Classification of renal tumor histology subtypes based on radiomics features of CT images

Yang Yi[1,2], Qian Xusheng[2], Zhou Zhiyong[2], Shen Junkang[3], Zhu Jianbing[4,5*], Dai Yakang[2]

1 University of Science and Technology of China, Hefei230027;

2 Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163;

3 Second Affiliated Hospital of Suzhou University, Suzhou 215163;

4 Suzhou Hospital Affiliated to Nanjing Medical University, Suzhou 215163;

5 Suzhou Science & Technology Town Hospital, Suzhou 215163 Corresponding authors:DAI Yakang(E-mail:daiyk@sibet.ac.cn) ;

ZHU Jianbin(E-mail:zeno1839@.126.com)

Abstract:

Objective Accurate preoperative differential diagnosis of fat-poor angiomyolipoma (fp-AML) and clear cell renal cell carcinoma (ccrcc) is essential for proper treatment planning. In order to increase the accuracy of discrimination of fp-AML from ccrcc, we develop a classification model based on radiomics technology. Methods The study retrospectively collected CT images of 18 cases with fp-AML and 42 cases with ccrcc from department of radiology, the Second Affiliated Hospital of Suzhou University. Firstly, 430 radiomics features were extracted from CT images. Then, the feature selection was carried by three steps: Pearson's correlation matrices were calculated to remove redundant features, Welch's t-test was utilized to determine the statistically significant features, and sequential forward floating selection method was used to select the discriminative features. Finally, k-nearest neighborhood, random forest, support vector machine and adaboost classifiers were built for classification. **Results** The model built by SVM classifier achieved the best classification performance, with accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curves of 91.67%, 88.89%, 92.86%, 84.21%, 95.12%, and 0.9418. **Conclusions** The proposed model can increase the classification accuracy of discrimination of fp-AML from ccrcc, and has great potential in helping radiologists to discriminate fp-AML from ccrcc.

Keywords: Fat-poor angiomyolipoma; Clear cell renal cell carcinoma; Radiomics; Machine learning

DOI: 10 3969/j.issn. 1002-3208 2020 01 003.

CLC classification No.: R318 04 Document mark code A

Article No.: 1002-3208 (2020) 01-0015-06

Fund Project: Funded by National Natural Science Foundation of China (61501452, 61801476), national key R & D plan (2017YFB1103602, 2017YFC0114304, 2018YFC0116904), Suzhou Minsheng science and technology project (SYS2018010), Suzhou high tech Zone health and Family Planning Bureau (2017Z005)

Author unit: 1 University of science and technology of China (Hefei 230027)

2 Suzhou Institute of biomedical engineering technology, Chinese Academy of Sciences (Suzhou215163)

3 the Second Affiliated Hospital of Suzhou University (Suzhou 215163)

4 Suzhou Hospital Affiliated to Nanjing Medical University (Suzhou 215163)

5 Suzhou science and Technology City Hospital (Suzhou 215163)

Corresponding author:

Daiyakang, researcher, e - mail:daiyk@ sibet AC CN;

Zhujianbing, chief physician, e - mail:zeno1839@ 126 com

Description format of this article Yang Yi, qianxusheng, zhouzhiyong, et al Classification of histological subtypes of renal tumors based on CT imaging features [j] Beijing Biomedical Engineering, 2020, 39 (1): 15-20 YANG Yi , QIAN Xusheng, ZHOU Zhiyong, et al. Classification of renal tumor histology subtypes based on radiomics features of CT images [J] . Beijing Biomedical Engineering, 2020, 39(1) :15-20.

