An Ultrasound-Based Computer-Aided Diagnosis Tool for Steatosis Detection
Ricardo T. Ribeiro, Rui Tato Marinho, and J. Miguel Sanches, Senior Member, IEEE

Abstract—Liver steatosis is a common disease usually associated with social and genetic factors. Early detection and quantification is important since it can evolve to cirrhosis. In this paper, a new computer-aided diagnosis (CAD) system for steatosis classification, in a local and global basis, is presented. A Bayes factor is computed from objective ultrasound textual features extracted from the liver parenchyma. The goal is to develop a CAD screening tool to help in the steatosis detection. Results showed an accuracy of 93.33%, with a sensitivity of 94.59% and specificity of 92.11%, using the Bayes classifier. The proposed CAD system is a suitable graphical display for steatosis classification.

Index Terms—Classification, computer-aided diagnosis (CAD), steatosis, textual features, ultrasound (US).

I. INTRODUCTION

Liver steatosis (HS) is a liver disease that occurs when hepatocytes fat content increases [1]. Predominantly associated with obesity, insulin resistance, and alcohol [2]–[4], it is considered the hepatic pandemic of the XXI century [4]. Early detection is important, since HS is a biomarker for the development of liver cirrhosis or even hepatocellular carcinoma [3].

The high prevalence of HS creates the need for a noninvasive screening tool aimed to diagnose and quantify HS. Liver biopsy (LB) remains the gold standard in HS diagnosis [3], [4]. However, its invasive nature, low social acceptance, and sampling errors do not make LB suited for screening. New noninvasive techniques have been proposed, where imaging methods are commonly baseline tests [3]. In clinical practice, ultrasound (US) is the first method used for HS diagnosis [4]. This imaging technique shows appreciable advantages such as its noninvasiveness, low price, accessibility, and nonionizing nature.

Visual detection of HS from US images is based on a gradual increase in liver echogenicity and US beam attenuation [5]. However, US-based diagnosis is often related to a subjective interpretation of US findings, which may also be present in other liver disorders [6]. As reported by [3], the majority of US studies, focused to detect HS, are based on subjective visual criteria.

Several strategies [7]–[12] have been proposed to detect and quantify HS. In this paper, a multifrequency approach is adopted, where a monogenic decomposition (MD) is used for feature extraction. The performance of the proposed algorithm is assessed by several feature sets described in the literature [7], [8], [11]. Bayes classifier is used in the binary classification and the underlying Bayes factor (BF), computed in a local basis, and is overlapped on the US image to provide local evidence of the disease. A computer-aided diagnosis (CAD) tool is designed and implemented to display this evidence map and help in the diagnosis and quantification of HS. This interactive tool is particularly useful for the appropriate selection of the region of interest (ROI) for which the method is sensitive.

The remainder of this paper is organized as follows. Section II introduces the problem formulation and is divided in a preprocessing step and feature extraction procedure (see Section II-A), the global detection of HS (see Section II-B), feature selection and the classification procedures; and, subsequently (Section II-C), in the local detection approach as well as the framework used to build the CAD system. Section III describes the used dataset, experimental tests, and the classification results. In Section IV, a discussion of the results from the proposed algorithm is given and conclusions are provided.

II. PROBLEM FORMULATION

Here, two different formulations of the HS detection problem are presented: 1) Global approach, where five different sets of features are presented to classify HS. A brief theoretical motivation is given, regarding feature selection and the Bayes classifier. 2) Local characterization that creates a confidence map, built from the BF results, related to the disease evidence.

A normalization and decomposition procedure is adopted before analysis. This procedure minimizes the results dependence on the specific US parameters used during data acquisition and on the subjective criteria of the operator.

A. Preprocessing and Feature Selection

To guaranty the reproducibility of the results, the procedure described in [13] is used to separate the textural and intensity information of the US images where the radio frequency (RF)
envelope, \( rf(i, j) \), is estimated from the observed B-mode image, \( y(i, j) \). The estimation of this envelope image is performed based on physical considerations about the data generation process, namely, by taking into account the US scanner parameters tuned by the clinician during the US exam.

