Applied Soft Computing 25 (2014) 399–413

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Liver fibrosis staging using CT image texture analysis and soft computing

Ömer Kayaaltı[a], Bekir Hakan Aksebzeci[b], İbrahim Ökkeş Karahan[c], Kemal Deniz[d], Mehmet Öztürk[c], Bülent Yılmaz[e], Sadık Kara[f], Musa Hakan Asyalı[g,*]

[a] Develi Hüseyin Şahin Vocational College, Erciyes University, Develi, Kayseri, Turkey
[b] Department of Biomedical Engineering, Abdullah Gül University, 38039 Kayseri, Turkey
[c] Department of Radiology, Medical School, Erciyes University, 38039 Kayseri, Turkey
[d] Department of Pathology, Medical School, Erciyes University, 38039 Kayseri, Turkey
[e] Department of Electrical and Electronics Engineering, Abdullah Gül University, 38039 Kayseri, Turkey
[f] Institute of Biomedical Engineering, Fatih University, 34500, Büyükçekmece, İstanbul, Turkey
[g] Department of Electrical and Electronics Engineering, Antalya International University, 07190, Antalya, Turkey

ARTICLE INFO

Article history:
Received 18 June 2014
Received in revised form 18 August 2014
Accepted 29 August 2014
Available online 8 September 2014

Keywords:
Liver fibrosis staging
Texture features
Feature selection
k-Nearest neighbor
Support Vector Machines

ABSTRACT

Liver biopsy is considered to be the gold standard for analyzing chronic hepatitis and fibrosis; however, it is an invasive and expensive approach, which is also difficult to standardize. Medical imaging techniques such as ultrasonography, computed tomography (CT), and magnetic resonance imaging are non-invasive and helpful methods to interpret liver texture, and may be good alternatives to needle biopsy. Recently, instead of visual inspection of these images, computer-aided image analysis based approaches have become more popular. In this study, a non-invasive, low-cost and relatively accurate method was developed to determine liver fibrosis stage by analyzing some texture features of liver CT images. In this approach, some suitable regions of interests were selected on CT images and a comprehensive set of texture features were obtained from these regions using different methods, such as Gray Level Co-occurrence matrix (GLCM), Laws' method, Discrete Wavelet Transform (DWT), and Gabor filters. Afterwards, sequential floating forward selection and exhaustive search methods were used in various combinations for the selection of most discriminating features. Finally, those selected texture features were classified using two methods, namely, Support Vector Machines (SVM) and *k*-nearest neighbors (*k*-NN). The mean classification accuracy in pairwise group comparisons was approximately 95% for both classification methods using only 5 features. Also, performance of our approach in classifying liver fibrosis stage of subjects in the test set into 7 possible stages was investigated. In this case, both SVM and *k*-NN methods have returned relatively low classification accuracies. Our pairwise group classification results showed that DWT, Gabor, GLCM, and Laws' texture features were more successful than the others; as such features extracted from these methods were used in the feature fusion process. Fusing features from these better performing families further improved the classification performance. The results show that our approach can be used as a decision support system in especially pairwise fibrosis stage comparisons.

1. Introduction

Chronic liver diseases (CLD) is a major public health problem worldwide. Some of CLD conditions are caused by the infection with B, C or delta hepatic viruses, whereas other cases include non-alcoholic fatty liver disease, alcoholic hepatopathy and other rare genetic, metabolic or immunological conditions. Liver fibrosis can be defined as the accumulation of collagen, proteoglycans, and other macromolecules within the extracellular matrix, which is a common feature of almost all causes of CLD. Hepatocarcinoma is one of the most common malignancies in patients that are affected by these diseases [1]. The fibrosis is the scarring response formed in the chronic injury of any cause. It is a dynamic process, with a possibility of reversibility. Hepatic dysfunction, portal hypertension, and cirrhosis are detected respectively due to progressing liver

* Corresponding author. Tel.: +90 242 245 00 50.
E-mail addresses: musahakan.asyali@antalya.edu.tr, asyali@gmail.com (M.H. Asyalı).

http://dx.doi.org/10.1016/j.asoc.2014.08.065
1568-4946/

fibrosis. Therefore, in our study we have focused on the assessment of fibrosis in medical images, as it is the unchanging marker for the diagnosis, prognosis, and treatment follow-up of all CLD conditions.

At present, liver biopsy is considered to be the gold standard in evaluating fibrosis. Using the liver biopsy one can establish the diagnosis with some certainty, assess the severity of necro-inflammation and fibrosis, and indicate the existence of simultaneous liver diseases, if there is any. On the other hand, it is an invasive procedure with possible side effects such as pain in 30–40% of the cases, hemorrhage, biliary peritonitis, penetration of abdominal vessels, pneumothorax (3/1000) or even death (2/10,000) [2,3]. Furthermore, sampling errors (either by the fragmentation of the biopsy specimen or by removing an inadequate volume) can be seen in 24% of the cases [2]. Fibrosis distribution is not always homogenous inside the liver, hence biopsy specimen has usually a very small volume (in average it has the shape of a cylinder with a diameter of 1 mm and a length of 1.5 cm) when compared to the whole liver volume. In addition, intra- and inter-operator variability may be seen. Such situations are reported in 10–20% of the cases. The two combined factors (lack of representativity of the biopsy specimen and the variability in the assessment of fibrosis) lead to a cirrhosis false negative rate of 24%. Therefore, in a significant percentage of cases, biopsy does not yield precise diagnosis results for liver fibrosis.

Clinical assessment of fibrosis can also be made using non-invasive tests (e.g., serologic markers or elastography) and medical imaging techniques, e.g., ultrasonography (US), computed tomography (CT), and magnetic resonance imaging (MRI). However, the sensitivity and the specificity of non-invasive tests still remain insufficient [3]. Although imaging tests may detect manifestations associated with fibrosis, e.g., portal hypertension, splenomegaly or cirrhosis, they currently fail to diagnose parenchymal fibrosis itself particularly in early stages of the disease [4,5]. Nevertheless, there are various studies in the literature about staging of the liver fibrosis using image analysis, mostly applied on ultrasound images [6–11]. Therefore, we believe that the appeal of non-invasive techniques in characterization of liver tissue will be remain to be an active area of research in the coming years as well.

Some of the suggested approaches have been based on different image features such as the first-degree and second-degree gray level statistics and the texture features measured from fractal size estimator in these studies [6–9]. Zhang et al. tried to determine the stage of the fibrosis by using 2 liver shape features and 5 gray level co-occurrence matrix texture features (contrast, angular second moment, entropy, mean and inverse difference moment), and two first-order statistics (mean and standard deviation) features [10]. Sun et al. defined the cirrhosis disease type by using the pyramid plexus and discrete wavelet transform from liver ultrasound images. They classified the disease into 4 classes: Healthy, stage 1, stage 2, and "very serious" [11].

In this study, using several supervised machine learning algorithms, we have tried to classify liver fibrosis stages using various texture features obtained from CT liver images. The reasons behind using/analyzing 'CT liver images' will be discussed in Section 2. Our study is comprised of two parts, namely data collection/preparation and analysis (classifier design) phases.

The first phase involved building a database of CT liver images and associated fibrosis stage information obtained through pathological analysis performed by the expert physicians in our team. Further, in this phase we have selected some suitable 'Regions of Interest' (ROIs) or sub-images from the CT images (i.e., slices of volume data) for each subject, again with the help of physician colleagues.

In the second phase of the study, we have extracted and selected texture features to be used in designing classifiers for predicting 'fibrosis stage.' We have used the following approaches to extract some suitable texture features to be used in classifier design:

- Gray Level Co-occurrence Matrix (GLCM) [12],
- Gray Level Run Length Matrix (GLRLM) [13],
- Gray Tone Difference Matrix (GTDM) [14],
- Laws' Texture Features (LTF) [15],
- Discrete Wavelet Transform (DWT) [16],
- Discrete Fourier Transform (DFT),
- Gabor filters [17], and
- First Order Statistics (FOS).

We have employed sequential floating forward selection (SFFS) [18] and exhaustive search methods for feature selection. As for classification methods, *k*-nearest neighbor (*k*-NN) [19] and Support Vector Machines (SVM) methods [20] were utilized. Further details of our 'classification experiment design' will be given in 'Materials and Methods' (Section 2.6).

In a preliminary study [21], we have used the features obtained by GLCM, DWT, and DFT methods in classifying liver fibrosis stage. After determining classification performance individual features, we have combined or fused better performing features to be used in the final classifier design. We have obtained relatively better results using the features obtained through GLCM and DWT methods. In pairwise group classification, we were able to reach classification accuracy rates of 86% and 90%, using *k*-NN and SVM classifiers respectively. Different from our previous study [21] and other studies in the literature [6–11], in this study we have included a very comprehensive set of texture features obtained through different approaches. These texture feature extraction approaches will be described briefly in the next section; detailed explanation of these methods is given in Appendix. Specifically, in order to raise classification accuracy figures to clinically acceptable levels, we have included features based on GLRLM, GTDM, LTF, Gabor filters, and FOS methods in classifier design, in addition to previously employed feature extraction methods. Furthermore, in this current study we have also investigated the possible effects of gray level scaling and quantization in obtaining GLCM features and the way partitioning of the sample set into test and train sets (i.e., fold value in the cross validation) on the classification performance.

2. Materials and methods

2.1. Image acquisition and histological analysis

In parallel with the noticeable development of CT technology over time, CT imaging of liver has gained importance and speed. CT is a standard method for liver imaging today and it is used in detecting and characterizing the primary and secondary liver masses and in reviewing the diffuse liver diseases [22].

Although MRI could be more advantageous in soft tissue imaging, its use in liver imaging is not common practice yet, mostly due to long data acquisition times that lead to motion (respiration) artifacts in the images and economic considerations (MRI is still an expensive procedure and is not available at every center). As for the US imaging, although it is practical (easy to operate/use) and less harmful (it does not involve any radiation), it suffers from problems such as somewhat subjective (operator-dependent) imaging quality and interpretation, degraded imaging quality in obese patients and in cases where there is intensive abdominal gas. As such, US is not a viable alternative to CT either in evaluating or staging liver fibrosis.

In CT imaging/evaluation of liver, an uncontrasted screening may cause misdiagnosis of hypodense vascular structures as lesions. Therefore, the main principle includes the optimum imaging of the vascular structures and review of the liver parenchyma

Table 1
Ishak fibrosis staging system.

Score	Architectural changes, fibrosis/cirrhosis
0	No fibrosis
1	Fibrous expansion of some portal areas, with or without short fibrous septa
2	Fibrous expansion of most portal areas, with or without short fibrous septa
3	Fibrous expansion of most portal areas with occasional portal-portal bridging
4	Fibrous expansion of portal areas with marked bridging (portal-portal or portal-central)
5	Marked bridging (portal-portal or portal-central) with occasional nodules (incomplete cirrhosis)
6	Cirrhosis, probable or definite

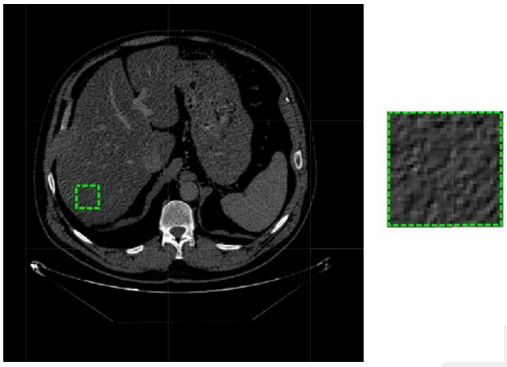

Fig. 1. Left: A sample liver CT image and a selected 32 × 32 pixels ROI on the image. Right: Blown up version of the ROI.

and surrounding soft tissues. This information can be obtained by using contrast agents in appropriate volume and concentration. The contrast agents in CT are extracellular and injected intravenously and they directly associated with radiographic attenuation rise [23].

