Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm.

ARTICLE *in* MEDICAL PHYSICS · JULY 2012

Impact Factor: 3.01 · DOI: 10.1118/1.4725759 · Source: PubMed

7 AUTHORS, INCLUDING:

Ricardo Ribeiro
Instituto Politécnico de Lisboa
15 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Rui Tato Marinho
University of Lisbon
127 PUBLICATIONS **585** CITATIONS
SEE PROFILE

João M Sanches
Technical University of Lisbon
135 PUBLICATIONS **447** CITATIONS
SEE PROFILE

Jasjit Suri
269 PUBLICATIONS **2,081** CITATIONS
SEE PROFILE

Data Mining Framework for Fatty Liver Disease Classification in Ultrasound: A Hybrid Feature Extraction Paradigm

U Rajendra Acharya , S Vinitha Sree, Ricardo Ribeiro, Luca Saba, Ganapathy Krishnamurthi, João Sanches, Jasjit S. Suri, *Sr. Member, IEEE, Fellow AIMBE*

***Abstract*—Fatty Liver Disease (FLD) is an increasing prevalent disease that can be reversed if detected early. Ultrasound is the safest and ubiquitous method for identifying FLD. Since expert sonographers are required to accurately interpret the liver ultrasound images, lack of the same will result in inter-observer variability. For more objective interpretation, high accuracy, and quick second opinions, Computer Aided Diagnostic (CAD) techniques may be exploited. In this paper, we present a CAD technique (a class of Symtosis, by Global Biomedical Technologies Inc., CA) that uses significant features such as texture, wavelet transform and higher order spectra in various supervised learning based classifiers in order to determine parameters that classify normal and FLD-affected abnormal livers. On evaluating the proposed technique using 20 abnormal and 15 normal liver ultrasound images, we were able to achieve a high classification accuracy of 93.3% using several classifiers. This high accuracy added to the completely automated classification procedure makes our proposed technique highly suitable for clinical deployment and use, once the technique is verified on a larger database of liver images. We have also proposed a novel *SteatosisClassificationIndex* (*SCI*) that uses a combination of the significant features to output a single number which can be used to determine the class of the tested image. This index can also be easily incorporated into a clinical setting at no extra cost and it can provide additional diagnostic confidence to the physician.**

***Index Terms*— Fatty Liver Disease, Steatosis, CAD, texture, Higher Order Spectra, Discrete Wavelet Transform, classifiers.**

Manuscript received May 31, 2011.

U. Rajendra Acharya is with the Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore (email: aru@np.edu.sg)
S Vinitha Sree is with Global Biomedical Technologies Inc., CA, USA (email: vinitha.sree@gmail.com)
Ricardo Ribeiro is with Institute for Systems and Robotics and Escola Superior de Tecnologia da Saúde de Lisboa
Luca Saba is with the Department of Radiology, Azienda Ospedaliero Universitaria di Cagliari, Cagliari, Italy
Ganapathy Krishnamurthi is with Case Western Reserve University in Cleveland, OH, USA
João Sanches is with Institute for Systems and Robotics and Instituto Superior Técnico, Lisbon, Portugal
Jasjit S. Suri, Fellow AIMBE, is a CTO with Global Biomedical Technologies, CA, USA and is also affiliated with Biomedical Engineering Department, Idaho State University, ID, USA (jsuri@comcast.net).

I. INTRODUCTION

Fatty Liver Disease (FLD) or hepatic steatosis is a condition which is characterized by the presence of vacuoles of triglyceride fat in liver cells. This accumulation of fat happens through a process called steatosis in which there is abnormal retention of lipids in the cells. Some of the key causes of FLD are chronic alcohol consumption, obesity due to insulin resistance, and metabolic syndrome [1, 2]. There are two major types of FLD based on the contribution of alcohol, namely alcoholic steatosis and Non-Alcoholic Fatty Liver Disease (NAFLD). NAFLD is progressively prevalent in Western countries and affects people of all ages and ethnicities [3, 4]. Both alcoholic and non-alcoholic FLD, if left undetected and untreated, will progress to advanced liver diseases like inflammation (steatohepatitis), cirrhosis, and liver cancer. However, if found and treated early, FLD may be reversible. Therefore, early detection is of utmost importance in order to save patients from unwanted anxiety and also to reduce costs associated with providing treatments for advanced liver diseases. Liver biopsy is currently the standard for the assessment of steatosis. It is, however, invasive, uncomfortable, and prone to sampling errors [5-7]. The noninvasive techniques include ultrasound, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI). Even though these methods have shown promise in detecting fatty infiltration in the liver, they are insensitive in detecting steatosis of less than 25% to 30% [8]. In the case of ultrasound, FLD causes increased echogenicity on ultrasound causing the liver to appear brighter than the cortex of the ipsilateral kidney [9]. Ultrasound has a sensitivity of around 82% to 94% and specificity greater than 82% for detecting a fatty liver [10-13]. Even though ultrasound is more sensitive than CT [14], it is less specific and also has poor visualization in obese patients.

In the case of CT imaging, hepatic steatosis produces an attenuation that is lower for the hepatic parenchyma than for the surrounding blood vessels, spleen and the kidneys, thereby enabling the visualization of the presence of the steatotic liver. CT is limited by the calibrations required for different scanners and inter-observer variabilities. MRI presents anatomical information of the imaged liver and Magnetic Resonance Spectroscopy (MRS) provides a biochemical component [15]. MR has demonstrated that it can detect small fat fractions of less than 33% [16]. However, MR techniques, especially MRS, can be technically challenging. A good MRS of the liver requires good spatial resolution, high SNR, and adequate compensation for or elimination of patient motion.

