CHRONIC LIVER DISEASE STAGING CLASSIFICATION BASED ON ULTRASOUND, CLINICAL AND LABORATORIAL DATA

Ricardo Ribeiro[1,2], *Rui Marinho*[3], *José Velosa*[3], *Fernando Ramalho*[3] *and J. Miguel Sanches*[1]

[1]Institute for Systems and Robotics / Instituto Superior Técnico
[2]Escola Superior de Tecnologia da Saúde de Lisboa
[3]Faculdade de Medicina da Universidade de Lisboa
Lisbon, Portugal

ABSTRACT

In this work the identification and diagnosis of various stages of chronic liver disease is addressed. The classification results of a support vector machine, a decision tree and a k-nearest neighbor classifier are compared. Ultrasound image intensity and textural features are jointly used with clinical and laboratorial data in the staging process.

The classifiers training is performed by using a population of 97 patients at six different stages of chronic liver disease and a leave-one-out cross-validation strategy.

The best results are obtained using the support vector machine with a radial-basis kernel, with 73.20% of overall accuracy. The good performance of the method is a promising indicator that it can be used, in a non invasive way, to provide reliable information about the chronic liver disease staging.
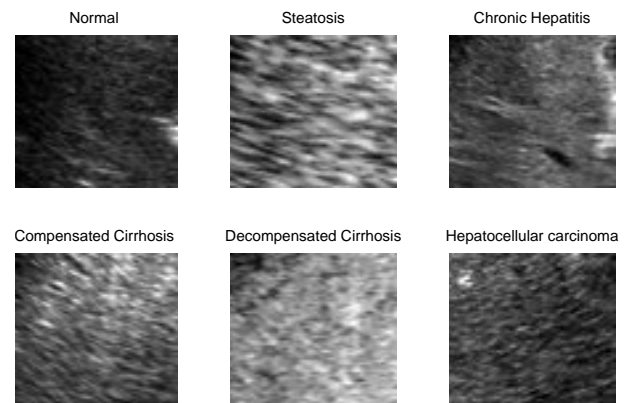
Index Terms— Chronic liver disease, Tissue characterization, Ultrasound, Classification

1. INTRODUCTION

Staging of liver disease is needed because it is progressive, most of the time asymptomatic and potentially fatal. An accurate characterization of the disease is difficult but crucial to prevent its evolution and avoid irreversible pathologies such as the *hepatocellular carcinoma*.

Fatty liver infiltration (*steatosis*) is the earliest stage of the liver disease and occurs when the fat content in *hepatocytes* significantly increase. It is asymptomatic and the progress of the hepatic injury to other conditions, more severe, is common. e.g., fibrosis. Pathologically, fibrosis appears during the course of tissue injury or organ damage and its progression rate strongly depends on the cause of liver disease, such as *chronic hepatitis* [1].

Cirrhosis is the end-stage of every chronic liver disease. It is characterized by an asymptomatic stage, known as *compensated cirrhosis*, followed by a rapidly progressive phase

Fig. 1. Examples of US images in different stages of the chronic liver disease

where liver dysfunction occurs, called *decompensated cirrhosis*. The most severe evolution condition of the cirrhosis is the *hepatocelllular carcinoma* (HCC), also called, primary liver cancer [1].

Liver biopsy is the most accurate method for diagnosis. However it is highly invasive and several medical image modalities are valuable alternative methods to detect and quantify this disease [1]. The non-ionizing and non-invasive nature of ultrasound (US) imaging and its widespread presence at almost all medical and clinical facilities makes it the preferred method for assessing steatosis and other liver diseases such as cirrhosis. Nevertheless, the echo pattern observed in the US liver images in different chronic liver disease stages is difficult to distinguish [2], as shown in Fig. 1.

In [3] it is referred that the pulse echo data from different grain types contain distinguishable statistical regularities. In addition, the study in [4] proposes a quantitative tissue characterization to increase the usefulness of US for evaluating the diffuse liver disease. Thus, the characteristic granular pattern affecting US images, called *speckle*, may contain useful information for diagnosing different pathologies.