For an ideal liver CT imaging process, it is absolutely necessary to use intravenous contrast. Multiphasic study gains importance because of the vascular hemodynamic of liver. Approximately 60–75% of blood build-up of liver is derived from portal vein, and the remaining 25–40% is derived from hepatic artery. Multiphasic liver CT process consists of arterial, portal venous and late venous phases. Liver parenchyma shows opacification in portal venous phase mostly [24].

In this retrospective study, the cases of patients who have undergone needle biopsy since 2006 were examined by pathologists and radiologists in our team at Erciyes University Medical School Hospital. The fibrosis in liver needle biopsy specimens were histopathologically evaluated on the hematoxylin-eosin-stained and mason trichrome-stained parts and fibrosis stage was determined (or scored) in accordance with Ishak Staging System [25], ranging from 0 (no fibrosis) to 6 (cirrhosis) (see Table 1).The images of 116 patients were acquired by using the multi-slice CT scanner (GE Light Speed 16, Milwaukee, Wisconsin, USA) at the Radiology Department of Erciyes University Medical School, in Kayseri, Turkey. Considering the possible change/progress in patients' fibrosis stages, patient imaging dates were arranged or limited to be within ±2 months from the date of the biopsy.The patients were administered nonionic 300/100 cc contrast substance to and images of the venous phase upper abdominal sections were taken with 512 pixels by 512 pixels resolution at 1.25 mm thickness (along longitudinal axis). A sample liver CT image and a selected ROI are shown in Fig. 1. Number of subjects in each fibrosis stage is given in Table 2.

Table 2
Fibrosis stages and number of subjects recruited in each stage.

Stage	# of Subjects
0	21
1	16
2	12
3	16
4	13
5	13
6	25
Total	116

2.2. Texture analysis

Texture is a measure of intensity change of an image and associated with features such as smoothness, roughness and regularity. For an excellent discussion of this somewhat controversial issue, we refer the reader to [26].

Different families of approaches exist for measuring texture features/characteristics. The first-degree texture features do not consider the pixel neighborhood relations, however, the second-degree texture features generally take these relationships into account. In this study, we have utilized many of the 'texture feature extraction' methods approaches that have been utilized in the image processing literature. These include GLCM, GLRLM, GTDM, LTF, 2-D DWT, 2-D DFT, Gabor filters, and FOS. In order to increase the readability of the paper, we have found it suitable to move the cumbersome mathematical formulations of these texture features to the end of the paper, into Appendix; aforementioned methods are respectively discussed in Sections A.1–A.8. Here in this section, we will suffice by shortly describing these methods.

Gray Level Co-occurrence Matrix (GLCM) defines the relationship between the neighbor pixels, and shows the occurrence frequency of the brightness levels on the image in a definite distance and direction [12,27–29]. Gray Level Run Length Matrix (GLRLM) is one of the ways to detect the second-degree statistical texture features. The set of sequential pixels that have the same gray level value in the same direction forms the gray level run. Run length is the pixel quantity in the run and run length value is the occurrence number of the runs on the image [13,30–32]. The existence of many neighbor pixels with the same gray level represents a rough texture, and the lower quantity of such neighbor pixels represents a thinner texture with a faster change. Gray Tone Difference Matrix (GTDM) has been suggested by Amadasun and King by which visual properties of the texture can be extracted [14]. Laws' texture features (LTF), which were developed by Kenneth Ivan Laws [15,33,34], were used in many different practices. For these features, first, small convolution kernels are applied onto the image, and then they are extracted using a nonlinear windowing operation. In Discrete Wavelet Transform (DWT) of an image is generally calculated by convoluting the image with two filters, which are low-pass and high-pass, throughout the rows and columns [16]. In this approach we use some suitable transform coefficients as texture features. The 2-D Discrete Fourier Transform (DFT) is relatively well known but for a good refresher we still refer the reader to the book by Demirkaya, Asyali, and Sahoo [35]. In this approach, we compute 2-D DFT of ROIs and use the relative power in certain frequency bands (corresponding to low and high frequencies in horizontal and vertical directions) as features. In Gabor filtering, in order to accentuate features that could lie in certain orientations, the image is filtered by a suitable linear 2D filter which is a Gaussian kernel function modulated by a sinusoidal plane wave. Depending on the variances (width) of the Gaussian kernel in horizontal and vertical directions and the frequency of the sinusoidal, differed pattern are emphasized at the output image and then this

filtered image is averaged to produce a final single texture feature [17,36]. By using several Gabor filters with different orientations, multiple features can be obtained.

2.3. Classification methods

In this study, *k*-nearest neighbor (*k*-NN) and Support Vector Machines (SVM) methods were used in the classification stage. The *k*-NN method is one of the most commonly known and used classification methods. Though it lacks modeling capability, it is frequently used as a benchmark method, due to its simplicity and relatively high performance. In contrast, the SVM method is a powerful, state of the art, and relatively complex method with superior learning/modeling capability. Both of these classifier design approaches are nonparametric, i.e., there is not any assumption about the underlying statistical model (or distribution) for the data. In our case, since the number of samples was limited, employing a parametric classifier (such as a Bayesian classifier [37]) would not be a sensible choice, as estimating model/distribution parameters from a limited number of samples would result in parameter estimates with high bias and variance.

In the *k*-NN method, a test data/feature vector **x** is simply assigned to class of most occurring data among **x**'s *k*-nearest neighbors, using a proper distance metrics such as Euclidean distance [19]. SVM is a supervised learning method that was developed by Vapnik [20]. SVM is commonly used in various pattern recognition applications such as object, handwriting, speech and face recognition [20,38–40]. The main objective of SVM is to find a hyperplane that can separate the two classes in the most suitable way (i.e., with the possible largest margin on both sides) in the feature space. SVM optimizes the position of the hyperplane in order to maximize the distance between the hyperplane and data samples in both sides during learning. SVM carries out this process by defining the samples (support vectors) that determine the boundary of each class [19].

Although SVM method is mainly a binary classifier, several of them can be combined using generally *one*-versus-*one* (1-1) or *one*-versus-*rest* (1-r) approaches, in order to handle multi-class cases. In this study, SVM (1-1) method was preferred to classify the multi-category data. SVM (1-1) is the method where the classes in data set are compared in a pair-wise fashion. In this method, for a data set with c class, $c(c-1)/2$ binary classification processes are carried out with SVM. A test data is assigned to the most voted class by analyzing the results of these binary classifiers. If there are multiple classes that have the most votes, the test data is assigned to the class with the lowest index or to the class with the highest prior probability density function among these classes [41,42].

2.4. Assessment of classification performance

The assessment of classifier performance is an important issue and it deserves a long and involved elaboration. For a detailed discussion of this topic, we refer the reader to an extensive review by Asyali et al. [43]. However, we will suffice by noting that empirical prediction of classification accuracy (i.e., the ratio of correct decisions to the total number of cases studied) remains to be the popular and practical performance measure, especially for nonparametric classifiers. The estimate of 'classification accuracy' is a random variable as it depends on the particular training and test samples used. If there are enough number of data samples available, one can partition this set into two subsets and use one for 'training' and the other for 'testing.' Literature [44,45] shows that if 'classification accuracy' is estimated by randomly splitting the set of all available samples into two parts, its bias decreases and variance increases as the size of training (testing) set increases (decreases).

A common technique to assess classifier performance is to *m*-fold cross validation (CV), where the overall set of *n* data samples is randomly divided into *m* approximately equal size and balanced (i.e., the distribution of samples into different classes is similar) subsets Then *m* classification trials are carried, each time one of these subsets is excluded from the overall set and used as a test set (for the classifier that is trained based on the rest of the samples). This is repeated over the *m* subsets and the resultant 'classification accuracies' are averaged to obtain the so-called *m*-fold CV accuracy rate. Typical choices for *m* are 5 or 10, as these correspond to a reasonable tradeoff between the variance and bias in estimation of classification accuracy rate and computational burden. As an extreme case, *m* can also be set equal to *n*, size of the overall data set, in which case we will have what is known as 'leave one-out cross-validation' (LOOCV). Although the LOOCV accuracy rate estimator will have a lower bias, it will have a relatively higher variance (as the size of the test set is just 1) and computationally it be very costly. In our case, for the estimation classification accuracy, we chose to try 2- or 3-fold CV only due to our limited sample size within each group, especially for the case of pairwise group (liver fibrosis stage) classification trials.

2.5. Feature selection

The objective of the selection methods is to choose the best subset that will represent the original data set. Feature selection not only minimizes the search space, but also increases the quality of classification process [46]. While there are many studies in the literature about feature selection methods, we used exhaustive search (ES) and sequential floating forward selection (SFFS) methods for their advantages in our study.

In the Exhausted Search (ES) method, when an original set with M elements is to be represented with a subset with N elements, $\binom{M}{N}$ subsets are formed and the subset yielding the best classification performance is selected as the optimal subset. Because it checks all possible combinations, ES is guaranteed to find the best subset with N elements to represent the original set. However, it is very costly computationally, i.e., when M and N are relatively large, ES becomes unacceptably slow. In the sequential forward selection (SFS) method, the optimal feature subset starts with an empty set and each iteration/trial the relatively better performing features are added to the set [47]. In SFS, the features included in the optimal set cannot be removed is subsequent steps. In contrast, the Sequential Forward Floating Selection (SFFS) method allows removal of included features. Hence, it is possible to obtain comparatively better performing optimal subsets representing the original feature set. Nevertheless, neither SFS nor SFSS is guaranteed to the find the global optimum [18].

2.6. Classification experiment design

Our overall classification approach is summarized in detail in Fig. 2. All algorithms were implemented using Matlab R2010a (Mathworks Inc., Natick, MA) software [48]. Further details of the processing steps will be discussed next.