The estimated RF image is decomposed in despeckled and speckle fields according to the following model [13]:

\[
rf(i, j) = s(i, j)\eta(i, j)
\]

where \( s(i, j) \) and \( \eta(i, j) \) are called despeckled and speckle fields, respectively.

The de-speckled field, \( s(i, j) \), containing mainly intensity and echogenicity information about tissues and organs, is not used because it was concluded that it is not relevant for steatosis detection.

The speckle field, on the contrary, contains useful information for steatosis detection. The pixels \( \eta(i, j) \) are assumed independent and identically distributed (i.i.d.) random variables with Rayleigh distribution [14]. This noise model described in (1) is truly multiplicative, in the algebraic sense, where the variance is signal dependent [13]. The speckle field is used for extracting the following textural features that are related to the microstructure of the hepatic parenchyma:

1) first-order statistics—mean (\( \mu \)), standard deviation (\( \sigma \))), and skewness (\( \kappa \)), and kurtosis (\( K \)),
2) second-order statistics—first-order two-dimensional (2-D) autoregressive (AR) coefficients [15], \( a_{0,1} \), \( a_{1,0} \), \( a_{1,1} \), are computed as suggested by [16].

These statistics are calculated from the image decomposition fields obtained with two different approaches: Haar wavelet transform (WT) [17] [vertical (LH), horizontal (HL), and diagonal (HH) detail fields] and image MD (magnitude, \( A \), and two phase fields, \( \theta \) and \( \psi \)).

A monogenic representation of a 2-D signal [18], [19], \( f(x) \), \( x \in \mathbb{R}^2 \), is defined by the following triplet:

\[
f_m(x) = ((f(x), \text{Re}(\mathcal{R}(x))), \text{Im}(\mathcal{R}(x))) = (f, f_1, f_2)
\]

where \( \mathcal{R}(\cdot) \) denotes the Riez transform [18], [19].

The local amplitude, \( A \), local orientation, \( \theta \), and phase, \( \psi \), of \( f_m \), are defined as follows:

\[
A = \sqrt{f^2 + f_1^2 + f_2^2}
\]

\[
f = A \cos \psi
\]

\[
f_1 = A \sin \psi \cos \theta,
\]

\[
f_2 = A \sin \psi \sin \theta.
\]

The following feature sets, some proposed in the literature, are tested for optimization purposes:

1) (Wavelet based)—\( \{a_{0,1}, a_{1,0}, a_{1,1}, \epsilon, \mu\} \) extracted from the first and second WT detail decomposition levels, \( \{HL_1, LH_1, HH_1\} \), in a total of 30 feature scalars.
2) (Monogenic based)—\( \{a_{0,1}, a_{1,0}, a_{1,1}, \epsilon, \mu\} \) features extracted from the three fields of the first three MD decomposition levels, \( \{(A, \psi, \theta)_{1,2,3}\} \), in a total of 45 features.

3) (Wavelet and monogenic based)—Features A and B plus four first-order statistic features extracted directly from the speckle field in a total of 79 scalars.

4) (Features from [8])—median, standard deviation, and interquartile range of the first- and second-level decomposition details of the Daubechies 3 WT, in a total of 18 scalars.

5) (Acoustic attenuation coefficient [7])—slope of the linear regression of the mean image intensity along the depth direction (rows).

Automatic feature selection from sets A, B, and C was performed, using the stepwise regression analysis [20] (criterion to add: \( p < 0.05 \); to remove: \( p > 0.1 \)) to minimize the peaking phenomenon, also known as overfitting. Four optimal features for set A and six for sets B and C were obtained, after this step, to feed a Bayes classifier.

