

Effective staging of fibrosis by the selected texture features of liver: Which one is better, CT or MR imaging?



Xuejun Zhang^{a,b,c,*}, Xin Gao^{c,*}, Brent J. Liu^b, Kevin Ma^b, Wen Yan^d, Long Liling^d, Huang Yuhong^e, Hiroshi Fujita^f

^a School of Computer and Electronic Information, Guangxi University, Nanning 530004, Guangxi, PR China

^b IPI Lab, Biomedical Engineering Department, University of Southern California, Los Angeles, CA 90033, USA

^c Department of Medical Imaging, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, Jiangsu, PR China

^d Department of Radiology, First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, PR China

^e No. 1 People's Hospital of Nanning, Nanning 530022, Guangxi, PR China

^f Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, Gifu 501-1193, Japan

ARTICLE INFO

Article history:

Received 30 October 2014

Received in revised form 22 June 2015

Accepted 2 September 2015

ABSTRACT

Purpose: Texture patterns of hepatic fibrosis are one of the important biomarkers to diagnose and classify chronic liver disease from initial to end stage on computed tomography (CT) or magnetic resonance (MR) images. Computer-aided diagnosis (CAD) of liver cirrhosis using texture features has become popular in recent research advances. To date, however, properly selecting effective texture features and image parameters is still mostly undetermined and not well-defined. In this study, different types of datasets acquired from CT and MR images are investigated to select the optimal parameters and features for the proper classification of fibrosis.

Methods: A total of 149 patients were scanned by multi-detector computed tomography (MDCT) and 218 patients were scanned using 1.5 T and 3 T superconducting MR scanners for an abdominal examination. All cases were verified by needle biopsies as the gold standard of our experiment, ranging from 0 (no fibrosis) to 5 (cirrhosis). For each case, at least four sequenced phase images are acquired by CT or MR scanners: pre-contrast, arterial, portal venous and equilibrium phase. For both imaging modalities, 15 texture features calculated from gray level co-occurrence matrix (GLCM) are extracted within an ROI in liver as one set of input vectors. Each combination of these input subsets is checked by using support vector machine (SVM) with leave-one-case-out method to differentiate fibrosis into two groups: non-cirrhosis or cirrhosis. In addition, 10 ROIs in the liver are manually selected in a disperse manner by experienced radiologist from each sequenced image and each of the 15 features are averaged across the 10 ROIs for each case to reduce the validation time. The number of input items is selected from the various combinations of 15 features, from which the accuracy rate (AR) is calculated by counting the percentage of correct answers on each combination of features aggregated to determine a liver stage score and then compared to the gold standard.

Results: According to the accuracy rate (AR) calculated from each combination, the optimal number of texture features to classify liver fibrosis degree ranges from 4 to 7, no matter which modality was utilized. The overall performance calculated by the average sum of maximum AR value of all 15 features is 66.83% in CT images, while 68.14%, and 71.98% in MR images, respectively; among the 15 texture features, mean gray value and entropy are the most commonly used features in all 3 imaging datasets. The correlation feature has the lowest AR value and was removed as an effective feature in all datasets. AR value tends to increase with the injection of contrast agency, and both CT and MR images reach the highest AR performance during the equilibrium phase.

Conclusions: Comparing the accuracy of classification with two imaging modalities, the MR images have an advantage over CT images with regards to AR performance of the 15 selected texture features, while 3 T MRI is better than 1.5 T MRI to classify liver fibrosis. Finally, the texture analysis is more effective during equilibrium phase than in any of the other phased images.

© 2015 Elsevier Ltd. All rights reserved.

* Corresponding author at: School of Computer, Electronics and Information, Guangxi University, 100, Daxue Dong Road, Nanning 530004, Guangxi, PR China. Tel.: +86 771 3231260; fax: +86 771 3232214.

E-mail addresses: xjzhang@gxu.edu.cn (X. Zhang), xingaosam@yahoo.com (X. Gao).

<http://dx.doi.org/10.1016/j.compmedimag.2015.09.003>

0895-6111/© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Patients suffering from chronic liver diseases are at a severe risk from complications such as hepatocellular carcinoma (HCC) and liver failure [1]. In the diagnosis and therapeutic assessment of cirrhosis of the liver and chronic hepatitis, it is necessary to stage the degree of hepatic fibrosis as an important indicator of cirrhosis as well as a critical predictive factor for the occurrence of HCC [2], which is one of the most common malignancies in patients that are affected by these diseases [3]. Fibrosis is caused by excessive deposition of extracellular matrix owing to histological and molecular reshuffling of various components such as collagens, glycoproteins, proteoglycans, and other macromolecules within the extracellular matrix. These features, common to almost all patients with chronic liver disease, lead to the changes in the hepatic morphology, texture pattern, and degree of liver stiffness. Accurate assessment of hepatic fibrosis is crucial for the determination of the appropriate treatment because fibrosis is potentially a reversible process in the early stages.

Liver biopsy, which is used for histological scoring and is still used as a reference test for fibrosis staging, is considered the gold-standard method for the assessment of the degree of fibrosis. However, liver biopsy is an invasive procedure with possible side effects such as pain in 30–40% of the cases, pneumothorax (3%) or even death (2/10,000) [4,5]. To decrease the need for painful biopsies, non-invasive methods using MR and ultrasound imaging have been proposed to obtain images of the liver. With the development of high-speed imaging devices, highly precise medical imaging services are now widely available. Recently, radiological assessments of hepatic fibrosis by magnetic resonance elastography (MRE) [6–8], gadolinium- or superparamagnetic iron oxide-enhanced MR imaging [9,10], diffusion-weighted MR imaging [11], and real-time ultrasonographic elastography [12] have been reported, and their feasibility, usefulness, and limitations have been shown. Wang et al. [13] have reported on real-time elastography with a new quantitative technology for diffusing histological lesions as a new and promising sonography-based non-invasive method for assessing liver fibrosis in patients with chronic hepatitis B. These new imaging technologies have greatly impacted the traditional diagnostic methods. However, the interpretation of the numerous diseases from different types of medical images is not an easy task, especially for inexperienced residents or general radiologists. In the last decade, different types of computer-aided detection/diagnosis (CAD) systems have been developed to ease the workload of radiologists. There are several promising reports on the CAD of liver fibrosis on MRI, which have been obtained by analyzing morphology changes [14,15] and the texture pattern of fibrosis [16] through images.

