The construction of the model of pulmonary adenocarcinoma classification based on radiomics and random forest

He Ren[#a], Zhengguang Xiao[#b], Yijie Li[#e], Mengting Tu[c], Qi Sun[a], Ping Lu[d], Yejun Cao[d], Hanqing Chen[d], Ping Li*[a]

[a]Faculty of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, Shanghai, China 201318; [b]Department of Radiology, Shanghai Tongren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China 200050; [c]School of Materials Science and Engineering, Shanghai DianJi University, Shanghai, China 201306; [d]Tangqiao Community Health Service Center, No.131 Pujian Road, Pudong New District, Shanghai, China 200127; [e]College of Information Technology Shanghai Ocean University, Shanghai, China 201306

[#]Equally contributed author

*Corresponding author: lip@sumhs.edu.cn

ABSTRACT

Purpose: The present study aimed to construct classification models for pulmonary adenocarcinoma using computed tomography (CT)-based radiomics features and random forest method.

Methods: A total of 289 patients with 295 lung adenocarcinomas were included in this study. A total of 1066 CT images were extracted. The final data set was randomized into the training set and validation set at the ratio of 80%:20%. A total of 1082 features were captured from a semi-automatic segmentation method segmented lesion of a CT image. 9 optimal radiomic features obtained from root mean squared error (REMS) through cross validation and 14 radiographic characteristic features were selected to construct a random forest classification model. At the same time, compared with the results of the Support Vector Machine (SVM), Logistic Regression and C5.0 algorithm.

Results: The area under the curve (AUC) scores of training feature set, radiographic characteristics feature set, and the optimal radiomic feature set for testing dataset were 0.974, 0.483, and 0.835, respectively, and the corresponding AUC values for validation dataset were 0.964, 0.915, and 0.841, separately.

Conclusion: The developed random forest-based classification models using radiomics features and radiographic features of CT showed a relatively acceptable performance in lung adenocarcinoma and could assist clinical rapid diagnosis and triage.

Keywords: Radiomics feature, pulmonary adenocarcinoma, computed tomography, random forest

1. INTRODUCTION

Lung cancer, particularly non-small cell lung cancer (NSCLC), remains a major global health challenge due to its high incidence and mortality rates [1, 2]. NSCLC includes subtypes such as squamous carcinoma, pulmonary adenocarcinoma, and large cell carcinoma, accounting for approximately 85% of all lung cancer cases [3-5]. Among these, pulmonary adenocarcinoma can be further classified into adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IAC), each requiring different treatment approaches. Computed tomography (CT) imaging is a primary diagnostic tool for lung adenocarcinoma. However, distinguishing between the subtypes of pulmonary mini-nodules can be challenging due to limited resolution and the sheer volume of images needing interpretation [6]. Therefore, a non-invasive, repeatable, and quantitative method for the classification of these subtypes would be highly beneficial in clinical practice.

Despite the life-saving benefits of early CT detection, this imaging modality has several limitations, including high rates of detection of indeterminate pulmonary nodules [7-10]. Radiomics, which involves the extraction of quantitative features from medical images, offers a promising approach to this challenge [10-14]. This approach allows doctors to gain a more comprehensive view of lesions and enables quantitative, high-throughput analyses of lesion development

International Conference on Future of Medicine and Biological Information Engineering (MBIE 2024), edited by Yudong Yao, Xiaoou Li, Xia Yu, Proc. of SPIE Vol. 13270, 132700C · © 2024 SPIE · 0277-786X · doi: 10.1117/12.3048034

and progression [15-19] which can reflect both biological and medical image information of lesions, providing valuable insights for diagnosis, prognosis, and disease prediction[20-25].

This study aims to develop a model for the classification of pulmonary adenocarcinoma subtypes using radiomics and a random forest algorithm. The integration of advanced feature selection techniques with a robust classification model is anticipated to significantly improve diagnostic accuracy

2. MATERIALS AND METHODS

The study was approved by the Ethics Board of Shanghai University of Medicine & Health Sciences. The Ethics Committee gave up the written informed consent of the patient because it was a retrospective experiment and did not involve patient privacy.

2.1 Data collection and preprocessing

A total of 295 lesions were detected in 289 patients at the Shanghai Public Health Clinical Center (hospital 1). Additionally, in a recent examination, 127 patients with 138 lesions were re-selected at the same hospital (hospital 1.1). The basic patient information is detailed in Table 1. The selection criteria included: (1) patients who had undergone thin-slice chest CT imaging, (2) patients who had undergone an unenhanced chest CT exam, and (3) patients diagnosed with lung adenocarcinoma by doctors. Among these patients, one pGGN was detected in 187 patients and two in 6 patients. All lesions were pathologically diagnosed as 174 MIA, 54 AIS, and 67 IAC.

