036) of the fusion classifier showed the optimal null hypothesis performance (*Table 4*).

Discussion

At present, the clinical analysis of pulmonary nodule images is limited to qualitative and preliminary quantitative analysis of the lesions, including observation of the overall and marginal shape of the lesion, the uniformity of internal density, the relationship with the surrounding structure, and a rough measurement of the nodule's long and short diameters, while no in-depth or detailed analysis of the images is performed. In this study, the benign and malignant nodules overlapped on multiple imaging signs, that is, the so-called same shadow of different diseases. When the volume of the nodule is small, the images would not manifest malignant signs, so it is difficult to characterize the pulmonary nodules simply based on image signs. Radiomics is the application of digital image processing and machine learning techniques to medical image analysis (9). It entails extracting hundreds of quantitative image features from the ROI in the images, followed by screening and analyzing these features to describe the biological characteristics and

Table 1 Baseline data and imaging signs of 342 patients with solid pulmonary nodules

Baseline variables	Benign (N=171)	Malignant (N=171)	Total (N=342)	χ^2/t	P
Sex				0.187	0.665
Male	91 (53.2%)	87 (50.9%)	178 (52.0%)		
Female	80 (46.8%)	84 (49.1%)	164 (48.0%)		
Age	56.63±13.26	60.92±10.28	58.77±12.04	3.342	0.001*
Longest diameter (cm)	1.61±0.60	1.71±0.54	1.66±0.57	1.705	0.089
Shape				3.092	0.079
Irregular	45 (26.3%)	60 (35.1%)	105 (30.7%)		
Round or semiround	126 (73.7%)	111 (64.9%)	237 (69.3%)		
Vacuole sign				13.049	<0.001*
No	159 (93.0%)	136 (79.5%)	295 (86.3%)		
Yes	12 (7.0%)	35 (20.5%)	47 (13.7%)		
Bronchial inflation sign				2.974	0.085
No	122 (71.3%)	107 (62.6%)	229 (67.0%)		
Yes	49 (28.7%)	64 (37.4%)	113 (33.0%)		
Lobulation sign				2.784	0.095
No	73 (42.7%)	58 (33.9%)	131 (38.3%)		
Yes	98 (57.3%)	113 (66.1%)	211 (61.7%)		
Spicule sign				11.987	0.001*
No	99 (57.9%)	67 (39.2%)	166 (48.5%)		
Yes	72 (42.1%)	104 (60.8%)	176 (51.5%)		
Clear boundary				2.886	0.089
No	44 (25.7%)	31 (18.1%)	62 (18.1%)		
Yes	127 (74.3%)	140 (81.9%)	280 (81.9%)		
Pleural indentation sign				3.805	0.051
No	100 (58.5%)	82 (47.4%)	182 (53.2%)		
Yes	71 (41.5%)	99 (52.6%)	160 (46.8%)		
Vessel convergence sign				29.321	<0.001*
No	115 (67.3%)	65 (38.0%)	180 (52.6%)		
Yes	56 (32.7%)	106 (62.0%)	162 (47.4%)		

*, P<0.05.

heterogeneity of the lesions. Therefore, radiomics can identify information that is not visible to the naked eye in conventional imaging images, and it is not limited by lesion size or morphology (10,11).