Among all these modalities, ultrasound is the most commonly used modality due to its widespread availability in current clinical practice. In order to improve the specificity of ultrasound and also to address the inter-observer variability issue that is common in medical image interpretation, Computer Aided Diagnostic (CAD) or data mining techniques can be developed to more objectively and accurately detect the presence or absence of FLD in ultrasound images of liver. These techniques use the acquired ultrasound images to extract meaningful and discriminative features that are capable of adequately distinguishing a normal liver from an abnormal liver that is affected by FLD. These features are fed to supervised-learning based classifiers to train the classifiers to enable them to learn the parameters that effectively differentiate the patterns belonging to either of the classes. Thereafter, these learned parameters are used to classify new liver images into normal and abnormal categories. FLD affects the entire liver or

a lobe of the liver, and hence, causes changes in the texture of the liver in the B-mode US images. As indicated earlier, due to failure in fat metabolism, there is an increase in the deposition of fat in the liver which in turn gives rise to an increased brightness in the ultrasound and results in changes in the image texture [2]. Therefore, texture of the image has been studied as one of the key distinguishing features in this work. We have also extracted Higher Order Spectral (HOS) features that characterize the non-linearity in the images and Discrete Wavelet Transform (DWT) based features that quantify the changes occurring in the time-frequency domain of the images.

The block diagram of the proposed Symtosis technique is shown in Fig.1. In the offline training system, the acquired ultrasound liver images in the training set are pre-processed by computing the three set of features: HOS, Texture and DWT. In order to reduce the dimensionality of the extracted feature set and to select only unique and highly discriminating features, the extracted features are subjected to the Student's *t-test* and only significant features are selected to form the final feature set. The significant feature set and the ground truth of whether the images belong to normal or abnormal cases (as predicted by doctors or by lab results) are used as inputs to several supervised-learning based classifiers in order to train them to determine appropriate parameters for differentiating both classes based on the features. In the online real-time system, the test images are pre-processed and significant features (similar to offline) are determined. The training parameters from the offline system are used on the features to determine the class of the images. The resultant class labels are used to determine five performance measures, namely, accuracy, sensitivity, specificity, and Positive Predictive Value (PPV), and Area under the Receiver Operating Characteristic curve (AUC).

The key contributions of this paper are the (a) development of a completely automated CAD technique for detection of FLD in ultrasound liver images; (b) high classification accuracy using very less number of representative features; and (c) development of a novel integrated single-valued index made up of the significant features for more objectively determining the class of the image quickly and accurately.

The paper is organized as follows. In Section II, we describe the data, extracted features, statistical techniques, and the classifiers. In Section III, we present the significant features and classification results, and propose the novel integrated index and show its range for both the normal and abnormal classes. Section IV summarizes key findings in related studies in literature and compares and discusses the results obtained in this work. We conclude the paper in Section V.

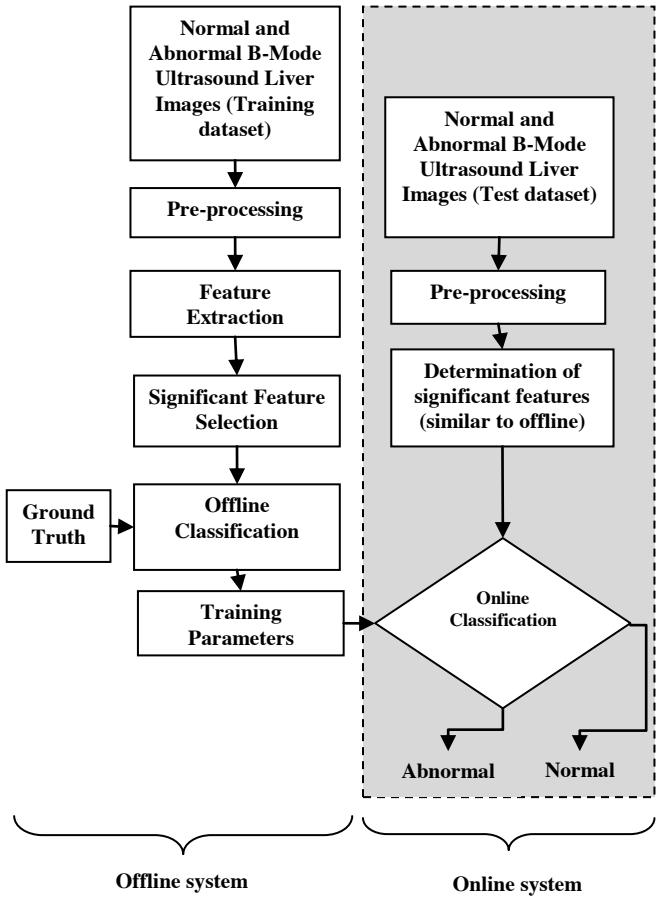

Fig. 1 Block diagram of the proposed Symtosis system for fatty liver disease detection; The blocks outside the dotted shaded rectangular box represent the flow of offline training system, and the blocks within the dotted box represent the online real-time system.

II. Materials and Methods

In this section, we describe the data used in this work, and present brief descriptions of the features extracted, statistical techniques used, and the classifiers evaluated.

A. Patient Data

Thirty five ultrasound liver images were used for classifier development and evaluation in this work. Among these 35 cases, 20 were abnormal (affected by FLD) and 15 were normal images. The ultrasound images were acquired by expert operators in a hospital facility, where one image was extracted from the hepatic parenchyma from each patient and stored in DICOM format. The ground truth as to whether each image was normal or abnormal was determined manually by the operators and confirmed by indicators obtained from laboratory analysis. Typical images of normal and abnormal liver are shown in Fig. 2.