For the study, 236 lesions were allocated to the training set, while 59 were included in the testing set. Two to eight images with clear features were selected from the CT scans of each lesion, resulting in 853 CT images in the training set and 213 in the testing set. Additionally, 138 lesions diagnosed as 75 MIA, 36 AIS, and 27 IAC from two other hospitals were included. One to three images with clear features were selected from the CT scans of each lesion in this new dataset, yielding 267 CT images in the validation set. The patient selection flowchart is depicted in Figure 1.

Table 1. Demographic characteristics

	Full Dataset	Training & Testing Dataset	Validation Dataset	p value
Age	52.26±1.54	52.46±0.64	51.78±3.02	*
Long diameter	9.69±0.15	9.73±0.18	9.60±0.25	*
Short diameter	7.62±0.12	7.65±0.15	7.56±0.21	*
Sex	176M,240F	113M,176F	63M 64F	

Ages and size are shown as mean ± standard deviation;
M male, F female
*p value < 0.05

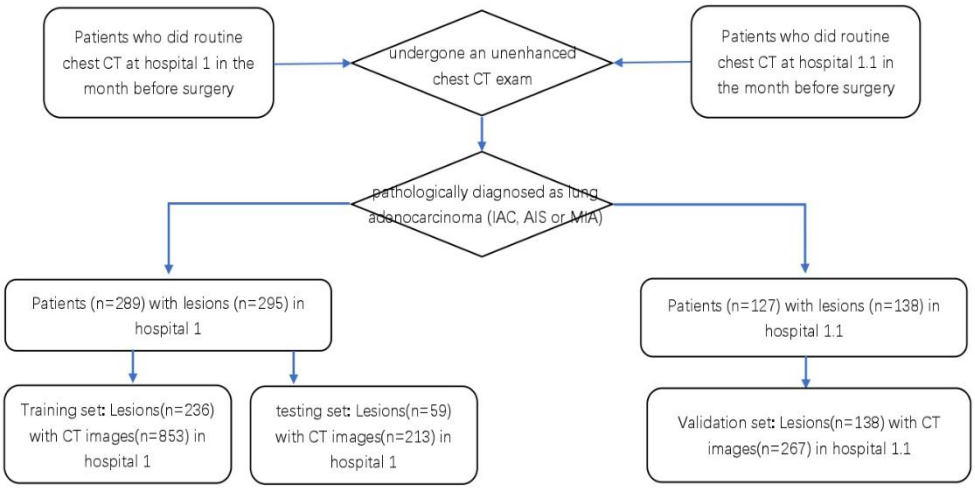

Figure 1. Flow chart of case sample selection

2.2 CT scanning

An unenhanced chest CT exam was performed during the entire lung scan of each patient with a United-Imaging 760 CT device (42-126 mA 120 kV, slice thickness of 1 mm), a Siemens Emotion 16 CT device (34-123 mA, 130 kV, 1 mm) and a Siemens Perspective 77427 CT device (155mA, 110kV, 1mm), under a 512×512 resolution.

2.3 Feature extraction

A semi-automated, unsupervised method was employed to segment the region of interest (ROI) for each CT image[26]. This segmentation was subsequently confirmed by a radiologist with a decade of experience in chest CT interpretation.

Radiographic characteristics were independently evaluated by two experienced thoracic radiologists, with 6 and 13 years of experience in chest CT interpretation, respectively, who were blinded to the pathological results. Discrepancies between the observers were resolved through consensus. The radiographic characteristics analyzed for each lesion included: (1) margin (clear, blurred), (2) lobulation (absent, present), (3) spiculation (absent, present), (4) pleural attachment, including pleural tag and indentation (absent, present), (5) air bronchogram (absent, present), (6) vessel change (absent, present), and (7) bubble lucency (absent, present) (Table 2).

A total of 1,068 features were extracted from the ROI using commercial software (PyRadiomics 3.0.1)[27]. These features included tumor size, shape, first-order statistical descriptors (histogram features), and high-order texture features (gray level co-occurrence matrix and gray level run length). The original images were normalized prior to feature extraction.