This study extracted 450 radiomics features from 342 cases of primary pulmonary solid nodules and proposed a multiclassifier fusion method based on radiomics to predict benign and malignant nodules. The main findings include the following. The top 25 features according to weight after feature screening played a major role in the correct classification of the two groups of patients, which included texture features (NGLDM, GLRLM, GLCM, NGTDM,
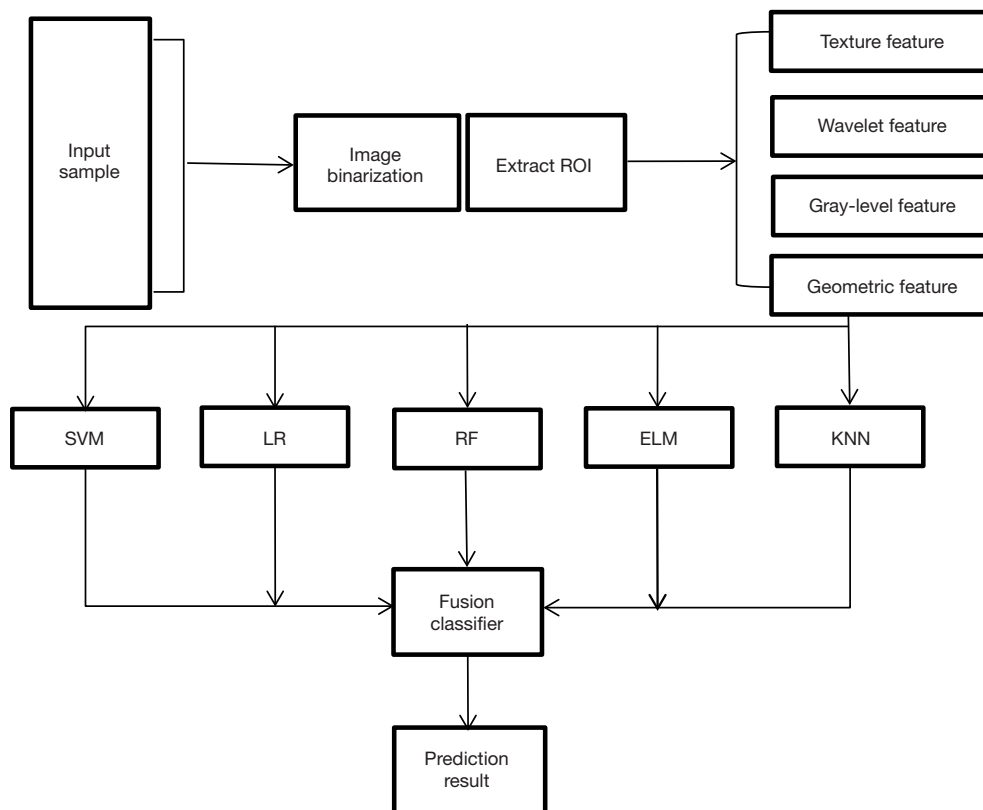
Page 6 of 11

Shen et al. Multiclassifier fusion predicts primary pulmonary solid nodules

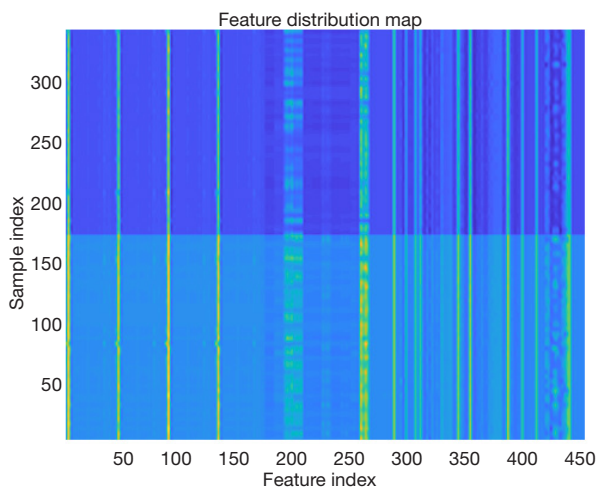**Figure 2** Flow chart of the whole experiment.

Figure 3 Distribution of radiomics features between the two groups of patients. The x-axis shows the feature index, and the y-axis shows the sample (pulmonary nodule) index. Sample indices 1–171 are benign nodules, and 172–342 are malignant nodules. Different colors represent different categories of features. The feature distribution chart indicates that the extracted features are conducive to distinguishing benign and malignant lesions.

and LBP), wavelet features, and gray-level features. Moreover, the prediction performance of the fusion classifier was better than that of any single classifier.

