

RESEARCH ARTICLE

Open Access



# Preoperative diagnosis and prediction of hepatocellular carcinoma: Radiomics analysis based on multi-modal ultrasound images

Zhao Yao<sup>1†</sup>, Yi Dong<sup>2†</sup>, Guoqing Wu<sup>1</sup>, Qi Zhang<sup>2</sup>, Daohui Yang<sup>2</sup>, Jin-Hua Yu<sup>1\*</sup> and Wen-Ping Wang<sup>2\*</sup>

## Abstract

**Background:** This study aims to establish a radiomics analysis system for the diagnosis and clinical behaviour prediction of hepatocellular carcinoma (HCC) based on multi-parametric ultrasound imaging.

**Methods:** A total of 177 patients with focal liver lesions (FLLs) were included in the study. Every patient underwent multi-modal ultrasound examination, including B-mode ultrasound (BMUS), shear wave elastography (SWE), and shear wave viscosity (SWV) imaging. The radiomics analysis system was built on sparse representation theory (SRT) and support vector machine (SVM) for asymmetric data. Through the sparse regulation from the SRT, the proposed radiomics system can effectively avoid over-fitting issues that occur in regular radiomics analysis. The purpose of the proposed system includes differential diagnosis between benign and malignant FLLs, pathologic diagnosis of HCC, and clinical prognostic prediction. Three biomarkers, including programmed cell death protein 1 (PD-1), antigen Ki-67 (Ki-67) and microvascular invasion (MVI), were included and analysed. We calculated the accuracy (ACC), sensitivity (SENS), specificity (SPEC) and area under the receiver operating characteristic curve (AUC) to evaluate the performance of the radiomics models.

**Results:** A total of 2560 features were extracted from the multi-modal ultrasound images for each patient. Five radiomics models were built, and leave-one-out cross-validation (LOOCV) was used to evaluate the models. In LOOCV, the AUC was 0.94 for benign and malignant classification (95% confidence interval [CI]: 0.88 to 0.98), 0.97 for malignant subtyping (95% CI: 0.93 to 0.99), 0.97 for PD-1 prediction (95% CI: 0.89 to 0.98), 0.94 for Ki-67 prediction (95% CI: 0.87 to 0.97), and 0.98 for MVI prediction (95% CI: 0.93 to 0.99). The performance of each model improved when the viscosity modality was included.

**Conclusions:** Radiomics analysis based on multi-modal ultrasound images could aid in comprehensive liver tumor evaluations, including diagnosis, differential diagnosis, and clinical prognosis.

**Keywords:** Shear wave dispersion, Viscoelasticity, Radiomics approach, Ultrasound, Hepatocellular carcinoma

\* Correspondence: [jhyu@fudan.edu.cn](mailto:jhyu@fudan.edu.cn); [puguang61@126.com](mailto:puguang61@126.com)

<sup>1</sup>Department of Electronic Engineering, Fudan University, No. 220, Handan Road, Yangpu District, Shanghai 200433, China

<sup>2</sup>Department of Ultrasound, Zhongshan Hospital, Fudan University, 180 Fenqilin Road, Shanghai 200032, China



## Background

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer and the most common cause of death in people with liver cirrhosis [1]. Early and accurate diagnosis of HCC is of vital importance in clinical decision-making and treatment. Currently, although various treatments have been proven effective in the treatment of HCC, recurrence remains an important clinical challenge, with its aggressive biological behaviour and negative impact on overall patient survival. Conventional B-mode ultrasound (BMUS), as a non-invasive, easy and safe procedure, is currently the first-line imaging modality for the diagnosis of HCC. However, BMUS has a limited role in the clinical diagnosis of focal liver lesions (FLLs) and of complicated recurrent lesions. Recent technical advances in shear wave elastography (SWE) and viscosity ultrasound increase the diagnostic efficiency of ultrasound and allow it to evaluate liver stiffness with the aim of assessing hepatic fibrosis and cirrhosis. To date, only a few studies have focused on the quantification of SWE stiffness in FLLs [2–4].

More recently, as an emerging method for medical image processing, radiomics is used to convert medical images into high-dimensional, mineable features that reflect underlying pathophysiological information [5]. Radiomics employs a variety of state-of-the-art machine learning or deep learning techniques to complete a variety of clinical tasks, which greatly pushed the development of precision medicine [6]. Microvascular invasion (MVI) and antigen Ki-67 (Ki-67) are regarded as high-risk factors for HCC recurrence. As an immunotherapy target, programmed cell death protein 1 (PD-1) has also become increasingly meaningful for the treatment of patients with HCC.

According to previous research, radiomics has great potential for the diagnosis and treatment of liver diseases. In a study by Virmani et al. [7], 48 features were extracted from gray-scale ultrasound images to differentiate normal livers, cirrhotic livers and HCC. A genetic algorithm and support vector machine (SVM) were used as feature selection and classification methods. Owjimehr et al. [8] performed a wavelet packet transform on the gray-scale ultrasound image and extracted 61 features to differentiate normal, fatty and heterogeneous livers. SVM and k-nearest neighbour classifiers were applied to classify the images into three groups. Furthermore, some studies applied artificial neural networks to diagnose abnormal livers [9], chronic liver disease [10] and HCC malignancy [11]. These studies demonstrate the feasibility of ultrasound imaging in liver disease diagnosis and imply the great potential of radiomics analysis.

Current radiomics methods have several limitations when analysing the data of our study. First, traditional

engineered features (intensity, shape, margin, calcification, wavelet, etc.) are designed for different diseases and are poorly adaptable for HCC. Second, the deep learning algorithms are easily over-fitted when dealing with data with a small sample size. Finally, most of the above studies used a single modal ultrasound imaging, without utilizing comprehensive information provided by multi-modal ultrasound images.

Due to its good performance in signal representation and reconstruction, sparse representation (SR) is widely used in feature selection [12, 13] and image classification [14, 15]. By training the optimal texture to represent images, SR can adaptively learn image features with a small amount of imaging data. Furthermore, as a nonparametric model, SR can effectively avoid over-fitting and has strong robustness [16].

A radiomics analysis system based on SR and SVM was proposed in our study. We trained the model with multi-modal ultrasound images and used histology as the gold standard measure. Our study aims to evaluate the feasibility of ultrasound radiomics models in the differential diagnosis and characterization of histologically proven FLLs and to determine an initial prognosis of HCC.

## Methods

### Patients and materials

Between July 2017 and June 2018, 177 consecutive patients (102 women and 75 men; age range: 15–91 years, mean:  $55.5 \pm 10.4$  years) who were referred to our institution for FLL SWE assessment were included in the prospective study. For patients under the age of 16 years, informed consent was obtained from a parent and/or legal guardian. All patients underwent multi-modal ultrasound examination, including BMUS, shear wave elastography (SWE), and shear wave viscosity (SWV) imaging before surgery. The final diagnoses for all 177 patients were based on histopathological results obtained from liver biopsy during surgery.

Data collected included the patient's age, gender, and focal liver lesion location. Among all 177 patients included in this study, 66 were excluded, and the exclusion criteria were as follows: (1) missing important pathology results; (2) poor imaging quality; (3) accompanied with other diseases, including cirrhosis and fatty liver. Characteristics of the 111 FLL patients enrolled are summarized in Table 1, which include patient gender and age. Statistics show that the gender and age of patients are related to the benign and malignant classification of tumors ( $p < 0.05$ ). In addition, patient age was statistically related to the type of malignant tumor ( $p < 0.05$ ).