0 Introduction

According to statistics, more than 90% of renal cell carcinomas are renal cell carcinoma (RCC), of which clear cell renal cell carcinoma (ccrcc) accounts for about 70% of RCC. It is the most common and fatal subtype of RCC. The treatment scheme is mainly surgical resection [1-2]. Angiomyolipoma (AML) is the most common benign renal tumor, which is mainly diagnosed by regional fat in CT images. However, fat - poor angiomyolipoma (FP - AML) does not contain or only a small amount of scattered adipose tissue, and its imaging findings are very similar to ccrcc, so it is very easy to be misdiagnosed [3]. Most amls only need follow-up. However, the statistical results show that up to 65% of patients receiving surgical treatment are finally diagnosed as FP - AML by pathology [4]. Therefore, it is of great clinical significance to accurately identify FP - AML and ccrcc before operation.

At present, the classification research of FP - AML and ccrcc mainly combines texture analysis technology with machine learning to build classification models [5-8]. For example, Feng et al. [7] extracted 14 texture features from CT images of each patient, used Mann - Whitney U test and support vector machine recursive feature elimination (SVM - RFE) to select features, and established a support vector machine (SVM) classifier for classification. After extracting 64 texture features, Lee et al. [8] combined three feature selection algorithms and four classifiers to explore the combination that can obtain the best classification performance. It can be seen that the number of features extracted by texture analysis technology is extremely limited. However, there is a risk of losing useful tumor information by extracting only a small number of features, which is difficult to describe the characteristics of tumors comprehensively. Therefore, more effective methods are needed to promote the classification research of FP - AML and ccrcc. Imageomics technology is a new method in the field of medical image analysis. Its essence is to extract quantitative features from medical images to describe the characteristics of tumors in order to improve the accuracy of diagnosis [9]. Previous studies have confirmed that the imaging omics model constructed by the combination of imaging omics technology and machine learning can greatly improve the accuracy of non-invasive diagnosis and has guiding significance for clinical decision-making [10-13]. However, it has not been applied to the classification of FP - AML and ccrcc.

In addition, the initial feature set extracted from the image often has many irrelevant features, weak correlation features and redundant features, which brings unnecessary computational overhead and affects the performance of the classifier [14]. However, most of the existing classification studies of FP - AML and ccrcc only use one feature selection algorithm for feature selection, which can not effectively complete feature selection [5-6, 8]. Feng et al. [7] carried out feature selection in two steps, and proved through experiments that

further feature selection can improve the accuracy of classification.

Therefore, in order to improve the classification accuracy of FP - AML and ccrcc, this paper will use the image omics technology to extract the features of CT images with high throughput, then select the features in three steps, and finally use four classifiers to build a model for classification.

1 Research object

The experimental data set in this paper is the CT images of 18 patients with FP - AML and 42 patients with ccrcc provided by the Second Affiliated Hospital of Suzhou University. The diagnostic results of all patients were based on biopsy pathological data, and the interval between CT image acquisition time and pathological diagnosis time was not more than 4 weeks. The imaging equipment used is Ge multi row spiral CT in the United States. The scanning parameters are: tube voltage 120 kv, automatic milliampere technology, section thickness 5 mm, reconstruction slice thickness 1 mm. All the included cases are single, and the basic information of the patients is shown in Table 1.

Table 1 Summary of patient characteristics

Clinical features	Fp-AML(n=18)	Ccrcc(n=42)
Number of cases (male/female)	7/11	26/16
Age/year	47.6±141	57.6±135
Tumor diameter/cm	1.49~930	1.50~925

Before feature extraction, a radiologist with 25 years of clinical experience outlined the region of interest (ROI). After selecting the largest section of the tumor, the doctor uses a rectangular box to extract ROI according to the size of the tumor according to the following rules: a. Extract ROI in the tumor area; b. Select the most obvious and uniform area; c. Avoid areas of necrosis, cyst, hemorrhage and calcification. In order to reduce the computational complexity of the algorithm, the maximum size of the rectangular box is set to 23 × 23 pixels. An example of ROI is shown in Figure 1.