Gray-level features can quantitatively reflect the amplitude and frequency of pixel-value distribution in the ROI. Kamiya *et al.* (12) found that compared with benign nodules, malignant nodules showed higher skewness and lower kurtosis in gray-level features. Petkovska *et al.* (13) reported that comprehensive use of shape, size, and gray-level features could improve the AUC from 0.79 to 0.84 for distinguishing benign from malignant nodules. Chi *et al.* (14) also found that the skewness and kurtosis demonstrated statistical significance in the identification of benign and malignant nodules in their study of 110 cases of pulmonary solid nodules. Texture features can quantify the subtle differences in image pixel values and their arrangement. Compared with gray-level features, texture features have the advantage of retaining the spatial features of the lesions (15,16). In a study on mediastinal lymph nodes in patients with lung cancer, Bayanati *et al.* (17) found that entropy, gray-level nonuniformity (GLNU), and running length

Table 2 Distribution of the top 25 radiomics features according to weight after relief feature screening

Weight rank	Feature category	Feature name	Feature weight
1	NGLDM	Contrast	0.0317
2	NGLDM	Correlation	0.0310
3	NGLDM	Cluster Shade	0.0290
4	Wavelet feature	Ed (Energy Diagonal)	0.0290
5	Wavelet feature	Ev (Energy Vertical)	0.0257
6	GLRLM	SRLGLE (Short Run Low Gray-Level Emphasis)	0.0248
7	GLRLM	GLNU (Gray-Level Nonuniformity)	0.0248
8	GLRLM	HGLRE (High Gray-Level Run Emphasis)	0.0246
9	GLRLM	SRHGLE (Short Run High Gray-Level Emphasis)	0.0233
10	GLRLM	RP (Run Percentage)	0.0230
11	GLRLM	LRLGLE (Long Run Low Gray-Level Emphasis)	0.0223
12	GLRLM	LRHGLE (Long Run High Gray-Level Emphasis)	0.0223
13	GLRLM	LGLRE (Low Gray-Level Run Emphasis)	0.0223
14	GLCM	Correlation	0.0222
15	Gray-level feature	Entropy	0.0220
16	Gray-level feature	Skewness	0.0218
17	Gray-level feature	Kurtosis	0.0208
18	Gray-level feature	Variance	0.0207
19	GLCM	Homogeneity	0.0207
20	GLCM	Cluster shade	0.0207
21	Gray-level feature	Mean	0.0202
22	GLCM	Inverse difference	0.0197
23	NGLDM	Dissimilarity	0.0194
24	GLRLM	RLNU (Run Length Nonuniformity)	0.0192
25	GLRLM	SRE (Short Run Emphasis)	0.0192

NGLDM, neighborhood gray-level difference matrix; GLRLM, gray-level run-length matrix; GLCM, gray-level cooccurrence matrix; GLRLM, gray-level run-length matrix.

nonuniformity (RLNU) of the texture features could correctly distinguish the benign and malignant mediastinal lymph nodes in patients with primary lung cancer. In another study on texture features, Chi *et al*. (18) reported that the contrast, correlation, entropy, and homogeneity had value in the qualitative diagnosis of pulmonary nodules.