Liver parenchyma that does not contain big blood vessels is selected from the CT images as the ROI. In order to avoid selecting the wrong area, 2 ROIs of size 32 by 32 pixels were manually selected from 5 images (slices) of each patient as shown in Fig. 1. Therefore, in a way, we have thoroughly scanned available imaged liver area in 3-D to obtain some good representative areas of liver fibrosis condition. In subsequent feature extraction steps, we have used these selected 10 ROIs for each patient. Before feature extraction process (the first step in Fig. 2), the Hounsfield values on the

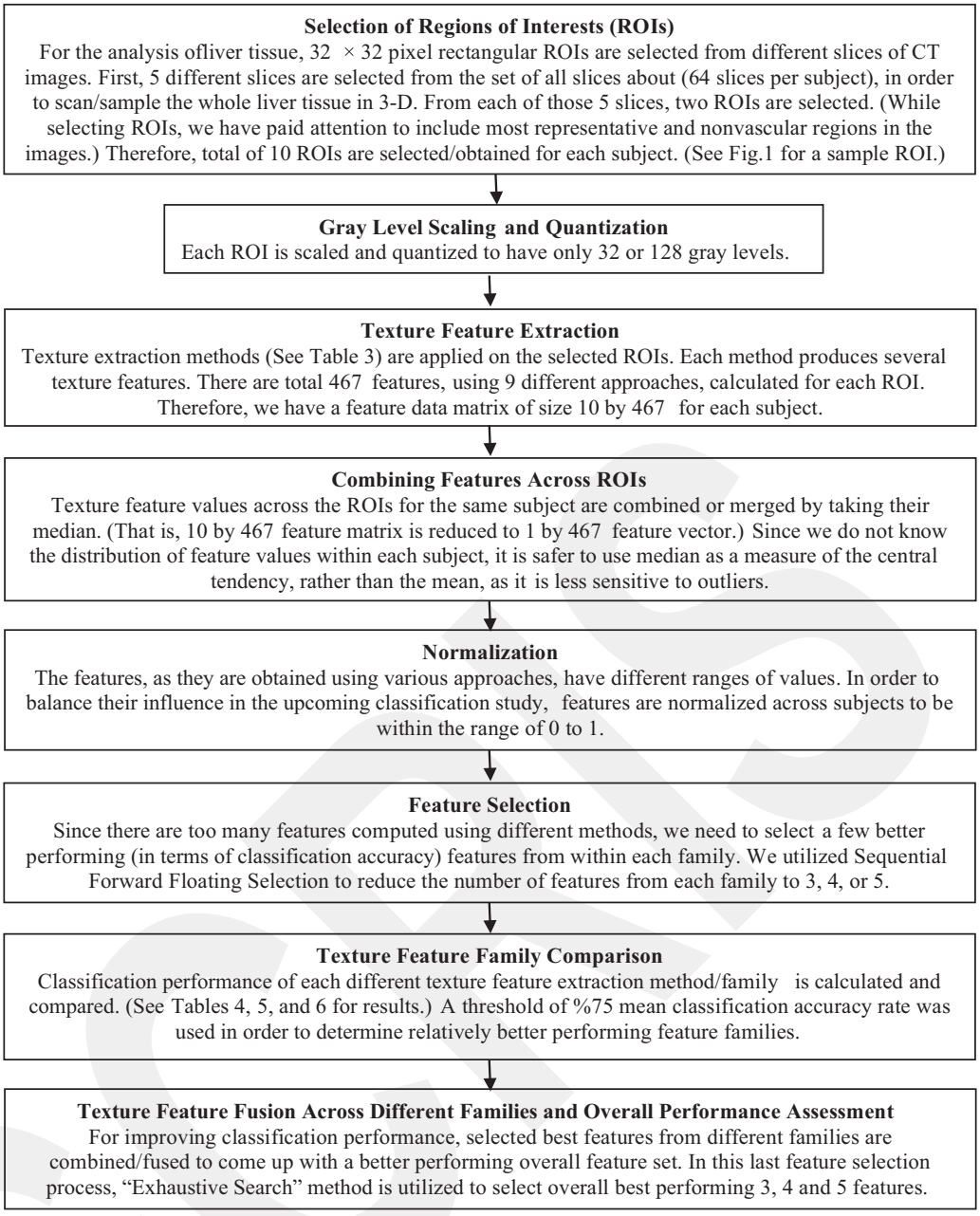

Fig. 2. Summary of processing steps followed from ROI selection to classification.

ROIs were linearly scaled/normalized and quantized to be 32 and 128 gray levels. For 32 (128) level scaling, the minimum and maximum pixel values were assigned as 1 (1) and 32 (128), respectively. Scaling was necessary to moderate the cost of GLCM and GLRLM computations. In order to assess the effect of gray level quantization on the final classification performance, we have repeated the whole scheme (feature extraction, selection, and classification) at these two quantization levels. Different studies in the literature selected/used various quantization levels such as 8, 16, 32, 64, 128, and 256. We tested these levels and decided to omit 8, 16 and 256 levels due to either low performance or high computational cost, and set lower and upper bound as 32 and 128 respectively.

In the second step, GLCMs of 10 ROIs were computed symmetrically at 0, 45, 90, and 135 degrees up to 8-pixel distance. Thus, we obtained 32 GLCMs (4 orientations times 8 distances) for one ROI and 320 for 10 ROIs for each patient. Later, 20 features mentioned in Section 2.3 were extracted from these GLCMs using the approach followed in [49]. Same features with the same distance in different angles were averaged, and thus 20 features were obtained for each distance getting a total of $20 \times 8 = 160$ features using the GLCM method for each ROI.

A single representative feature to be used in the further analysis was obtained by taking the median of the features extracted from 10 ROIs. When the underlying distribution of the parameters is not known it is safer to use the median as the measure of central tendency, rather than the mean, as it is more immune to outliers.

Similarly, 11 GLRLM texture features were extracted at 0, 45, 90, and 135 degrees up to runs in length of 8 using GLRLM Toolbox v1.0 [50]. In order to make the features rotation invariant, the feature values were determined by taking the means of the same features with the same run length in different angles as we have done in finding GLCM features. Totally, $11 \times 8 = 88$ features were obtained in runs in length of 8 using this particular method.

Table 3
List of feature extraction methods/families and the number of features coming from each family.

Feature family	Number of features
GLCM	160
GLRLM	88
GTDM	5
LTF_3[a]	25
LTF_5[b]	70
Gabor filters	30
DWT[c]	80
DFT	4
FOS	5
Total	467

[a] Laws' Texture Features with of Kernel Length of 3.
[b] Laws' Texture Features with of Kernel Length of 3.
[c] DWT Coefficients/features for each family of filter banks.

Using the LTF method by means of applying 3×3, 5×5, and 7×7 windows with length-5 kernels, 14 texture images were obtained for each window. The window size with the best performance among these 3 windows determined the features of Laws' texture feature family. Therefore, $14 \times 5 = 70$ features were extracted for each window by extracting the five FOS features (mean, standard deviation, average energy, skewness, and kurtosis; see Appendix A.8 for details) out of these images. This process was also carried out for the length-3 kernels, and $5 \times 5 = 25$ features were obtained for each window. The window size features with the highest classification success were also selected in this method.

Another feature extraction approach we used was the Gabor filters. We created Gabor filters to be applied on the ROI images by using variances as 1, 2, 3, 4, and frequencies as 0, 2, 4, 8, 16, 32, at the angles 45, 90, 135, and 180 degrees in combination (see Appendix for the formulation of Gabor filters). For each resultant/filtered image, 5 FOS features were extracted. The features were made rotation invariant by averaging the features obtained through the use of the filters with the same variance and frequency combination in different angles. For each variance combination (16 combination) $6 \times 5 = 30$ features (5 FOS features for each frequency) were obtained using Gabor filters.

The DWT was applied as another feature extraction method. Matlab's wavelet functions db1, db2, db3, coif1, coif2, coif3, sym2, sym3, sym4, dmey, bior1.1, bior2.2, bior3.1, bior4.4, rbio1.1, rbio2.2, rbio3.1, and rbio4.4 were used as the filter banks [48]. DWTs were computed using these different wavelet filter banks and decomposition coefficients were obtained at 4 levels. Then, 5 FOS of the DWT coefficients were obtained for each level.

In addition to these features, 5 features were extracted by using the GTDM method. Furthermore, using DFT 4 texture features, corresponding to the relative power in certain bands, were obtained. Finally, 5 FOS features were directly computed from each ROI. For ease of referral, Table 3 lists/summarizes all of the feature extraction methods that we have utilized, along with the number of features computed in each category.

Once we have obtained all the features, we have normalized them to be between 0 and 1, in order to balance their influence in the subsequent fusing step. After feature extraction and normalization steps, we have performed the "dimension reduction" or "feature selection" step. In this step, we have basically reduced the number of features by selecting the most discriminative features from the set of all texture features. In order to include best possible features in the analysis, we have started with a very extensive/comprehensive set of features and eliminated less performing features at this critical step. By doing so, we have also overcome the "curse of dimensionality" problem [51]. This problem refers to the fact that, when the number of features is too many compared to the number of samples in a classification problem/setup, due to increasing sparsity of matrices that are involved in the analysis, numerical problems may arise and statistical power may be lost, which in effect hinder the significance of the estimated parameters [37]. In our case, as Table 3 shows, the total number of features (obtained using different texture measurement approaches) is 467, which is too many compared to the number of samples (116 samples from 7 different classes). In most practical applications, the number of features should be around one tenth of the number of samples per class [37].

We did not want to resort to feature combination methods such as principal component analysis [37], because such methods combine features in somewhat inexplicable ways. In our case, we wanted to identify/pinpoint texture features that are relatively highly relevant to fibrosis characterization. As such, combining texture features that indicate totally different texture characteristics in an uncontrolled way would not be suitable for our purposes.

In feature selection (i.e., dimension reduction), as performance metric we have used 'classification accuracy rates' that are returned by k-NN and SVM classifiers at each trial/step, as these two were the methods of our choice in the final classification step. The dimension reduction was carried out in two phases. In the first phase, the best features were selected from each family using the SFFS method (see Section 2.5). In this phase, exhaustive search was not preferred as it is computationally costly. For example, to reduce 160 features obtained from GLCM family to 3 features requires $\binom{160}{3} = 665,680$ trials. Because the feature values were not monotonic the branch and bound method [52] could not be applied. Since the cost of selecting few features out of such high number of features with the backward selection methods [18,53] is higher than with the forward selection, backward selection methods were not preferred either. In the second phase of feature selection, that is, after the best features were obtained from each feature family, the resultant features were fused and "exhaustive search" was performed on the fused set of features, in order to determine the best performing features in the final classification stage.

We have proceeded with selection of 3, 4, or 5 best performing features from each family according to their classification accuracy. As we will discuss in the next section, GLCM, DWT, GF and LTF methods consistently yielded a mean classification accuracy of larger than 75%. Therefore, only the features coming from these methods were used in the fusing stage. Otherwise, for the case of 4 features, for instance, if we had included 4 best features from all of the 9 feature families, we would have total of $4 \times 9 = 36$ features in the fused set and an exhaustive search on this set for overall 4 best features would require almost 60,000 trials. On the other hand, if we just include aforementioned 4 feature families in the final search (in the fused feature space), again for the case of 4 features for instance, we will have $\binom{16}{4} = 1820$ trials, which is a reasonable number.

After fusing, a final set of 3, 4, or 5 features were selected from the set of fused features by using exhaustive search where all feature combinations are examined in terms of classification performance. Although not reported here, we have also experimented with the case of 6 features (i.e., included 6 features in the feature fusing step and in the final round of feature selection) and noticed that performance did not improve significantly whereas computation time increased significantly. For this reason, we have decided not to include the case of 6 features in the analysis.

3. Results

In the classification phase of the study, the k-NN classifier was used with $k = 3$ and $k = 5$ neighbors. However, we have noticed that classification performance of k-NN classifier with 3 or 5 neighbors

Table 4
Pairwise classification accuracy results (%) with the **3 best features** for each feature family, using *k*-NN and SVM classifiers.