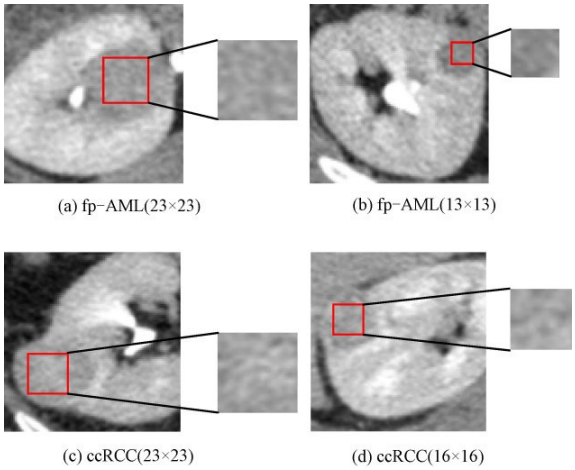


Figure 1 Examples of ROI

2 Research methods

2.1 Extraction of image omics features

In this paper, we use image omics to extract 430 features from each ROI, including 18 gray features, 68 texture features and 344 wavelet features. The calculation method of all features is consistent with the definition of features in the imaging biomarker Standardization Initiative (IBSI). For detailed feature list and calculation formula, please refer to ibsi[15].

The gray-scale features can describe the gray-scale value distribution of the image without considering the spatial relationship [16]. The 18 gray-scale features extracted in this paper include energy, entropy, skewness and kurtosis [15].

Texture features provide objective and quantitative evaluation of tumor heterogeneity by analyzing the distribution and relationship of pixel gray values in the image [16]. In this paper, 22 gray level co-occurrence matrix features, 16 gray level run length matrix features, 16 gray level region size matrix features and 14 gray

level dependency matrix features are extracted [15].

The first-order discrete wavelet transform (DWT) can decompose the original ROI image into four sub images: ll, LH, HL and HH, as shown in Figure 2. The above gray and texture features are extracted from each sub image, and a total of 344 wavelet features are obtained. In this paper, DB4 wavelet is used for wavelet transform, because it has good performance in medical image classification [17].

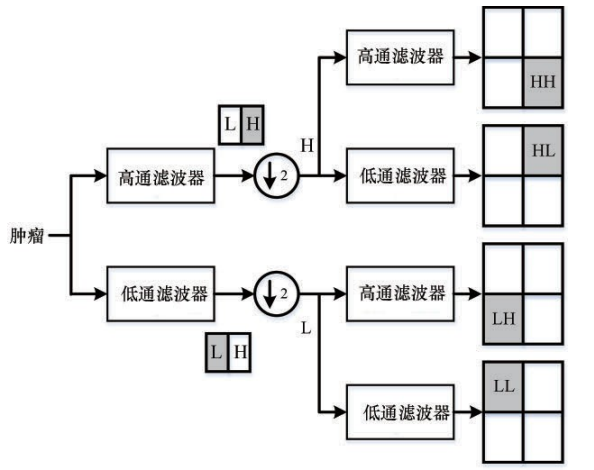


Figure 2 Illustration of the discrete wavelet transformation

2.2 Feature selection

Before feature selection, first preprocess the features and normalize them to [0,1] to avoid the features in the large value range over dominating the features in the small value range [11]. Then, feature selection is carried out in three steps. The specific process is as follows.

(1) The Pearson's correlation matrices (PCM) [18] are used to eliminate redundant features. In this paper, the threshold of the average absolute correlation coefficient is set to 0.6.

(2) Welch's t test [19] was used to determine the characteristics with statistically significant difference. In this paper, the threshold of P value was set to 0.05.

(3) The sequential forward floating selection (SFFS) algorithm [20] is used to select feature subsets with good identification ability, which can reduce the possibility of falling into local optimization and improve the accuracy of classification [21].

In this paper, the feature evaluation index in SFFS algorithm is set as the classification accuracy obtained by linear discriminant analysis [22], and the number of features selected by the algorithm is set as 10, 20, 30, 40, 50 and 60 to determine the most appropriate number of features.