Our research not only confirmed that the gray-level features and texture features were important in the classification of pulmonary nodules, it also employed the relief feature selection algorithm to rank the weight of the four categories of features and selected the top 25 features as the input features of the classifiers. The results showed that the gray-level features and texture features had the greatest weights, especially the texture features, which made up the majority of the high-weight features. In contrast to principal component analysis (PCA), the relief algorithm uses the characteristics of pulmonary nodules as the evaluation indices and uses the clustering method for internal calculations. Therefore, this algorithm contains both the external and internal characteristics of
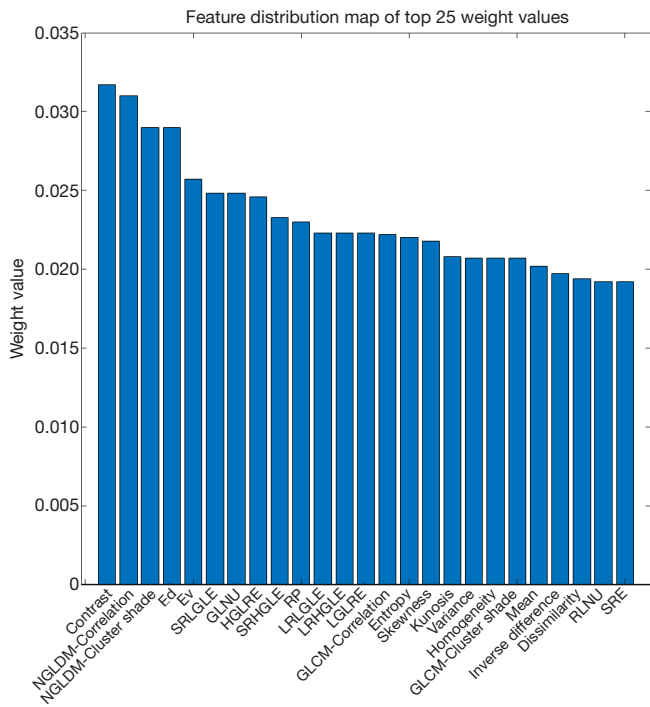
Page 8 of 11

Shen et al. Multiclassifier fusion predicts primary pulmonary solid nodules

Figure 4 Distribution of the top 25 radiomics features according to weight after relief feature screening. Distribution of top 25 radiomics features according to weight after relief feature screening. Features with high weights are mainly clustered in texture features, wavelet features, and gray-level features.
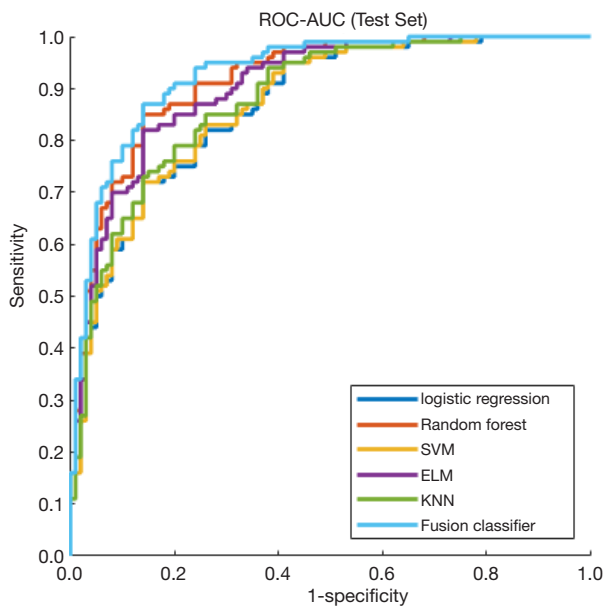

Figure 6 ROC curves of single classifiers and the fusion classifier in the verification set. The fusion classifier shows the largest AUC. ROC, receiver operating characteristic; AUC, area under the ROC curve.