Table 2. The Parameters of lesions in datasets

		Training dataset		Testing dataset		Validation dataset	
Category	**MIA**	473		127		150	
	AIS	130		40		49	
	IAC	250		46		68	
Lobe	**Left**	345		85		115	
	Right	508		128		152	
Segment	**Left**	**S1+2** 124	**S6** 48	**S1+2** 20	**S6** 17	**S1+2** 42	**S6** 14
		S3 73	**S8** 13	**S3** 22	**S8** 3	**S3** 22	**S8** 3
		S4 20	**S9** 29	**S4** 3	**S9** 4	**S4** 4	**S9** 14
		S5 15	**S10** 23	**S5** 4	**S10** 12	**S5** 5	**S10** 11
	Right	**S1** 98	**S6** 58	**S1** 24	**S6** 16	**S1** 22	**S6** 21
		S2 121	**S7** 10	**S2** 34	**S7** 1	**S2** 38	**S7** 5
		S3 88	**S8** 54	**S3** 16	**S8** 9	**S3** 22	**S8** 14
		S4 28	**S9** 35	**S4** 13	**S9** 8	**S4** 17	**S9** 7
		S5 10	**S10** 6	**S5** 4	**S10** 3	**S5** 6	**S10** 0
Margin	**Blurred**	786		195		243	
	Clear	67		18		24	
Lobulation	**Present**	651		162		192	
	Absent	202		51		75	
Spiculation	**Present**	637		156		191	
	Absent	216		57		76	
Pleural Attachment	**Present**	236		50		61	
	Absent	617		163		206	
Air Bronchogram	**Present**	427		114		134	
	Absent	426		99		133	
Vessel Change	**Present**	706		187		227	
	Absent	147		26		40	
Bubble Lucency	**Present**	190		49		48	
	Absent	663		164		219	

Right lung, superior lobe	Right lung, inferior lobe	
Apical segment [S1]	Superior segment (Fowler) [S6]	Left lung, superior lobe
Posterior segment [S2]	Medial basal segment [S7]	Apicoposterior segment [S1+2]
Anterior segment [S3]	Anterior basal segment [S8]	Anterior segment [S3]
Right lung, middle lobe	Lateral basal segment [S9]	Superior lingular segment [S4]
Lateral segment [S4]	Posterior basal segment [S10]	Inferior lingular segment [S5]
Medial segment [S5]		

2.4 Feature selection

Given that the initial number of features extracted was close to the number of patients, a systematic feature selection process was implemented to avoid overfitting. The feature set was divided into radiographic characteristics and radiomic features. All radiographic characteristic features were preserved due to their clinical relevance. For the radiomic features, an initial screening was conducted where features with low variance were removed as they provide minimal information and can introduce noise. Furthermore, highly correlated features (correlation coefficient > 0.9) were identified, and one feature from each pair was removed to reduce multicollinearity. Feature importance was then assessed using the

'RandomForest' [28] package to compute the 'MeanDecreaseGini' metric, ranking features in decreasing order of their importance scores. To determine the optimal subset of radiomic features, cross-validation was employed, calculating the root mean squared error (RMSE) for various feature subsets and identifying the subset that minimized RMSE. This iterative process ensured the retention of the most predictive features. Additionally, the correlation of each feature in the optimal radiomic feature set was analyzed to ensure that the selected features were not redundant and provided unique information. The final selected features included original_shape2D_Perimeter, squareroot_firstorder_90Percentile, original_shape_Maximum2DDiameterRow, wavelet.LL_glcm_ClusterShade, wavelet.HL_glszm_SmallAreaLowGrayLevelEmphasis, original_shape2D_Sphericity, diagnostics_Image.original_Mean, wavelet.LL_firstorder_Minimum, and logarithm_firstorder_90Percentile.

The optimal radiomic feature subset was then combined with the radiographic characteristics feature set to form a comprehensive training feature set, which was used for model training. This thorough feature selection process ensured that the final model included only the most relevant and non-redundant features, thereby enhancing its predictive performance and robustness.