Texture patterns of hepatic fibrosis are one of the important features to diagnose the chronic liver disease from initial to end stage on computed tomography (CT) or magnetic resonance (MR) images. However, quantification and classification of hepatic fibrosis with only texture patterns in liver is an extremely difficult task for both the radiologist and computer, resulting in a low accuracy rate of diagnosis. However, subtle differences of tiny structures of fibrosis within a small size of ROI provide the computer with an advantage to the reader as compared with the clinical interpretation using other image features such as shape, volume, surface irregularity, etc. Some work has been developed utilizing intelligent algorithms to stage the fibrosis by feeding the features calculated from 1st and 2nd order texture matrixes into a classifier, and stating their promising results on different kinds of datasets. Gobert et al. [17] reported the use of six features extracted from texture features and two statistical measures as input to an ANN classifier with genetic algorithm, resulting in 73% correct classification rate for the characterization of cirrhotic and non-cirrhotic

in MRI images. Li et al. [18] found that integrating the texture features into shape features with ANN classification of cirrhosis may improve the overall performance of cirrhosis detection. However, these studies used samples of a normal and a cirrhotic case and the selected features were based on experience. Other improved classification methods state their effective findings of parameters [19], but did not provide a “real true” selection of features that should be recommended to doctors as their investigation always works on individual sequenced images or modality. In addition, the optimal number and effectiveness of input vectors have not been addressed or identified. These may be due to the limitation of case numbers and large time consumption needed to optimize each combination of features. Although some efficient training methods were proposed to reduce the time in training and testing procedures, their effects are difficult to compare and evaluate as there was no evaluation against the gold standard until now.

China has a high incidence rate of hepatic cancer along the coastal areas, especially in Fushui county of the Guangxi autonomous region, where high liver cancer incidence rate of 56.45/100,000 in average has been reported and for male 91.11/100,000 and female 21.80/100,000, respectively [20]. Therefore, collecting a large amount of fibrotic cases is feasible in our local hospitals. Several hundred patients with different grade distribution of fibrosis were scanned in at least four different sequenced phases using various CT and MRI modalities. In order to find out the “real true” features that impact the classification of texture patterns of hepatic fibrosis utilized in computer-aided diagnosis, we chose the most accurate validation method of leave-one-case-out to check each combination of features. Considering that $2^{15} - 1$ different combinations are obtained in a single phased image, the iteration is large and time consuming which would estimate the overall time length to perform the experiment at over 3 years. The accuracy rates outputted from different modalities, sequenced phased images and texture features could not only tell us the “true” of optimal set of fibrotic texture patterns guiding the radiologists to analyze fibrosis efficiently from numerous medical images, but also help to clarify how many and what kind of texture features should be utilized for CAD [21].

This study is focused on the effective selection of texture features for fibrosis classification. We utilized 15 texture features calculated from all of the CT datasets with slice thickness from 0.625 mm to 5 mm, and over four different sequenced CT or MR phased images are investigated. Each combination of features is evaluated using support-vector-machine (SVM) with leave-one-case-out method to select the optimal feature subsets according to their performance. In this paper, Section 2 describes the acquirement of experimental datasets by CT and MR modalities with their imaging protocols. Section 3 describes the methods of the ROI selection, feature extraction and optimization by SVM and leave-one-case-out approach. In Section 4, classification results from the different imaging modalities are demonstrated and discussed. The accuracy rate of each combination of feature subsets is compared to find out the most informative features as well as the optimal number of texture features from the experimental results. The paper concludes in Section 6 with further discussion of the proposed approach and future works.

2. Experimental materials

To evaluate and compare the contribution of texture features derived from CT and MR images towards the diagnosis of liver fibrosis, a total of 367 patients with or without hepatic fibrosis are scanned by 3 different modalities from two different hospitals, including 149 cases from CT and 218 cases from MR.

Table 1
Patients statistic with five fibrous stages according to VHPTA system [22].

Stage of fibrosis	Score	Description of hepatic fibrosis degree	Number of MRI cases (hospital no. 1)	Number of MRI cases (hospital no. 2)	Number of CT cases (hospital no. 1)	Groups
Normal	0	No fibrosis	40	5	36	Non-cirrhosis (Group 0)
Mild fibrosis	1	Fibrous expansion of some portal areas, with or without short fibrous septae	30	4	26	
	2	Fibrous expansion of most areas and bridging fibrosis appeared	23	10	13	Cirrhosis (Group 1)
Severe fibrosis	3	Most fibrous septum and lobular structure disorder	18	8	19	
	4	Early cirrhosis, diffuse fiber hyperplasia	22	16	19	
Cirrhosis	5	Cirrhosis	40	2	36	

2.1. Computed tomography (CT) image

A total of 149 patients were scanned by CT (GE Lightspeed VCT) from June 2009 to March 2012 at the Department of Radiology, First Affiliated Hospital of Guangxi Medical University (Hospital no.1). The data set consists of 36 normal cases, 39 mild fibrosis cases, 38 severe fibrosis cases and 36 typical cirrhosis cases. The imaging protocol is as follows: quad-phase scans are made at 120 kV tube voltage, 250 mA tube current; the image size is 512×512 pixel. The arterial, portal venous and equilibrium phases are acquired at 25, 60, and 120 s after contrast injection, respectively. The injection speed is 3.0 ml/s, concentration is 320 mg/ml, dosage is 85–90 ml. For each case, five different CT datasets with different slice thickness are acquired with interval of 0.625 mm, 1.25 mm, 2.5 mm, 3.75 mm and 5 mm, respectively. For a case with one thickness, there are four sequenced phase images: pre-contrast, arterial, portal venous and equilibrium phase. Therefore, a total of 20 sequence/series types were acquired in one CT case. All cases have been verified by needle biopsies as the gold standard of our experiment. Surgical specimens were retrospectively examined by a pathologist who was blinded to patient histories and radiology and surgery reports. The patients' fibrosis stages are evaluated in accordance with the Chinese Viral Hepatitis Prevention and Treatment Plan (VHPTA) [22], ranging from 0 (no fibrosis) to 5 (cirrhosis). In this paper, the entire dataset is divided into two groups: non-cirrhosis (containing normal cases S0 and mild liver fibrosis cases S1, S2); and cirrhosis (containing severe liver fibrosis cases S3, S4 and typical cirrhosis cases S5), as illustrated in Table 1.

2.2. Magnetic resonance (MR) image

From February 2011 to March 2012, a total of 173 patients had an abdominal examination performed using a 3-T superconducting MR scanner (Intera Achieva Quasar Dual; Philips Medical Systems, Netherlands) with a six-channel torso array coil at Hospital no. 1. The imaging protocols used include TR/TE, 2.3/82 ms in plan scanning phase; 7 mm slice thickness and 1 mm inter-slice gap; interpolated imaging matrix, 512×512 ; slice thickness; 4 mm and 2 mm inter-slice gap. In dynamic contrast scanning phase: 7 mm slice thickness and 1 mm inter-slice gap; contrast injection, Gd-DTPA; the dosage, 0.2 mmol/kg; injection speed, 2–3 ml/s. Since contrast injection after 16 s, 50–60 s and 2–3 min arterial, portal and equilibrium scanning are performed, respectively. Among the datasets, there are 18 patients without liver disease history and hepatic dysfunction, and the fibrosis stage of other patients is confirmed by the result of the liver biopsy procedure.