This paper uses a set of features from *speckle* and *de-*

speckled image fields, computed from US images, to detect chronic liver disease. The commonest features described in literature for the diagnosis of diffuse liver diseases include first order statistics, co-occurrence matrices, wavelet transform, attenuation and backscattering parameters and coefficients [5, 6, 7, 8].

Spatial resolution plays a central role in pattern analysis because US images typically exhibit anisotropic pattern characteristics that can be used for diagnosis.

In this sense it is proposed a method to detect *chronic liver disease* based on an image processing procedure that decomposes the US images of the liver parenchyma into two fields: the *speckle* image containing textural information and the *de-speckled* image containing intensity and anatomical information of the liver. Features extracted from these fields, as well as clinical and laboratorial features, are used to train a supervised classifier to detect the disease.

The chronic liver disease stages are considered and several classifiers are used to assess the discriminative power of the selected features: (i) the *support vector machine* (SVM), (ii) the *decision tree* and (iii) the *k-nearest neighbor* (kNN).

Several *figures of merit* (FOM) were computed to assess and compare the performance of each classifier.

This paper is organized as follows. Section 2 formulates the problem and describes the pre-processing procedures, the extraction and selection of features and classifiers. Section 3 presents the experimental tests and reports the classification results. In addition, the correlation between US features and disease profile is discussed. Section 4 concludes the paper.
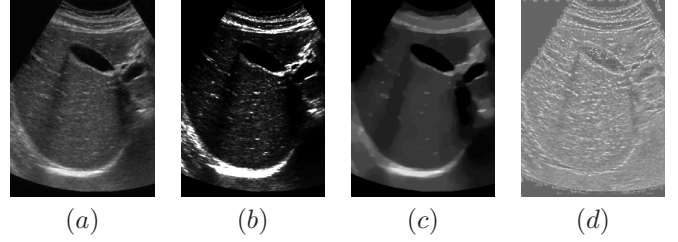
2. PROBLEM FORMULATION

The decomposition procedure described in [9] to separate the textural and intensity information within US images is here adopted. In this approach an estimation of the *radio frequency* (RF) raw data is firstly done based on physical considerations about the data generation process, namely, by taking into account the dynamic range compression performed by the ultrasound equipment over the signal generated by the US probe. The observation model, in this approach, also considers the brightness and contrast parameters tuned by the medical doctor during the exam which changes from patient to patient.

The estimated RF image is decomposed in *de-speckled* and *speckle* fields according to the following model [10]

$$y(i,j) = x(i,j)\eta(i,j), \qquad (1)$$

where $\eta(i,j)$ are considered independent and identically distributed (i.i.d.) random variables with *Rayleigh* distribution. This image describes the noise and textural information and is called *speckle* field. In this model, the noise is multiplicative in the sense that its variance, observed in the original image, depends on the underlying signal, $x(i,j)$. Fig.2 illustrates an example of the decomposition methodology.

In clinical practice, the commonest criteria to assess *chronic liver disease* are the increase of the echo intensity

(a)   (b)   (c)   (d)

Fig. 2. Decomposition procedure of US liver parenchyma. a) Observed *B-mode* US image. Estimated b) envelope RF image, c) *de-despeckled* and d) *speckle* image fields.

decay along the depth, changes of the hepatic parenchyma texture and changes in the liver boundaries [2]. Here the following extend set of US image features are considered from the *de-speckled* field:

1. **Acoustic attenuation coefficient**, m, measured by the slope coefficient of the linear regression of intensities along the depth/lines [6],
2. **First-order statistics**, including the mean (μ) and variance (σ^2) of the pixel intensities;

and from the *speckle* field:

1. **Co-occurrence matrix**, which enables to derive different statistics such as the contrast ($cont$) [5], correlation (ϕ) that measures the joint probability occurrence of specific pixel pairs, energy (E) of the image (obtained by summing of squared elements of the image) and homogeneity (H) which quantifies the closeness of the distribution of matrix elements to its diagonal,
2. **Wavelet energies**, measures by the vertical (E_V) and horizontal (E_H) detail energies of the first Haar wavelet decomposition and
3. **Autorgressive** (AR) coefficients of a first order 2D model, $\{a_{0,0}, a_{1,0}, a_{0,1}\}$ (the first order model was adopted because it was confirmed by [8] that in this scope it leads to the minimum error probability).