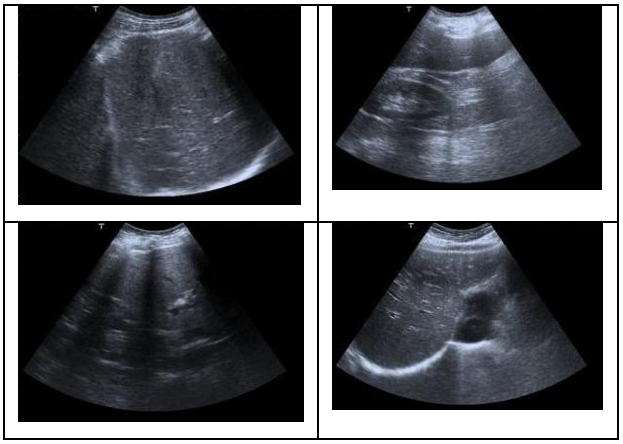

Fig. 2 Normal liver images (left) and abnormal liver images (right)

B. Feature Extraction

Higher Order Spectra (HOS)-based features

Higher Order Spectra based features quantify the nonlinear behavior of a process [17]. Higher order statistics deal with higher order moments (order greater than two) and nonlinear combinations of these higher order moments, called the higher order cumulants. The bispectrum, which is the spectrum of the third order cumulants, is one of the most commonly used HOS features. Prior to the calculation of the bispectrum, the pre-processed images were first subjected to Radon transform [18]. This transform determines the line integrals along many parallel paths in the image from different angles θ by rotating the image around its centre. Hence, the intensities of the pixels along these lines are projected into points in the resultant transformed signal. Thus, the Radon transform converts a 2D image into a 1D signal at various angles. This 1D signal is then used to determined the bispectrum, which is a complex valued function of two frequencies given by

$$B(f_1, f_2) = E\left[A(f_1)X(f_2)^* X(f_2 + f_2)\right] \quad (1)$$

where *X(f)* is the Fourier transform of the signal studied. As per the equation, the bispectrum is the product of the three Fourier coefficients. The function exhibits symmetry, and is computed in the non-redundant/ principal domain region Ω as shown in Fig. 3.

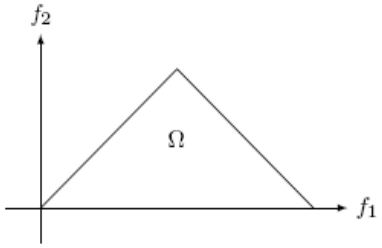

Fig. 3. Principal domain region (Ω) used for the computation of the bispectrum for real signals.

The bispectrum phase entropy [19-21] obtained from the bispectrum is used as one of the features in this work. This bispectrum phase entropy is defined as

$$ePRes = \sum_n p(\psi_n) log p(\psi_n) \tag{2}$$

where

$$p(\psi_n) = \frac{1}{L} \sum_{\Omega} l(\phi(b(f1, f2)) \in \psi_n) \tag{3}$$

$$\psi_n = \{\phi \mid -\pi + 2\pi n / N \le \phi < -\pi + 2\pi(n+1)/N\}, \quad n = 0,1,..., N-1 \tag{4}$$

where L is the number of points within the region Ω, ϕ is the phase angle of the bispectrum, and *l(.)* is an indicator function which gives a value of 1 when the phase angle is within the range depicted by ψ_n in equation (4). In this work, we calculated the Radon transformed signals for every 1º step size and then determined the phase entropy of these signals.

Texture-based features

In most images processing applications, assumptions are made regarding the uniformity of gray-level intensity values in the image. In real applications, most images have a variation in gray levels which are repetitive and these variations are characterized as the texture of the image [22]. The most commonly used texture matrices are the Gray Level Co-occurrence Matrix (GLCM) and the run length matrix. We have calculated one homogeneity feature from the GLCM [23] and three features from the run length matrix [24]. These features are described briefly below.

Homogeneity

The gray level co occurrence matrix defined as follows

$$C_d(i, j) = \left| \begin{Bmatrix} (p,q),(p+\Delta x, q+\Delta y) : I(p,q) = i \\ ,I(p+\Delta x, q+\Delta y) = j \end{Bmatrix} \right| \tag{5}$$

where *(p, q), (p+Δx, q+Δy)* belong to *m* x *n*, *d = (Δx, Δy)* and |...| denotes the set cardinality. The probability of a pixel with a gray level intensity value *i* having a pixel with a gray level intensity value *j* at a distance *(Δx, Δy)* away in an image is defined as

$$P_d(i, j) = \frac{C_d(i, j)}{\sum_{<i>} \sum_{<j>} C_d(i, j)} \tag{6}$$

The homogeneity of the image is now defined as

$$C_h = \sum_i \sum_j \left(\frac{P_d[i, j]}{1 + |i - j|} \right) \tag{7}$$

The homogeneity measures the closeness of the distribution of the co-occurrence matrix elements to the main diagonal. A homogenous image will give rise to a $P_d(i,j)$ clustered around the main diagonal. In other words, the similarity between two pixels that are $(\Delta x, \Delta y)$ apart is measured by the homogeneity feature.

Run Percentage (TexRL)

The run percentage is a texture property derived from the run length matrix of an image. The run length matrix P_θ contains all the elements where the gray level value i has the run length j continuous in direction θ [24]. Often the direction θ is set as 0^0, 45^0, 90^0, or 135^0. The run percentage is defined as the total number of runs in the image divided by the total number of pixels in the image as depicted in equation (8).

$$TexRL = \frac{\sum_{i=1}^{N_G}\sum_{j=1}^{N_r} P_\theta(i,j)}{N_p} \tag{8}$$

Run percentage has the lowest value for images with the most linear structure. Here $P_\theta(i,j)$ is the element of the run length matrix, N_p is the total number of pixels in the image, N_g is the number of gray levels in the image, and N_r is the number of different run lengths that occur.