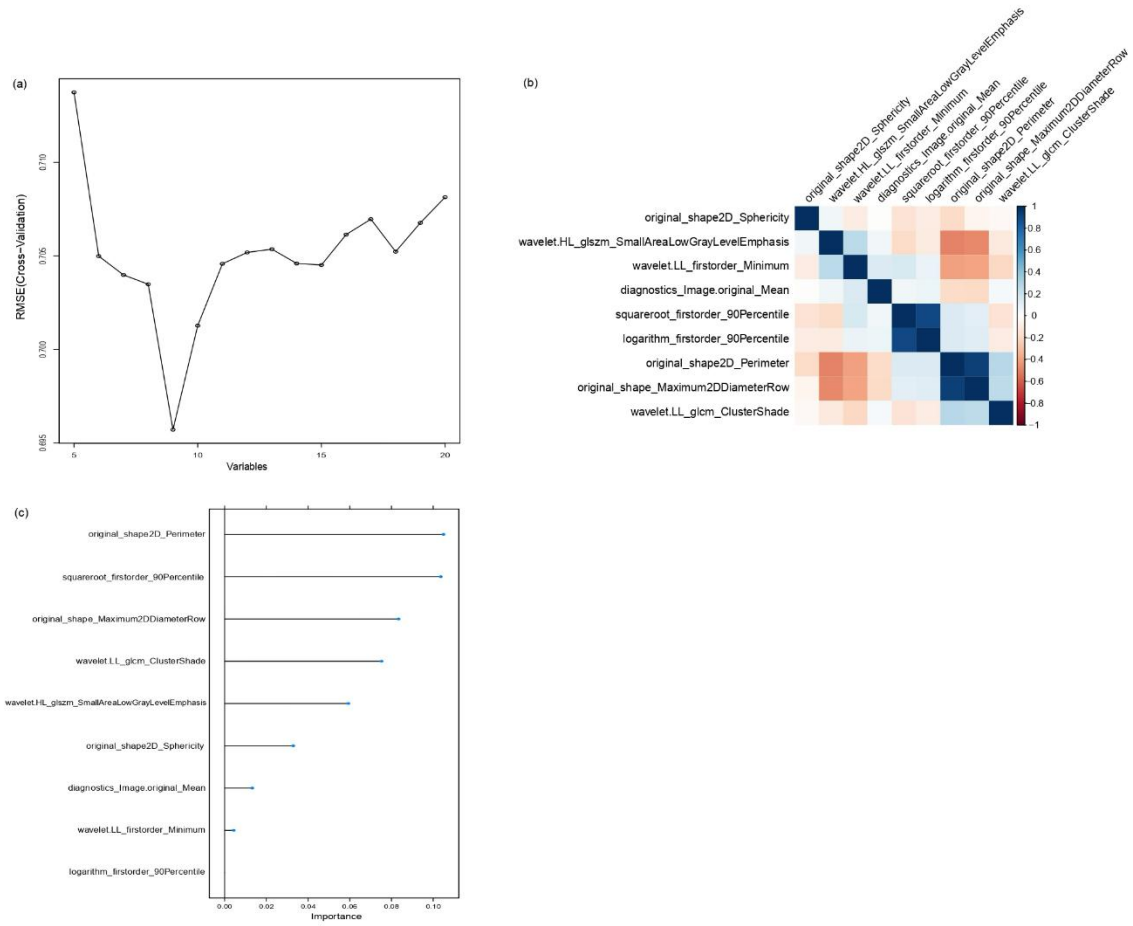

Figure 2. The optimal radiomic feature subset and its performance are obtained through cross validation. (a) represents the performance of cross validation. (b) represents the correlation diagram of 9 features. (c) represents the important value of 9 features.

2.5 Model training and validation

Based on the preliminary training of all features, the optimal values of parameters 'mtry' and 'ntree' required by the random forest model were calculated (mtry=20, ntree=900, Figure 3). Then, on the premise of the same parameters, the random forest model was trained with those three feature sets respectively (training feature set as model 1, radiographic

characteristics feature set as model 2 and the optimal radiomic feature set as model 3). At the same time, SVM, Logistic Regression and C5.0 algorithm were used to verify the results.