Figure 5 ROC curves of single classifiers and the fusion classifier in the test set. The fusion classifier shows the largest AUC. ROC, receiver operating characteristic; AUC, area under curve.
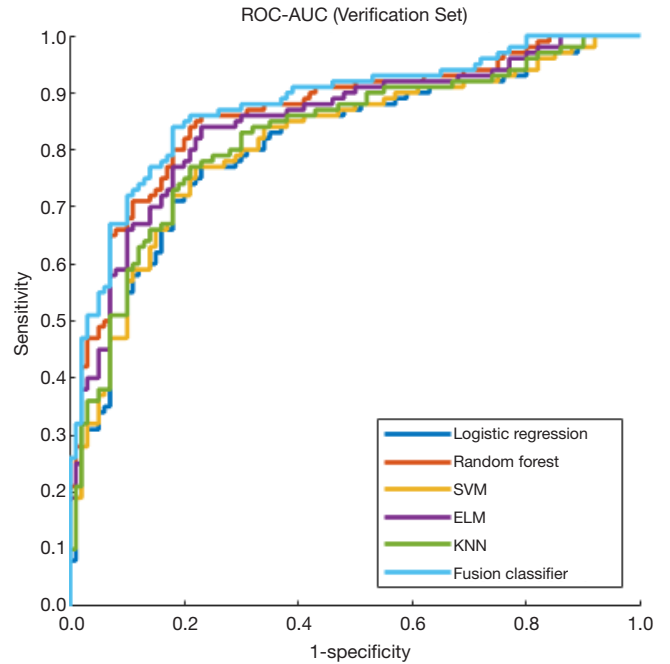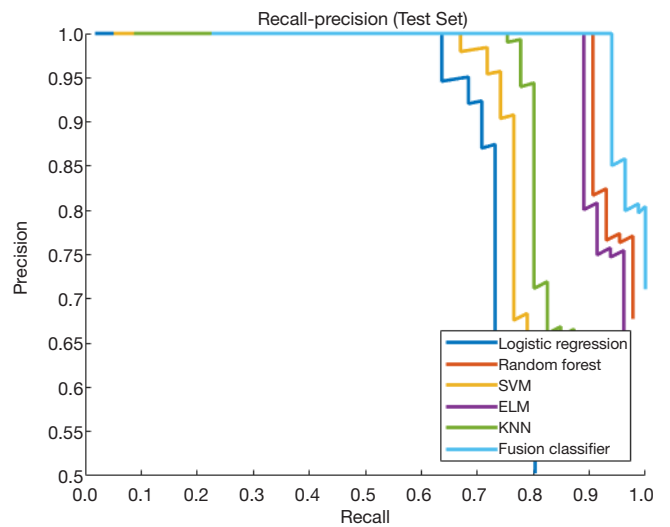

Figure 7 Recall-precision curves of single classifiers and the fusion classifier in the test set. The fusion classifier shows a higher classification precision than any single classifier.

pulmonary nodules, while the PCA method only analyzes the internal characteristics of the pulmonary nodule and thus is not representative or conducive to feature selection. Heterogeneity is a recognized feature of malignant tumors, reflecting changes in cell permeability, abnormal angiogenesis, and changes in tissue structure caused by mucus-like changes, necrosis, and fibrosis (19). Therefore, the images of malignant nodules demonstrate an uneven distribution of gray-level, complex texture, and disarranged local texture. To handle this heterogeneity, we introduced the LBP to analyze the local texture features of the images. The LBP features have the advantages of rotation

invariance, gray-level invariance, and strong resistance to image noise (20). Feature selection results (LBP feature weights ranked between 25 and 50) showed that LBP was valuable for the qualitative diagnosis of pulmonary nodules. The wavelet features are multiscale features obtained after wavelet transformation of the images, which integrates features at the boundary and vicinity of the lesions and reflects the change rate of the pixel value in the frequency domain (21,22). These features had high weights in our study, indicating that the local texture of malignant nodules was complex, the texture changed quickly, and the lesion boundary was irregular.

In terms of classifier selection, this study proposed a prediction method using multiclassifier fusion (23,24). This method weighted and fused the results of five classic classification methods to obtain the optimal prediction. The method of fusing classifiers first calculates the output weight of each single-classifier prediction to construct an objective function, then calculates the weight corresponding to the single classifier using a Lagrangian and QR decomposition method. Due to the different algorithms and working principles of each single classifier, the sensitivity of different classifiers to different data sets is different, which leads to differences in their predictive performance. The fusion classifier combines the excellent performance of each classifier and has a higher adaptation to the data and better generalization performance. This study used a simple 10-fold cross-validation method to statistically analyze the prediction performance of the fusion classifier and five single classifiers, and the fusion classifier demonstrated the best prediction performance, and the prediction precision, recall rate, and AOC fluctuated within a small range, indicating that the fusion algorithm had high robustness.
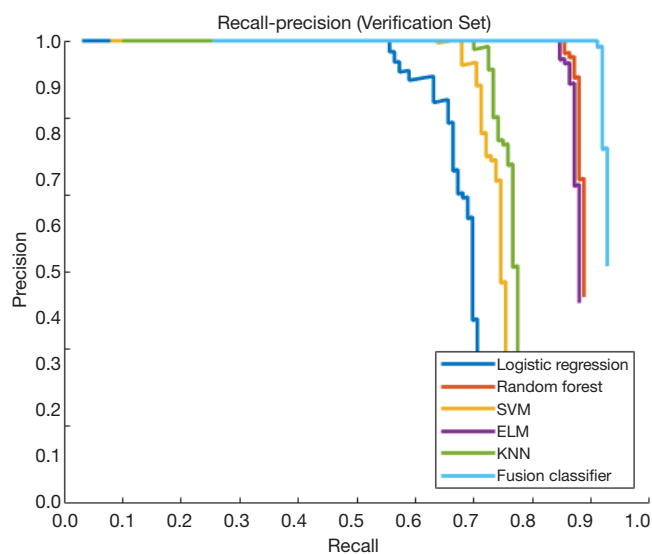