An additional 45 MR cases were acquired from the No.1 People's Hospital of Nanning (Hospital No. 2) in China using a 1.5 T (Signa Horizon, GE Healthcare) MR scanner and were included in our experiment. The gadolinium-enhanced T1-weighted images

(150/1.6; matrix, 512×512 ; flip angle, 90° ; signal acquisition, 1; slices, 18 per 26-second acquisition time) sequences are with section thickness 8 mm with a 2 mm intersection. Table 1 gives the detailed distribution of the 3 datasets. This study was approved by the institutional review board at both hospitals and informed consent was obtained from all patients.

3. Methods

In this section, we give a brief overview of the gray level co-occurrence matrix (GLCM), Support Vector Machine (SVM) and Leave-one-out cross-validation methods, including the guideline of ROIs selection and pre-processing procedures for feature extraction. We then describe a texture features optimizing method for the analysis of the hepatic fibrosis on CT/MR images.

3.1. Selecting ROIs for feature extraction

We have been developing a CAD system named *CirrhoView* for cirrhosis clinical research. Fig. 1 is the graphical user interface (GUI) of our software that enables the radiologists to open DICOM files on a server; manually select ROIs by moving and clicking the mouse button; automatically calculating texture features within the ROI and store to an excel file for further analysis. In this experiment, 10 ROIs in liver were manually picked in a disperse fashion from each sequenced image by L. Long, a diagnostic radiologist with over 20 years' experience of radiological images interpretation. The ROIs are shown as red squares in Fig. 1. Selecting ROIs relies on the guidelines illustrated in Table 2: 1 ROI selected in left medial segment, right anterior lateral segment and right posterior lateral segment, respectively, at the second porta hepatis; 2 ROIs selected in right anterior lateral segment and right posterior lateral segment, respectively, at the first porta hepatis; 1 ROI selected in right anterior lateral segment and right posterior lateral segment, respectively, below the first porta hepatis. Considering the diffuse distribution of liver, large blood vessels and focal liver lesions within the liver are excluded. Three sizes of the ROIs (32×32 , 20×20 , 16×16 pixel) are generated individually according to the

Table 2
Guideline of selecting ROIs relied on the location of liver.

Number of ROIs	Position of ROIs in hepar	Hepatis partition
2	Right anterior lateral segment	First porta hepatis
2	Right posterior lateral segment	
1	Right anterior lateral segment	Below the first porta hepatis
1	Right posterior lateral segment	
1	left medial segment	Second porta hepatis
1	Right anterior lateral segment	
1	Right posterior lateral segment	hepatis
1	Left lateral segment	

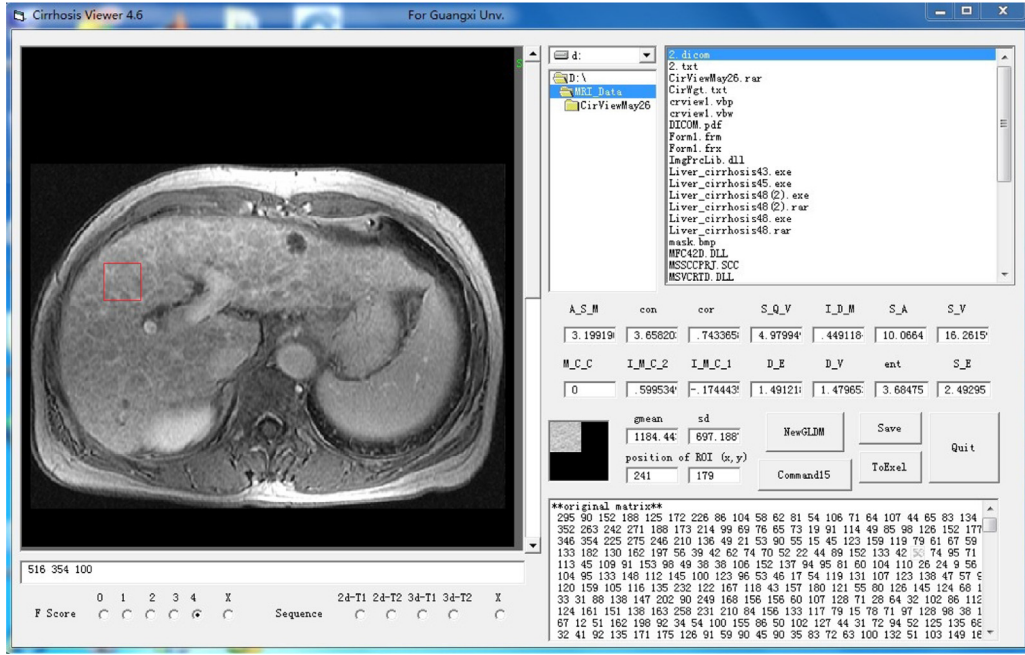


Fig. 1. The graphical user interface of our software CirrhoView, designed for selecting ROIs by radiologists and automatic feature extraction by CAD system. The red square of ROI indicates a texture pattern of liver cirrhosis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

start point of a square, and all ROIs are pre-processed with or without Sobel filter. Thus we have 10 ROIs with 3 sizes, processed by 2 different filtering procedures, and in 4 phases for calculating features in a single sequenced CT/MR image. Considering there are 5 available slice thicknesses in one CT case, the total number of ROIs reaches to 1200 as illustrated in Fig. 2. Due to the technical difficulty and the focus on this study, only one slice thickness is used in MRI.

3.2. Gray level co-occurrence matrix and texture features

The texture features are obtained by gray level co-occurrence matrix (GLCM) method. As the DICOM images have 12-bit gray levels, which is too large for establishing co-occurrence matrices, the gray levels of an image are reduced to 4-bit, which is accurate

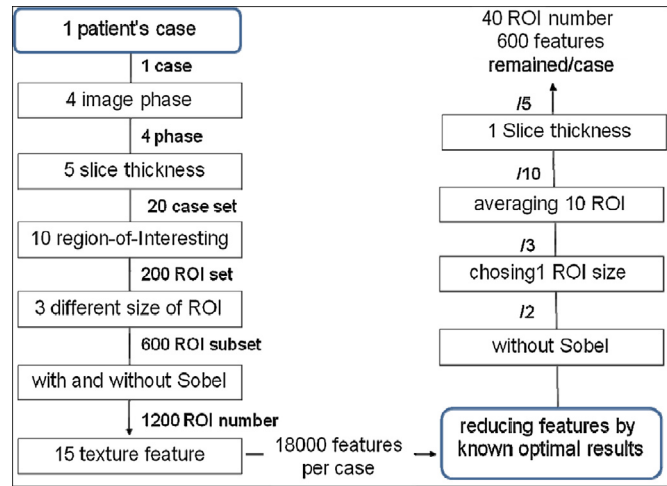


Fig. 2. The evolution of number of features in one CT case demonstrates the input features generated for initial optimization (left) and reduced by some previous investigations (right).

enough for this study [17], and the matrix element in row i and column j are denoted:

$$P(i, j, \delta, \theta) = \{ [(x, y), (x + \Delta x, y + \Delta y)] \} \quad (3-1)$$

$$f(x, y) = i, \quad f(x + \Delta x, y + \Delta y) = j; \quad x, y = 0, 1, \dots, N - 1$$

where $i, j = 16; \delta = 1; \theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$.