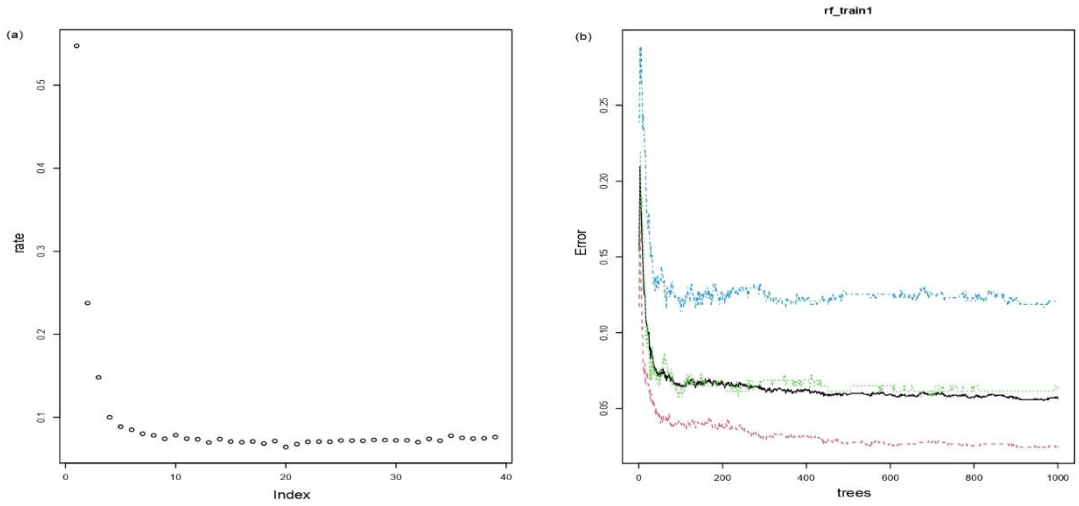

Figure 3. Optimal parameters of random forest model. (a) represents the performance of the 'mtry'. (b) represents the performance of the 'ntree'.

2.6 Statistical analysis

Quantitative data were presented as mean ± SD or median (25th−75th percentile), while qualitative data were described as counts (n). Comparisons between groups for qualitative variables were performed using Fisher's exact test, and comparisons for quantitative variables were performed using either a t-test or the Wilcoxon test. A p-value of less than 0.05 was considered statistically significant.

The Area Under the Curve (AUC) represents the degree of separability and serves as a summary measure of the accuracy of a quantitative diagnostic test. The Receiver Operating Characteristic (ROC) curve, a commonly used graph to summarize the performance of a classifier across all possible thresholds, is generated by plotting the true positive rate (y-axis) against the false positive rate (x-axis). The ROC curve and AUC were used to evaluate the diagnostic performance, precision, and discrimination accuracy of the models.

All statistical analyses were performed using R statistical software (http://www.Rproject.org, version 3.6.0) with the following packages: 'randomForest', 'ggplot2', 'caret', 'pROC', 'e1071', and 'reshape2'.

2.7 Results

The performance of the radiomics model constructed using random forest was evaluated by precisions and accuracies. The ROC curves of random forest models based on three feature set are shown in Figure 4. Because there are three types of lesions, the subsequent evaluation is to combine two of them into one and compare them with the remaining one (MIA represents the comparison between MIA and IAC & AIS, AIS represents the comparison between AIS and IAC &MIA, IAC represents the comparison between IAC and MIA & AIS).

For the testing dataset, the AUC scores of model 1, model 2, and model 3 were 0.974, 0.483, and 0.835, respectively. The sensitivity values of model 1, model 2, and model 3 at the optimum critical point were 0.873(MIA:0.967, AIS:0.825, IAC:0.826), 0.814(MIA:0.866, AIS:0.750, IAC:0.826), and 0.641(MIA:0.902, AIS:0.378, IAC:0.642). The specificity values of model 1, model 2, and model 3 at the optimum critical point were 0.936(MIA:0.837, AIS:0.994, IAC:0.976), 0.897(MIA:0.791, AIS:0.965, IAC:0.934), and 0.829(MIA:0.589, AIS:0.972, IAC:0.925).

For validation dataset, the AUC scores of model 1, model 2, and model 3 were 0.964, 0.915, and 0.841, respectively. The sensitivity values of model 1, model 2, and model 3 at the optimum critical point were 0.867(MIA:0.940, AIS:0.837, IAC:0.824), 0.825(MIA:0.880, AIS:0.756, IAC:0.838), and 0.648(MIA:0.927, AIS:0.245, IAC:0.772). The specificity values of model 1, model 2, and model 3 at the optimum critical point were 0.932(MIA:0.855, AIS:0.986, IAC:0.955),

0.906(MIA:0.803, AIS:0.949, IAC:0.965), and 0.798(MIA:0.521, AIS:0.986, IAC:0.886). The performance of the random forest models based on the three feature sets are shown in the Table 3.

At the same time, SVM, Logistic Regression and C5.0 algorithm used the same training dataset and training feature set to train the model. Evaluate the performance of the model on the testing dataset. For SVM model, the AUC score was 0.734. The sensitivity values were MIA:0.921, AIS:0.750, IAC:0.348. The specificity values were MIA:0.547, AIS:0.959, IAC:0.976. For Logistic Regression model, the AUC score was 0.780. The sensitivity values were MIA:0.850, AIS:0.750, IAC:0.587. The specificity values were MIA:0.709, AIS:0.971, IAC:0.892. For C5.0 model, the AUC score was 0.700. The sensitivity values were MIA:0.843, AIS:0.475, IAC:0.522. The specificity values were MIA:0.558, AIS:0.965, IAC:0.886. The details are shown in the table 4.