The flow chart of the proposed radiomics system is shown in Fig. 1. We first classified the benign and malignant cases. Then, we separated patients with HCC from

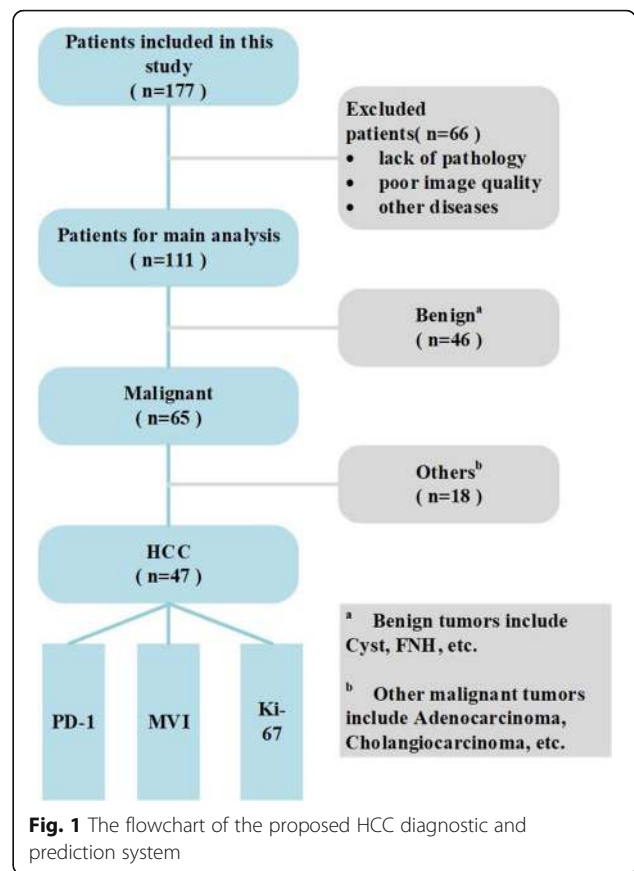
**Table 1** Baseline characters of patients

Parameters	All patients	Male (N; %)	Ages (mean ± variance)
<b>Tumor category</b>			
benign	46	21; 46%	50.5 ± 13.4
malignant	65	54; 83%	56.6 ± 8.3
P value	–	0.00004	0.0040
<b>Malignant subtyping</b>			
HCC	47	41; 87%	55.3 ± 8.4
others	18	13; 72%	60.2 ± 7.0
P value	–	0.1550	0.0267
<b>PD-1 prediction</b>			
PD-1 present	15	14; 93%	53.0 ± 8.8
PD-1 absent	24	20; 83%	56.2 ± 8.9
P value	–	0.3831	0.2782
<b>Ki-67 prediction</b>			
≤ 25	21	19; 90%	53.9 ± 9.6
> 25	23	19; 83%	56.6 ± 7.6
P value	–	0.4647	0.2441
<b>MVI prediction</b>			
MVI present	21	18; 86%	53.9 ± 8.0
MVI absent	22	19; 86%	56.0 ± 8.9
P value	–	0.9677	0.3810

the remaining 65 patients with malignancies. Finally, multi-modal ultrasound images of 47 patients with HCC were used to predict PD-1, Ki-67 and MVI indicators of HCC. Benign tumors in the study mainly include cyst and focal nodular hyperplasia (FNH). Other malignant tumors that differ from HCC include adenocarcinoma and cholangiocarcinoma.

Multi-modal ultrasound examinations were performed using Toshiba Aplio i900 ultrasound equipment (Canon Medical, Japan). A PV1-475BX convex array probe (1–8 MHz) was used. Patients lied in a supine position with the right arm in maximal extension. The transducer was positioned in a right intercostal space to visualize the right liver lobe. Large vessels were avoided. Optimally, patients were instructed to perform a transient breath hold in a neutral position. Regions of interest (ROIs) were placed a minimum of 1–2 cm and a maximum of 8 cm beneath the liver capsule [17]. An ROI was placed inside the lesion or surrounding the hepatic parenchyma at the same depth as the lesion. In the ROIs of lesions and the parenchyma, SWE and viscosity were measured.

A multi-modal ultrasound image includes four different modalities, as shown in Fig. 2, where the upper left panel is an elastography image, the lower left panel is a gray-scale ultrasound image and the lower right panel is a viscosity image. The propagation map in the upper right panel reflects the image quality (> 90% was



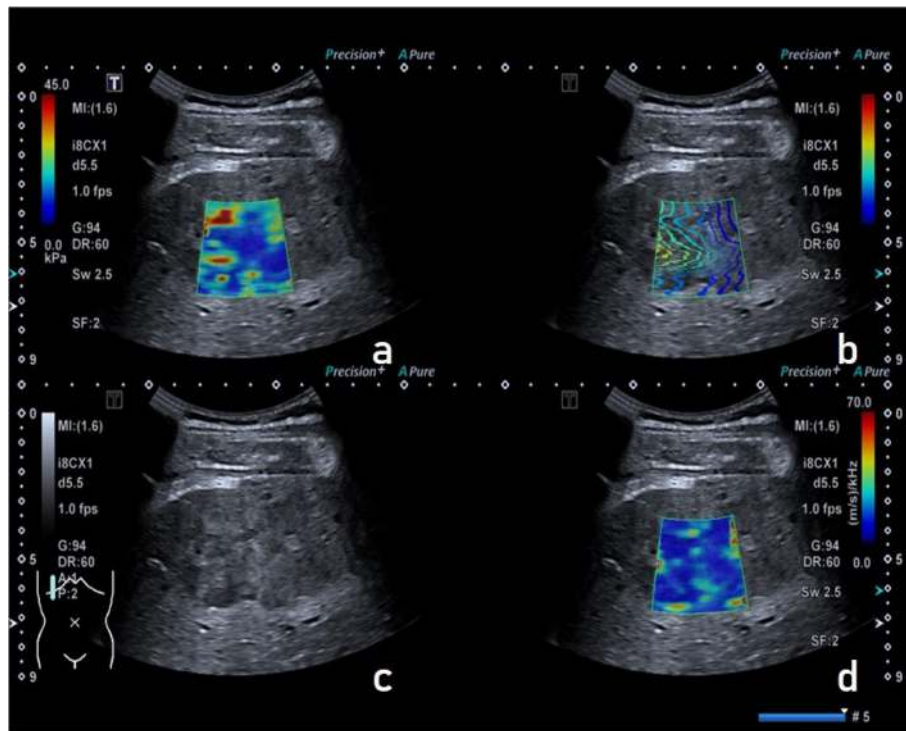
considered to be good quality for measurement). The regions with optimal and stable imaging qualities were manually selected by an experienced sonographer and marked as ROIs.

**Overall design**

The overall methods include three steps: feature extraction, feature selection and classification. First, the SR dictionary was trained to extract features. Then, an iterative algorithm based on SR was used for feature selection. Finally, we trained an SVM model with the selected features. We validated the model by leave-one-out cross-validation (LOOCV).

**Feature extraction**

We adopted an SR-based feature extraction method to extract image features. First, we used the K-singular value decomposition (KSVD) algorithm to learn the corresponding structural texture dictionary from each type of image [18]. Then, the various types of dictionaries were combined into a feature extraction dictionary (FED), and the FED was used to sparsely represent the test images. The representation coefficients reflect the relationship between the test images and each type of dictionary (each class), so the coefficients can be



**Fig. 2** Multi-modal colour ultrasound image. **a.** Elastography. **b.** Propagation map, which reflects the image quality. **c.** Gray-scale ultrasound. **d.** Viscosity modality

classified as the test image features. We used the orthogonal matching pursuit (OMP) algorithm to calculate the SR coefficients and extract the coefficients for features. The detailed process of feature extraction can be found in [Appendix: Feature extraction](#).