Short Run Emphasis (SRE)

Based on the run length matrix, the short run emphasis is defined as

$$SRE = \frac{\sum_{i-1}^{N_g}\sum_{j=1}^{N_r} \frac{P_\theta(i,j)}{j^2}}{\sum_{i-1}^{N_g}\sum_{j=1}^{N_r} P_\theta(i,j)} \tag{9}$$

where the index i runs over the gray level values in the image and the index j runs over the run length. Higher values of j i.e. long run lengths will contribute less to the sum in equation (9) and consequently higher sum emphasizes short runs.

Gray Level Non Uniformity (GLNU)

The gray level non uniformity is defined as

$$GLNU = \frac{\sum_{i=1}^{N_g}\left(\sum_{j=1}^{N_r} P_\theta(i,j)\right)^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P_\theta(i,j)} \tag{10}$$

The gray level non uniformity squares the run lengths for each gray value. Hence, longer run lengths will make significant contributions to the summation *i.e.* uniform images will have higher values of this sum as compared to images that are non-uniform in their gray levels.

DWT-based features

A wavelet transform is the representation of a function by wavelets, which are scaled and translated copies of a basic wavelet shape called the 'mother wavelet'. Mother wavelets are functions that are localized in both time and frequency and have varying amplitudes during a limited time period and very low or zero amplitude outside that time period. Wavelet transforms like Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), and Wavelet Packet Decomposition (DWT) determine a limited number of wavelets coefficients that adequately describe the image. 2D DWT was used in his work. DWT analyzes the image at different frequency bands with different resolutions by decomposing the image into coarse approximation and detail information. The approximation coefficients are obtained by passing the image through a Low Pass Filter (LPF), and the detail coefficients are obtained by filtering the image using a High Pass Filter (HPF). This decomposition is done recursively on the low pass approximation coefficients obtained at each level until the desired number of iterations is reached.

An illustration of DWT is given in Fig. 4. The rows of the image I are convolved using a LPF and the columns of the convolved output are down-sampled *i.e.*, only the even indexed columns are retained for further filtering. Next, the down-sampled columns are passed through another LPF, the output of which is again sampled to keep the even indexed rows alone. These are the approximation coefficients cA_1 at level 1. Similarly, the down-sampled columns are passed through a HPF, sampled to retain the even indexed rows alone to get the horizontal detail coefficients cH_1. In a similar fashion, the rows of the image I are high passed filtered and processed through a set of low pass and high pass filters to get the vertical detail coefficients cV_1 and diagonal detail coefficients cD_1, respectively. In our work, we calculated the averages of each set of coefficients cA_1, cH_1, cV_1 and cD_1 at level 1, and again found the average of these individual averages. This overall average value was used as a feature.

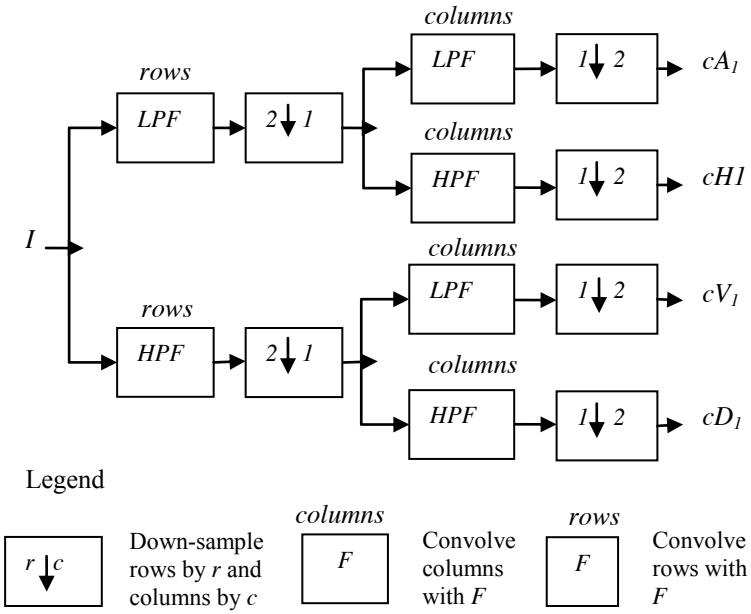

Fig. 4. Discrete Wavelet Transform (DWT) decomposition.

C. Classifiers

k-Nearest Neighbour (KNN)

The KNN classifier, one of the simplest classifiers, is based on determining the minimum distance from a test sample to the training samples. The *k*-nearest neighbours are first determined using this method. After determining these *k*-nearest neighbours, the test sample is assigned to the class that is most common amongst its *k*-nearest neighbors [25].

Radial Basis Probabilistic Neural Network (PNN)

PNN is a specific type of two layer radial basis network. The first layer of neurons has radial basis activation functions. When a test sample is presented, this layer computes the distance vector by calculating the distances between the sample and training vectors. The output of this layer will be a vector whose elements indicate how close the sample is to a training input. The second layer (competitive layer) sums the contributions of these vectors for each input classes and produces a vector of probabilities as its output. The so called *compete* transfer function, at the output of the second layer, selects the maximum of these probabilities and assigns a 1 for the selected class and a 0 for all other classes [26].

Gaussian Mixture Model (GMM)

In the GMM classifier, the conditional Probability Density Functions (PDF) of the feature vector with respect to the different classes is modeled as a linear mixture of multivariate Gaussian PDFs. The training data is used to determine the parameters of the GMM for each class, generally using Expectation Maximization algorithm [27]. The trained GMMs are then used to determine the class of new test samples.

Naive Bayes Classifier (NBC)

NBC is a simple probabilistic supervised learning based classifier that is based on Bayes theorem. Since this classifier strongly assumes that the features are independent random variables, it is easier to compute probabilities required by the Bayes formula from even a small training data [28].