Assume that P is the matrix of GLCM extracted from ROI, G is the gradation order of GLCM, μ is the mean value of the matrix,, $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the means and standard deviations of P_x and P_y , respectively. $P_x(i)$ is the sum of row i of the matrix P , while $P_y(j)$ is the sum of column j of P .

$$P_x(i) = \sum_{j=0}^{G-1} P(i, j)$$

$$P_y(j) = \sum_{i=0}^{G-1} P(i, j)$$

$$\mu_x = \sum_{i=0}^{G-1} i \sum_{j=0}^{G-1} P(i, j) = \sum_{j=0}^{G-1} i P_x(i)$$

$$\mu_y = \sum_{i=0}^{G-1} j \sum_{j=0}^{G-1} P(i, j) = \sum_{j=0}^{G-1} j P_y(j)$$

$$\sigma_x^2 = \sum_{i=0}^{G-1} (i - \mu_x)^2 \sum_{i=0}^{G-1} P(i, j) = \sum_{i=0}^{G-1} (P_x(i) - \mu_x(i))^2$$

$$\sigma_y^2 = \sum_{j=0}^{G-1} (j - \mu_y)^2 \sum_{i=0}^{G-1} P(i, j) = \sum_{j=0}^{G-1} (P_y(j) - \mu_y(j))^2$$

Thirteen texture features are calculated by the method introduced by Haralick [23] based on the co-occurrence matrix:

Angular second moment (ASM):

$$ASM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\}^2 \quad (3-2)$$

Contrast (CNT):

$$CNT = \sum_{n=0}^{G-1} n^2 P_{x-y}(n) \quad (3-3)$$

where

$$P_{x-y}(k) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \quad |i - j| = k \quad k = 0, 1, 2, \dots, G - 1$$

Correlation (COR):

$$COR = \frac{\sum_{i=1}^G \sum_{j=1}^G \{i \times j\} \times P(i, j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \quad (3-4)$$

Entropy (ENT):

$$ENT = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\} \log(P(i, j)) \quad (3-5)$$

Sum of squares: Variance (SQV):

$$SQV = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\} \log(P(i, j)) \quad (3-6)$$

Inverse Difference Moment (IDM):

$$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i + j)^2} P(i, j) \quad (3-7)$$

Sum Average (SUM_{AVE}):

$$SUM_{AVE} = \sum_{i=0}^{2G-2} iP_{x+y}(i) \quad (3-8)$$

where

$$P_{x+y}(k) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \quad i + j = k, \quad k = 0, 1, 2, \dots, 2(G - 1)$$

Sum Variance (SUM_{VAR}):

$$SUM_{VAR} = \sum_{i=0}^{2G-2} (i - SUM_{AVE})^2 P_{x+y}(i) \quad (3-9)$$

Sum Entropy (SUM_{ENT}):

$$SUM_{ENT} = \sum_{i=0}^{2G-2} P_{x+y}(i) \log(P_{x+y}(i)) \quad (3-10)$$

Difference Variance (DIF_{AVE}):

$$DIF_{AVE} = - \sum_{l=0}^{G-1} \left\{ i - \sum_{l=0}^{G-1} iP_{x+y}(i) \right\}^2 P_{x-y}(i) \quad (3-11)$$

Difference Entropy (DIF_{ENT}):

$$DIF_{ENT} = - \sum_{i=0}^{G-1} P_{x-y}(i) \log(P_{x-y}(i)) \quad (3-12)$$

Information Measures1 (IMC1):

$$IMCI = \frac{ENT - HXY1}{\max(ENT_x, ENT_y)} \quad (3-13)$$

where

$$ENT_x = - \sum_{i=0}^{G-1} P_x(i) \log\{P_x(i)\}$$

$$ENT_y = - \sum_{j=0}^{G-1} P_y(j) \log\{P_y(j)\}$$

$$HXY1 = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \log\{P_x(i) P_y(j)\}$$

$$HXY2 = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_x(i) P_y(j) \log\{P_x(i) P_y(j)\}$$

Information Measures2 (IMC2):

$$IMC2 = \sqrt{1 - \exp\{-2.0(HXY2 - ENT)\}} \quad (3-14)$$

In our experiment, two features based on first order statistics of the image: (14). Mean gray value (MGV) and (15). Standard deviation (SD) are also used in this study, thus there are total 15 texture features used in a ROI.

$$MGV = \frac{1}{N^2} \sum_{x=1}^N \sum_{y=1}^N f(x, y) \quad (3-15)$$

$$SD = \sqrt{\frac{1}{N} \sum_{x=1}^N \sum_{y=1}^N \{f(x, y) - MGV\}^2} \quad (3-16)$$

3.3. Support vector machine (SVM)

SVM originated on the basis of statistics by Vapnik et al. [24] is a current general learning method. The discrimination function for linear separable problem is:

$$f(X) = \sum_{i=1}^N y_i a_i * k(X_i, X) + b^* \quad (3-17)$$

where N is the number of support vector, X_i, y_i are the labels of corresponding support vectors, and a_i^* and b^* are parameter learned from training samples. The kernel $k(X_i, X)$ is significant for determining the behavior of the classifier. In this study, the Radial Basis Function (RBF) kernel function is used:

$$k(X, Y) = \exp\left\{-\frac{|X - Y|^2}{2\sigma^2}\right\} \quad (3-18)$$

The $|x-y|$ represents the distance between two vectors, where σ is a constant 1. The input vectors to SVM are 15 texture features in this study while the output is probability for the presence of cirrhosis.

3.3.1. Leave-one-out cross-validation (LOOCV) method

The practical goal of using the SVM model would be to determine which subset of the 15 texture features should be used to produce the best predictive model. For most modelling procedures, if we compare feature subsets using the in-sample error rates, the best performance will occur when all 15 features are used. However

under cross-validation, the model with the best fit will generally include only a subset of the features that are deemed truly informative.

The Leave-one-out (LOO) [25] method, as the name suggests, is to take out a single observation from the original datasets containing M samples as the validation data, and the remaining $M - 1$ samples as the training data to build a classification model of SVM. After validating this model, the sample will be moved back into the training data set and another sample is selected. The procedure is repeated such that each observation in the sample is used once as the validation data. After looping a total of M times of training and testing on M samples, all of the cases in one dataset are validated by the SVM model. In this study, the number of samples M is selected as 50, 50, 18, corresponding to CT set, 3.0T MR and 1.5T MR sets.