#### Feature selection

Redundant and irrelevant features can seriously affect the performance of the classification. Hence, we adopted an iterative SR method to select some crucial features for the classifier. We used sample features to sparsely represent sample labels, and the absolute value of the SR coefficient was the importance of the feature. To improve the stability of feature selection, we performed iterative SR for feature selection. We selected a partial sample for SR in each iteration and then averaged the results of multiple SRs to determine the final coefficients. Finally, we sorted the features according to the absolute value of the SR coefficients. Specific mathematical models for feature selection can be found in [Appendix: Feature selection](#).

#### Classification

There are many types of classifiers in radiomics, and SVM is widely used for stability and optimal performance. In this work, we used LibSVM for classification,

which can solve the problem of sample imbalance [19]. A specific mathematical model of LibSVM is shown in [Appendix: SVM model](#). By adjusting the penalty factor, we eliminated the effects of sample imbalance. A receiver operating characteristic (ROC) curve was used to show the overall performance of the model. We also calculated some indexes to evaluate the performance of the classifier, including accuracy (ACC), sensitivity (SENS), specific (SPEC) and area under the ROC (AUC).

#### Cross-validation

Each time LOOCV takes one sample as a test sample, and all the remaining samples are used as training sets. This process was repeated until all the samples were traversed. We used LOOCV to evaluate our model.

#### Statistical analysis

Descriptive statistics are summarized as the mean  $\pm$  SD. The Mann-Whitney U test was used to test whether a feature has discriminative power in different tasks, and  $p$  values less than 0.05 indicated statistical significance. SPSS statistics 20.0 software (SPSS, Chicago, IL, USA) and MedCalc software (V.11.2; 2011 MedCalc Software bvba, Mariakerke, Belgium) were used to perform the statistical analysis.

## Results

### Multi-modal ultrasound image feature extraction and feature analysis

Because the model establishment process was similar for the five radiomics models, we used benign and malignant differentiation as an example analysis. A schematic diagram of the dictionary training is shown in Fig. 3. Figure 3a shows a blank dictionary that has not been trained. Because the initial discrete cosine transform (DCT) dictionary cannot optimally represent the image information of each category simultaneously, it was necessary to train different dictionaries that include the texture structure features of each type based on the DCT dictionary. We use the KSVD algorithm to train the dictionary, and we finally obtained a dictionary with rich texture information, as shown in Fig. 3b.

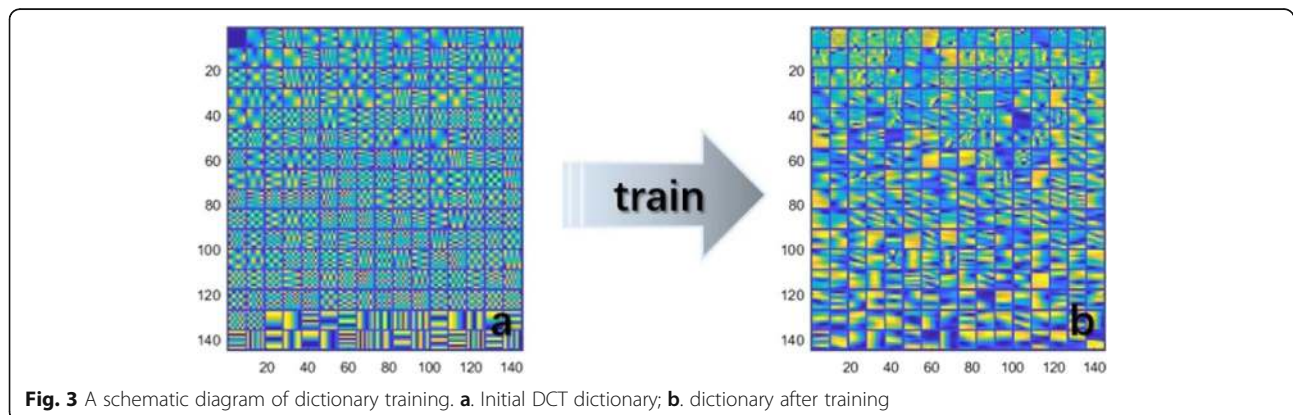
The overall flowchart of feature extraction is shown in Fig. 4. We manually selected three corresponding square measurement areas as the ROIs in the multi-modal images. The size of the dictionary we used in this study is  $64 \times 256$ . A dictionary contains 256 atoms, corresponding to 256 sparse coefficients, which can be taken as 256 features. In the case of using only gray-scale ultrasound images, two dictionaries need to be trained separately for the two categories, so a total of 512 features can be extracted. Particularly, when extracting features from elastography or viscosity modality images, because the image is three channels (RGB), we first performed HSV (hue, saturation, value) conversion on the RGB images. Then, we used the hue (H) and value (V) channels to train the dictionary separately. Hence, for elastography or viscosity images, we trained four dictionaries to obtain 1024 features. Finally, after multi-modal feature combination, the gray-scale modality (GM), gray-scale and elastography modality (GEM) and gray-scale, elastography and viscosity modality (GEVM) corresponded to 512, 1536 and 2560 features, respectively.