2.3 Establish classification model

This paper uses K - nearest neighbor (KNN), random forest (RF), support vector machine (SVM) and adaboost classifiers to classify FP - AML and ccrcc. The kernel function of SVM classifier is radial basis function, and the basis classifier of adaboost classifier is decision tree.

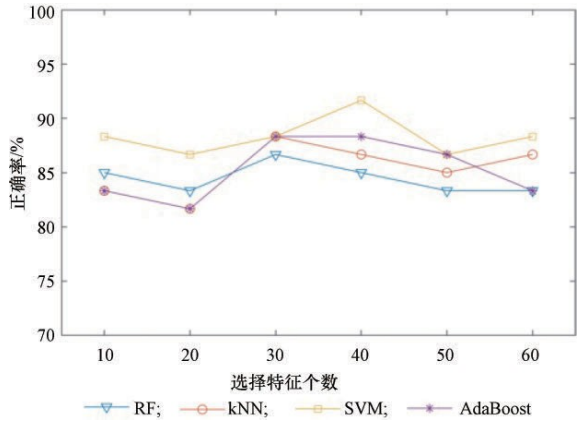
The left one method cross validation is used to evaluate the performance of the classifier. Because it is suitable for data detection of small samples, it can provide an almost unbiased estimate of the generalization ability of the classifier [23]. The calculated performance evaluation indicators include accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV) and area under the receiver operating characteristic curves (AUC).

3 Results and analysis

All the experiments in this paper are programmed in Spyder 328 development environment. Due to the use of "leave one method" cross validation, feature selection is carried out for each compromised training set

data. Therefore, this experiment will carry out 60 feature selection. After PCM selection, 297~333 features are retained; then the Welch's t test was used to select features, and 80~96 features were retained; Finally, SFFS algorithm is used to select 6 feature subsets with sizes of 10, 20, 30, 40, 50 and 60 from the remaining features of each fold. RF, KNN, SVM and adaboost classifiers are constructed using the selected feature subsets for classification.

Because there is a large difference in the number of samples between the two types in the experimental data set used in this paper (18 patients with FP - AML and 42 patients with ccrcc), which will affect the performance of the classifier, smote algorithm is used for data balancing. In each fold cross validation, smote algorithm is used for the training set samples, so that the ratio of the two types of training samples is 1:1, and then the classifier is trained. The accuracy obtained by each classifier is shown in Figure 3.



Accuracy

Number of selected features

Figure 3 Accuracy of four classifiers based on different numbers of selected features

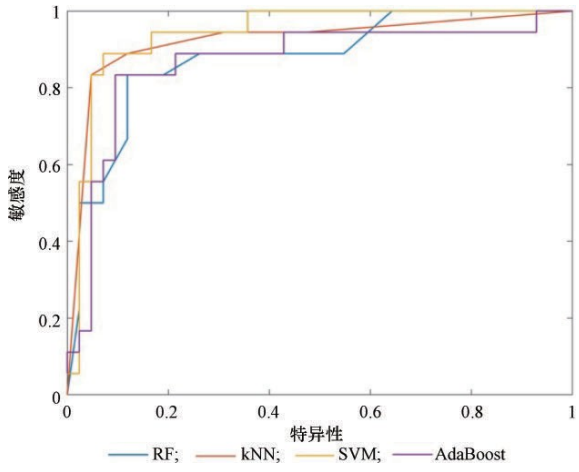
As can be seen from Figure 3, based on six feature subsets of different sizes, SVM classifiers have obtained the highest accuracy. RF, KNN, SVM and adaboost classifiers have the highest accuracy when using feature subsets with sizes of 30, 30, 40 and 30 respectively. The classification performance at this time is shown in Table 2, and the ROC curve is shown in Figure 4. The experimental results show that among RF, KNN, SVM and adaboost classifiers, the model constructed by SVM classifier achieves the best classification performance.