Group 1	Group 2	*k*-NN					
Stage	Stage	DWT (rbio4.4)	GLCM	LTF_5	LTF_3	Gabor filters ($s_{x'}=2, s_{y'}=3$)	FUSION
0	1	83.78	75.68	70.27	83.78	70.27	83.78
0	2	78.79	78.79	84.85	69.70	72.73	84.85
0	3	83.78	78.38	83.78	64.86	75.68	86.49
0	4	85.29	88.24	85.29	76.47	82.35	91.18
0	5	85.29	85.29	76.47	73.53	73.53	85.29
0	6	76.09	78.26	67.39	73.91	76.09	82.61
1	2	75.00	82.14	71.43	82.14	78.57	82.14
1	3	78.13	78.13	81.25	81.25	78.13	93.75
1	4	89.66	82.76	75.86	75.86	93.10	93.10
1	5	75.86	79.31	79.31	86.21	68.97	86.21
1	6	78.05	78.05	73.17	73.17	70.73	80.49
2	3	82.14	82.14	82.14	75.00	78.57	85.71
2	4	92.00	92.00	88.00	92.00	88.00	96.00
2	5	84.00	84.00	88.00	88.00	68.00	88.00
2	6	83.78	86.49	75.68	78.38	81.08	86.49
3	4	86.21	79.31	82.76	79.31	75.86	93.10
3	5	82.76	79.31	79.31	65.52	75.86	89.66
3	6	85.37	78.05	73.17	80.49	80.49	87.80
4	5	80.77	84.62	76.92	84.62	84.62	92.31
4	6	86.84	76.32	84.21	78.95	78.95	89.47
5	6	81.58	84.21	84.21	78.95	81.58	84.21
Mean:		82.63	81.50	79.21	78.20	77.77	87.74

Group 1	Group 2	SVM					
Stage	Stage	DWT (haar)	GLCM	LTF _5	LTF _3	Gabor filters ($s_{x'}=2, s_{y'}=3$)	Fusion
0	1	83.77	78.07	75.73	81.29	70.32	83.77
0	2	72.79	78.86	87.87	69.67	75.74	87.87
0	3	86.55	75.73	83.92	75.58	81.43	86.55
0	4	85.29	82.35	79.41	73.53	85.29	94.12
0	5	79.41	85.29	79.41	73.53	82.35	85.29
0	6	76.09	78.26	76.09	73.91	73.91	82.61
1	2	89.29	78.57	71.43	78.57	71.43	89.29
1	3	84.38	81.25	81.25	84.38	81.25	87.50
1	4	93.33	85.95	82.86	79.52	96.67	96.67
1	5	72.62	72.62	75.71	85.95	75.95	93.10
1	6	85.36	87.98	75.60	75.48	77.98	87.98
2	3	89.29	85.71	89.29	78.57	82.14	92.86
2	4	91.99	91.99	88.46	88.14	87.82	96.15
2	5	91.99	91.99	91.99	83.97	72.12	100.00
2	6	83.77	81.29	75.73	83.92	72.95	84.06
3	4	82.86	93.10	82.38	86.19	72.62	93.10
3	5	86.67	82.86	82.62	72.62	69.05	86.67
3	6	80.48	80.60	82.74	75.60	75.60	87.74
4	5	96.15	80.77	88.46	84.62	76.92	96.15
4	6	84.21	86.84	84.21	78.95	81.58	89.47
5	6	86.84	89.47	84.21	81.58	76.32	89.47
Mean:		84.91	83.31	81.87	79.31	78.07	90.02

did not change significantly. Therefore, for clarity, in what follows we present results for $k=3$ only, i.e., when we refer to a *k*-NN classifier, it is a '3 nearest neighbors' classifier. As for the SVM method, the radial basis function (RBF) was preferred as the kernel function.

The performance of the classifiers was assessed using 2- or 3-fold CV method, due to the limited number subjects in each class (fibrosis stage). Here, we must note that the same classification methods are used during the feature selection and actual data classification steps. We have observed that using 2 or 3-fold CV did not affect classification results significantly. This is not surprising as there only few samples (about 15) in each group. Therefore we have just reported results for the case 2-fold CV, for the sake of brevity.

Table 4 summarizes the accuracies of the pairwise group classifications for 3 best features using the aforementioned feature extraction approaches, namely DWT, GLCM, LTF_3, LTF_5, and Gabor filters, using *k*-NN and SVM classifiers. For DWT and Gabor filters respectively, name of the wavelet family and value of filter parameters that produced best performing features is shown in parenthesis. From Table 4 we deduce that GLCM, DWT, GF and LTF methods consistently yielded a mean classification accuracy of larger than 75%. Therefore, only the features coming from these methods/families were used in the fusing stage. At the fusing step we have included those best 3 features from the selected feature families and selected overall best performing 3 features out of this merged feature set using exhaustive search. The last column of the table gives the classification accuracies for the overall 3 best features out of the fused set. Tables 5 and 6 are very similar to Table 4 in the sense of summarizing the classification performances for the cases of 4 or 5 features, respectively.

As there are 7 different classes and $\binom{7}{2} = 21$ pairwise classification cases (corresponding to the rows of Table 4), there will be total of $21 \times 3 = \mathbf{63}$ different selected overall best features for case of fusing 3 best features from the selected families. Similarly, there will be total of $21 \times 4 = \mathbf{84}$ and $21 \times 5 = \mathbf{105}$ different selected overall best features, for the cases of fusing 4 and 5 best features from the

Table 5
Pairwise classification accuracy results (%) with the **4 best features** for each feature family, using *k*-NN and SVM classifiers.

Group 1	Group 2	*k*-NN					
Stage	Stage	DWT (rbio4.4)	GLCM	LTF_5	LTF_3	Gabor filters ($s_{x'}=2, s_{y'}=3$)	Fusion
0	1	83.78	81.08	72.97	81.08	72.97	83.78
0	2	87.88	81.82	75.76	72.73	75.76	87.88
0	3	83.78	83.78	86.49	62.16	81.08	86.49
0	4	82.35	88.24	82.35	82.35	79.41	97.06
0	5	88.24	85.29	79.41	73.53	76.47	91.18
0	6	78.26	78.26	67.39	76.09	76.09	84.78
1	2	82.14	85.71	67.86	85.71	75.00	85.71
1	3	81.25	81.25	84.38	81.25	78.13	90.63
1	4	89.66	82.76	75.86	72.41	89.66	96.55
1	5	86.21	79.31	79.31	86.21	68.97	89.66
1	6	80.49	80.49	80.49	75.61	75.61	82.93
2	3	78.57	89.29	85.71	75.00	78.57	92.86
2	4	96.00	92.00	88.00	92.00	88.00	100.00
2	5	84.00	100.00	88.00	88.00	72.00	100.00
2	6	89.19	86.49	81.08	81.08	81.08	89.19
3	4	86.21	86.21	72.41	82.76	75.86	93.10
3	5	82.76	79.31	82.76	65.52	75.86	89.66
3	6	85.37	78.05	73.17	78.05	78.05	90.24
4	5	84.62	80.77	80.77	88.46	80.77	96.15
4	6	92.11	84.21	84.21	84.21	81.58	92.11
5	6	78.95	84.21	84.21	81.58	81.58	92.11
Mean:		84.85	84.22	79.65	79.32	78.21	91.05

Group 1	Group 2	SVM					
Stage	Stage	DWT (haar)	GLCM	LTF_5	LTF_3	Gabor filters ($s_{x'}=2, s_{y'}=3$)	Fusion
0	1	81.14	78.07	78.36	81.29	70.32	89.04
0	2	72.79	78.86	84.93	69.67	75.74	90.81
0	3	91.81	78.22	86.55	78.51	84.06	91.81
0	4	79.41	82.35	82.35	76.47	88.24	94.12
0	5	79.41	88.24	79.41	76.47	82.35	88.24
0	6	86.96	89.13	82.61	71.74	73.91	89.13
1	2	89.29	85.71	71.43	89.29	71.43	89.29
1	3	87.50	84.38	81.25	81.25	81.25	90.63
1	4	93.33	85.95	82.86	75.95	96.67	96.67
1	5	86.43	72.38	82.86	85.95	79.52	93.10
1	6	90.12	85.48	78.10	75.60	82.74	90.12
2	3	92.86	85.71	92.86	78.57	82.14	96.43
2	4	100.00	96.15	92.31	96.15	87.82	100.00
2	5	100.00	91.99	91.99	91.99	68.27	100.00
2	6	86.40	86.55	78.51	83.92	73.10	89.33
3	4	82.86	93.10	92.86	79.52	82.86	93.10
3	5	86.67	86.19	82.62	72.38	72.62	89.76
3	6	82.86	80.36	80.48	77.98	75.60	90.12
4	5	96.15	84.62	88.46	84.62	73.08	96.15
4	6	86.84	86.84	89.47	78.95	81.58	92.11
5	6	84.21	86.84	84.21	78.95	81.58	89.47
Mean:		87.48	85.10	84.02	80.25	79.28	92.35

selected families. In order to identify the feature families that contribute to these final feature sets or pools more frequently, we have prepared Table 7, where we indicate the total number of features selected (across all pairwise classification cases) from each family.

In the pairwise group comparisons (i.e., Tables 4–6), we observe that

- The case of 5 best overall features performed more successfully than 3 and 4 features. With 5 features, minimum, maximum, and mean classification accuracies obtained were respectively 83.78%, 100%, and 93.17%, using *k*-NN method. Using SVM method, these figures were respectively 89.18%, 100%, and 94.00%.
- Without fusion, the features obtained using DWT performed best in terms of classification accuracy. After DWT, the performance of the extraction methods can be listed in descending order as the GLCM, LTF_5, LTF_3, and Gabor filters for both classification methods. (As one would realize, this is the order of the methods/columns in Tables 4–6).
- In the process of fusion, when we used *k*-NN classifier GLCM texture features turned out to be the least used/selected features for pairwise classifications, and when SVM was chosen as the classification methodology LTF_3 texture features turned out to be the least used/selected features for pairwise classifications (see Table 7).
- The difference between 32 or 128 levels of quantization was insignificant. The lower scaling level was preferred to optimize the cost of GLCM and GLRLM computation. Therefore, we presented the results of our approach for 32 scaling levels only.
- In general, SVM provided better classification accuracies compared to *k*-NN.

In addition to pairwise classification trials (i.e., stage 0 versus 1, 0 versus 2, etc.), we have also designed multi-category classifiers (using *k*-NN and SVM methods) to predict fibrosis stage (0, 1, . . ., 6) of subjects directly, using the same feature extraction, selection, and fusing approaches. Similar to the case of pairwise classifiers, GLCM, DWT, Gabor, LTF_5, and LTF_3 texture features were used

Table 6
Pairwise classification accuracy results (%) with the **5 best features** for each feature family, using *k*-NN and SVM classifiers.