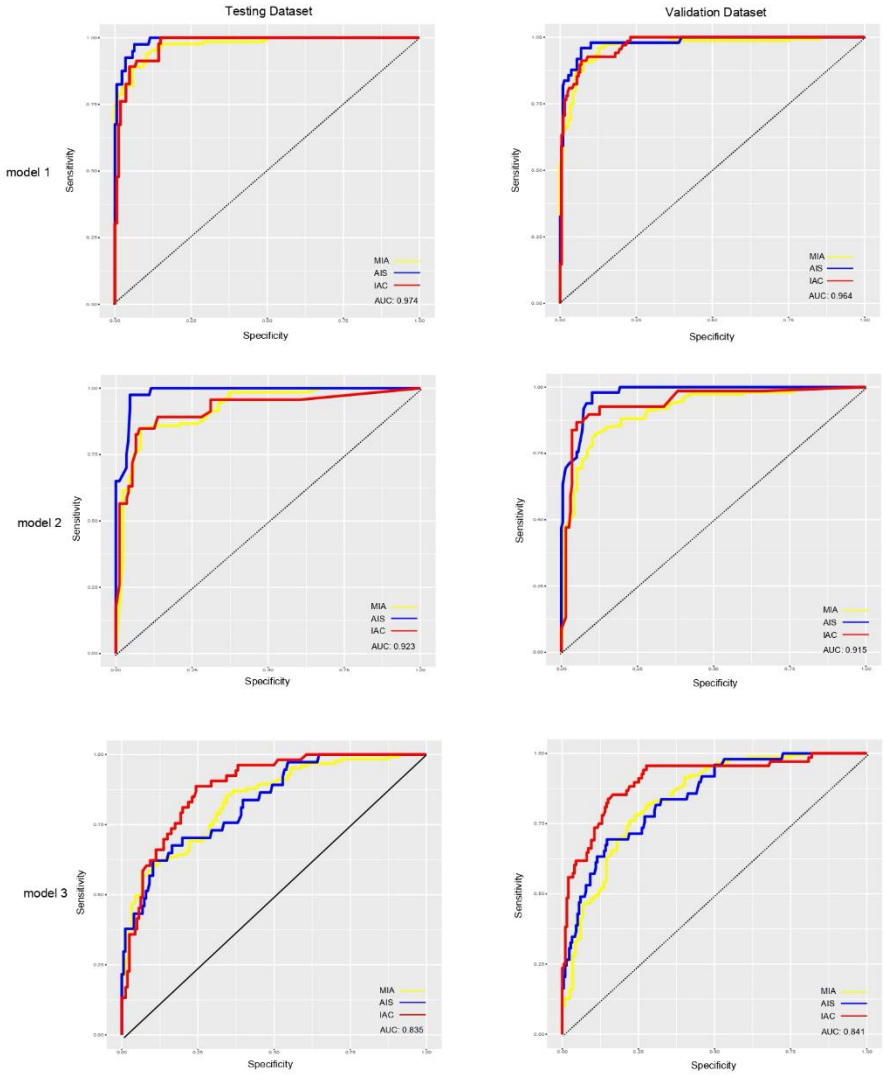

Figure 4. The performance of random forest model based on different feature sets on testing dataset and verification dataset. The yellow line represents the performance of the model in MIA compared with AIS&IAC. The blue line represents the performance of the model in AIS compared with MIA&IAC. The red line represents the performance of the model in IAC compared with MIA&AIS.

Table 3. The performance of random forest models in testing and validation datasets

		sensitivity		precision		specificity		f1	
		testing	validation	testing	validation	testing	validation	testing	validation
Model1	MIA	0.967	0.94	0.898	0.892	0.837	0.855	0.932	0.916
	AIS	0.825	0.837	0.971	0.932	0.994	0.986	0.891	0.882
	IAC	0.826	0.824	0.905	0.862	0.976	0.955	0.864	0.842
Model2	MIA	0.866	0.880	0.859	0.852	0.791	0.803	0.863	0.866
	AIS	0.750	0.756	0.833	0.771	0.965	0.949	0.789	0.763
	IAC	0.826	0.838	0.776	0.891	0.934	0.965	0.800	0.864
Model3	MIA	0.902	0.927	0.750	0.713	0.589	0.521	0.819	0.806
	AIS	0.378	0.245	0.737	0.800	0.972	0.986	0.500	0.375
	IAC	0.642	0.772	0.739	0.647	0.925	0.886	0.687	0.704