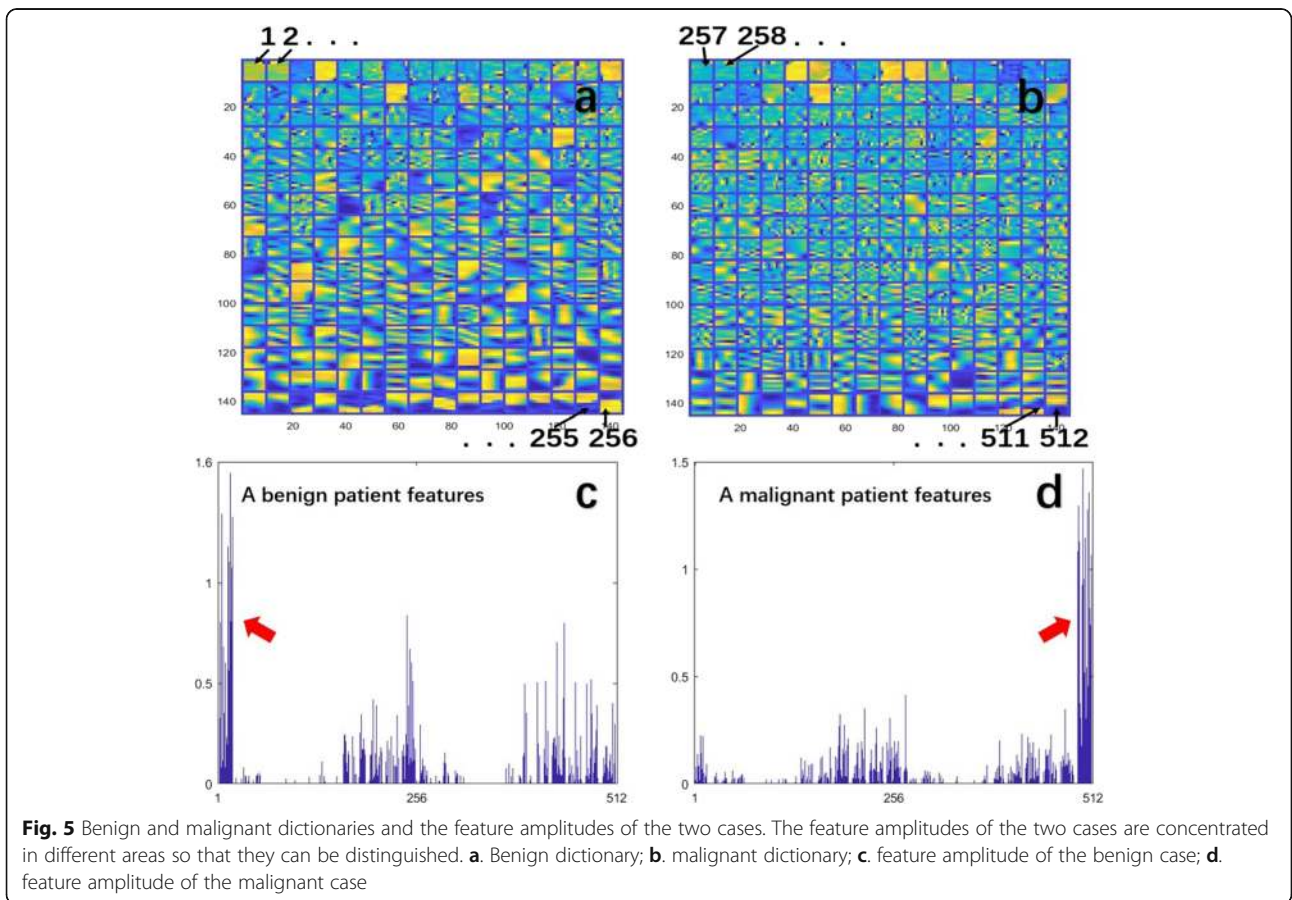
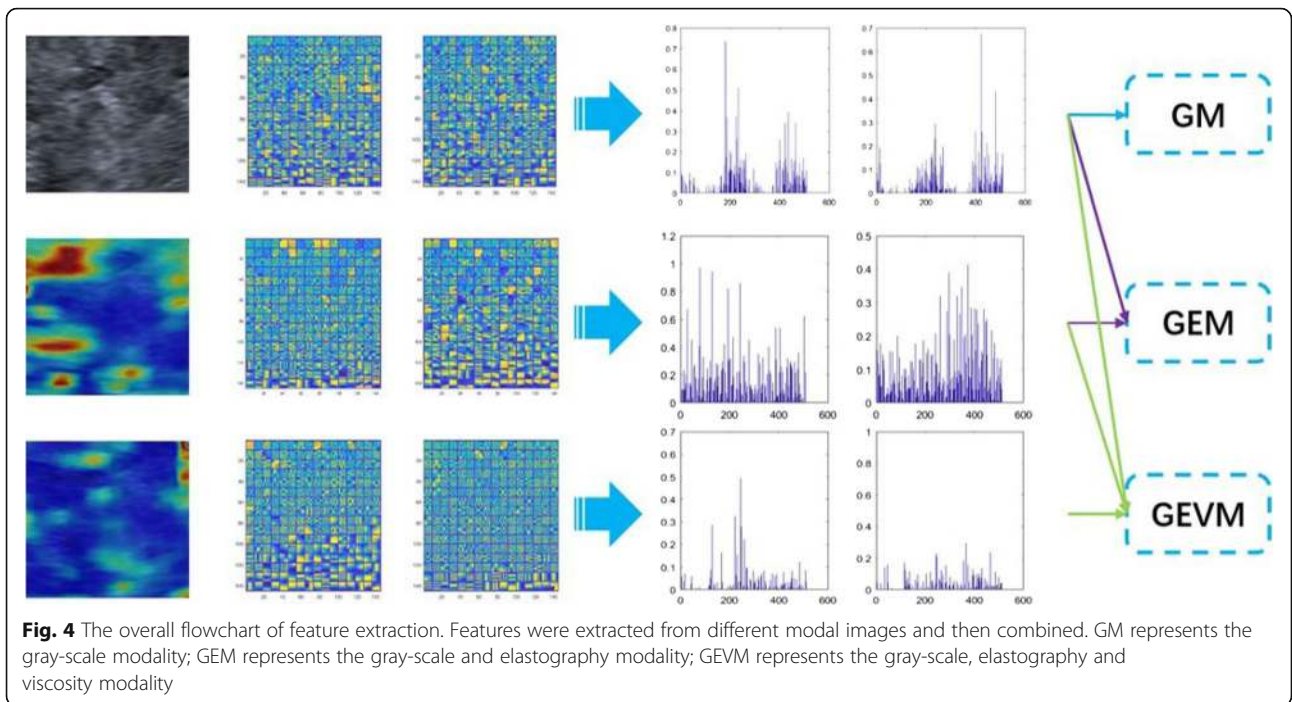
We randomly selected two cases (one benign and the other malignant) to analyse the features of their GM images. The feature amplitudes of the two cases and two

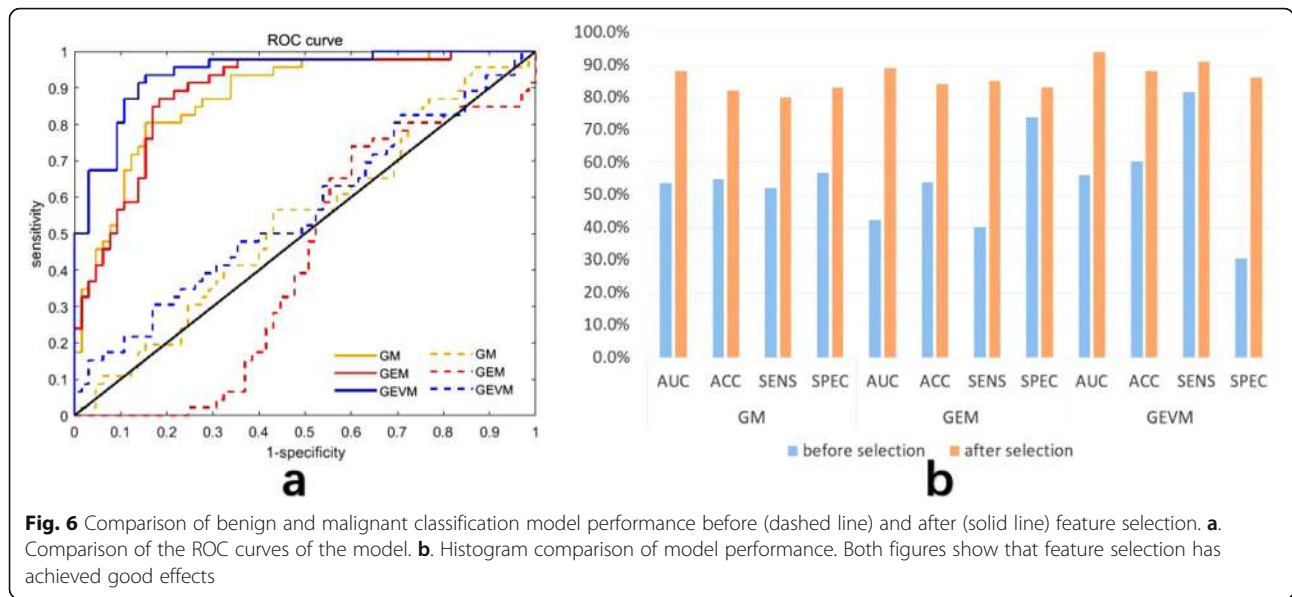
corresponding benign and malignant dictionaries are shown in Fig. 5. Figure 5a and b correspond to benign and malignant dictionaries, respectively, and they together form an FED. In the two dictionaries, 512 atoms correspond to 512 features of a case. It is obvious that the two dictionaries have quite different textures and that the malignant dictionary has more structural information. The linear combination of atoms in FED makes up the entire ROI, and the different feature magnitudes represent the different proportions of atoms. The special region in Fig. 5 is marked by a red arrow. The area with the highest amplitude of the benign patient is located in the feature interval corresponding to the benign dictionary (1 to 256), while the area with the highest amplitude of the malignant patient is located in the 257 to 512 feature interval, which corresponds to the malignant dictionary. This result indicates that the image of the benign case is mainly composed of textures from the benign dictionary, while the image of the malignant case is mainly composed of textures from the malignant dictionary. This significant difference can distinguish benign and malignant tumors effectively.