**B. Global Detection of Liver Steatosis**

The discriminant functions computed to quantify the evidence of steatosis, with the BF (\( \Lambda \)),

\[
\Lambda(x) = \log \left( \frac{g_{HS}(x)}{g_{N}(x)} \right)
\]

are defined under the assumptions of multivariate normal distribution of the features [21], [22] with means \( \mu_{HS} \) and \( \mu_N \) and covariance matrices, \( \Sigma_{HS} \) and \( \Sigma_N \), for HS and Normal (N) classes, respectively:

\[
g_{\tau}(x) = -\frac{1}{2}(x - \mu_{\tau})^T \Sigma_{\tau} (x - \mu_{\tau}) - \frac{1}{2} \ln |\Sigma_{\tau}| + \ln P(\omega_{\tau})
\]

with \( \tau \in \{HS, N\} \). \( P(\omega_{\tau}) \) is the prior probability of each class, computed as follows:

\[
P(\omega_{\tau}) = \frac{N_{\tau}}{N_S}
\]

where \( N_{\tau} \) is the number of samples within the class \( \omega_{\tau} \) and \( N_S \) is the number of samples of the population.

Positive values of \( \Lambda \) are interpreted as evidence of HS, whereas negative ones mean evidence of a normal tissue. Small absolute values, \( |\Lambda| \approx 0 \), indicated no evidence of HS or normal tissues.

Global characterization of the liver is obtained by selecting a representative region of the liver parenchyma from which the feature vector, \( x \), in (5), is computed. In this situation, evidence of steatosis or normal liver is easily obtained if \( \Lambda > 0 \) or \( \Lambda < 0 \), respectively.

The accuracy of the results depends on the location of the ROI. Better results are obtained with ROIs extracted from the vicinity of the medial axis of the image, as superficial as possible, to avoid beam distortions, representative of the hepatic parenchyma and should not contain major blood vessels and ligaments.

An ideal 128 \( \times \) 128 ROI (ROI\(_b\)) for global liver characterization is displayed in Fig. 1, in red, corresponding to a 45.7 mm \( \times \) 45.7 mm window size in a 3-pixel/mm image resolution.
Also in Fig. 1, two nonideal ROIs are displayed, in different locations, to illustrate this dependence:  
1) ROI\textsubscript{1}, extracted at the same depth of ROI\textsubscript{0} but with a right horizontal shift, in blue, and  
2) ROI\textsubscript{2}, extracted as deep as possible and with a left shift from the original ROI, in green.  

The two-sample t-tests \((p < 0.01)\) \cite{22} and the Bhattacharyya distance \cite{23} are used to compare and analyze the results obtained with ROI\textsubscript{0}, ROI\textsubscript{1}, and ROI\textsubscript{2} \footnote{The Bhattacharyya distance \((B)\) \cite{23} is defined as}.

C. Local Characterization

The same strategy, used for the global characterization, can be used to characterize the liver parenchyma in a local basis \cite{24, 25}.

Let us consider the following confidence map, \(S = \{s(i, j)\}\),

\[
s(i, j) = \Lambda(x(i, j)) \tag{9}
\]

where \(\Lambda(x(i, j))\), defined in (5), is computed with optimum set of features and parameters \((\mu_i, \Sigma_i)\) obtained in the training step. Feature vector, \(x(i, j)\), is computed from a \(N \times N\) sliding window centered at a given \((i, j)^{th}\) pixel.

This map aims at characterizing HS in a local basis by detecting foci of fat accumulation in the hepatic parenchyma. To make easy the interpretation and visualization of this information a new image, called diagnosis map (DM), is created. DM is obtained, DM = \{d(i, j)\}, by merging \(S\) with the original US data, as follows:

\[
d(i, j) = \begin{cases} 
\alpha s(i, j) + (1 - \alpha) y(i, j) & \text{if } s(i, j) > 0 \\
y(i, j) & \text{otherwise} 
\end{cases} \tag{10}
\]

where \(y(i, j)\) is the original US B-mode image and \(\alpha\) is a transparency parameter. In this merging process, the original US B-mode image is represented by using a gray-scale color map, whereas \(S\) is displayed with the color map represented in Fig. 2.