Leave-one-out cross-validation is typically used in the analysis of very small datasets. Although usually the LOOCV method is expensive to train since such training must be carried out repeatedly, it is still the most robust and accurate way to evaluate a predictive model. In some studies, a few modified methods such as the least squares and kernel regression, cross-validation can be sped up significantly by pre-computing certain values that are needed repeatedly in the training, or by using fast updating rules. However, careful attention is needed to preserve the “total blinding” of the validation set from the training procedure, otherwise bias may result. In order to develop the modelling procedures accurately, we choose the traditional LOOCV method in our experiment.

Among the testing results, the ratio between the number of correctly classified cases as cirrhosis and total number of cirrhosis cases is defined as true positive (TP) while the ratio between the number of correctly classified cases as non-cirrhosis and total non-cirrhosis cases is defined as true negative (TN). The average of TP and TN is defined as the accuracy rate (AR).

3.4. Calculation of texture features with different conditions

Texture features will present different values if calculated from varying pre-processing methods, image quality or ROI sizes. Therefore, we should optimize each of these conditions before analysing and optimizing feature subsets.

First, the accuracy rate detected w/o a Sobel Filter is tested in a sequenced dataset to determine the best input pattern of SVM, and half of the datasets will be excluded with this result [18]. Second, for CT image sets, accuracy rate calculated from 5 different slice thicknesses are obtained by testing the samples from portal venous phased images. Only one slice thicknesses with highest rank will be kept in further study and other data is excluded [19]. This step is skipped in MR image sets with single slice thickness. Third, one optimal size is chosen among the three different ROI sizes [26].

Since the numbers of ROIs selected are too large, 15×10 texture features in one sequenced case are reduced by averaging the features value from 10 selected ROIs, respectively, and turned into a new dataset with only 15 texture features. After all of the above processing steps, the total number of ROIs in a case is reduced from 1200 to 40 as shown in Fig. 2. This step may significantly decrease the number of input samples to ensure SVM efficient training.

3.5. Optimizing the features

Besides the above optimizing problem, there remains the challenge to optimize the 15 texture features, e.g. how many and which ones are the best features to represent fibrosis? In this paper, the data samples are divided into two categories in the SVM classification model: the mild liver fibrosis cases S0, S1, S2 as the negative samples group 0, while S3, S4, S5 severe liver fibrosis cases and typical cirrhosis cases as the positive samples group 1. Among sequenced datasets, each possible combination of 15 features is

performed by the SVM model with LOOCV method to compare their accuracy. The method is executed as the following steps in detail:

- Step 1: Non-cirrhosis and cirrhosis groups are read into Group 0 and Group 1, respectively, containing M samples in each group. 15 features are extracted by GLCM and averaged over 10 chosen ROIs. ($2M \times 15$ features as input candidates).
- Step 2: The number of input items n is selected from the combinations of 15 features. ($2^{15} - 1$ different combinations obtained, where $n \in [1, 15]$).
- Step 3: According to the selected n items in a subset from Step2, $2M \times n$ input features are picked up from Group 0 and Group 1, respectively. A single sample with n features is chosen as the validation data, and the remaining $2M - 1$ samples are used as the training data.
- Step 4: SVM classifier is modelled with $2M - 1$ training samples and validates the remained test sample k . If k belongs to Group 0 and is correctly classified, then $TN = TN + 1$; likewise, if Group 1 is correctly classified, then $TP = TP + 1$.
- Step 5: If there are undetermined samples within $2M$ then run Step 3; if not, calculate the accuracy rate (AR): $AR0 = TN/M$, $AR1 = TP/M$, and $AR = (AR0 + AR1)/2$.
- Step 6: If the ergodic process does not finish, then run Step 2, otherwise, run Step 7.
- Step 7: Complete. $2^{15} - 1 = 32,767$ numbers of AR are produced for further processing.

3.6. Evaluation of the performance of texture features

Each sequenced datasets has 32,767 numbers of AR values that can be divided into 15 groups $[AR^1, AR^2, \dots, AR^i, \dots, AR^{15}]$ according to the numbers of selected items in a subset. Each AR^i also has numbers of subsets by the combination of i items out of 15 features, among which a maximum value of AR^i is defined as $Max AR^i$. Since the CT and MR datasets contain 4 phased images, the maximum value of AR^i in a certain j th sequenced datasets is defined as $Max AR_j^i$. In addition, an average sum of maximum value in a whole m sequenced datasets acquired by one modality k can be calculated as

$$AR_k = \left[\frac{1}{m} \sum_{j=1}^m MaxAR_j^1, \frac{1}{m} \sum_{j=1}^m MaxAR_j^2, \dots, \frac{1}{m} \sum_{j=1}^m MaxAR_j^i, \dots, \frac{1}{m} \sum_{j=1}^m MaxAR_j^n \right] \quad (3-19)$$

where $k = 1$:CT, 2: MRI 3T, 3: MRI 1.5T.

4. Results and discussion

The grade accuracy rate varies with different imaging conditions and combinations of different number of texture features. In our previous study [18], the performance of classification of fibrosis on MRI w/o Sobel filter was investigated. The result showed a best variation accuracy of 0.76 without Sobel pre-processing vs. 0.64 with filter. Comparing with the high score of using raw data directly, Sobel filter is insignificant in pre-processing the MR or CT images. This may be due to the fact that the Sobel filter not only enhances the fibrosis pattern but also the background noises. We also concluded that different sizes of ROI may impact the classification, and 20×20 is the optimal size among the three ROI sizes [26]. Therefore, this experiment was performed by using 20×20 ROI raw data. For the CT datasets, the optimal size of CT image slice thickness is 1.25 mm, which demonstrates the fibrosis information

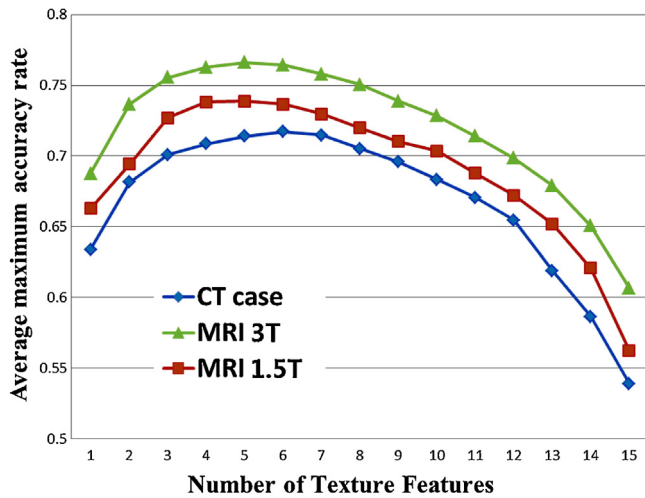


Fig. 3. Accuracy Curve by selection of number of features indicates the different performance on 3 different modalities. Where ▲ MR images from Hospital no. 1; ■ MR images from Hospital no. 2; and ◆ CT images from Hospital no. 1.

in this interval is the most adequate on CT image. Likewise, the 20 × 20 ROI raw data with 1.25 mm interval is the default setting for this experiment on CT datasets.