Table 2 Best performance comparison of four classifiers

Table 2 Comparison of the best performance of four classifiers

Project	RF	kNN	SVM	AdaBoost
ACC/%	86.67	88.33	91.67	88.33
SEN/%	83.33	88.89	88.89	83.33
SPE/%	88.10	88.10	92.86	90.48
PPV/%	75.00	76.19	84.21	78.95
NPV/%	92.50	94.87	95.12	92.68
AUC	0.8776	0.9226	0.9418	0.8677

The best classification performance obtained in this paper is compared with the results of existing studies, as shown in Table 3. The initial feature set and the selected feature set are the size of the initial feature set extracted from each ROI and the size of the feature subset selected in the feature selection link, respectively. It can be seen that the number of features extracted in this paper is much larger than the other two methods, which realizes a more comprehensive description of tumor features. Although Feng et al. [7] obtained the best classification performance, their experiments first



Sensitivity
Specificity

Figure 4 ROC curves of four classifiers

After feature selection of all samples, cross validation is used to evaluate the classification performance, which will make the evaluation result too optimistic.

Table 3 Comparison of the performance of this study and previous studies

Performance comparison	Lee et al. [8]	Feng et al. [7]	This paper
FP - number of AML cases	25	17	18
Number of ccrcc cases	25	41	42
Initial feature set	64	14	430
Select feature set	20	10	30
Feature selection algorithm	Relieff	Mann - Whitney U test + SVM - RFE	PCM + Welch's t-test + SFFS
Classifier type	Knn	SVM	SVM
Accuracy	72 3%	93 9%	91 67%
Sensitivity	73 0%	87 8%	88 89%
Specificity	71 7%	100%	92 86%
PPV	72 3%	—	84 21%
NPV	72 8%	—	95 12%
AUC	0 782	0 955	0 9418

The experiment of Lee et al. [8] is the same as the experiment method in this paper. In each fold cross validation, only after the feature selection of the training set, the training classifier tests the test set, which solves the problems existing in Feng et al. Compared with the classification results of Lee et al., the model constructed in this paper significantly improves the classification accuracy of FP - AML and ccrcc.

4 Discussion and conclusion

Compared with most classification studies of FP - AML and ccrcc, which use texture analysis technology to extract fewer texture features for analysis [5-8], this paper uses image omics technology to extract quantitative features of CT images in a high-throughput manner, realizing a more comprehensive description of tumor characteristics. Unlike most existing studies that use a single algorithm for feature selection [5-6, 8], this paper uses a combination of two filter feature selection algorithms and a wrapper feature selection algorithm to achieve effective feature selection. In addition, this paper also uses four widely used classifiers to build the model and compares their performance to determine the classifier that can obtain the best performance.

Image omics technology has attracted much attention in medical image analysis. In this paper, image omics technology is used to extract a total of 430 gray, texture and wavelet features from each CT image region of interest, and then PCM, Welch's t test and SFFS algorithm are used for effective feature selection in three steps. Finally, a model is built based on RF, KNN, SVM and adaboost classifiers for classification. The best classification performance is obtained by SVM classifier. The accuracy, sensitivity, specificity, PPV, NPV and AUC are 91.67%, 88.89%, 92.86%, 84.21%, 95.12% and 0.9418

respectively. The experimental results show that the method used in this study can improve the classification accuracy of FP - AML and ccrcc, and assist doctors to make decisions.

In the next study, we will collect more data to further verify the conclusions of this paper, and try to extend the proposed model to the classification of other kidney diseases.

References

[1] Hsieh JJ, Purdue MP, Signoretti S, et al. Renal cell carcinoma [J] . Nature Reviews Disease Primers, 2017, 3:17009.

[2] Wei J, Zhao J, Zhang X, et al. Analysis of dual energy spectral CT and pathological grading of clear cell renal cell carcinoma (ccrcc) [J] . Plos One, 2018, 13(5) :e0195699.