Group 1	Group 2	*k*-NN					
Stage	Stage	DWT (rbio4.4)	GLCM	LTF_5	LTF_3	Gabor filters ($s_{x'}=2$, $s_{y'}=3$)	FUSION
0	1	81.08	81.08	75.68	81.08	72.97	83.78
0	2	90.91	81.82	75.76	81.82	75.76	90.91
0	3	83.78	83.78	86.49	81.08	81.08	91.89
0	4	85.29	91.18	85.29	82.35	79.41	100.00
0	5	88.24	85.29	85.29	73.53	76.47	91.18
0	6	84.78	78.26	69.57	76.09	76.09	86.96
1	2	82.14	85.71	71.43	85.71	75.00	89.29
1	3	81.25	81.25	84.38	84.38	78.13	96.88
1	4	93.10	86.21	79.31	68.97	89.66	100.00
1	5	82.76	75.86	79.31	86.21	68.97	93.10
1	6	80.49	82.93	75.61	73.17	75.61	87.80
2	3	85.71	89.29	85.71	75.00	78.57	96.43
2	4	96.00	92.00	88.00	92.00	88.00	100.00
2	5	80.00	100.00	88.00	84.00	72.00	100.00
2	6	89.19	86.49	86.49	75.68	81.08	89.19
3	4	89.66	93.10	79.31	79.31	75.86	93.10
3	5	86.21	82.76	82.76	72.41	75.86	93.10
3	6	85.37	82.93	75.61	80.49	78.05	92.68
4	5	80.77	84.62	84.62	88.46	80.77	96.15
4	6	94.74	84.21	81.58	86.84	81.58	94.74
5	6	84.21	84.21	86.84	81.58	81.58	89.47
Mean:		85.98	85.38	81.29	80.48	78.21	93.17

Group 1	Group 2	SVM					
Stage	Stage	DWT (haar)	GLCM	LTF _5	LTF _3	Gabor filters ($s_{x'}=2$, $s_{y'}=3$)	FUSION
0	1	80.99	75.44	83.77	81.29	70.47	89.18
0	2	72.79	84.74	84.93	69.49	78.86	90.81
0	3	89.04	75.73	86.55	78.51	84.21	97.22
0	4	88.24	85.29	76.47	70.59	88.24	94.12
0	5	79.41	88.24	82.35	82.35	82.35	91.18
0	6	91.30	89.13	78.26	69.57	71.74	91.30
1	2	89.29	85.71	71.43	89.29	67.86	92.86
1	3	87.50	90.63	84.38	81.25	81.25	93.75
1	4	93.33	85.95	86.43	79.29	93.33	96.67
1	5	79.52	76.19	82.86	82.62	79.52	93.10
1	6	85.36	90.36	75.60	75.60	82.74	90.36
2	3	92.86	85.71	92.86	82.14	82.14	100.00
2	4	100.00	96.15	96.15	92.31	83.97	100.00
2	5	96.15	91.99	95.83	95.83	72.12	100.00
2	6	86.55	86.55	81.29	86.55	73.10	91.96
3	4	82.86	89.76	86.19	86.43	79.29	96.67
3	5	90.00	86.19	79.29	79.29	79.05	93.10
3	6	77.98	75.60	78.10	80.48	77.98	90.12
4	5	92.31	88.46	92.31	84.62	69.23	100.00
4	6	86.84	86.84	89.47	78.95	78.95	92.11
5	6	84.21	84.21	84.21	81.58	78.95	89.47
Mean:		86.98	85.66	84.23	81.33	78.83	94.00

Table 7
Frequency of total number of features selected for each extraction approach and used in the final fusing process. There were 21 pairwise classifications for 7 fibrosis stages.

	k-NN # of features			SVM # of features		
Feature family	3	4	5	3	4	5
GLCM	6	13	11	20	19	13
DWT	10	31	26	18	19	25
Gabor filters	11	13	22	8	17	21
Laws_3	20	14	21	5	10	19
Laws_5	16	13	25	12	19	27
Total # of features	63 (21×3)	84 (21×4)	105 (21×5)	63	84	105

here. With *k*-NN, it is possible to design a 7-class classifier with the selected features directly. However, SVM is a binary classifier and therefore repetitive use of binary SVM (1-1) method classifiers was used. (See Section 2.3 Classification Methods, for details.) Table 8 shows the classification accuracies obtained when we use 3, 4 or 5 best features to classify all stages at once.

4. Discussion and conclusion

Liver fibrosis is a common feature of almost all causes of chronic liver disease and therefore it needs to be staged carefully. Several approaches based on non-invasive imaging techniques have been suggested for the characterization of liver tissue. For instance, some

Table 8
Classification accuracy (%) for the staging of liver fibrosis; 3, 4 or 5 best features were used to classify all 7 stages at once (not pairwise comparison, multi-category classification).

# of features	*k*-NN					
	GLCM	LTF_3	LTF_5	DWT (rbio4.4)	Gabor filters ($s_{x'} = 2, s_{y'} = 3$)	Fusion
3	30.17	28.45	22.41	27.59	28.45	33.62
4	33.62	32.76	23.28	29.31	30.17	36.21
5	32.76	37.93	28.45	29.31	30.17	40.52
# of features	SVM					
	GLCM	LTF_3	LTF_5	DWT (haar)	Gabor filters ($s_{x'} = 2, s_{y'} = 3$)	FUSION
3	31.29	34.52	32.74	26.63	30.16	34.52
4	36.34	34.52	35.6	30.46	30.33	36.34
5	37.65	34.52	34.29	37.12	34.16	38.86

studies focused on staging of liver fibrosis and diagnosis of other liver conditions using image analysis techniques [6–11].

In this study, in addition to the features used in a preliminary work [21], we have tried new features and feature inclusion strategies, and improved classification performance significantly. As for the difference or novelty of our study compared to the previous studies in the literature, we can underline three points. Firstly, unlike other studies, which mostly use US and MRI images, we have employed CT images due to the advantages that were described previously, in Section 2.1. Secondly, other studies used or tested only a small number of features extraction methods or families (usually 2 or 3), whereas we have studied many (around 10) different feature extraction methods that are found in the literature. Thirdly, the 'gold standard' or the reference dataset, with which we have developed our algorithms, has CT images of about 16–17 cases or subjects from each of the fibrosis stages, as such this dataset by itself very comprehensive and valuable. We should note here that a significant portion of the effort that was involved in this study went into the meticulous preparation of the dataset. We are willing to share this dataset with interested parties/colleagues; please write to us to learn about the details.

In this study, fibrosis staging using CT liver images was carried in 4 steps: (1) Determination of ROIs on the images obtained from 116 patients, (2) Extraction of texture features using several different techniques, (3) Reduction of number of extracted features and selection of feature families that have provided more than 75% classification accuracy for the next (fusion) step, (4) Fusion of the best performing features coming from different feature families, reduction of fused texture features, and use of overall few best features in pairwise or multi-class classification of fibrosis stages.

GLCM, GLRLM, GTDM, LTF, Gabor filters, DWT, DFT, and FOS approaches were used in the feature extraction step of our study. Gabor filters, DWT, GLCM, and LTF approaches were found to be capable of generating relatively better performing features in our classification trials. SFFS and exhaustive search techniques worked successfully in the feature selection part of the study. Finally, accuracies of *k*-NN and SVM classifiers were analyzed; it turned out that the SVM method outperforms the *k*-NN method to some extent, in the classification step.

While extracting the Gabor and DWT texture features, different filters were created and filters giving rise to better performing features were selected and used. (The corresponding parameters indicated in Tables 4–6.) The classification accuracies of features obtained with Gabor filters created at 16 different variances and that of DWT coefficients obtained through the use 18 different filter banks for were calculated. Although the classification accuracies for different filters were relatively close to each other, it was observed that Gabor filter with $S_{X'} = 2, S_{y'} = 3$ variances for *k*-NN and SVM classifiers; *rbio4.4* DWT kernel for *k*-NN classifier; and *haar* DWT kernel for SVM classifier were the most successful combinations. As Gabor, DWT, GLCM, and LTF texture features were more effective than other techniques, the features extracted from these methods were used in the feature fusion process.

As the results shown in Tables 4–6 indicate, fusion of the best performing features from different feature families improves the classification accuracy levels considerably. For instance, in the pairwise comparison of 21 healthy (stage 0) subjects and 16 'stage 1' liver fibrosis patients, using 5 best features extracted from the same texture feature family, we were able to achieve classification accuracy of approximately 83.77% using SVM classifier (Table 6, 1st row of lower panel with SVM results, LTF_5 column). Whereas, when the best 5 features were selected out of the fused feature set the classification accuracy increased to 89.18% (Table 6). Along these lines, the results also show that fusing the features and then selecting the best ones from them leads to at least the same or better accuracy levels than the features obtained using only one method.

In pairwise classifications with the *k*-NN method using 5 features, DWT, LTF_5, and Gabor filter texture features are the most commonly used features in the fusion step. With the SVM method, DWT, LTF_5, and Gabor filter feature families contributed more features than other approaches (see Table 7).

As for the results of pairwise group classifications (in terms of mean classification accuracies), the accuracy (%) figures for 3, 4, and 5 'fused features' were approximately 88 (90), 91 (92), and 93 (94) for the SVM (*k*-NN) classifiers respectively (see Tables 4–6). It was interesting to note that a simple and fast approach like *k*-NN performed at a comparable accuracy with a more complex and costly approach, namely SVM.

In addition, we investigated the performance of these feature extraction, feature selection, and classification approaches in classifying the liver fibrosis stage of the subjects in the test set into 7 possible stages, i.e., multiclass classification. In this case, both SVM and *k*-NN methods have yielded a classification accuracy of approximately 40% (see Table 8). If we consider the class prior probabilities are the same, choosing one stage/class out of 7 stages randomly gives rise to a classification/guessing accuracy of 1/7 (approximately 14%). Compared to this figure, a figure of 40% indicates that there is some value in our multi-category classification effort. Even though this result is insufficient and needs to be improved considerably, it still conveys some insight about the difficulty of classification problem that we have tackled.

In the light of previous discussions and assessments, we can argue that the value of our study mostly lies in the area of 'stage to stage comparisons,' at least for the time being. After determining of liver fibrosis stage of a patient using needle biopsy, in the follow-ups, needle biopsy is again routinely performed except when the fibrosis reaches stage 6. (Stage 6 can be recognized clinically, without a need for biopsy.) Therefore, instead of a needle biopsy, our approach can be used on CT images as a means (i.e., as a decision support tool) for the follow-up of fibrosis stage. Liver fibrosis can be rarely cured with a suitable treatment, it is generally an irreversible, chronic, and diffuse liver disease, however follow-ups are still very

important. As we have discussed earlier, while explaining the motivation behind our study, liver biopsy is a difficult, invasive, and expensive process. It also requires opinion/assessment of a senior pathologist specialized in the subject. Moreover, there are cases for which performing biopsy could be very difficult. Hence, instead of a needle biopsy, our approach can be instrumental in following fibrosis stage progress. We may obtain the current stage of the liver using a biopsy and in the later examinations some suitable features extracted from the CT images can be fed into to the proper SVM pairwise classifiers, in order to determine whether the fibrosis stage has advanced or stayed the same.

If a researcher wants to determine the fibrosis stage of his/her patient using the approach suggested in this study, first the CT images of the patient's liver should be taken. Then, 10 ROIs (sub-images or windows) corresponding to the liver parenchyma tissue will be selected from the stack of the images and texture features will be extracted the using the methods described here. Although many features were extracted in our study, not all of them were used in the final classification phase. Therefore, it is enough to extract only the selected features (mainly Gabor, DWT, GLCM, and LTF) from the subject's CT images.