### Feature selection results

Eliminating redundant and invalid features is critical to the performance of the classifier. As an example, we analysed the importance of feature selection in benign and malignant tumor classification. Figure 6 shows a comparison of the performance of the features before and after feature selection. Under all imaging modalities, each evaluation indicator of the model has been improved by feature selection. Figure 6a shows a comparison of the ROC curves of the models. The dashed line and solid line correspond to the results before feature selection and after feature selection, respectively. The histogram in Fig. 6b describes the AUC before and after feature selection. The blue bar represents the AUC before feature selection, while the yellow bar corresponds to AUC after feature selection. The results clearly show







that our feature selection strategy has achieved good results. The detailed statistical results are shown in Table 2.

**Classification of benign and malignant liver tumors**

A total of 111 cases were used in this experiment, of which 65 were malignant cases. We compared the performance of GM, GEM and GEVM in the classification of benign and malignant liver tumors. Some indicators of the model are summarized in Table 3.

The AUCs of GEVM and GEM reach 0.94 (95% confidence interval [CI]: 0.88 to 0.98) and 0.89 (CI: 0.81 to 0.94), respectively, which are 0.06 and 0.01 higher than that of GM (CI: 0.80 to 0.93). The AUC of GEVM is 0.05 higher than that of GEM. The ROC curves of these models are shown in Fig. 7. We calculated the statistical significance level of the AUCs for GM and GEVM ( $p = 0.14$ ). Although the application of multi-modal images increased the AUCs, multi-modal images do not exhibit significant differences from single BMUS in terms of differentiation between benign and malignant tumors.

**Malignant liver tumor subcategories**

A total of 47 HCC and 18 other malignant tumor cases (11 adenocarcinoma cases and 7 cholangiocarcinoma

cases) were studied in this experiment. The AUC of GM reached 0.90 (CI: 0.85 to 0.96). The AUCs of GEM and GEVM are slightly greater than that of GM, reaching 0.92 (CI: 0.86 to 0.97) and 0.97 (CI: 0.93 to 0.99), respectively. The ROC curves of these models are shown in Fig. 8. The calculation results show that there are significant differences between GM and GEVM ( $p = 0.04$ ). The application of multi-modal images achieved better results in distinguishing the subtypes of malignant tumors. The results for classification of the subtypes of malignant liver tumors are shown in Table 4.

**PD-1, Ki-67, and MVI indicator prediction**

The classification criterion of PD-1 is whether or not the indicator is expressed. The Ki-67 indicator is classified by a 25% threshold value ( $\leq 25\%$  or  $> 25\%$ ). The MVI indicator is divided into two categories according to low risk and high risk. The prediction results of the three indicators are summarized in Table 5. The ROC curves of each indicator are shown in Fig. 9. GEVM resulted in significant differences in the AUCs of the three predictive indicators ( $p = 0.02$  for PD-1,  $p = 0.04$  for Ki-67,  $p = 0.0006$  for MVI) relative to those of GM. Better performance can be obtained by predicting three indicators using multi-modal ultrasound images.

**Table 2** Performance comparison of models before and after feature selection

	GM				GEM				GEVM			
	AUC	ACC	SENS	SPEC	AUC	ACC	SENS	SPEC	AUC	ACC	SENS	SPEC
BF	54	55	52	57	42	54	40	74	56	60	82	30
AF	88	82	80	83	89	84	85	83	94	88	91	86

AUC area under the receiver operating characteristic curve, ACC accuracy, SENS sensitivity, SPEC specificity, GM gray-scale modality, GEM gray-scale and shear wave elastography modality, GEVM gray-scale, shear wave elastography and viscosity modality, BF before selection, AF after selection. The auc, acc, sens and spec are expressed as a percentage

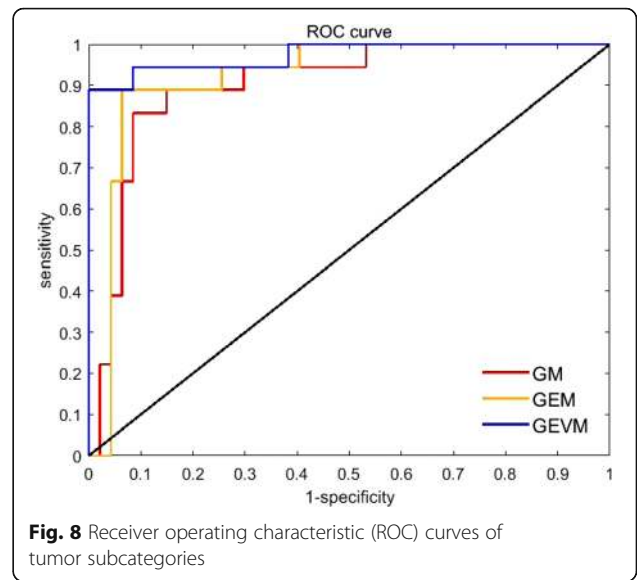
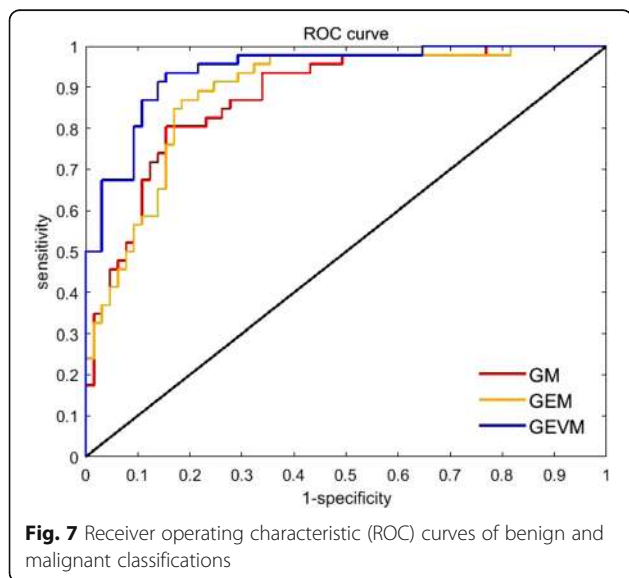
**Table 3** Diagnostic performance of GM,GEM and GEVM for classifying benign and malignant tumors

	AUC(%)	ACC(%)	SENS(%)	SPEC(%)
GM	88	82	80	83
GEM	89	84	85	83
GEVM	94	88	91	86

AUC area under the receiver operating characteristic curve, ACC accuracy, SENS sensitivity, SPEC specificity, GM gray-scale modality, GEM gray-scale and shear wave elastography modality, GEVM gray-scale, shear wave elastography and viscosity modality