D. Statistical Analysis

In order to select unique and highly discriminating features, the Student's *t-test* was used to select the features that were significantly different between the normal and abnormal cases. In this test, initially, for each feature, the null hypothesis is assumed to consider that the mean of the feature from the normal class is equal to the mean of the feature from the abnormal class. Subsequently, the *t-statistic*, which is the ratio of difference between the means of two classes to the standard error between class means, and the corresponding *p-value* are calculated. The *p-value* is the probability of rejecting the null hypothesis given that the null hypothesis is true. A low *p-value* (less than 0.01 or 0.05) indicates rejection of null hypothesis, which implies that the means are not equal in both classes and are significantly different, and hence, the feature is significant.

Sensitivity, specificity, positive predictive value, and accuracy were calculated to evaluate the performance of the classifiers. Another important performance measure is the Area under the Receiver Operating Characteristic (ROC) curve, called AUC. The ROC curve is obtained by calculating the sensitivity and specificity of a classifier at different cut-off values and plotting sensitivity vs. (1-specificity) [29]. (1-specificity) is called the False Positive Rate (FPR). A classifier that perfectly discriminates between the two classes would yield a curve that coincides with the left and top sides of the plot. This means that sensitivity is high and the FPR is low. A classifier that is completely useless would give a straight line that follows a diagonal path from the bottom left corner to the top right corner. Generally, the curve will lie somewhere between these extremes because of the overlap of the values in the two classes. The goodness of a classifier is assessed by determining the Area under the ROC curve (AUC). For an ideal test, the AUC would be 1. For a useless classifier, which follows the diagonal ROC curve, the AUC would be 0.5 which is equivalent to having sensitivity and specificity of 0.5 (50%). Hence, in practice, the closer the AUC is to 1.0, the better the classifier is, and the closer the AUC is to 0.5, the worse the classifier is [30].

III. RESULTS

A. Significant Features

As shown in Table I, all the three selected features had statistically significant differences between the abnormal and normal classes, as indicated by the low *p-value* (< 0.01). The table also presents the mean and standard deviation of all the features. In the case of HOS-based features, one phase entropy based feature obtained for Radon transform angle $\theta = 93^0$, denoted in Table I as *ePRes*(93^0), was found to be significant. In the case of texture features, only the Short Run Emphasis (*SRE*) was found to be significant. To obtain the DWT features, around 54 mother wavelets were studied to find the mean value of the level 1 coefficients. Among them, the mean of the coefficients obtained at

level one of decomposition using the *sym4* mother wavelet was found to be significantly different between the two classes. In the case of abnormal images, both the *SRE* and DWT feature ($DWTMean1_{sym4}$) have registered lower values compared to that of the normal cases. Moreover, in the case of phase entropy ($ePRes(93^0)$), the abnormal cases have higher values than normal cases. This higher value is due to the fact that in abnormal images, the variation among pixel intensities is higher, and hence the entropy is higher due to the complex texture of the images.

TABLE I

MEAN ± STANDARD DEVIATION (SD) VALUES OF THE SIGNIFICANT FEATURES FOR THE NORMAL AND ABNORMAL CLASSES

Features	Normal (Mean ± SD)	Abnormal (Mean ± SD)	*p-value*
SRE	0.920 ± 1.007E-02	0.901± 1.023E-02	< 0.0001
$ePRes(93^0)$	2.42 ± 0.504	2.90 ± 0.468	0.0058
$DWTMean1_{sym4}$	19.5± 5.57	14.6 ± 4.81	0.0089

B. Classification Results

In view of the low sample size, three-fold stratified cross validation was employed to obtain robust classifiers. In this resampling technique, the entire dataset is split into three equal parts, each part containing the same proportion of samples from both the classes. In the first fold, two parts of the data are used for training the classifier, and the remaining one part is used for testing the trained classifier and to obtain the performance measures. This procedure is repeated twice, using a new test set each time. The average of the performance measures obtained during each fold is taken to be the final values of the performance measures. To be specific, approximately four normal and six or seven abnormal cases are used in each fold. Classification accuracy, sensitivity, specificity, PPV, and AUC were used as the performance measures to select the optimal classifier for this work. Table II presents the classification results obtained using all features except the HOS feature. Table III presents the classification results obtained using all features including the HOS feature.

TABLE II

SYMTOSIS CLASSIFICATION RESULTS OBTAINED USING ALL FEATURES EXCEPT THE HOS FEATURE (THE LISTED VALUES ARE AVERAGE OF VALUES OBTAINED IN THE THREE FOLDS) TN-TRUE NEGATIVES, FN - FALSE NEGATIVES, TP - TRUE POSITIVES, FP - FALSE POSITIVES, A – ACCURACY, PPV - POSITIVE PREDICTIVE VALUE, SN – SENSITIVITY, SP - SPECIFICITY

	TN	FN	TP	FP	A (%)	PPV (%)	Sn (%)	Sp (%)
GMM	3	1	5	1	76.7	83.3	77.8	75
KNN	4	2	4	0	80	93.3	72.2	91.7
NBC	3	1	5	1	76.7	84.9	77.8	75
PNN	4	2	4	0	80	93.3	72.2	91.7

TABLE III

SYMTOSIS CLASSIFICATION RESULTS OBTAINED USING ALL FEATURES INCLUDING THE HOS FEATURE (THE LISTED VALUES ARE AVERAGE OF VALUES OBTAINED IN THE THREE FOLDS) TN-TRUE NEGATIVES, FN - FALSE NEGATIVES, TP - TRUE POSITIVES, FP - FALSE POSITIVES, A – ACCURACY, PPV - POSITIVE PREDICTIVE VALUE, SN – SENSITIVITY, SP – SPECIFICITY