Besides image based features, several other clinical data and biochemical tests are useful for evaluating and managing patients with hepatic dysfunction. The clinical study of the disease, conducted in [1], reported the following meaningful **clinical information** to be used:

1. Cause of disease ($diagnose$), which include none (0), alcohol (1), hepatitis B (2), hepatitis C (3), alcoholic hepatitis B (4) and C (5) and others (6), and the following binary indicators:
2. Tumor (T),
3. Ascites (A) which is the presence of free fluid within the peritoneal cavity; encephalopathy ($Ence$),
4. Gastro-Intestinal bleeding (GIB) infection (Inf) and alcoholic habits (Alc).

The **laboratorial features** related with the liver function [1] are: i) total bilirubin (Bil), ii) prothrombin time (INR), iii) albumin (Al), iv) creatinine ($Crea$), v) aspartate transaminase (AST), vi) alanine transaminase (ALT), vii) gamma glutamyl transpeptidase (gGT), viii) glycemia (Gly) and ix) sodium (Na).

All these features, organized in a 28 length vector, are used to train each tested classifier and a forward selection method with the criterion of 1 - Nearest Neighbor *leave-one-out cross-validation* (LOOCV) performance was used to select the most significant features and increase the discriminative power of the classifier. Three different classifiers were implement and tested: i) the SVM, ii) decision tree and iii) *k-nearest neighbor* (kNN). A short description of each one is provided.

The aim of SVM is to find a decision plane that has a maximum distance (margin) from the nearest training pattern [7]. Given the training data $\{(x_i, \omega_i)|\omega_i = 1 \ or \ -1, \ i = 1, ..., N\}$ for a two-class classification (where x_i is the input feature; ω_i is the class label and N is the number of training sample), the SVM maps the features to a higher-dimensional space. Then, SVM finds a hyperplane to separate the two classes with the decision boundary set by the support vectors [7]. In this paper, a multiclass SVM classifier was adopted, using a Gaussian radial-basis kernel function and a polynomial kernel.

Decision tree is a supervised classifier that has found application in medical-image analysis. A node in a tree represents a test on a particular feature, and each branch from that node represents the possible outcome to the test. A path in the tree, from the root of the tree to an end leaf, specifies the classification, with the end leaf representing an object class. Several methods have been proposed for the tree structure design. In this work, the algorithm proposed by [11] is adopted.

The non-parametric kNN classifier is also tested in this paper. It classifies a test sample to a class according to the majority of the training neighbors in the feature space by using the minimum Euclidean distance criterion [4].

All classifiers were implemented using the algorithm proposed by [11].

3. EXPERIMENTAL RESULTS

Ninety seven data samples were obtained from 97 patients. The patients were selected from a gastroenterology department of the Santa Maria Hospital, in Lisbon, with known diagnosis. The samples were labeled in six classes; *Normal*, ω_N, *Steatosis*, ω_S, *Chronic Hepatitis without cirrhosis*, ω_{CHC}, *Compensated Cirrhosis*, ω_{CC}, *Decompensated Cirrhosis*, ω_{DC}, and *Hepatocelular Carcinoma*, ω_{HCC}. Among them, 34 belong to ω_N, 4 to ω_S, 9 to ω_{CHC} and ω_{CC}, 35 to ω_{DC} and 6 patients to ω_{HCC}.

A ROI of 128×128 pixels along the medial axis was extracted from each image. No acquisition protocol was used because the RF estimation procedure attenuates differences between images acquired in different conditions. The LOOCV strategy was used in the training and testing process, due to the number of observations [12].

Using the feature selection method proposed for evaluating the feature set, six optimal features were obtained. The features, displayed in a decreasing performance order, are:

INR, $diagnose$, m, T, $Ence$ and E_V. It is important to stress that the optimal set of features is composed of three clinical features ($diagnose$, T and $Ence$), two US image features (m and E_V) and one laboratorial feature (INR), which support the strategy of including features taken from different sources (image, clinical, etc) as it is used in medical practice.