**Discussion**

Multi-modal ultrasound technology increases the diagnostic efficiency of ultrasound and makes it possible to diagnose FLL before surgery. In contrast to the evaluation of diffuse parenchymal liver disease, little is known about FLL characterization using SWE or SWV technology. Here, we investigate the value of multi-modal ultrasound technology for the differential diagnosis of benign and malignant FLLs using radiomics analysis. Previously, Dong et al. [20] applied ElastPQ measurements for differential diagnosis of benign and malignant FLLs and successfully found the optimal threshold of shear wave speed. Ozmen et al. [21] used the optimal threshold of SWE to differentiate benign and malignant liver tumors and obtained an AUC of 0.77. However, the cut-off values of measurement for differentiating benign and malignant liver tumors tend to show great variability. In our study, innovative multi-modal ultrasound images were used to diagnose liver tumors. By converting the images into high-throughput features, radiomics was used to mine the rich texture information in the patient images in order to classify the images. We found that malignant tumor images have more complex textures and more



structural information. The experimental results also show that the model has achieved good results on the classification of benign and malignant liver tumors (0.94 AUC for differentiating between benign and malignant liver tumors).

The most common type of histology of primary liver cancer is HCC, which represents 90% of cases [22, 23]. Difficulties in treatment and poor prognosis make it important to accurately detect HCC. In addition, early diagnosis of HCC is also crucial for optimizing treatment options. In a study by Thomas et al., alpha-fetoprotein (AFP) was used to detect HCC [24]. However, AFP is only a supplement to the ultrasound image information, and the accuracy of detecting HCC is not satisfactory. In our experiments, multi-modal ultrasound images were used to directly distinguish between HCC and other malignancies noninvasively, and the model performed well (0.97 AUC for liver tumor subtyping). This result illustrates the great potential of ultrasound images for tumor diagnosis.

Patients with HCC have a poor prognosis due to a high recurrence rate. It has been reported that the 5-year recurrence rate of primary liver cancer is as high as 45%~ 60% [25]. We mainly studied two factors that

**Table 4** Diagnostic performance of GM,GEM and GEVM for liver tumor subtyping

	AUC(%)	ACC(%)	SENS(%)	SPEC(%)
GM	90	89	83	91
GEM	92	92	89	94
GEVM	97	97	89	100

AUC area under the receiver operating characteristic curve, ACC accuracy, SENS sensitivity, SPEC specificity, GM gray-scale modality, GEM gray-scale and shear wave elastography modality, GEVM gray-scale, shear wave elastography and viscosity modality



**Table 5** Performance of GM,GEM and GEVM for indicators prediction

	PD-1				Ki-67				MVI			
	AUC	ACC	SENS	SPEC	AUC	ACC	SENS	SPEC	AUC	ACC	SENS	SPEC
GM	84	85	80	88	86	84	86	83	85	84	86	81
GEM	94	90	93	88	92	89	86	91	95	93	91	95
GEVM	97	92	100	88	94	93	95	91	98	95	91	100

AUC area under the receiver operating characteristic curve, ACC accuracy, SENS sensitivity, SPEC specificity, GM gray-scale modality, GEM gray-scale and shear wave elastography modality, GEVM gray-scale, shear wave elastography and viscosity modality, PD-1 programmed cell death protein 1, Ki-67 antigen Ki 67, MVI micro vascular invasion. The auc, acc, sens and spec are expressed as a percentage

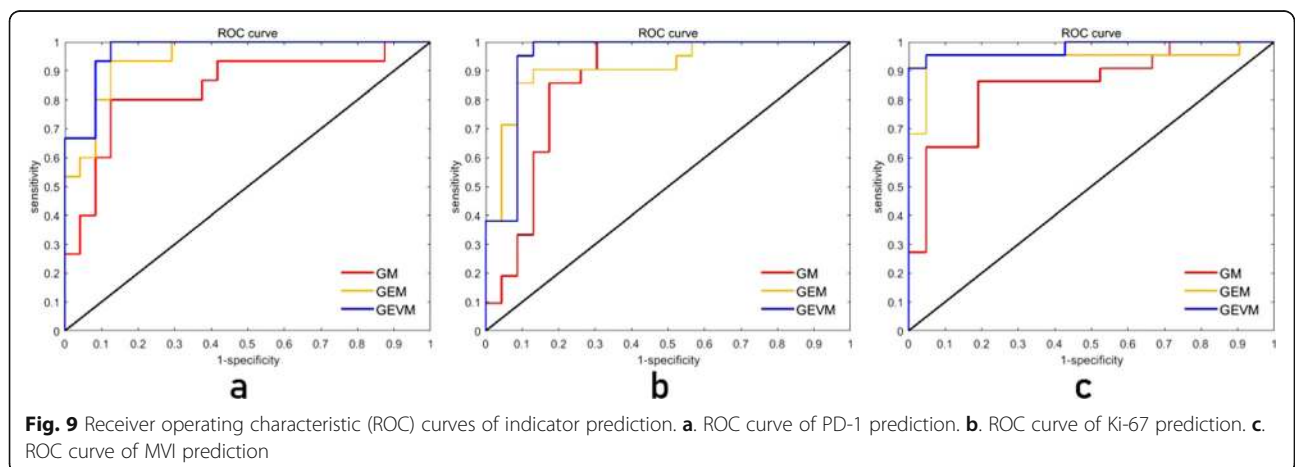
affect the recurrence of liver cancer. One of the factors is MVI. MVI has been reported as one of the major risk factors related to HCC recurrence and represents a poor prognosis [26, 27]. Many previous studies have focused on identifying radiologic features (such as tumor size, tumor margin, and number of lesions) in various types of medical images for the preoperative prediction of MVI [28–30]. However, the best predictive feature of MVI in HCC remains controversial. In addition, another study used a radiomics nomogram to predict MVI pre-operatively, resulting in a C-index of 0.84 [31]. However, the results of these studies are not satisfactory. Our radiomics-based model achieved better results (0.98 AUC) in predicting MVI than did previous studies using multi-modal ultrasound images.

Another factor we studied that has an effect on HCC recurrence is Ki-67. A previous study suggested that a higher Ki-67 index confers poor prognosis in patients with HCC [32–34]. Clinically, immunohistochemistry is needed to detect the Ki-67 index. Studies have analysed the correlation between the expression of other proteins (such as PDIA3) and Ki-67 [35]. However, to the best of our knowledge, no study has applied medical images to predict Ki-67 noninvasively. Our results (0.94 AUC for Ki-67 prediction) demonstrated that it is feasible to noninvasively predict Ki-67 based on radiomics. In our study, we successfully determined MVI and Ki-67 for HCC prognosis by applying multi-modal ultrasound images.

Recent studies have shown that immunotherapy is a promising approach for HCC treatment and that PD-1 is crucial for tumor immunity [36]. Accurate assessment of PD-1 can be useful in assessing the range of applications of PD-1/PD-L1 blockers in liver cancer patients. In addition, an increase in PD-1 predicts a poorer prognosis for HCC [37]. The prediction of PD-1 is important for the progression and postoperative recurrence of HCC. The model we built for PD-1 prediction has achieved good results (0.97 AUC for PD-1 prediction). By integrating multi-modal ultrasound image information, the radiomics model can determine PD-1 noninvasively.