Table 4. The performance of validation models in testing datasets

		sensitivity	precision	specificity	f1
SVM	MIA	0.921	0.75	0.547	0.827
	AIS	0.75	0.811	0.959	0.779
	IAC	0.348	0.8	0.976	0.485
LOGISTIC	MIA	0.85	0.812	0.709	0.831
	AIS	0.75	0.857	0.971	0.8
	IAC	0.587	0.6	0.892	0.593
C5.0	MIA	0.843	0.738	0.558	0.787
	AIS	0.475	0.76	0.965	0.585
	IAC	0.522	0.558	0.886	0.539
Model1	MIA	0.967	0.898	0.837	0.932
	AIS	0.825	0.971	0.994	0.891
	IAC	0.826	0.905	0.976	0.864

3. DISCUSSION

This study demonstrates the potential of using a random forest-based classification model incorporating radiomics and radiographic features for the differentiation of pulmonary adenocarcinoma subtypes. The key findings indicate that the combination of radiomic features with traditional radiographic characteristics significantly enhances the accuracy of lung adenocarcinoma classification compared to using either feature set alone.

The random forest model achieved high AUC scores for both training and validation datasets, highlighting its robustness and reliability. The model's ability to effectively integrate radiomic features, which capture complex imaging patterns not visible to the naked eye, with conventional radiographic features, underscores the importance of a multifaceted approach in medical imaging analysis. This combined feature set resulted in a more comprehensive representation of the lesions, improving the model's diagnostic performance.

Radiomics offers a transformative approach by converting medical images into high-dimensional data that reflect underlying tumor biology. The study's findings support the utility of radiomics in clinical settings, providing a non-invasive and reproducible method for lesion characterization. The enhanced ability to classify subtypes of pulmonary adenocarcinoma can lead to more personalized treatment strategies, potentially improving patient outcomes.

Despite the promising results, several limitations were noted. The study was conducted at a single center with a relatively small sample size, which may limit the generalizability of the findings. Potential biases in the dataset due to retrospective data collection and selection criteria may affect the model's performance. Additionally, the semi-automatic segmentation method, while effective, still involves manual steps that could introduce variability. The generalizability of the model to

different populations or imaging protocols remains uncertain due to the specific CT imaging protocols used in this study. Future research should aim to develop fully automated segmentation techniques to further standardize the analysis and include multi-center data to validate the model across diverse patient cohorts and imaging conditions.

Furthermore, the study's feature selection process, while rigorous, still resulted in a large number of features that require manual assessment by clinicians. Incorporating additional clinical data and expanding the radiomics feature set could enhance the model's accuracy and clinical applicability.

To address these limitations and improve the model's utility, future studies should focus on multi-center collaborations to validate the findings across diverse populations and imaging equipment. Increasing the dataset size and diversity would enhance the model's robustness and applicability in different clinical scenarios. Additionally, integrating clinical data such as patient history and genetic markers could provide a more holistic approach to lung adenocarcinoma classification. This study highlights the significant potential of radiomics and machine learning in enhancing lung cancer diagnostics. The developed model represents a step forward in the integration of advanced imaging techniques with artificial intelligence, paving the way for more precise and personalized cancer care. Further research and technological advancements will be crucial in fully realizing the benefits of this innovative approach in clinical practice.

4. CONCLUSIONS

In conclusion, we developed an automatic pulmonary adenocarcinoma grading model utilizing the most significant radiomics features through a random forest method. This model demonstrates potential in aiding rapid clinical diagnosis and triage. Future research should focus on extracting additional radiomics parameters and incorporating patient clinical characteristics to enhance model accuracy.

ETHICS DECLARATIONS

Ethics approval and consent to participate. This article does not contain any studies with human participants or animals performed by any of the authors. Ethical approval for this study was obtained from Shanghai Public Health Clinical Center.

REFERENCES

[1] McGuire, and S. "Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015." Advances in Nutrition An International Review Journal 7.2(2016):418-419.

[2] El Kinany K, Huybrechts I, Kampman E, et al. Concordance with the World Cancer Research Fund/American Institute for Cancer Research recommendations for cancer prevention and colorectal cancer risk in Morocco: A large, population-based case-control study. Int J Cancer. Oct 1 2019;145(7):1829-1837.

[3] Donahue JM, Morse CR, Wigle DA, et al. Oncologic efficacy of anatomic segmentectomy in stage IA lung cancer patients with T1a tumors. Ann Thorac Surg. Feb 2012;93(2):381-387; discussion 387-388.