Table 1 shows that the *steatosis* group, ω_S, had the lowest mean value, 1.053, of INR and the compensated cirrhosis group, ω_{CC}, had the highest mean value, 1.608, with highest variance, 1.494. The importance of this feature for classification is in accordance with the findings reported by [13] which stated that the INR can be used to quantify coagulopathy and hepatic synthetic dysfunction in this liver disease.

As expected, the feature $diagnose$ is higher for increasing disease severity. Nevertheless, in the classes $\omega_C C$, $\omega_D C$ and $\omega_H CC$, we can observe a dispersion of the causes/diagnosis that led to that disease stage. The same behavior can be verified in the encefalopathy, ($Ence$) and tumor (T) features.

The slope index (m) which characterizes tissue attenuation also showed complete differentiation between classes, specially for separating normal liver (ω_N) from other more severe stages ($\omega_C HC$, $\omega_C C$ and $\omega_D C$). This finding is in accordance with the study of [6]. Similar results can also be reported in the vertical detail energy (E_V) of the wavelet transform, where differences in the mean values are observed from class to class. These facts suggest the usefulness of attenuation and textural parameters for identifying different types of liver disease.

The classification technique significantly affects the final diagnosis [4]. Using the LOOCV method, the same data set was tested with different types of classifiers, namely a kNN, a decision tree classifier and a SVM classifier with polynomial (SVM$_P$) and radial-basis (SVM$_R$) kernels.

To determine the optimal parameters for the classifiers the following procedures were done. The kNN algorithm was implemented for values of k=1,2,3,5,7 and 9. The SVM$_P$ was trained with a degree range of $[1 : 5]$ and the SVM$_R$ was implemented with a radius close to 1 ([0.1,0.2,...,2]). The best performance of the kNN classifier was achieved with $k = 5$, which resulted in an error rate of 29.90%. Considering the SVM classifiers, the optimal SVM$_P$ corresponds to a degree of 3 while the optimal SVM$_R$ has a radius of 0.6 (both with error rates of 26.80%). In the case of the decision tree classifier, the computation was performed using a purity binary splitting criterion, with no pruning, as suggested by [11].

The overall accuracies achieved using the referred classifiers are summarized in Table 2. The best overall result was obtained using the SVM classifier, SVM$_P$ and SVM$_R$. Moreover, the results obtained using the SVM$_R$, showed a better capability for differentiating normal liver (ω_N) and a decompensated cirrhosis (ω_{DC}) from the other stages, with an accuracy of 94.12% and 97.14%, respectively. This last result is particularly relevant since ω_{DC} is characterized by major complications with lower survival rate.

Feature	ω_N		ω_S		ω_{CHC}		ω_{CC}		ω_{DC}		ω_{HCC}	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
INR	1.206	0.355	1.053	0.009	1.218	0.033	1.608	1.494	1.432	0.072	1.105	0.021
$diagnose$	0.00	0.00	0.00	0.00	1.429	3.619	2.714	6.571	2.429	4.032	2.250	0.917
m	-0.034	0.049	-0.081	0.008	0.053	0.011	0.218	0.055	-0.454	4.956	-0.025	0.024
T	0.147	0.129	0.00	0.00	0.00	0.00	0.143	0.143	0.036	0.036	1	0
$Ence$	0.00	0.00	0.00	0.00	0.00	0.00	0.143	0.143	0.393	0.247	0.00	0.00
E_V	0.134	0.031	0.082	0.012	0.081	0.005	0.146	0.006	0.280	1.089	0.140	0.025

Table 1. Mean (μ) and variance (σ^2) of features belonging to each class.

Although the best overall, ω_N and ω_{DC} accuracies were achieved with the SVM classifier, a similar behavior was not observed for the other classes. The best individual accuracy regarding the classification of ω_{HCC} was achieved with the kNN classifier, with a diagnostic yield of 83.33%. In the case of ω_{CC} and ω_{CHC} the best individual accuracy was obtained with the decision tree classifier, with the result of 44.44% and 11.11%, respectively.

In the case of ω_S, the classification process failed to distinguish this class from the other ones. The main reason for this result is related to the low number of patients (n=4) that characterize the class.