[3] Prasad SR, Surabhi VR, Menias CO, et al. Benign renal neoplasms in adults: cross-sectional imaging findings [J] . American Journal of Roentgenology, 2008, 190(1) :158-164.

[4] Lane BR, Aydin H, Danforth TL, et al. Clinical correlates of renal angiomyolipoma subtypes in 209 patients: classic, fat poor, tuberous sclerosis associated and epithelioid [J] . The Journal of Urology, 2008, 80(3) :836-843.

[5] Yan L, Liu Z, Wang G, et al. Angiomyolipoma withminimal fat : differentiation from clear cell renal cell carcinoma and papillary renal cell carcinoma by texture analysis on CT images [J]. Academic Radiology,2015,22(9):1115.

[6] Xu Y. Preliminary study on differential diagnosis of renal angiomyolipoma and renal clear cell carcinoma based on CT image texture analysis [D]. Shenyang:China Medical University, 2017.

[7] Feng Z, Rong P, Cao P, et al. Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma [J]. European Radiology, 2018,28(4):1625-1633.

[8] Lee HS, Hong H, Jung DC, et al. Differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma in contrast-enhanced MDCT images using quantitative feature classification [J]. Medical Physics,2017,44(7):3604-3614.

[9] Lambin P,Leijenaar RTH,Deist TM,et al. Radiomics:the bridge between medical imaging and personalized medicine [J]. Nature Reviews Clinical Oncology,2017,14:749.

[10] Bernardi ED,Buda A,Guerra L, et al. Radiomics of the primary tumour as a tool to improve 18F -FDG-PET sensitivity in detecting nodal metastases in endometrial cancer [J]. EJNMMI Research, 2018,8(1):86.

[11] Ferreira Junior JR, Koenigkam-Santos M, Cipriano FEG, et al. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases [J]. Computer Methods and Programs in Biomedicine, 2018,159:23-30.

[12] Sapate SG, Mahajan A, Talbar SN, et al. Radiomics based detection and characterization of suspicious lesions on full field digital mammograms [J]. Computer Methods and Programs in Biomedicine,2018,163:1-20.

[13] Meng X,Xia W,Xie P,et al. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer [J]. European Radiology,2019,29(6):3200-

3209.
[14] Inza I,Larran~aga P,and Saeys Y. A review of feature selection techniques in bioinformatics [J]. Bioinformatics,2007,23(19): 2507-2517.
[15] Zwanenburg A, Leger S, Vallières M, et al. Image biomarker standardisation initiative -feature definitions [D/ OL]. [2016-12-21]. Http://arxiv. Cry/ pdf/0612.07003. Pdf.
[16] Lubner MG, Smith AD, Sandrasegaran K, et al. CT texture analysis: definitions, applications, biologic correlates, and Challenges [J]. Radiographics, 2017,37(5):1483-1503.
[17] Harikumar R,Vinoth kumar B. Performance analysis of neural networks for classification of medical images with wavelets as a feature extractor [J]. International Journal of Imaging Systems and Technology,2015,25(1):33-40.
[18] Pearson K. Note on regression and inheritance in the case of two parents [J]. Proceedings of the Royal Society of London,1895, 58:240-242.
[19] Derrick B,White P. Why Welch' s test is Type I error robust [J]. The Quantitative Methods in Psychology,2016,12 (1): 30-38.
[20] Aha DW, Bankert RL. A comparative evaluation of sequential feature selection algorithms [M]// Learning from data:artificial intelligence and statistics. Fisher VD,Lenz HJ,Eds. New York: Springer New York,1996:199-206.
[21] Bugdol MD, Bugdol MN, Lipowicz AM, et al. Prediction of menarcheal status of girls using voice features [J]. Computers in Biology and Medicine,2018,100:296-304.
[22] Fallahpour S, Lakvan EN, Zadeh MH. Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem [J]. Journal of Retailing and Consumer Services,2017,34:159-167.
[23] Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines [J]. Machine Learning, 2002,46(1):131-159.