[4] Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: international multidisciplinary classification of lung adenocarcinoma: executive summary. Proc Am Thorac Soc. Sep 2011;8(5):381-385.

[5] He J, Hu Y, Hu M, Li B. Development of PD-1/PD-L1 Pathway in Tumor Immune Microenvironment and Treatment for Non-Small Cell Lung Cancer. Sci Rep. Aug 17 2015; 5:13110.

[6] Sun Z, Ng K.H, & Sarji S.A. Is utilisation of computed tomography justified in clinical practice? Part IV: applications of paediatric computed tomography. Singapore medical journal,2010;51(6), 457-463.

[7] Alahmari SS, Cherezov D, Goldgof D, Hall L, Gillies RJ, Schabath MB. Delta Radiomics Improves Pulmonary Nodule Malignancy Prediction in Lung Cancer Screening. IEEE Access. 2018; 6:77796-77806.

[8] Choi W, Nadeem S, Alam SR, Deasy JO, Tannenbaum A, Lu W. Reproducible and Interpretable Spiculation Quantification for Lung Cancer Screening. Comput Methods Programs Biomed. Nov 13 2020:105839.

[9] Balagurunathan Y, Schabath MB, Wang H, Liu Y, Gillies RJ. Quantitative Imaging features Improve Discrimination of Malignancy in Pulmonary nodules. Sci Rep. Jun 12 2019;9(1):8528.

[10] Xu Y, Lu L, E LN, et al. Application of Radiomics in Predicting the Malignancy of Pulmonary Nodules in Different Sizes. AJR Am J Roentgenol. Dec 2019;213(6):1213-1220.
[11] Yang X, He J, Wang J, et al. CT-based radiomics signature for differentiating solitary granulomatous nodules from solid lung adenocarcinoma. Lung Cancer. Nov 2018; 125:109-114.
[12] Ortiz-Ramon R, Larroza A, Arana E, Moratal D. A radiomics evaluation of 2D and 3D MRI texture features to classify brain metastases from lung cancer and melanoma. Annu Int Conf IEEE Eng Med Biol Soc. Jul 2017; 2017:493-496.
[13] Kalpathy-Cramer J, Mamomov A, Zhao B, et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. Tomography. Dec 2016;2(4):430-437.
[14] Zhao W, Zhang W, Sun Y, et al. Convolution kernel and iterative reconstruction affect the diagnostic performance of radiomics and deep learning in lung adenocarcinoma pathological subtypes. Thorac Cancer. Oct 2019;10(10):1893-1903.
[15] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. Jun 3 2014; 5:4006.
[16] Constanzo J, Wei L, Tseng HH, El Naqa I. Radiomics in precision medicine for lung cancer. Transl Lung Cancer Res. Dec 2017;6(6):635-647.
[17] Kalra M, Wang G, Orton CG. Radiomics in lung cancer: Its time is here. Med Phys. Mar 2018;45(3):997-1000.
[18] 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. Dec 2017;10(6):886-894.
[19] Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res. Feb 2017;6(1):86-91.
[20] Gillies RJK, Paul E, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-77.
[21] Nardone V, Boldrini L, Grassi R, Franceschini D, & Desideri I. Radiomics in the setting of neoadjuvant radiotherapy: a new approach for tailored treatment. Cancers.2021;13(14), 3590.
[22] Pinto Dos Santos D, Dietzel M, Baessler B. A decade of radiomics research: are images really data or just patterns in the noise? Eur Radiol. Jan 2021;31(1):1-4.
[23] Bae JM, Jeong JY, Lee HY, et al. Pathologic stratification of operable lung adenocarcinoma using radiomics features extracted from dual energy CT images. Oncotarget. Jan 3 2017;8(1):523-535.
[24] Thawani R, McLane M, Beig N, et al. Radiomics and radiogenomics in lung cancer: A review for the clinician. Lung Cancer. Jan 2018; 115:34-41.
[25] Yang F, Simpson G, Young L, Ford J, Dogan N, Wang L. Impact of contouring variability on oncological PET radiomics features in the lung. Sci Rep. Jan 15 2020;10(1):369.
[26] Ren H, Zhou L, Liu G, et al. An unsupervised semi-automated pulmonary nodule segmentation method based on enhanced region growing. Quant Imaging Med Surg. Jan 2020;10(1):233-242.
[27] van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. Nov 1 2017;77(21): e104-e107.
[28] Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. Nov-Dec 2003;43(6):1947-1958.