4.1. Selecting optimal number of features

In order to determine a subset of features that are deemed truly informative, LOOCV is performed on each of the sequenced imaging datasets. In this modelling procedure, every combination of 15 features is generated to provide the best AR from 1 to 15 feature subsets in Table 3. As shown in Fig. 3, the optimal number of texture features to classify liver fibrosis degree is from 4 to 7, no matter what modalities are used. The best testing average sum of maximum value of AR are 71.72% and 76.61%, obtained by 6 features from CT and MRI (3.0T) images at hospital no. 1, respectively, and 73.22% in case of 5 features from MRI (1.5 T) at hospital no. 2. The overall performance calculated by the average sum of the maximum value of all 15 features is 66.83% in CT images, with 68.14% (1.5 T), and 71.98% (3 T) in MR images, respectively. It is obvious that the AR of MR images is better than that of CT images by comparing the three curves in Fig. 3. The results also imply that 3 T MR has better image quality for classifying fibrosis than that of 1.5 T, as both average sum of maximum value of AR's in MR are higher than CT. We should notice that the fibrotic texture is better observed on MR than on CT images in clinical practice and the results reflect this trend. In addition, the lowest average sum of maximum value of AR indicated in Table 3 occurs when using all of the 15 features together to build the predictive SVM model, which implies the importance of selecting informative features.

4.2. Distribution of informative features

From the results of the average sum of maximum values of AR, we have concluded that the number of feature subset ranging from 5 to 7 features has better performance than other numbers. Another interesting point to investigate is what these 5 to 7 features actually are. A threshold value of AR value is set to 0.7 so as to pick the subset features having excellent performance. If an AR value is >0.7, then the number of each of the items in this AR is plus 1. Based on this statistics, a discriminative power histogram of every feature contributed to the classification is generated in Tables 4a and 4b and shown in Fig. 4 after normalizing the scale into a range of [0–1].

Table 3
AR values of MRI and CT by using different number of features as input vectors of SVM.

Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
MRI3T	0.687	0.737	0.755	0.763	0.766	0.764	0.758	0.750	0.739	0.728	0.714	0.699	0.679	0.651	0.607
MRI1.5T	0.663	0.694	0.727	0.738	0.739	0.737	0.730	0.720	0.710	0.704	0.688	0.672	0.652	0.621	0.562
CT	0.634	0.682	0.701	0.709	0.714	0.717	0.715	0.705	0.696	0.683	0.670	0.655	0.619	0.587	0.539

Table 4a
Value of feature importance in MRI and CT modalities (1).

Feature	ASM	CNT	COR	SQV	IDM	SUM _{AVE}	SUM _{VAR}	SUM _{ENT}
MRI3T	0.644	0.034	0.000	0.070	0.647	0.736	0.006	0.484
MRI 1.5T	0.213	0.568	0.000	0.491	0.476	0.315	0.502	0.698
CT	0.164	0.375	0.000	0.347	0.474	0.147	0.253	0.243

Table 4b
Value of feature importance in MRI and CT modalities (2).

Feature	ENT	DIF _{AVE}	DIF _{ENT}	IMC1	IMC2	MGV	SD
MRI 3T	0.581	0.026	0.061	0.452	0.007	1.000	0.336
MRI 1.5T	0.596	0.225	0.253	0.227	0.282	0.614	1.000
CT	0.386	0.399	0.330	0.284	0.163	1.000	0.475

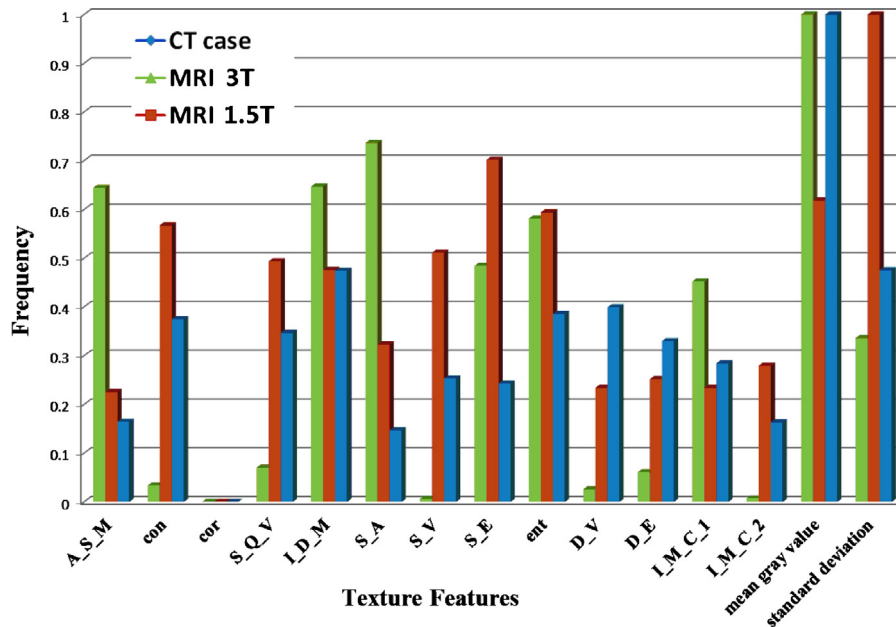


Fig. 4. From the discriminative power histogram for all of the 15 characteristics extracted from 3 different datasets, the ability of each feature in the differentiation of fibrous texture is investigated by counting the frequency of high scores appeared in AR values.

Fig. 5 indicates the top 7 texture features extracted in CT images and MR images in the discrimination power histogram. The red circle contains the top 7 features extracted in CT images: mean gray value (MGV), standard deviation (SD), inverse difference moment (IDM), difference variance (DV), entropy (ENT), contrast (CON), and

sum of squares variance (SQV). The blue circle contains the top 7 features extracted in MR images from hospital no. 2: SD, sum entropy (SE), MG, ENT, CON, sum average (SV), SQV. On the other hand, the green circle contains the top 7 features extracted in MR images from hospital no. 2: MG, sum average (SA), IDM, angular second moment (ASM), ENT, SE, and information measures of correlation1 (IMC1). It should be noted that among the 15 texture features, mean gray value and entropy are the most commonly included as high performing in all 3 imaging datasets. With the development of fibrosis, more and more fibrous pattern appear in liver region. This phenomenon significantly changes the intensity and the complexity of an image/ROI, which corresponds to the features of MGV and ENT, respectively. Other features such as standard deviation, contrast, sum of squares variance, inverse difference moment, and SE are selected by 2 datasets. Correlation (COR) has the lowest AR value and is removed the feature set in all image datasets. Selecting features based on this ranking will be more efficient in future studies.