In the DM image, yellow to red color regions indicate high probability of fat presence, the green region indicates no evidence to support the classification, and the light to dark blue regions indicate normal liver tissue, as shown in Fig. 2.

The CAD system, used to process and visualize this information, contains ROI selection tools. This interactive framework allows the physician to specify the location and size of a liver region.

III. EXPERIMENTAL RESULTS

Results obtained with real data are presented in this section. A total of 74 US liver images, acquired by expert operators, were obtained from 36 patients with HS and 38 control subjects, enrolled in the experiments. All patients were outpatients in the Gastroenterology Department of the Santa Maria Hospital, Lisbon, with known diagnosis, based on the clinical and US criteria \cite{1}. The study protocol was approved by the hospital ethic committee and informed consent was obtained from each subject.

The acquisition conditions were predefined to guarantee the reproducibility of the results. For each patient, an US B-mode image was acquired using a general-purpose clinical US equipment (CX50, Philips, Amsterdam, the Netherlands), coupled with a convex array transducer (Philips C5-1, Amsterdam, the Netherlands). The US acquisition protocol was set with 3.5 MHz (transmitted frequency), depth of 18 cm, focal zone at 9 cm, dynamic range at 75 dB, and only the gain was left variable, according to the patient morphotype. US images were acquired from the right liver lobe and the same anatomical landmarks were used.

The results were assessed with a cross-validation-based strategy, since when the number of samples is small, this is the preferred approach to assess the performance of a classifier \cite{26}. In the present study, a five-fold leave-one-out cross-validation method is used. Here, the performance criteria are the overall
TABLE II
MEAN AND STANDARD DEVIATION OF THE SELECTED FEATURES AND THE INFLUENCE OF DIFFERENT ROIS

<table>
<thead>
<tr>
<th>Feature</th>
<th>Normal</th>
<th></th>
<th>HS</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \mu )</td>
<td>( \sigma )</td>
<td>( \mu )</td>
<td>( \sigma )</td>
<td>( B_{(ROI_0,ROI_1)} )</td>
<td>( B_{(ROI_0,ROI_2)} )</td>
</tr>
<tr>
<td>( a_{1,1} ) HL(_1)</td>
<td>0.0884</td>
<td>0.0937</td>
<td>0.3049</td>
<td>0.5051</td>
<td>0.0148</td>
<td>0.0091</td>
</tr>
<tr>
<td>( a_{1,1} ) LH(_1)</td>
<td>0.8649</td>
<td>0.7594</td>
<td>2.7930</td>
<td>0.7745</td>
<td>0.0408**</td>
<td>0.0457**</td>
</tr>
<tr>
<td>( E ) HL(_2)</td>
<td>0.5201</td>
<td>0.3402</td>
<td>0.2386</td>
<td>0.1036</td>
<td>0.041**</td>
<td>0.0426</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.8496</td>
<td>0.1177</td>
<td>0.7567</td>
<td>0.0534</td>
<td>0.0373</td>
<td>0.0038</td>
</tr>
<tr>
<td>mean ( \theta_1 )</td>
<td>0.0017</td>
<td>0.0182</td>
<td>-0.0175</td>
<td>0.0401</td>
<td>0.1912**</td>
<td>0.2672**</td>
</tr>
<tr>
<td>( a_{0,1} ) ( A_3 )</td>
<td>-0.1253</td>
<td>0.0519</td>
<td>-0.1239</td>
<td>0.0735</td>
<td>0.0433</td>
<td>0.0543**</td>
</tr>
<tr>
<td>( a_{1,1} ) ( \theta_3 )</td>
<td>0.5465</td>
<td>0.3394</td>
<td>0.8531</td>
<td>0.5404</td>
<td>0.1042</td>
<td>0.0102</td>
</tr>
</tbody>
</table>

** Statistical differences \( p < 0.01 \) observed.

Fig. 3. BF probability density function for the normal and HS class, in ROI\(_0\).