Figure 8 Recall-precision curves of single classifiers and the fusion classifier in the verification set. The fusion classifier shows a higher classification precision than any single classifier.

Table 3 Prediction performance of each classifier in the test set and verification set

Classifier	Test set			Verification set		
	Precision	Recall	AUC	Precision	Recall	AUC
KNN	77.1%±1.71%	76.8%±1.12%	0.702±0.017	76.5%±1.81%	76.6%±1.62%	0.695±0.019
SVM	75.3%±1.75%	75.9%±1.5%	0.740±0.012	75.2%±1.24%	75.7%±1.47%	0.738±0.012
ELM	87.0%±1.28%	88.7%±1.96%	0.830±0.015	87.2%±1.93%	87.4%±1.35%	0.823±0.017
RF	89.1%±1.68%	89.9%±1.34%	0.855±0.017	88.8%±1.35%	88.1%±1.83%	0.850±0.017
LR	68.4%±1.66%	70.3%±1.59%	0.681±0.018	69.2%±1.20%	69.6%±1.89%	0.685±0.013
Fusion classifier	92.0%±1.16%	92.2%±1.22%	0.915±0.019	92.1%±1.25%	92.3%±1.55%	0.921±0.015

The fusion classifier shows the best prediction performance in both the test set and the verification set. AUC, area under the ROC curve. KNN, K-nearest neighbor; SVM, support vector machine; ELM, extreme learning machine; RF, random forest; LR, logistic regression.

Page 10 of 11

Shen et al. Multiclassifier fusion predicts primary pulmonary solid nodules

Table 4 The t-test and F-test results of each classifier in the test set and verification set

Classifier	Test set		Verification set	
	t-test	F-test	t-test	F-test
KNN	0.041	0.038	0.041	0.038
SVM	0.040	0.039	0.040	0.039
ELM	0.037	0.039	0.037	0.039
RF	0.039	0.039	0.039	0.039
LR	0.037	0.038	0.037	0.038
Fusion classifier	0.035	0.036	0.035	0.036

The fusion classifier shows the best null hypothesis performance in the t-test and the F-test in both the test set and the verification set. KNN, K-nearest neighbor; SVM, support vector machine; ELM, extreme learning machine; RF, random forest; LR, logistic regression.

To determine whether the prediction performance of the classifier on the experimental samples could be generalized to the whole population, this study used the t-test and F-test to statistically analyze the mean and variance of the prediction performance indicators. The results showed that the performance of this fusion classifier had a probability of 0.965 to represent the mean of the entire population and a probability of 0.964 to represent the variance of the entire population, indicating that this fusion algorithm had strong generalization. Moreover, this fusion classifier integrated the excellent performance of individual classifiers. When the optimal hyperparameters of the classifier and the data set distribution are unknown, a fusion classifier can further simplify the parameter adjustment process.

This study had certain limitations: (I) it was a retrospective study and thus had certain biases. (II) All radiomics features were extracted from manually segmented images. It was difficult to exclude small blood vessels and small bronchi in or around the nodules, which might have affected the precision of the features. (III) The prediction precision of radiomics is affected by the choice of the classifiers, and most of the parameter optimization of the classifiers was done based on experience or experimental adjustment, without the theoretical support of parameter adjustment optimization, which may not guarantee that the parameters reach or approach the optimal performance.