	ω_N	ω_S	ω_{CHC}	ω_{CC}	ω_{DC}	ω_{HCC}	Overall
kNN (k=5)	94.12	0	0	0	88.57	83.33	70.1
Dec. Tree	73.53	0	11.11	44.44	77.14	0	58.76
SVM_P	94.12	0	0	33.33	91.43	66.67	73.20
SVM_R	94.12	0	0	22.22	97.14	50.0	73.20

Table 2. Overall and individual class accuracies (%) obtained with different classifiers.

4. CONCLUSIONS

Results shown throughout this paper suggest the usefulness of combining US image features with laboratory and clinical parameters for accurately identifying different stages of chronic liver disease.

A novelty introduced in this work is a pre-processing step which enables to obtain a de-speckled (intensity) and a speckle (texture) image field, from where distinct intensity and textural features can be extracted.

An optimal feature set, obtained by means of forward selection, was found to be: INR, $diagnose$, m, T, $Ence$ and E_V. This observation reinforces the relevance of combining features from various sources to improve classification.

The SVM classifier outperformed the kNN and the decision tree classifiers, showing the highest overall and individual class accuracies for classifying ω_N and ω_{DC} (97.14%) groups. Moreover, the kNN classifier yields the best performance for correctly classifying in ω_{HCC} while in ωCC and ωCHC the decision tree classifier performs better.

In the future the authors intend to (i) expand the data set in order to obtain an equitable number of samples in each class, (ii) include other features to increase diagnostic accuracy and (iii) perform a more exhaustive analysis in terms of classifiers, such as using a combination of classifiers.

5. REFERENCES

[1] S. Sherlock and J.Dooley, *Diseases of the liver and Biliary System*, Blackwell Science Ltd, 11 edition, 2002.

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

[3] T. Wang, J. Saniie, and X. Jin, "Analysis of low-order autoregressive models for ultrasonic grainsignal characterization," *IEEE transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 38, no. 2, pp. 116–124, March 1991.

[4] Y. Kadah, 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, pp. 466–478, 1996.

[5] Kazuo Maeda, Masaji Utsu, and Paul E. Kihaile, "Quantification of sonographic echogenicity with grey-level histogram width: A clinical tissue characterization," *Ultrasound in Medicine & Biology*, vol. 24, no. 2, pp. 225 – 234, 1998.

[6] D. Gaitini, Y. Baruch, E. Ghersin, E. Veitsman, H. Kerner, B. Shalem, G. Yaniv, C. Sarfaty, and H. Azhari, "Feasibility study of ultrasonic fatty liver biopsy: Texture vs. attenuation and backscatter," *Ultrasound in Medicine & Biology*, vol. 30, no. 10, pp. 1321 – 1327, 2004.

[7] W. Yeh, Y. Jeng, C. Li, P. Lee, and P. Li, "Liver fibrosis grade classification with b-mode ultrasound," *Ultrasound in Medicine & Biology*, vol. 29, pp. 1229–1235, 2003.

[8] J.S. Bleck, U. Ranft, M. Gebel, H. Hecker, M. Westhoff-Bleck, C. Thiesemann, S. Wagner, and M. Manns, "Random field models in the textural analysis of ultrasonic images of the liver," *Medical Imaging, IEEE Transactions on*, vol. 15, no. 6, pp. 796–801, Dec 1996.

[9] J. Seabra and J. Sanches, "Modeling log-compressed ultrasound images for radio frequency signal recovery," *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008.

[10] José C. Seabra and João M. Sanches, "On estimating de-speckled and speckle components from b-mode ultrasound images," in *Proceedings of the 2010 IEEE international conference on Biomedical imaging: from nano to Macro*. 2010, ISBI'10, pp. 284–287, IEEE Press.

[11] Ferdinand van der Heijden, Robert Duin, Dick de Ridder, and David M. J. Tax, *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*, Wiley, 1 edition, 2004.

[12] Anil K. Jain, Robert P. Duin, and Jianchang Mao, "Statistical pattern recognition: A review," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, pp. 4–37, 2000.

[13] Steven R Deitcher, "Interpretation of the international normalised ratio in patients with liver disease," *The Lancet*, vol. 359, no. 9300, pp. 47 – 48, 2002.