Fig. 4. BF probability density function at different ROI positions (ROI\(_1\) and ROI\(_2\)).

accuracy (OA):

\[
OA = \frac{TP + TN}{TP + FN + FP + TN}
\]  

the sensitivity (sens),

\[
sens = \frac{TP}{TP + FN}
\]  

and specificity (spec),

\[
spec = \frac{TN}{FP + TN}
\]

where TP, FP, FN, and TN denote the true-positives, false-positives, false-negatives, and true-negatives, respectively.

The performances of the Bayes classifier with the five sets of features are listed in Table I. Comparison between the performances of the classifier is focused on the sensitivity, because it reveals the capability of correctly detecting HS. Feature set C
outperforms feature sets A and B, as well as the sets proposed in the literature, D [8] and E [7]. This set, C, achieves an OA of 93.3% and a detection rate of 92.11% and 94.59% for normal and HS class, respectively.

The mean BF value for the normal class was $-11.17 \pm 5.25$ and the mean BF value for the HS class was $10.70 \pm 7.75$. Data from the feature set C, extracted in the global detection step, are used to train the classifier in the local analysis. This feature vector is composed of three wavelet coefficients, $a_{1,1}$ (HL1), $a_{1,1}$ (LH1), and $\epsilon$ (HL2), kurtosis, and three features from the MD, $\mu(\theta_1)$, $a_{0,1}$ ($A_3$), and $a_{1,1}$ ($\theta_3$), all extracted from the speckle field.

Table II summarizes the mean and standard deviation of each selected feature, at different ROIs. Relevant differences ($p < 0.01$) in feature values are observed for different ROIs, mainly between central and peripheral locations. For both peripheral ROIs (ROI1, ROI2), BF shows smaller confidence in the classification result. In ROI1, normal and HS class attained $-0.611 \pm 1.01$ and $0.25 \pm 1.04$, respectively. Similar results are observed in ROI2, with $-0.95 \pm 1.36$ and $0.58 \pm 1.39$, for normal and HS class, respectively. This result is not surprising because the classifier was trained with data from ROI0 region.

Fig. 3 presents the histograms of the BF values for each class, obtained with ROI0. In this location, relevant interclass separability is observed, indicating high confidence in the classification. To illustrate that the ROI location influences the performance of the classifier, in Fig. 4, similar histograms are shown. In this case, BF histograms are highly overlapped and with lower BF values, revealing low confidence.

Figs. 5 and 6 display examples of the CAD system interface for normal and HS samples, respectively. BF and its graphical representation are useful tools for the quantification and detection of fat accumulation within the liver parenchyma, in the sense that it allows a graphical correlation between the classification evidence and its localization in the US image.

IV. DISCUSSION AND CONCLUSION

In this study, a CAD tool for the detection and quantification of HS in a global and local basis using US images is described. The proposed algorithm is able to quantify the subjective visual criteria, used in clinical practice, providing an objective measure of HS. The results showed an OA of 93.3% and a sensitivity of 92.1% and 94.6% for normal and HS class, respectively.

Similar experiments, described in [6], report a sensitivity and specificity ranging from 60.0% to 94.0% and 84% to 95%, respectively. Other studies report similar results [7], [8], [11], [27], which support our method.

An important result from this study is that the AR coefficients, extracted from the multiscale Haar wavelet decomposition and MD, are relevant US features in HS discrimination. These features are mathematical and objective characterizations of the visual pattern recognition used to classify steatosis.

The study of [28] describes that textural features, based on power spectrum analysis, allows good discrimination between normal liver and steatosis. The differences were found in the region of high frequencies. This result is in accordance with our study. Wavelet-based features have been proposed in the studies of [8], [29] and [30] for this problem. A novelty of this study is the inclusion of the MD as a feature extractor for the US tissue characterization. The results show that the amplitude ($A$) and local orientation ($\theta$) contain key information for this discrimination problem. It should be stressed that all the features were extracted from the speckle field of the US image.