4.3. Performance of different phased images

AR values tend to increase with the injection of contrast agency, and both CT and MR images reach highest performance in the equilibrium phase as shown in Fig. 6. MR can demonstrate fibrotic

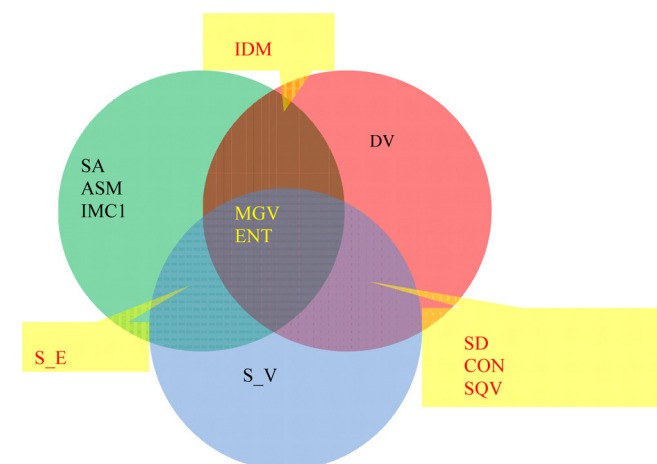


Fig. 5. The distribution of features used in 3 datasets with top ranking 7 from Fig. 3.

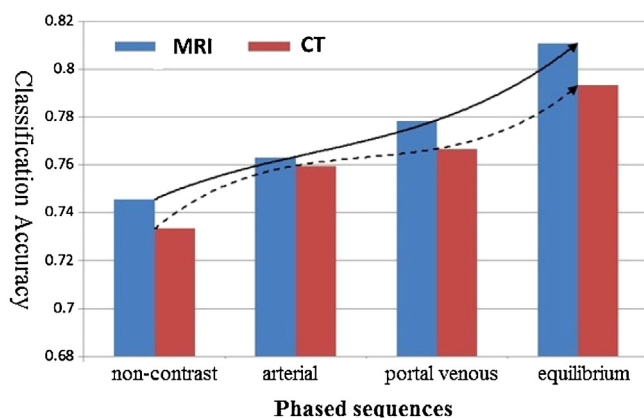


Fig. 6. The AR value in 4 phased images show that the texture analysis is effective in equilibrium phase than in other phased images.

texture efficiently and the equilibrium phase image is recommended as a main tool for interpretation of cirrhosis.

Exhaustive validation processing by LOO with SVM method is very time consuming because the number of combination of 15 texture features is very large. In our experiment, a 40 group 0+40 group 1 dataset in one sequenced image dataset utilizing the loop procedure of training and testing, for example, takes an average of 6 to 9 h on a workstation (4.3G CPU with 4G RAM, Win7 64bit version). However, once the modeling is completed, we can use this predictive model to diagnose a new dataset within a second. Although some studies announced their effective methods for training samples with improved algorithms that may greatly reduce the validation processing time, the robustness and accuracy still has problems. Furthermore, it is hard to determine a gold standard to verify their results. Therefore, in this study we chose the most critical validation method of LOOCV to ensure the objectivity of our results.

It will be a great challenge to include a same procedure of human's observation experiment to check if its conclusion is identical with CAD's. One of our former studies [27] demonstrated some preliminary results on the comparison of the human's observation with computer-aided diagnosis. Although radiologists' performance was significantly lower than that of the computer algorithm with only the texture pattern of fibrosis, the trend of accuracy changing is the same on both sides.

There are some limitations to our study, and these should be improved upon in our future work. First, the datasets are only divided into two groups. This is due to the unbalanced distribution of each of the fibrous stages and the purpose of decreasing the complexity of the validation. In the next step, our studies will extend to the quantitative assessment of fibrosis into 5 stages with the increasing datasets. Second, this study does not combine images from multiple phases due to the computational complexity. Top features from each phase could be combined to get up to 60 (15×4) features per case, and the performance of such a future work will tell us the variable importance from all phases. Third, this study was conducted as preliminary research, and considerable time and effort was expended to manually define the ROIs and record the values for validation. We have now started to develop an automated algorithm to select ROI within the extracted liver region, as well as integrate the GLCM, LOOCV and SVM method into one program. Finally, the overall AR's classified by texture analysis are not high (less than 0.8) enough to be used as an individual model to stage fibrosis in clinical practice. This may due the fact that the fibrous pattern is not obviously different from different stages, and even the experienced radiologist cannot interpret fibrosis well only with texture on images. It is possible to differentiate individual fibrosis

stages by using texture features with combination with other effective features such as shape, volume, elasticity, and so on [28].

5. Conclusion

This study demonstrated that SVM and leave-one-case-out method are effective in feature selection from a large number of datasets. The iteration test on all of the subsets indicates that 4 to 7 features are the optimal number of features to classify liver fibrosis into two groups. Comparing the accuracy of classification on two modalities, we have shown that MR images have an advantage over CT images, while 3 TMR is better than 1.5 TMR to classify liver fibrosis, and mean gray value and entropy are the most useful features. The texture analysis is more effective in equilibrium phase than in other phased images and thus is recommended as a main tool for interpretation of cirrhosis.

Acknowledgments

This work was supported in part by two research support from the National Natural Science Foundation of China (Nos. 61262027 & 45627390), and in part by a financial support from China Scholarship Council (201308455018), and by JSPS Grant-in-Aid for Scientific Research on Innovative Areas (grant number 26108005). The authors are grateful to Yufan Zeng, Hiroshi Kato and members at Fujita lab in Gifu University, for their early efforts and discussions on this research.