To investigate the effects of feature selection on classifier performance, we compared the performance of models before and after feature selection in benign and malignant tumors. Feature selection truncates redundant and invalid features, so the model becomes robust. The experimental results show that the performance of the model after feature selection is better than that before feature selection (significant level in ROC curves,  $p < 0.0001$ ).

There are some limitations to our research. It should be mentioned that our study lacks multi-centre validation, which would provide more convincing results. In addition, more samples should be collected to build a more robust model. Furthermore, we employed only the image information from diseased livers, and some text



**Fig. 9** Receiver operating characteristic (ROC) curves of indicator prediction. **a.** ROC curve of PD-1 prediction. **b.** ROC curve of Ki-67 prediction. **c.** ROC curve of MVI prediction

descriptions of the cases and biomarkers were not applied.

**Conclusions**

In summary, we successfully established an HCC diagnosis and prognosis system based on ultrasound radiomics and proved its potential feasibility and effectiveness. Simultaneously, we demonstrated the potential value of multi-modal ultrasound-based radiomics analysis in computer-aided diagnosis (CAD).

**Appendix**

**Feature extraction**

The SR method can adaptively learn and extract texture features of images. First, we exploited the KSVD algorithm to train the dictionary corresponding to different categories. This algorithm trains different categories of dictionaries by iteratively updating each atom in the dictionary. We denote  $i \in \{1, 2, \dots, I\}$  as all sample categories.  $D_i$  is the corresponding  $i$ -class dictionary. Then, the process of feature extraction can be written as:

$$\hat{\alpha} = \arg \min_{\alpha} \|y - D\alpha\|_2^2 + \mu \|\alpha\|_p \tag{1}$$

where  $y$  is classifier label;  $D = [D_1, D_2, \dots, D_I]$  is a collection of all SR dictionaries;  $\alpha$  is the SR coefficients, which can be considered as features of the samples;  $\hat{\alpha}$  is the estimated value of  $\alpha$ ;  $\|\cdot\|_p$  represents the  $L_p$  norm;  $\mu$  is the regularization parameter.  $\mu \|\alpha\|_p$  can be regarded as the error term that can be discarded. We used the OMP algorithm to solve (1) to obtain the image features.

**Feature selection**

Different from the traditional feature selection method, the SR method adopts the strategy of the sliding window, so it can comprehensively utilize the information of all samples in the window. An iterative process can be expressed as:

$$\hat{d}^{(k)} = \operatorname{argmin}_d \|s^{(k)} - F^{(k)}d\|_2^2 + \varepsilon \|d\|_0 \tag{2}$$

where  $s^{(k)}$  is the label used for the  $k$ -th iteration;  $F^{(k)}$  is the feature selected for the  $k$ -th iteration;  $\varepsilon$  is a small constant;  $\hat{d}^{(k)}$  is the coefficient calculated by the  $k$ -th iteration. Then, we calculated the average of  $\hat{d}^{(k)}$  for the  $k$ -th iteration:

$$d^{(k)} = \frac{1}{k} \sum_{i=1}^k \hat{d}^{(i)} \tag{3}$$

The  $d^{(k)}$  was used for feature selection. After the iteration, each feature obtained a score that combines all the sample information due to the averaging operation. The higher the score, the higher the importance of

the feature. In this way, the feature selection results are obtained.

**SVM model**

The LibSVM model can solve the sample imbalance problem by adjusting different penalty coefficients. The improved SVM mathematical model can be written as:

$$\min_{w,b,\xi} \frac{1}{2} \omega^T \omega + C_+ \sum_{y_i=1} \xi_i + C_- \sum_{y_i=-1} \xi_i \tag{4}$$

$$\text{subject to } y_i(\omega^T \phi(x_i) + b) \geq 1 - \xi_i \tag{5}$$

$$\xi_i \geq 0, i=1, \dots, l$$

where  $\omega$  is the hyperplane normal vector and  $b$  is the bias, which collaboratively determines the hyperplane;  $\phi(x_i)$  is the feature vector mapped by  $x_i$ ;  $y_i$  is the sample label;  $\xi_i$  is a small constant;  $C_{+(-)}$  is the penalty parameter, which assigns weights to different proportions of samples. By assigning an appropriate  $C$ , we can eliminate the sample imbalance problem.

**Abbreviations**

ACC: accuracy; AFP: alpha-fetoprotein; AUC: area under the receiver operating characteristic curve; BMUS: B-mode ultrasound; CAD: computer-aided diagnosis; CI: confidence interval; DCT: discrete cosine transform; FED: feature extraction dictionary; FLLs: focal liver lesions; FNH: focal nodular hyperplasia; GEM: gray-scale and elastography modality; GEVM: gray-scale, elastography and viscosity modality; GM: gray-scale modality; HCC: hepatocellular carcinoma; Ki-67: antigen Ki-67; KSVD: K-singular value decomposition; LOOCV: leave-one-out cross-validation; MVI: microvascular invasion; OMP: orthogonal matching pursuit; PD-1: programmed cell death protein 1; ROC: receiver operating characteristic; SENS: sensitivity; SPEC: specificity; SR: sparse representation; SRT: sparse representation theory; SVM: support vector machine; SWE: shear wave elastography; SWV: shear wave viscosity

**Acknowledgements**

Not applicable.

**Funding**

This work was financially supported by the National Natural Science Foundation of China (61471125, 81571676, 81501471). The funding body had no role in the design of the study, in the collection, analysis, and interpretation of data or in the writing of the manuscript.

**Availability of data and materials**

The datasets used and analysed during the current study are available from the corresponding author(s) on reasonable request.

**Authors' contributions**

ZY and YD designed the methods, carried out the experiments and wrote the paper; GQW designed the methods and assisted in the experimental design; QZ assisted in the clinical design, data collection and writing the paper; DHY was responsible for the clinical design and data collection; JHY and WPW proposed the project, directed the methods and experimental design, and revised the paper. All authors read and approved the manuscript.

**Ethics approval and consent to participate**

This study was approved by the ethics committee of the Zhongshan Hospital Affiliated with Fudan University. Written informed consent was obtained from all individual participants included in the study. For patients under the age of 16 years, written informed consent was obtained from a parent and/or legal guardian.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