	TN	FN	TP	FP	A (%)	PPV (%)	Sn (%)	Sp (%)	AUC
GMM	4	1	5	0	86.7	94.4	83.3	91.7	0.917
KNN	4	0	6	0	93.3	95.2	94.4	91.7	1.000
NBC	4	1	5	0	90	94.4	88.9	91.7	0.917
PNN	4	0	6	0	93.3	95.2	94.4	91.7	1.000

The maximum accuracy that could be achieved using all the features except the HOS feature (*ePRes(93^0)*) was only 80% using the KNN and PNN classifiers (Table II). This accuracy substantially improved to 93.3% on including the HOS feature for training and testing the classifiers KNN and PNN (Table III). Because of the variations in the number of False Positives (FP) and False Negatives (FN), many classifiers having the same accuracy can have different values for sensitivity and specificity. Sensitivity denotes the proportion of abnormal samples classified as abnormal, and specificity represents the proportion of the normal samples classified as normal. Even though achieving a 100% value for these measures is highly difficult, both these measures must at least be high for a classifier to be termed as an optimal one. The KNN and PNN classifiers have recorded high accuracy (93.3%),

and equally high values for both sensitivity (94.4%) and specificity (91.7%). These classifiers are not biased towards accurate detection of one class alone. Moreover, the average AUC of these classifiers is 1, which indicates the excellent performance of these classifiers. The ROC curves are also depicted in Fig 5. Both the PNN and KNN classifiers have the ROC curves that coincide with the left and top corners of the plot indicating the high AUC.

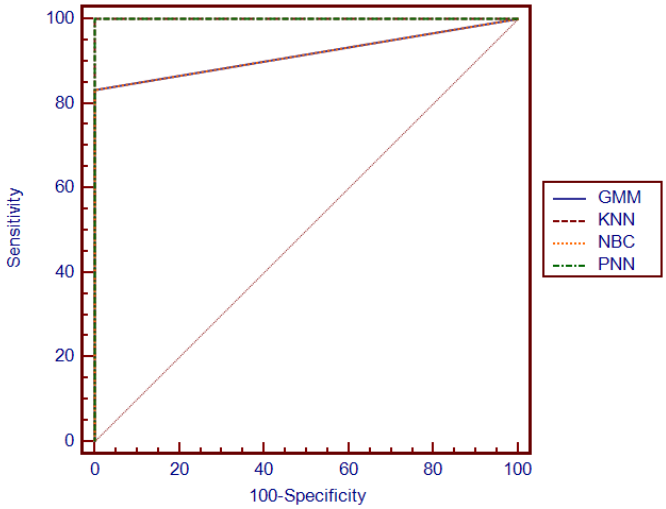

Fig. 5. ROC curves of the classifiers

C. *Integrated Index*

It is difficult to monitor and interpret the variations in each individual feature listed in Table I. Therefore, we have attempted to formulate a novel index, called the *SteatosisClassificationIndex (SCI)*, that integrates all the three significant features into one function (shown in equation (11)), such that the combination of these features results in an almost unique range for both the classes. Such indices were earlier developed for various medical issues [31, 32]. In equation (11), β=20, and δ=10.

$$SCI = \beta \times SRE - \frac{1}{\delta} log_{10}(\chi) \tag{11}$$

where $\chi = DWTMean1_{sym4} \times ePRes93$ (12)

This integrated index is a single-valued index that can be used to more objectively, quickly, and efficiently detect the normal and abnormal cases. Another key advantage of this index is that it can be easily incorporated into even commonly available software like Microsoft Excel, and the physicians would just have to input the values of the three features in order to calculate the index. Thus, it can be used to quickly give a second opinion to the physician. The ranges of this index summarized in Table IV show significant statistical differences between the normal and abnormal cases. This difference is also depicted in Fig. 6. It is evident from Table IV and Fig. 6 that the normal class

has higher value for this index than the abnormal images. The reason for this difference may be attributed to that fact that the SRE feature has a higher value for the normal class as SRE is generally high for images with fine texture like the normal B-mode images.

TABLE IV

RANGE OF *SCI* FOR NORMAL AND ABNORMAL CASES

	Normal (Mean ± SD)	Abnormal (Mean ± SD)	*p-value*
SCI	18.2 ± 0.191	17.9 ± 0.194	< 0.0001

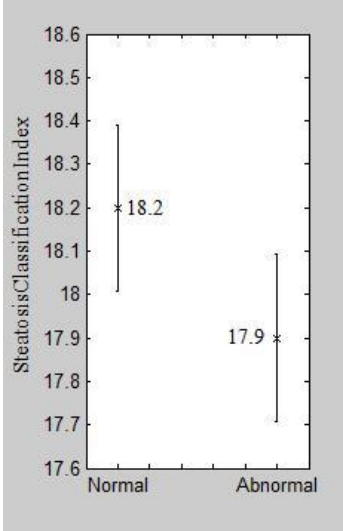

Fig. 6. Plot of the range of *SteatosisClassificationIndex* for both normal and abnormal classes

IV. DISCUSSION

Very few studies have been carried out to automatically classify diffuse liver diseases [33, 34]. Kadah *et al*. [33] performed a study where they evaluated the relative performance of several classifiers in classifying normal, fatty, and cirrhotic liver conditions. They concluded that KNN classifier and functional link networks showed the best results. The study was done on a large patient population (120 cases) with equal prevalence of diseased (fatty and cirrhotic) and normal livers. Moreover, they analyzed a three-class problem, while in this work we focused on a two-class problem. Ribeiro *et al*. [34] used acoustic attenuation coefficients, wavelet coefficients and AR model coefficients as features in a Bayes classifier for classifying steatosis and normal livers. Their data set comprised of 44 ultrasound liver images obtained from 22 patients. Among them, 12 had steatotic liver and 10 had normal liver. Using all the features, they obtained very high sensitivity (~95%), specificity (~95%), and accuracy (~95%). Using only the AR model coefficients, the accuracy, sensitivity, and specificity dropped to 90%, 87%, and 95%,

respectively. Using the same features, the KNN classifier showed poor results compared to the Bayes classifier. All the texture features in this case were derived from the speckle image obtained from envelope RF images, whereas in our work we used the B-mode ultrasound images directly for feature extraction. This reduces the computational complexity of the algorithm. In [35], three types of liver images, namely, normal, hepatoma, and cirrhosis, were classified using fractal feature vector based on M-band wavelet transform. Having tested their methodology using various classifiers, they observed that a hierarchical classifier was least 96.7% accurate in the classifying normal and abnormal liver images. The limitation of this work is that prior to the image analysis, the region of interest covering the liver parenchyma without major blood vessels, acoustic shadowing, or any type of distortion was chosen manually by a physician. Hence, the process is not completely automated.