References

- [1] El-Serag HB, Mason AC. Rising incidence of hepatocellular carcinoma in the United States. *N Engl J Med* 1999;340(March):745–50.
- [2] Yatsushashi H, Yano M. Natural history of chronic hepatitis C. *J Gastroenterol Hepatol* 2000;15(May (Suppl. s2)):111–6.
- [3] Wynn TA. Cellular and molecular mechanisms of fibrosis. *J Pathol* 2008;214(Jan (2)):199–210.
- [4] Saadeh S, Cammell G, Carey WD, Younossi Z, Barnes D, Easley K. The role of liver biopsy in chronic hepatitis C. *Hepatology* 2001;33(1):196–200.
- [5] Poynard T, Ratziu V, Bedossa P. Appropriateness of liver biopsy. *Can J Gastroenterol* 2000;14(6):543–8.
- [6] Muthupillai R, Lomas DJ, Rossman PJ, Greenleaf JF, Manduca A, Ehman RL. Magnetic resonance elastography by direct visualization of propagating acoustic strain waves. *Science* 1995;269(Sep (5232)):1854–7.
- [7] Rouvière O, Yin M, Dresner MA, et al. MR elastography of the liver: preliminary results. *Radiology* 2006;240(Aug (2)):440–8.
- [8] Huwart L, Sempoux C, Salameh N, Jamart J, Annet L, Sinkus R, et al. Liver fibrosis: noninvasive assessment with MR elastography versus aspartate aminotransferase-to-platelet ratio index. *Radiology* 2007;245(Nov):458–66.
- [9] Lucidarme O, Baleston F, Cadi M, Bellin MF, Charlotte F, Ratziu V, et al. Non-invasive detection of liver fibrosis: Is superparamagnetic iron oxide particle-enhanced MR imaging a contributive technique? *Eur Radiol* 2002;13(Sep (3)):467–74.
- [10] Aguirre DA, Behling CA, Alpert E, Hassanein TI, Sirlin CB. Liver fibrosis: noninvasive diagnosis with double contrast material-enhanced MR imaging. *Radiology* 2006;239(May):425–37.
- [11] Taouli B, Tolia AJ, Losada M, Babb JS, Chan ES, Bannan MA, et al. Diffusion-weighted MRI for quantification of liver fibrosis: preliminary experience. *AJR* 2007;189(Oct (4)):799–806.
- [12] Friedrich-Rust M, Ong MF, Herrmann E, Dries V, Samaras P, Zeuzem S, et al. Real-time elastography for noninvasive assessment of liver fibrosis in chronic viral hepatitis. *AJR* 2007;188(3):758–64.
- [13] Wang J, Guo L, Shi XY, Pan WQ, Bai YF, Ai H. Real-time elastography with a novel quantitative technology for assessment of liver fibrosis in chronic hepatitis B. *Eur J Radiol* 2012;81(1):31–6.
- [14] Zhang XJ, Li WG, Fujita H, Kanematsu M, Hara T, Zhou XR, et al. Automatic segmentation of hepatic tissue and 3D volume analysis of cirrhosis in multi-detector row CT scans and MR imaging. *IEICE Trans Inf Syst* 2004;E87-D(8):2138–47.
- [15] Goshima S, Kanematsu M, Kobayashi T, Furukawa T, Zhang X, Fujita H, et al. Staging hepatic fibrosis: computer-aided analysis of hepatic contours on gadolinium ethoxybenzyl diethylenetriaminepentaacetic acid-enhanced hepatocyte-phase magnetic resonance imaging. *Hepatology* 2012;55(1):328–9.
- [16] Li WD, Zeng YF, Zhang XJ, Huang Y, Long LL, Fujita H. Staging the hepatic fibrosis on CT images: optimizing the slice thickness and texture features. In: *IEEE 2011 international symposium on bioelectronics and bioinformatics*. 2011. p. 267–70.

- [17] Lee GN, Zhang X, Kanematsu M, et al. Classification of cirrhotic liver on MR images using texture analysis. *Int J Comput Assist Radiol Surg* 2006;1(Suppl. 1):379–81.
- [18] Li W, Zhang X, Kanematsu M, Hara T, Zhou X, Fujita H, et al. Development of an automated method for differentiation of cirrhotic liver in abdominal MR images. *Med Img Inf Sci (MII)* 2004;21(2):137–43.
- [19] Wu C, Chen Y. Texture features for classification of ultrasonic liver images. *IEEE Trans Med Imag* 1992;11(2):141–52.
- [20] Wei Z, Li Y, He M. Analysis of incidence rate of hepatic cancer in Fushui county from 2004 to 2008. *Chin J Oncol Prev Treat* 2011;3(2):137–8 (in Chinese).
- [21] Zhang X, Zeng Y, Fujita H, Wen L, Long L, Huang Y. Effective staging of fibrosis by the selected texture features of liver: which one is better, CT or MR imaging? Radiological Society of North America scientific assembly and annual meeting program, Radiological Society of North America (RSNA) SSG08-03. 2013 (abstract).
- [22] Viral Hepatitis Prevention and Treatment Plan, Chinese Society of Hepatology. *Chin J Infect Dis* 2001;19(1):56–62.
- [23] Haralick R. Statistical and structural approaches to texture. *Proc IEEE* 1979;67(5):786–804.
- [24] Vapnik VN. The nature of statistical learning theory. Springer Verlag: Heidelberg, DE; 1995.
- [25] Schapire R. Leave one out error stability, and generalization of voting combinations of classifiers. *Mach Learn* 2004;55:71–97.
- [26] Kato Hiroki, Zhang Xuejun, Kondo Hiroshi, et al. Computer-aided diagnosis of chronic liver disease using 3 Tesla MRI artificial intelligence: optimization of the region of interest in texture analysis. In: *Proceeding of 38th Japan magnetic resonance medicine conference*. 2010 (in Japanese).
- [27] Kato Hiroki, Kanematsu Masayuki, Zhang Xuejun, et al. Computer-aided diagnosis of hepatic fibrosis: preliminary evaluation of MRI texture analysis using the finite difference method and an artificial neural network. *Am Roentgen Ray Soc (AJR)* 2007;189:117–22.
- [28] Zhang Xuejun, Goshima Satoshi, Zhou Xiangrong, Hara Takeshi, Kanematsu Masayuki, Fujita Hiroshi, et al. Quantitative staging the hepatic fibrosis: computer-aided diagnosis by shape, texture, volume, and elasticity analysis. In: *Proc. of Radiological Society of North America Scientific Assembly and annual meeting program* Radiological Society of North America (RSNA), LL-INE1211. 2011.

Xuejun Zhang (1969–, Nanning, Guangxi, China) received the B.S. degree in Physics from Guangxi University, PR China, in 1991; the M.S. degree in Electronics and Information Systems Engineering from Gifu University, Japan, in 2001; and the Ph.D. degree in Electronics and Information Systems Engineering from Gifu University, Japan, in 2004. From 2004 to 2007, he continued his research in Computer-Aided Diagnosis as a postdoctoral researcher at Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Science, Graduate School of Medicine, Gifu University. He was invited as a visiting professor of Center for Medical Diagnostic Systems and Visualization, University of Bremen, Bremen, Germany at 2006. Since 2007, he was a Professor of School of Computer and Electronic Information, Guangxi University, Nanning, Guangxi, China. Now he is a visiting scholar in University of Southern California. His research interests include computer-aided diagnosis system, image segmentation, pattern recognition, visualization in medicine. He has published over 70 papers in Journals, Proceedings, Book chapters and Scientific Magazines.

Hiroshi Fujita received the B.S. and M.S. degrees in Electrical Engineering from Gifu University, Japan, in 1976 and 1978, respectively, and Ph.D. degree from Nagoya University in 1983. He was a Research Associate at University of Chicago, USA, from 1983 to 1986. He is currently a Chairman and a Professor in the Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, Japan. His research interests include computer-aided diagnosis system, image analysis and processing, and image evaluation in medicine. He has published over 500 papers in Journals, Proceedings, Book chapters and Scientific Magazines. He is currently President of the Japan Society of Medical Imaging and Information Sciences.