If a different investigator/clinician uses different CT/instrument settings than the ones we used in our study, in that case, we may argue or envision that one can produce images with (i) different Hounsfield values, (ii) different resolution, or (iii) a slight tilt. Liver fibrosis changes the tissue structure of the liver and the texture on the CT image. In our approach the Hounsfield values of the area selected from a liver image are linearly scaled between to be 1 and 32. Further, the features extracted from the images are standardized or normalized in our approach; thus images obtained through different systems/settings will not cause a complication. As for the selection of ROIs, if the images have a different resolution, we may still propose or suggest the use of a ROI window size which is close to the actual/physical size of the window that we have used in this study. A length of 32 pixels wide approximately corresponds to 3 cm at a spatial resolution of 50 cm/512 pixels $\cong$ 0.1 cm/pixel, on the transverse plane. Consequently any ROI size that is close to a physical window size of 3 cm $\times$ 3 cm could be used. Finally, since the selected features are rotation invariant, a slight tilt in the image does not affect the feature extraction phase. Therefore, when it is fully developed, our approach can be put in a stand-alone software/application from to be used a decision support tool in the clinic.

Numerous studies show that different serum markers, physiological parameters [54], and features extracted from other imaging methods such as ultrasonography [55] might be valuable in the diagnosis of cirrhosis (or liver fibrosis). In a future study, we are planning to investigate the possible utilization of serum markers and physiological parameters that may be related to fibrosis stage, such as the values obtained using liver function tests, e.g., aspartate transaminase (AST), alkaline phosphatase (ALP), gamma glutamyl transpeptidase (GGT), bilirubin level, etc. in classifier design. We are also planning to investigate different texture analysis tools for feature extraction and novel classifiers in order to increase the classification accuracy in the prediction of fibrosis stages using CT liver images.

We should also note that increasing number of subjects or samples will also be very helpful in developing a more reliable system. Considering that there are 7 classes/stages and around at least 30 samples per class are required in order to draw statistically sound inferences, we need to have total of about 210 samples. Currently, we are about 100 subjects short of this figure; as such we are planning to increase our sample size. Because of this sample size limitation, we were not able to test the performance of model based classifiers, such as a quadratic discriminant analysis classifier that fits multivariate normal densities with covariance estimates stratified by group. To be more specific, when the sample size is too small, some numerical problems arise in the estimation of covariance matrices and it becomes troublesome to invert those ill conditioned matrices.

Although not reported here, we have also explored the idea of using each ROI as a separate case/sample. In that case our sample size would be 10 fold, as we have obtained 10 ROIs from each subject. (As explained in Section 2.6, we were processing each ROI separately to obtain the features and then merging those features through the median operator to obtain a single representative value.) However, in that case we have run into the problem of correlated samples, which diminishes the data quality drastically and consequently adversely affects the classification performance.

Appendix.

A.1. Gray Level Co-occurrence Matrix

Gray Level Co-occurrence Matrix (GLCM) defines the relationship between the neighbor pixels and shows the occurrence frequency of the brightness levels on the image in a definite distance and direction. GLCMs matrices (shown with $P(i, j,$ distance, angle)) are calculated at multiples of 45-degree angle and defined with the formulas [12]:

$$P(i,j,d,0^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) | k-m=0, |l-n|=d, I(k,l)=i, I(m,n)=j\}$$
$$P(i,j,d,45^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) | (k-m=d, l-n=-d) \text{ or } (k-m=-d, l-n=d), I(k,l)=i, I(m,n)=j\}$$
$$P(i,j,d,90^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) | k-m=0, |l-n|=d, I(k,l)=i, I(m,n)=j|\}$$
$$P(i,j,d,135^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) | (k-m=d, l-n=d) \text{ or } (k-m=-d, l-n=-d), I(k,l)=i, I(m,n)=j\}$$

where $L_x = \{1, 2, \ldots, M\}$ and $L_y = \{1, 2, \ldots, N\}$ respectively are the horizontal and vertical spatial domains of the image, and # shows the number of elements in the set. After P matrices are computed, they are normalized and gray-tone spatial dependence matrices $p(i, j)$ are obtained. Then, Haralick's texture features of this matrix (matrices) are defined:

$$\text{Angular Second Moment: } f_1 = \sum_i \sum_j \{p(i,j)\}^2$$

$$\text{Contrast: } f_2 = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{\substack{j=1 \\ |i-j|=n}}^{G} p(i,j) \right\}$$

$$\text{Correlation: } f_3 = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

$$\text{Sum of Squares: Variance: } f_4 = \sum_i \sum_j (i-\mu)^2 p(i,j)$$

$$\text{Inverse Difference Moment: } f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$$

$$\text{Sum Average: } f_6 = \sum_{i=2}^{2G} i p_{x+y}(i)$$

$$\text{Sum Variance: } f(k,l)$$

$$\text{Sum Entropy: } f_8 = -\sum_{i=2}^{2G} p_{x+y}(i) \log\{p_{x+y}(i)\}$$

Entropy: $f_9 = -\sum_i \sum_j p(i,j)\log(p(i,j))$

Difference Variance: f_{10} = variance of p_{x-y}

Difference Entropy: $f_{11} = \sum_{i=0}^{G-1} p_{x-y}(i)\log\{p_{x-y}(i)\}$

Information Measures of Correlation: $f_{12} = \frac{HXY-HXY1}{\max\{HX,HY\}}$

$f_{13} = (1 - \exp[-2.0(HXY2 - HXY)])^{1/2}$

Maximal Correlation Coefficient: f_{14} = (second largest eigenvalue of $Q)^{1/2}$ where G is the number of distinct gray levels in the quantized image, μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of p_x and p_y, respectively, and

$$HXY = -\sum_i \sum_j p(i,j)\log(p(i,j)),$$

$$HXY1 = -\sum_i \sum_j p(i,j)\log\{p_x(i)p_y(j)\},$$

$$HXY2 = -\sum_i \sum_j p_x(i)p_y(j)\log\{p_x(i)p_y(j)\},$$

$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$$

In addition to Haralick's texture features, two cluster parameters were introduced by Conners et al. [27] in 1984. Autocorrelation and Dissimilarity texture features are defined as [28]:

Autocorrelation: $f_{15} = \sum_i \sum_j (ij)p(i,j)$

Dissimilarity: $f_{16} = \sum_i \sum_j |i-j|p(i,j)$

Cluster Shade: $f_{17} = \sum_i \sum_j (i+j-\mu_x-\mu_y)^3 p(i,j)$

Cluster Prominence: $f_{18} = \sum_i \sum_j (i+j-\mu_x-\mu_y)^4 p(i,j)$

Clausi described and applied following texture features in 2002 [29]:

Inverse Difference Normalized (INN): $f_{19} = \sum_{i,j=1}^{G} \frac{p(i,j)}{1+|i-j|/G}$

Inverse Difference Moment Normalized (IDN): $f_{20} = \sum_{i,j=1}^{G} \frac{p(i,j)}{1+(i-j)^2/G^2}$.

A.2. Gray Level Run Length Matrix

Gray Level Run Length Matrix (GLRLM) is one of the ways to detect the second-degree statistical texture features. The set of sequential pixels that have the same gray level value in the same direction forms the gray level run. Run length is the pixel quantity in the run and run length value is the occurrence number of the runs on the image [13]. While the existence of many neighbor pixels in the same gray level represents a thick-script rough texture, lower quantity of neighbor pixels that have the same gray level represents a thinner texture with a faster change.

GLRLM is a two-dimensional matrix and each $P(i,j,\theta)$ elements show how many times the i gray level runs in length of j in the θ direction occurred. Different texture features have been found by using GLRLM. The most common GLRLM features are below. Galloway [30] proposed the following features for each θ direction:

Short Run Emphasis (SRE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{j^2}$

Long Run Emphasis (LRE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}P(i,j)j^2$

Gray-Level Nonuniformity (GLN): $\frac{1}{S}\sum_{i=1}^{M}\left(\sum_{j=1}^{N}P(i,j)\right)^2$

Run Length Nonuniformity (RLN): $\frac{1}{S}\sum_{j=1}^{N}\left(\sum_{i=1}^{M}P(i,j)\right)^2$

Run Percentage (RP): $\frac{1}{n}\sum_{j=1}^{N}\sum_{i=1}^{M}P(i,j)$.

Chu et al. [31] proposed the following two features:

Low Gray-Level Run Emphasis (LGRE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{i^2}$

High Gray-Level Run Emphasis (HGRE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}P(i,j)\cdot i^2$.

In addition to above seven texture features, Dasarathy and Holder [32] described and applied:

Short Run Low Gray-Level Emphasis (SRLGE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{i^2\cdot j^2}$

Short Run High Gray-Level Emphasis (SRHGE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)\cdot i^2}{j^2}$

Long Run Low Gray-Level Emphasis (LRLGE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)\cdot j^2}{i^2}$

Long Run High Gray-Level Emphasis (LRHGE): $\frac{1}{S}\sum_{i=1}^{M}\sum_{j=1}^{N}P(i,j)\cdot i^2\cdot j^2$

where S is the total number of runs in the image $\left(\left(S = \sum_{i=1}^{M}\sum_{j=1}^{M}P(i,j)\right)\right)$, and n is the number of pixels in the image.

A.3. Gray Tone Difference Matrix

Gray Tone Difference Matrix (GTDM) has been suggested by Amadasun and King [14] and it is a column matrix. Given that $f(k, l)$ has i gray level value in any (k, l) position.

$$\bar{A}_i = \bar{A}(x,y) = \frac{1}{W-1}\sum_{m=-d}^{d}\sum_{n=-d}^{d} f(k+m, l+n), \quad (m,n) \neq (0,0).$$

In the above equation, the average gray level in a definite neighborhood except (k, l) is found, and d shows the neighborhood level

and $W=(2K+1)^2$. The GTDM value at level i is calculated following by [14]:

$$s(i)=\begin{cases}\sum|i-\bar{A}_i| & \text{if } N_i\neq 0\, i\in N_i\\ 0 & \text{otherwise}\end{cases},$$

where N_i is the set of all pixels that have i gray level (except circumferential areas in d width). By using GTDM matrix $s(i)$, the features below are found [14].

$$\text{Coarseness } (C_{cos}):\left(\varepsilon+\sum_{i=0}^{G}p_i s(i)\right)^{-1},\quad p_i=\frac{N_i}{n^2}$$

$$\text{Contrast } (C_{con}):\left[\frac{1}{N_t(N_t-1)}\sum_{i=0}^{G}\sum_{j=0}^{G}p_i p_j(i-j)^2\right]\left[\left(\frac{1}{n^2}\sum_{i=0}^{G}s(i)\right)\right]$$

$$\text{Busyness } (C_{bus}):\frac{\sum_{i=0}^{G}p_i s(i)}{\sum_{i=0}^{G}\sum_{j=0}^{G}|ip_i-jp_j|},\quad p_i\neq 0, p_j\neq 0$$

$$\text{Complexity } (C_{com}):\sum_{i=0}^{G}\sum_{j=0}^{G}\frac{|i-j|}{n^2(p_i+p_j)}[p_i s(i)+p_j s(j)],\ p_i\neq 0, p_j\neq 0$$

$$\text{Texture strength } (C_{str}):\frac{\sum_{i=0}^{G}\sum_{j=0}^{G}(p_i+p_j)(i-j)^2}{\varepsilon+\sum_{i=0}^{G}s(i)},\ p_i\neq 0, p_j\neq 0,$$

where G is the highest gray tone value, N_t is the total number of different gray levels in the image and ε is a small number to prevent C_{cos} becoming infinite.