In our work, on using all the features including the HOS features, the KNN and PNN classifiers registered similar accuracies of 93.3% and balanced sensitivity (94.4%) and specificity (91.7%) values. The classification results indicate that the classification accuracy is influenced not only by the choice of features (number and type) but also on the choice of the classifier. We believe that by adding more relevant features we can improve the overall performance of our classifier. Our study is also limited by the small number of samples used for classifier development and evaluation. The separation index ranges for both the classes have a slight overlap primarily because of the less number of images used to develop the index. In future, a larger dataset from a multi-ethnic population would be studied. We also intend to incorporate, in future, the information about the aggressiveness of the disease in the abnormal cases in order to more clearly understand how the features discriminate the normal and abnormal cases. We plan to use the speckle images obtained from the envelope Radio Frequency (RF) images to extract the features to investigate if the accuracy may be improved further at the expense of a slighter higher computational cost.

In spite of these limitations, the following are the key features of the proposed Symtosis - a CAD-based technique: (a) The technique is fully automated and does not require any segmentation to select the region of interest. Traditional ultrasound liver images are the only input required, (b) This is the first study that has exploited the HOS features for FLD detection. We have demonstrated the utility and power of these features by evaluating the performance of the classifiers by training them without (Table II) and with (Table III) the HOS feature. It is evident that the accuracy significantly increased from 80% to 93.3% on including the HOS feature for classifier development and evaluation, (c) A high classification accuracy has been obtained (93.3%) with a small sample size. This emphasizes the discriminating capability of the significant features used. To account for the small sample size, we have employed the cross-validation data resampling technique in order to build robust classifiers, (d) The high classification accuracy has been achieved using only three features, making the entire process computationally less complex and cost-effective, (e) No additional cost is needed to incorporate the built classifier into a physician's computer. An executable software can be written and it can be downloaded from the internet easily, (f) No expert training is necessary to operate the software. The user has to only input the acquired liver ultrasound image, and the software will output the class label, and (g) The proposed *SteatosisClassificationIndex* also has the same advantages:

easy implementation, no cost, objective results, faster diagnosis, and no pre-requisite expert knowledge.

V. CONCLUSION

In this paper, we explored the possibility of a CAD based technique (Symtosis - Global Biomedical Technologies Inc., CA, USA) for the classification of normal and liver affected by fatty liver disease (abnormal cases). Several features that were based on the image texture, Higher Order Spectra, and Wavelet Packet Decomposition were extracted from the speckle images of the liver ultrasound images. Among the extracted features, highly discriminatory significant features alone were used to train and build several supervised learning based classifiers. Using only three features, the KNN, and PNN classifiers presented a high accuracy of 93.3%. They recorded almost similar values for both sensitivity and specificity (94.4% and 91.7%, respectively). All these classifiers have given significant performance measures using a small dataset. We believe that with the inclusion of more representative features, and study of other classifiers using a larger dataset, it should be possible to improve the current accuracy of the technique. Since it is fully automated and highly user friendly, the technique can be easily used in clinical practice once larger dataset is evaluated. We have also proposed a novel *SteatosisClassificationIndex* that has almost unique ranges for both the normal and abnormal cases, and therefore, it can be easily used by physicians for a second opinion on the patient condition at no extra cost.

REFERENCES

[1] S. Sherlock and J. Dooley, Diseases of the liver and Biliary system. Malden: Blackwell Science Ltd, 2002.

[2] V. Droga and D. Rubens, Ultrasound Secrets. Hanley and Belfus, 2004.

[3] J. E. Lavine and J. B. Schwimmer, "Nonalcoholic fatty liver disease in the pediatric population," *Clin. Liver Dis.,* vol. 8, no. 3, pp. 549–558, viii–ix, Aug. 2004.

[4] G.C. Farrell and C. Z. Larter, "Nonalcoholic fatty liver disease: from steatosis to cirrhosis," *Hepatology*, vol. 43, no. 2 suppl 1, pp. S99–S112, Feb. 2006.

[5] D. Joy, V. R. Thava, and B. B. Scott, "Diagnosis of fatty liver disease: is biopsy necessary?," *Eur. J. Gastroenterol. Hepatol.,* vol. 15, no. 5, pp. 539–543, May 2003.

[6] P. Thampanitchawong and T. Piratvisuth, "Liver biopsy: complications and risk factors," *World J. Gastroenterol.*, vol. 5, no. 4, pp. 301–304, Aug. 1999.

[7] V. Ratziu, F. Charlotte, A. Heurtier, S. Gombert, P. Giral, E. Bruckert, A. Grimaldi, F. Capron, and T. Poynard; LIDO Study Group. "Sampling variability of liver biopsy in nonalcoholic fatty liver disease," *Gastroenterology*, vol. 128, no. 7, pp. 1898–1906, Jun. 2005.