The increased interest in liver evaluation by noninvasive methods has led to the development of new diagnostic modalities, such as gradient echo magnetic resonance imaging (DGE-MRI) [31], acoustic radiation force impulse (ARFI) elastography, [32] or transient elastography (Fibroscan) [4]. Although these methods have been used for the assessment of liver fibrosis, they may also be used to characterize HS.

DGE-MRI shows high accuracy in HS detection, attaining a sensitivity of 90.9% and a specificity of 94% [31]. However, this technique is limited when fibrosis is present [4] and has low accessibility [31].

ARFI parameters, shear wave velocity, and attenuation are strongly correlated with the degree of HS in animal studies [33], [34]. With human samples, studies have shown that HS has no statistical influence [35] or correlation [32] with ARFI results, in contrast to Fibroscan [32], [35], [36]. Fibroscan allows the extraction of the controlled attenuation parameter (CAP), a
relatively new parameter, that measures objectively liver attenuation (expressed as dB/m) [4]. CAP achieves a sensitivity of 90% in the detection of HS, with a cut-off value of 215 dB/m [37]. Despite being the most promising tool to monitor HS, Fibroscan still faces consistency issues [4] and low availability.

US is still the best way to assess HS in clinical practice [4]. Thus, the proposed CAD tool could lead to a more accurate patient surveillance or to the need of further diagnostic exams. Also, the proposed local characterization of liver parenchyma helps the detection of HS in cases of nonhomogeneous distribution of fat infiltration, that may be deposited in one well-circumscribed region (focal fatty infiltration) or, alternatively, discrete areas of liver parenchyma remain uninvolved when the remainder of the liver is diffusely infiltrated with fat (fatty sparing) [6].

To maintain the reproducibility and accuracy of the method, ROI should be always selected from the central part of the US image without artifacts.

REFERENCES

nosis of hepatic steatosis: A feasibility study,” IEEE Trans. Ultra-
Ricardo T. Ribeiro received the B.Sc. degree in radiology from the Lisbon School of Health Technology, Polytechnic Institute of Lisbon, Lisbon, Portugal, and the Ph.D. degree in biomedical engineering from the Institute for Systems and Robotics, Instituto Superior Técnico, Technical University of Lisbon, Lisbon, in collaboration with the University of Lisbon Medical School, Lisbon. 
He is currently an Adjunct Professor at the Lisbon School of Health Technology. His current research interests include the use of machine learning, computer vision, and tissue characterization for the classification of ultrasound imaging. He is also involved in clinical research in ultrasound imaging.

Rui Tato Marinho received the MD degree from the Faculty of Medical Sciences, Nova University of Lisbon, Portugal and the Ph.D. degree from Medical School of Lisbon, Lisbon University, Portugal.
He is currently a Gastroenterologist and Hepatologist at the Medical School of Lisbon, Hospital Santa Maria, Lisbon, Portugal.
Dr. Marinho is an Editor-in-Chief of Acta Médica Portuguesa, the Scientific Journal of Portuguese Medical Association (Ordem dos Médicos), Adviser of the Viral Hepatitis Prevention Board, and President of Portuguese College of Hepatology of the Portuguese Medical Association.

J. Miguel Sanches received the E.E., M.Sc., and Ph.D. degrees from the Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal, in 1991, 1996, and 2003, respectively.
He is a Professor at the Department of Bioengineering, IST and Researcher at the Institute for Systems and Robotics. His research interests include biomedical engineering, namely, in biomedical signal and image processing and physiological modeling of biological systems.
Dr. Sanches is a senior member of the IEEE Engineering in Medicine and Biology Society and member of the Bio Imaging and Signal Processing Technical Committee of the IEEE Signal Processing Society and president of the Portuguese Association of Pattern Recognition, affiliated to the International Association of Pattern Recognition.