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

Received: 28 August 2018 Accepted: 28 October 2018

Published online: 12 November 2018

**References**

- Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018; 391(10127):1301–14.
- Samir AE, Dhyani M, Vij A, Bhan AK, Halpern EF, Mendez-Navarro J, Corey KE, Chung RT. Shear-wave Elastography for the estimation of liver fibrosis in chronic liver disease: determining accuracy and ideal site for measurement. *Radiology*. 2015;274(3):888–96.
- Conti CB, Cavalcoli F, Fraquelli M, Conte D, Massironi S. Ultrasound elastographic techniques in focal liver lesions. *World J Gastroenterol*. 2016; 22(9):2647–56.
- Xie LT, Yan CH, Zhao QY, He MN, Jiang TA. Quantitative and noninvasive assessment of chronic liver diseases using two-dimensional shear wave elastography. *World J Gastroenterol*. 2018;24(9):957–70.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, They Are Data. *Radiology*. 2016;278(2):563–77.
- Acharya UR, Hagiwara Y, Sudarshan VK, Chan WY, Ng KH. Towards precision medicine: from quantitative imaging to radiomics. *J Zhejiang Univ Sci B*. 2018;19(1):6–24.
- Virmani J, Kumar V, Kalra N, Khandelwal N. SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors. *J Digit Imaging*. 2013;26(3):530–43.
- Owjimehr M, Danyali H, Helfroush MS. An improved method for liver diseases detection by ultrasound image analysis. *J Med Signals Sens*. 2015;5(1):21–9.
- Kalyan K, Jakhia B, Lele RD, Joshi M, Chowdhary A. Artificial neural network application in the diagnosis of disease conditions with liver ultrasound images. *Adv Bioinforma*. 2014;2014:708279.
- Bharti P, Mittal D, Ananthasivan R. Preliminary study of chronic liver classification on ultrasound images using an ensemble model. *Ultrasound Imaging*. 2018;161734618787447.
- Sugimoto K, Shiraishi J, Tanaka H, Tsuchiya K, Aso K, Kobayashi Y, Iijima H, Moriyasu F. Computer-aided diagnosis for estimating the malignancy grade of hepatocellular carcinoma using contrast-enhanced ultrasound: an ROC observer study. *Liver Int*. 2016;36(7):1026–32.
- Lin DD, Cao HB, Calhoun VD, Wang YP. Sparse models for correlative and integrative analysis of imaging and genetic data. *J Neurosci Methods*. 2014;237:69–78.
- Cao HB, Duan JB, Lin DD, Shugart YY, Calhoun V, Wang YP. Sparse representation based biomarker selection for schizophrenia with integrated analysis of fMRI and SNPs. *Neuroimage*. 2014;102:220–8.
- Yuan XT, Liu XB, Yan SC. Visual classification with multitask joint sparse representation. *IEEE Trans Image Process*. 2012;21(10):4349–60.
- Zhang HC, Nasrabadi NM, Zhang YN, Huang TS. Multi-observation visual recognition via joint dynamic sparse representation. 2011 IEEE Int Conf Comput Vision (iccv) 2011:595–602.
- Wu GQ, Chen YS, Wang YY, Yu JH, Lv XF, Ju X, Shi ZF, Chen L, Chen ZP. Sparse representation-based Radiomics for the diagnosis of brain tumors. *IEEE Trans Med Imaging*. 2018;37(4):893–905.
- Dietrich CF, Bamber J, Berzigotti A, Bota S, Cantisani V, Castera L, Cosgrove D, Ferraioli G, Friedrich-Rust M, Gilja OH, et al. EFSUMB guidelines and recommendations on the clinical use of liver ultrasound Elastography, update 2017 (long version). *Ultraschall Med*. 2017;38(4):e16–47.
- Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process*. 2006;54(11):4311–22.
- Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2011;2(3):13–5.
- Dong Y, Wang WP, Xu Y, Cao J, Mao F, Dietrich CF. Point shear wave speed measurement in differentiating benign and malignant focal liver lesions. *Med Ultrason*. 2017;19(3):259–64.
- Ozmen E, Adaletli I, Kayadibi Y, Emre S, Kilic F, Dervisoglu S, Kurugoglu S, Senyuz OF. The impact of shear wave elastography in differentiation of hepatic hemangioma from malignant liver tumors in pediatric population. *Eur J Radiol*. 2014;83(9):1691–7.
- Lewis RB, Lattin GE Jr, Makhlof HR, Levy AD. Tumors of the liver and intrahepatic bile ducts: radiologic-pathologic correlation. *Magn Reson Imaging Clin N Am*. 2010;18(3):587–609 xii.
- Ippolito D, Inchingolo R, Grazioli L, Drago SG, Nardella M, Gatti M, Faletti R. Recent advances in non-invasive magnetic resonance imaging assessment of hepatocellular carcinoma. *World J Gastroenterol*. 2018;24(23):2413–26.
- Bird TG, Dimitropoulou P, Turner RM, Jenks SJ, Cusack P, Hey S, Blunsum A, Kelly S, Sturgeon C, Hayes PC, et al. Alpha-fetoprotein detection of hepatocellular carcinoma leads to a standardized analysis of dynamic AFP to improve screening based detection. *PLoS One*. 2016;11(6):e0156801.
- Janevska D, Chaloska-Ivanova V, Janevski V. Hepatocellular carcinoma: risk factors, diagnosis and treatment. *Open Access Maced J Med Sci*. 2015;3(4):732–6.
- Unal E, Idilman IS, Akata D, Ozmen MN, Karcaaltincaba M. Microvascular invasion in hepatocellular carcinoma. *Diagn Interv Radiol*. 2016;22(2):125–32.
- Renzulli M, Buonfiglioli F, Conti F, Brocchi S, Serio I, Foschi FG, Caraceni P, Mazzella G, Verucchi G, Golfieri R, et al. Imaging features of microvascular invasion in hepatocellular carcinoma developed after direct-acting antiviral therapy in HCV-related cirrhosis. *Eur Radiol*. 2018;28(2):506–13.
- Lee S, Kim SH, Lee JE, Sinn DH, Park CK. Preoperative gadoxetic acid-enhanced MRI for predicting microvascular invasion in patients with single hepatocellular carcinoma. *J Hepatol*. 2017;67(3):526–34.
- Banerjee S, Wang DS, Kim HJ, Sirlin CB, Chan MG, Korn RL, Rutman AM, Siripongsakun S, Lu D, Imanbayev G, et al. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology*. 2015;62(3):792–800.
- Renzulli M, Brocchi S, Cucchetti A, Mazzotti F, Mosconi C, Sportoletti C, Brandi G, Pinna AD, Golfieri R. Can current preoperative imaging be used to detect microvascular invasion of hepatocellular carcinoma? *Radiology*. 2016; 279(2):432–42.
- Peng J, Zhang J, Zhang Q, Xu Y, Zhou J, Liu L. A radiomics nomogram for preoperative prediction of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma. *Diagn Interv Radiol*. 2018;24(3):121–7.
- Luo Y, Ren F, Liu Y, Shi Z, Tan Z, Xiong H, Dang Y, Chen G. Clinicopathological and prognostic significance of high Ki-67 labeling index in hepatocellular carcinoma patients: a meta-analysis. *Int J Clin Exp Med*. 2015;8(7):10235–47.
- Stroescu C, Dragnea A, Ivanov B, Pechianu C, Herlea V, Sgarbura O, Popescu A, Popescu I. Expression of p53, Bcl-2, VEGF, Ki67 and PCNA and prognostic significance in hepatocellular carcinoma. *J Gastrointest Liver Dis*. 2008;17(4):411–7.
- Shi W, Hu JF, Zhu SZ, Shen XY, Zhang XY, Yang CQ, Gao HJ, Zhang H. Expression of MTA2 and Ki-67 in hepatocellular carcinoma and their correlation with prognosis. *Int J Clin Exp Pathol*. 2015;8(10):13083–9.
- Takata H, Kudo M, Yamamoto T, Ueda J, Ishino K, Peng WX, Wada R, Taniai N, Yoshida H, Uchida E, et al. Increased expression of PDIA3 and its association with cancer cell proliferation and poor prognosis in hepatocellular carcinoma. *Oncol Lett*. 2016;12(6):4896–904.
- Xu F, Jin T, Zhu Y, Dai C. Immune checkpoint therapy in liver cancer. *J Exp Clin Cancer Res*. 2018;37(1):110.
- Shi F, Shi M, Zeng Z, Qi RZ, Liu ZW, Zhang JY, Yang YP, Tien P, Wang FS. PD-1 and PD-L1 upregulation promotes CD8(+) T-cell apoptosis and postoperative recurrence in hepatocellular carcinoma patients. *Int J Cancer*. 2011;128(4):887–96.