[8] S. Saadeh, Z. M. Younossi, E. M. Remer, T. Gramlich, J. P. Ong, M. Hurley, K. D. Mullen, J. N. Cooper, and M. J. Sheridan, "The utility of radiological imaging in nonalcoholic fatty liver disease," *Gastroenterology*, vol. 123, no. 3, pp. 745-750, Sep. 2002.

[9] S. F. Quinn and B. B. Gosink, "Characteristic sonographic signs of hepatic fatty infiltration," *AJR Am. J. Roentgenol.,* vol. 145, no. 4, pp. 753-755, Oct. 1985.

[10] K. J. Foster, A. H. Griffith, K. Dewbury, C. P. Price, and R. Wright, "Liver disease in patients with diabetes mellitus," *Postgrad. Med. J.*, vol. 56, no. 661, pp. 767-772, Nov. 1980.

[11] Y. Yajima, K. Ohta, T. Narui, R. Abe, H. Suzuki, and M. Ohtsuki, "Ultrasonographical diagnosis of fatty liver: significance of the liver-kidney contrast," *Tohoku. J. Exp. Med.*, vol. 139, no. 1, pp. 43-50, Jan. 1983.

[12] S. H. Saverymuttu, A. E. Joseph, and J.D. Maxwell, "Ultrasound scanning in the detection of hepatic fibrosis and steatosis," *Br. Med. J. (Clin Res Ed)*, vo. 292, no. 6512, pp. 13-15, Jan. 1986.

[13] U. L. Mathiesen, L. E. Franzen, H. Aselius, M. Resjö, L. Jacobsson, U. Foberg, A. Frydén, and G. Bodemar, "Increased liver echogenicity at ultrasound examination reflects degree of steatosis but not of fibrosis in asymptomatic patients with mild/moderate abnormalities of liver transaminases," *Dig. Liver Dis.*, vol. 34, no. 7, pp. 516-522, Jul. 2002.

[14] M. H. Mendler, P. Bouillet, A. Le Sidaner, E. Lavoine, F. Labrousse, D. Sautereau, and B. Pillegand, "Dual-energy CT in the diagnosis and quantification of fatty liver: limited clinical value in comparison to ultrasound scan and single-energy CT, with special reference to iron overload," J. Hepatol., vol. 28, no. 5, pp. 785-794, May 1998.

[15] S. R. Mehta, E. L. Thomas, J. D. Bell, D. G. Johnston, and S.D. Taylor-Robinson, "Non-invasive means of measuring hepatic fat content," *World J. Gastroenterol.*, vol. 14, no. 22, pp. 3476-3483, Jun. 2008.

[16] A. Qayyum, "MR spectrocopy of the liver: principles and clinical applications," *RadioGraphics*, vol. 29, pp. 1653-1664, Oct. 2009.

[17] C. Nikias and A. Petropulu, Higher-Order Spectral Analysis, Englewood Cliffs, NJ: Prentice-Hall, 1997.

[18] A. Ramm and A. Katsevich, The radon transform and local tomography. CRC Press, 1996.

[19] K. C. Chua, V. Chandran, R. Acharya, and C. M. Lim, "Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: A comparative study," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pp. 3824-3827, 2008.

[20] K. C. Chua, V. Chandran, U. R. Acharya, and C. Lim, "Application of higher order spectra to identify epileptic EEG," *J. Med. Syst.*, pp. 1–9, 2010, 10.1007/s10916-010-9433-z.

[21] U. R. Acharya, K. C. Chua, T.-C. Lim, Dorithy, and J. S. Suri, "Automatic identification of epileptic EEG signals using nonlinear parameters," *J. Mech. Med. Biol.*, vol. 9, no. 4, pp. 539-553, 2009.

[22] M. Mirmehdi, X. Xie, and J. S. Suri, Handbook of Texture Analysis. London, UK, UK: Imperial College Press, 2009.

[23] J.-H. Tan, E. Ng, U. R. Acharya, and C. Chee, "Study of normal ocular thermogram using textural parameters," *Infrared Phys. Technol.*, vol. 53, no. 2, pp. 120-126, 2010.

[24] M. M. Galloway, "Texture analysis using grey level run lengths," *NASA STI/Recon Technical Report N*, vol. 75, pp. 18, Jul. 1974.

[25] B. V. Dasarathy, Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. Los Alamitos: IEEE Computer Society Press, 1990.

[26] D. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109-118, 1990.

[27] J. A. Bilmes, A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models. International Computer Science Institute, 1998.

[28] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2005.

[29] C. E. Metz, "Basic principles of ROC analysis," *Semin. Nucl. Med.*, vol. 8, no. 4, pp. 283-298, Oct. 1978.

[30] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561-577, Apr. 1993.

[31] D. N. Ghista, Physiological systems' numbers in medical diagnosis and hospital cost effective operation," *J. Mech. Med. Biol.*, vol. 4, no. 4, pp. 401-418, 2004.

[32] D. N. Ghista, "Nondimensional physiological indices for medical assessment," *J. Mech. Med. Biol.*, vol. 9, no 4, pp. 643-669, 2009.

[33] Y. M. Kadah, A. A. Farag, J. M. Zurada, A. M. Badawi, and A. M. Youssef, "Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images," *IEEE Trans. Med. Imaging*, vol. 15, no. 4, pp. 466-478, 1996.

[34] R. Ribeiro and J. Sanches, "Fatty Liver Characterization and Classification by Ultrasound", *Pattern Recognition And Image Analysis, Lecture Notes in Computer Science*, vol. 5524/2009, pp. 354-361, DOI: 10.1007/978-3-642-02172-5_46.

[35] W. L. Lee, Y. C. Chen, and K. S. Hsieh, "Ultrasonic liver tissues classification by fractal feature vector based on M-Band wavelet transform,", *IEEE Trans. Med. Imaging*, vol. 22, no. 3, pp. 382-392, Mar. 2003.