In conclusion, the fusion classifier based on radiomics features can provide a noninvasive, fast, low-cost, and repeatable method to predict benign and malignant pulmonary solid nodules, which will be conducive to clinical treatment.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All experimental procedures were approved by the Medical Ethics Committee of Yinzhou Hospital affiliated with the School of Medicine of Ningbo University, Ningbo, China (ID: 2019-48).

References

1. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-6.
2. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30:1234-48.
3. Felix A, Oliveira M, Machado A, et al. Using 3D Texture and Margin Sharpness Features on Classification of Small Pulmonary Nodules. 29th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE, 2016.
4. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
5. Shen C, Liu Z, Guan M, et al. 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. Transl Oncol 2017;10:886-94.

6. Fan L, Fang M, Dong D, et al. Subtype discrimination of lung adenocarcinoma manifesting as ground glass nodule based on radiomics. Chin J Radiol 2017;51:912-7.

7. Zhou T, Lu H, Zhang J, et al. Pulmonary Nodule Detection Model Based on SVM and CT Image Feature-Level Fusion with Rough Sets. Biomed Res Int 2016;2016:8052436.

8. Li XX, Li B, Tian LF, et al. Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. IET Image Processing 2018;12:1253-64.

9. Peng J, Qi X, Zhang Q, et al. A radiomics nomogram for preoperatively predicting prognosis of patients in hepatocellular carcinoma. Transl Cancer Res 2018;7:936-46.

10. Yang M, Ren Y, She Y, et al. Imaging phenotype using radiomics to predict dry pleural dissemination in non-small cell lung cancer. Ann Transl Med 2019;7:259.

11. Xu X, Huang L, Chen J, et al. Application of radiomics signature captured from pretreatment thoracic CT to predict brain metastases in stage III/IV ALK-positive non-small cell lung cancer patients. J Thorac Dis 2019;11:4516-28.

12. Kamiya A, Murayama S, Kamiya H, et al. Kurtosis and skewness assessments of solid lung nodule density histograms: differentiating malignant from benign nodules on CT. Jpn J Radiol 2014;32:14-21.

13. Petkovska I, Shah SK, McNitt-Gray MF, et al. Pulmonary nodule characterization: a comparison of conventional with quantitative and visual semi-quantitative analyses using contrast enhancement maps. Eur J Radiol 2006;59:244-52.

14. Chi S. Density histogram analysis of CT scan in the differential diagnosis of solid pulmonary nodule. Radiol Practice 2016;31:866-9.

15. Andersen MB, Harders SW, Ganeshan B, et al. CT texture analysis can help differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer. Acta Radiol 2016;57:669-76.

16. Ng F, Ganeshan B, Kozarski R, et al. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. Radiology 2013;266:177-84.

17. Bayanati H, E Thornhill R, Souza CA, et al. Quantitative CT texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? Eur Radiol 2015;25:480-7.

18. Chi S. Differentiation between malignant and benign pulmonary nodules by CT-based texture analysis. J Pract Radiol 2016;32:1789-92.

19. Ganeshan B, Goh V, Mandeville HC, et al. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. Radiology 2013;266:326-36.

20. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002;24:971-87.

21. Tu SJ, Wang CW, Pan KT, et al. Localized thin-section CT with radiomics feature extraction and machine learning to classify early-detected pulmonary nodules from lung cancer screening. Phys Med Biol 2018;63:065005.

22. Zhang R, Shen J, Wei F, et al. Medical image classification based on multi-scale non-negative sparse coding. Artif Intell Med 2017;83:44-51.

23. Ding M, Antani S, Jaeger S, et al. Local-Global Classifier Fusion for Screening Chest Radiographs. SPIE Medical Imaging, 2017.

24. Shankar K, Elhoseny M, Lakshmanaprabu SK, et al. Optimal feature level fusion based ANFIS classifier for brain MRI image classification. Concurrency Computat Pract Exper 2018;32:e4887.