A.4. Laws' Texture Features

Laws' Texture Features are also used in many different applications [15]. In this approach, first small convolution kernels are applied onto the image, and then features are calculated by following several steps, including a nonlinear windowing operation. Two dimensional (2-D) texture detecting convolution kernels are produced from 1-D kernels known as "Level, Edge, Spot, Wave, and Ripple" abbreviated as L_5, E_5, S_5, W_5, and R_5, of length 5:

$$L_5=\begin{bmatrix}1&4&6&4&1\end{bmatrix},\ E_5=\begin{bmatrix}-1&-2&0&2&1\end{bmatrix},\ S_5=\begin{bmatrix}-1&0&2&0&-1\end{bmatrix},$$
$$W_5=\begin{bmatrix}-1&2&0&-2&1\end{bmatrix},\ R_5=\begin{bmatrix}1&-4&6&-4&1\end{bmatrix}$$

By convolving these 1-D kernels horizontally and vertically with each other, 25 different 2-D convolution kernels are produced:

L5L5	E5L5	S5L5	W5L5	R5L5
L5E5	E5E5	S5E5	W5E5	R5E5
L5S5	E5S5	S5S5	W5S5	R5S5
L5W5	E5W5	S5W5	W5W5	R5W5
L5R5	E5R5	S5R5	W5R5	R5R5

Each of these kernels is separately applied onto the image and 25 filtered images are obtained. Then, each pixel of 25 images is replaced with the Texture Energy Measurement (TEM) value which is obtained by adding the absolute values of the surrounding pixels, i.e., in a square neighborhood centered around the pixel. Since our ROIs were 32 by 32 pixels, we have used assuming neighborhood window of size 5×5 in this step. We call the resultant images as TEM images and denote them with a TEM subscript, as in $L5E5_{TEM}$.

Next, a normalization step is carried out on TEM images to eliminate the possible effect of contrast. All of the above listed 2-D kernels has zero mean except L5L5. Thus, the TEM image obtained from L5L5 kernel is used as normalization image, i.e., all other TEM images are pixel by pixel divided by the L5L5 TEM image. We designate the normalized images as with a TEM_NR subscript, as in $L5E5_{TEM\_NR}$.

In our case, direction of texture features was not important. Therefore we have combined pairing normalized TEM images are to obtain some rotationally invariant texture energy measurements (denoted by TEM_RI) as suggested by Rachidiet al. [33]. For instance, L5E5 and E5L5 images are sensitive to vertical and horizontal edges respectively. By adding these two TEM images, we will have a single measure sensitive to "edge feature". Following a similar argument for other similar features, we obtain our final set of Laws' features as:

$$E5L5_{TEM\_RI}=(E5L5_{TEM\_NR}+L5E5_{TEM\_NR})/2$$
$$S5L5_{TEM\_RI}=(S5L5_{TEM\_NR}+L5S5_{TEM\_NR})/2$$
$$W5L5_{TEM\_RI}=(W5L5_{TEM\_NR}+L5W5_{TEM\_NR})/2$$
$$R5L5_{TEM\_RI}=(R5L5_{TEM\_NR}+L5R5_{TEM\_NR})/2$$
$$S5E5_{TEM\_RI}=(S5E5_{TEM\_NR}+E5S5_{TEM\_NR})/2$$
$$W5E5_{TEM\_RI}=(W5E5_{TEM\_NR}+E5W5_{TEM\_NR})/2$$
$$R5E5_{TEM\_RI}=(R5E5_{TEM\_NR}+E5R5_{TEM\_NR})/2$$
$$W5S5_{TEM\_RI}=(W5S5_{TEM\_NR}+S5W5_{TEM\_NR})/2$$
$$R5S5_{TEM\_RI}=(R5S5_{TEM\_NR}+S5R5_{TEM\_NR})/2$$
$$R5W5_{TEM\_RI}=(R5W5_{TEM\_NR}+W5R5_{TEM\_NR})/2$$
$$E5E5_{TEM\_RI}=E5E5_{TEM\_NR}$$
$$S5S5_{TEM\_RI}=S5S5_{TEM\_NR}$$
$$W5W5_{TEM\_RI}=W5W5_{TEM\_NR}$$
$$R5R5_{TEM\_RI}=R5R5_{TEM\_NR}$$

where TEM_RI subscript denotes final rotational invariant text feature image. Hence, we have obtained a stack of 14 images from the original image, i.e., for each pixel in the original image we have generated total of 14 new values or measurements indicating different texture features at that pixel location.For this method, also the kernels of length-3 can be used [15,34], in which case our starting 1-D filters would be $L_3=\begin{bmatrix}1&2&1\end{bmatrix}$, $E_3=\begin{bmatrix}1&0&-1\end{bmatrix}$, and $S_3=\begin{bmatrix}1&-2&1\end{bmatrix}$.

A.5. Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) of an image matrix is generally calculated by convoluting the image with two filters, which are low-pass and high-pass, throughout its rows and columns. Two-dimensional discrete wavelet transform processes the image with separable filter bank as in the following equation:

$$A_n=[H_x*[H_y*A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}\quad V_n=[G_x*[H_y*A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}$$
$$H_n=[H_x*[G_y*A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}\quad D_n=[G_x*[G_y*A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2},$$

where the * symbol shows the convolution operation, $\downarrow 2$, 1 and $\downarrow 1$, 2 are the down-sampling throughout rows and columns, and $A_0=I(x,y)$ is the original image. H and G respectively show the low-pass and band-pass filters. The approximation image A is called as the low-resolution image in n scale obtained by the low-pass filter. H_n, D_n and V_n are the images in n scale obtained by applying the filters onto the image in specific directions [16].

For each level wavelet transform, an image is decomposed into four sub-bands (LL, LH, HL and HH) which represent the approximation (A), horizontal (H), vertical (V), and diagonal (D) information respectively, and wavelet coefficients are obtained. As each sub-band will reveal different features about the original image, different features can be obtained. Fig. 3 shows DWT computation for one level. Using this technique sub-bands can be produced for different levels. Fig. 4 shows n-level of wavelet decomposition of the image $f(x,y)$.

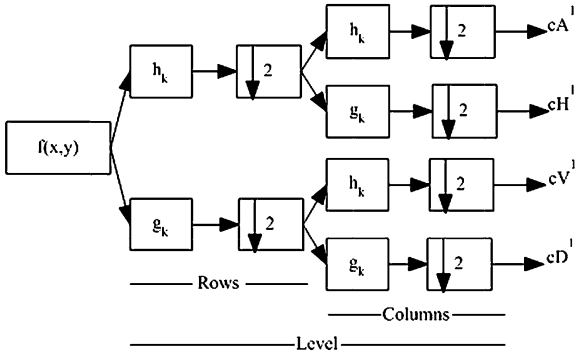

Fig. 3. Decomposition step for one level.

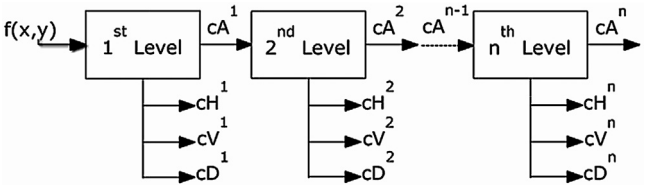

Fig. 4. Block diagram for n-level wavelet decomposition.

A.6. Discrete Fourier transform

2D Discrete Fourier Transform (DFT) of an $M \times N$ image $f(x, y)$ is defined as:

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-2\pi i((ux/M)+(vy/N))}$$

$$u = 0, 1, 2, \ldots, M-1, \quad v = 0, 1, 2, \ldots, N-1, \tag{5}$$

where $F(u, v)$ is the component at frequency (u, v) in the $M \times N$ frequency rectangle. Low frequencies (small gray-filled squares) represent the four corners of matrix F, and the highest frequencies are in the middle of the matrix (see Fig. 5a). As we are interested in the power in certain frequency regions/bands, we first take the magnitude of the 2-D DFT and apply a simple shift operation known as *fftshift*. (Because of the symmetries of the spectrum, the entire quadrant positions can be replaced or shifted diagonally.)

After the *fftshift*, high/low frequency components go to the corners/middle of the matrix respectively (see Fig. 5b). Therefore, four frequency bands LL, LH, HL and HH, where the first and second letters correspond to the horizontal and vertical low/high frequencies respectively, can be obtained as shown in Fig. 5c.We then compute the relative power in these bands, as the ratio of sum of magnitude of DFT coefficients in these particular regions to the grand total in the upper right quadrant (i.e., LL + LH + HL + HH) and use these ratios as features.

Some sample CT images (ROIs) from different fibrosis stages and their corresponding DFT magnitude spectrums obtained from the patients are shown in Fig. 6. In order to suppress or remove possible

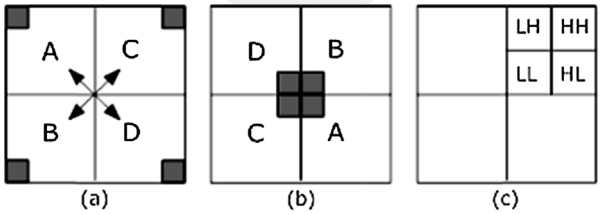

Fig. 5. DFT of an image (a) original spectrum, (b) replaced spectrum, and (c) replaced frequency bands.

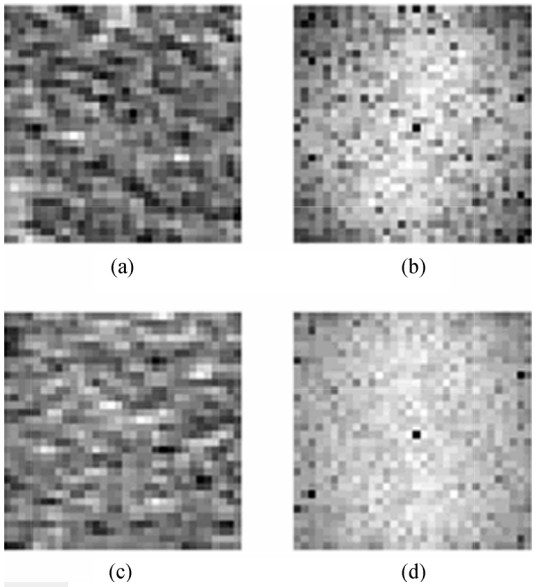

Fig. 6. Sample CT images (left) and their 2-D DFT magnitude spectrums (right). (a) CT liver image of a patient with fibrosis stage 0 and its DFT (b); (c) CT liver image of a patient with fibrosis stage 6 and its DFT (d).

distorting effect of the DC frequency term at $(u, v) = (0, 0)$ position, this component is replaced with a 0, before the relative power computations. This modification is visible in the magnitude spectrums (Fig. 6, panels b and d), where there is a black pixel at the center position.

A.7. Gabor filters

By using the kernels of Gabor filters in various directions and sizes, rotation and size-independent features can be acquired for texture analysis in pattern recognition. Physiological studies show that the cells in the primary cortex of the eye in the human vision system can be modeled with Gabor filters [17], which is the inspiration behind this approach. Gabor filters are obtained by using the following equation [36]:

$$G(x, y, f, \theta) = \exp\left[-\frac{1}{2}\left((x'/s_{x'})^2 + (y'/s_{y'})^2\right)\right]\cos(2\pi f x),$$

$$x' = x\cos(\theta) + y\sin(\theta), \quad y' = y\cos(\theta) - x\sin(\theta)$$

In the above equation, $s_{x'}$ and $s_{y'}$ respectively show the variances in the direction of x' and y' axis (i.e., θ degrees rotated coordinate system), f shows the frequency of sinusoidal function, and θ shows the orientation of Gabor filter. These filters can be produced in any angle and frequency.

In this approach, an image $I(x, y)$ is processed by 2-D convolution operation with a Gabor filter of desired parameters and a filtered (output) image $O(x, y, \theta, s_{x'}, s_{y'})$ is obtained, that is, $O = I * G$, where $*$ denotes the convolution operation. Then, the output image is averaged to produce a single texture feature $t(\theta, s_{x'}, s_{y'}) = \sum_x \sum_y O(x, y, \theta, s_{x'}, s_{y'})$.

A.8. First-Order Statistics

First-order texture measurements are calculated with the original image pixels. They do not take into consideration the relationships with the neighborhood pixel. First-order texture features

such as mean and standard deviation, average energy, skewness, and kurtosis are defined by the equations below:

$$\text{Mean } (\mu): \frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y),$$

$$\text{Standard Deviation } (\sigma): \sqrt{\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-\mu)^2},$$

$$\text{Energy } (e): \frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N} I^2(x,y),$$

$$\text{Skewness: } \frac{1}{M\times N\times \sigma^2}\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-\mu)^3,$$

$$\text{Kurtosis } (k): \frac{1}{M\times N\times \sigma^2}\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-\mu)^4 - 3,$$

where M and N respectively denote image height and width in pixels.

References

[1] T.A. Wynn, Cellular and molecular mechanisms of fibrosis, J. Pathol. 214 (2008) 199–210.
[2] S. Saadeh, G. Cammell, W.D. Carey, Z. Younossi, D. Barnes, K. Easley, The role of liver biopsy in chronic hepatitis C, Hepatology 33 (2000) 196–200.
[3] T. Poynard, V. Ratziu, P. Bedossa, Appropriateness of liver biopsy, Can. J. Gastroenterol. 14 (2000) 543–548.
[4] D.C. Rockey, Noninvasive assessment of liver fibrosis and portal hypertension with transient elastography, Gastroenterology 134 (2008) 8–14.
[5] J.A. Talwalkar, M. Yin, J.L. Fidler, S.O. Sanderson, P.S. Kamath, R.L. Ehman, Magnetic resonance imaging of hepatic fibrosis: emerging clinical applications, Hepatology 47 (2008) 332–342.
[6] Y.N. Sun, M.H. Homg, X. Lin, J.Y. Wang, Ultrasonic image analysis for liver diagnosis, IEEE Eng. Med. Biol. 15 (1996) 93–101.
[7] H. Suiana, S. Swamamani, S. Suresh, Application of artificial neural networks for the classification of liver lesions by image texture parameters, Ultrasound Med. Biol. 22 (1996) 1177–1181.
[8] Y.M. Kadah, A.A. Frag, J.M. Zurada, A.M. Badawi, A.-B.M. Youssef, Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images, IEEE Trans. Med. Imaging 15 (1996) 466–478.
[9] C.M. Wu, Y.C. Chen, K.S. Hsieh, Texture features for classification of ultrasonic liver images, IEEE Trans. Med. Imaging 1 (1992) 141–151.
[10] X. Zhang, H. Fujita, M. Kanematsu, X. Zhou, T. Hara, H. Kato, R. Yokoyama, H. Hoshi, Improving the classification of cirrhotic liver by using texture features, in: IEEE 27th Annual Conference of the Engineering in Medicine and Biology, Shanghai, 2005.
[11] Y. Sun, J. Lu, A. Kobayashi, T. Yahag, Neural network ultrasonographic diagnosis system of cirrhosis using DWT for preprocessing, IEEE Int. Sympos. Circ. Syst. Kobe 3 (2005) 2783–2786.
[12] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. 3 (1973) 610–621.
[13] F. Albregtsen, Statistical Texture Measures Computed from Gray Level Run Length Matrices, Tech. Rep. University of Oslo, 1995.
[14] M. Amadasun, R. King, Textural features corresponding to textural properties, IEEE Trans. Syst. Man Cybern. 19 (1989) 1264–1274.
[15] K. Laws, Textured Image Segmentation, PhD Thesis, University of Southern California, 1980.
[16] G. Van de Wouwer, P. Scheunders, D.V. Dyck, Statistical texture characterization from discrete wavelet representations, IEEE Trans. Image Process. 8 (1999) 592–598.
[17] J.G. Daugman, Two dimensional spectral analysis of cortical receptive field profile, Vision Res. 20 (1980) 847–856.
[18] P. Pudil, J. Novovicova, J. Kittler, Floating search methods in feature selection, Pattern Recognit. Lett. 15 (1994) 1119–1125.
[19] R. Polikar, Pattern recognition, in: Wiley Encyclopedia of Biomedical Engineering, John Wiley & Sons, 2006.
[20] V. Vapnik, The Nature of Statistical Learning Theory, 2nd edn, Springer, 2000.
[21] O. Kayaalti, B.H. Aksebzeci, I.O. Karahan, K. Deniz, M. Ozturk, B. Yilmaz, S. Kara, M.H. Asyali, Staging of the liver fibrosis from ct images using texture features, in: 7th International Symposium on Health Informatics and Bioinformatics (HIBIT), 2012, pp. 47–52.
[22] M. Prokop, M. Galanski, Spiral and Multislice Computed Tomography of the Body, 1st edn., Thieme, New York, 2003.
[23] B.T. Kyongtae, Test bolus versus Bolus–Tracking techniques for CT angiographic timing, Radiology 236 (2005) 369–370.
[24] T. Murakami, T. Kim, S. Takahashi, H. Nakamura, Hepatocellular carcinoma multi-detector row helical CT, Abdom. Imaging 27 (2002) 139–146.
[25] K. Ishak, A. Baptista, L. Bianchi, F. Callea, J. De Groote, F. Gudat, H. Denk, V. Desmet, G. Korb, R.N. MacSween, et al., Histological grading and staging of chronic hepatitis, J. Hepatol. 22 (1995) 696–699.
[26] C.H. Chen, L.F. Pau, P.S.P. Wang, The Handbook of Pattern Recognition and Computer Vision, 2nd edition, World Scientific Publishing Co., 1998, pp. 207–248.
[27] R.W. Conners, M.M. Trivedi, C.A. Harlow, Segmentation of a high-resolution urban scene using texture operators, Comput. Vision Graph. Image Process. 25 (1984) 273–310.
[28] L. Soh, C. Tsatsoulis, Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices, IEEE Trans. Geosci. Remote Sens. 37 (1999) 780–795.
[29] D.A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. Remote Sensing 28 (2002) 45–62.
[30] R.M.M. Galloway, Texture analysis using gray level run lengths, Comput. Graph. Image Process. 4 (1975) 172–179.
[31] A. Chu, C.M. Sehgal, J.F. Greenleaf, Use of gray value distribution of run lengths for texture analysis, Pattern Recognit. Lett. 11 (1990) 415–420.
[32] B.R. Dasarathy, E.B. Holder, Image characterizations based on joint gray-level run-length distributions, Pattern Recognit. Lett. 12 (1991) 497–502.
[33] M. Rachidi, A. Marchadier, C. Gadoi, E. Lespessailles, C. Chappard, C.L. Benhamou, Laws' masks descriptors applied to bone texture analysis: an innovative and discriminant tool in osteoporosis, Skeletal Radiol. 37 (2008) 541–548.
[34] K. Laws, Rapid texture identification SPIE Image Processing for Missile Guidance, vol. 238, 1980, pp. 376–380.
[35] O. Demirkaya, M.H. Asyali, P.K. Sahoo, Image Processing with MATLAB®: Applications in Medicine and Biology, CRC Press, 2008.
[36] L. Hong, Y. Wan, A. Jain, Fingerprint image enhancement: algorithm and performance evaluation, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 777–789.
[37] R.O. Duda, P.E. Hart, D. Stork, Pattern Classification, 2nd edition, Wiley and Sons, 2001.
[38] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, in: Data Mining and Knowledge Discovery, Kluwer Academic Publishers, Boston, 1998, pp. 121–167.
[39] M.A. Hearst, et al., Support vector machines, IEEE Intell. Syst. Appl. 13 (1998) 18–28.
[40] B. Schölkopf, A.J. Smola, Learning with Kernels, MIT Press, 2002.
[41] C.W. Hsu, C.J. Lin, A comparison of methods for multi-class support vector machines, IEEE Trans. Neural Netw. 13 (2002) 415–425.
[42] E.I. Karaç, Model Selection for Multi-Class Support Vector Machines, MS thesis, Boğaziçi University, 2005.
[43] M.H. Asyali, D. Colak, O. Demirkaya, et al., Gene expression profile classification: a review, Curr. Bioinform. 1 (2006) 55–73.
[44] A. Jain, P. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. PAMI 22 (2000) 4–37.
[45] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 252–264.
[46] P. Baim, A method for attribute selection in inductive learning systems, IEEE Trans. Pattern Anal. Mach. Intell. 10 (1988) 888–896.
[47] A.W. Whitney, A direct method of nonparametric measurement selection, IEEE Trans. Comput. C-20 (1971) 1100–1103.
[48] The MathWorks, 2014, http://www.mathworks.com (accessed 15.6.14).
[49] A Uppuluri, GLCM Texture Features, 2008, http://www.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features (accessed 15.6.14).
[50] X. Wei, Gray Level Run Length Matrix Toolbox v1.0. Software, Beijing Aeronautical Technology Research Center, 2007.
[51] R.E. Bellman, Adaptive Control Processes, Princeton University Press, 1961.
[52] P.M. Narendra, K. Fukunaga, A branch and bound algorithm for feature subset selection, IEEE Trans. Comput. C-26 (1977) 917–922.
[53] T. Marill, D.M. Green, On the effectiveness of receptors in recognition system, IEEE Trans. Inform. Theory 9 (1963) 11–17.
[54] L.G. Lu, M.D. Zeng, M.B. Wan, C.Z. Li, Y.M. Mao, J.Q. Li, D.K. Qiu, A.P. Cao, J. Ye, X. Cai, C.W. Chen, J.Y. Wang, S.M. Wu, J.S. Zhu, X.Q. Zhou, Grading and staging of hepatic fibrosis, and its relationship with noninvasive diagnostic parameters, World J. Gastroenterol. 9 (2003) 2574–2578.
[55] Y. Chen, B.E. Wang, J.D. Jia, L.X. Qian, T.L. Wang, M.H. Chen, G.Y. Chen, W. He, H.G. Ding, S.S. Yin, Y. Zhang, Z. Dong, Noninvasive evaluation of liver fibrosis in chronic hepatitis B patients, Chin. J. Hepatol. 11 (2003) 354–357.