Diffuse liver disease classification from ultrasound surface characterization, 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]Liver Unit, Department of Gastroenterology and Hepatology, Hospital de Santa Maria, Medical School of Lisbon
Lisbon, Portugal

Abstract. In this work liver contour is semi-automatically segmented and quantified in order to help the identification and diagnosis of diffuse liver disease. The features extracted from the liver contour are jointly used with clinical and laboratorial data in the staging process. The classification results of a *support vector machine*, a *Bayesian* and a *k-nearest neighbor* classifier are compared. A population of 88 patients at five different stages of diffuse liver disease and a leave-one-out cross-validation strategy are used in the classification process. The best results are obtained using the *k-nearest neighbor* classifier, with an overall accuracy of 80.68%. The good performance of the proposed method shows a reliable indicator that can improve the information in the staging of diffuse liver disease.

Keywords: Liver Cirrhosis, Contour Detection, 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 this 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. 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 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

* Corresponding author: Ricardo Ribeiro (ricardo.ribeiro@estesl.ipl.pt). This work was supported by project the FCT (ISR/IST plurianual funding) through the PIDDAC Program funds.

progressive phase 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. Due to its highly invasive nature, medical image modalities have been 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 facilities makes it the preferred method for assessing diffuse liver diseases such as cirrhosis.

Using US, cirrhosis is suggested by liver surface nodularity, portal vein mean flow velocity and the enlargement of the caudate lobe [1]. The study in [2] refer that nodular liver surface is a reliable sign in the detection of liver cirrhosis and can have a diagnostic accuracy of 70% or more. The authors in [3] showed that the observed liver contour irregularities directly correlated with the gross appearance of the cirrhotic liver as seen at laparoscopy. Liver surface nodularity in US sign can be appreciated when ascites is present or when a high-frequency transducer (7.5 - 12 MHz) is used [3]. In [2] the results, using a low-frequency transducer (3.5 -5 MHz), also showed that liver surface is a significantly parameter associated with the histopathological diagnosis of liver cirrhosis.

Nevertheless, as reported by [4], the validity of the different methods to detect changes in the liver surface are very subjective, since the segmentation and contour of such surface is operator-dependent. These fact leads to a subjective and non reproducible method to study the liver surface and consequently to a poor aid of an accurate liver diagnosis.

In this sense, it is proposed a semi-automatic method for the liver surface detection, 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 the liver contour detected in the *de-speckled* field, as well as clinical and laboratorial features, are used to train supervised classifiers to detect the disease.

Diffuse 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 *Bayesian classifier* 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, by reporting and comparing the classification results obtained with the features extracted from the liver contour, with the clinical and laboratorial features and with the total set of features. Section 4 concludes the paper.

2 Problem Formulation

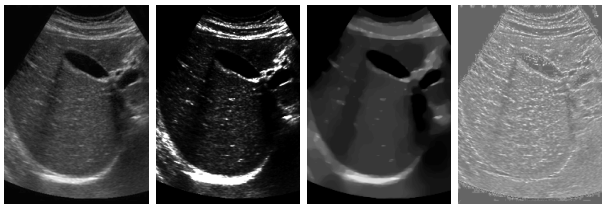
In the practice of US the perceived liver capsule and the adjacent overlying membranous structures (peritoneum, transverse fascia, pre-peritoneal fat) are not always clear and irregularities due to subfascial or sub-peritoneal pathology may be falsely described as abnormalities of the liver surface [3].

The decomposition procedure described in [5] 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 US 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 [6]

$$y(i,j) = x(i,j)\eta(i,j), \tag{1}$$

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.1 illustrates an example of the decomposition methodology.



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

Fig. 1. 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.

In the estimation of both images, RF envelope and *de-speckled* information, the use of total variation techniques allows the preservation of major transitions, as seen in the case of liver capsule and overlying structures.

Using the *de-despeckled* image, the liver surface contour is obtain using a snake technique, proposed by [7], which computes one iteration of the energy-minimization of active contour models. To initialize the snake, the operator needs to select four points of the liver surface.

Based on the detected **liver contour**, the following features were extracted:

1. root mean square of the different angles produced by the points that characterize the contour, rms_α, where the first point was assumed as the reference point,

2. root mean square of the variation of the points of the contour in the y axis, rms_y,
3. the mean and variance of the referred angles, m_α and v_α,
4. the variance of the y axis coordinates at each point, v_y, and
5. the correlation coefficient of the y axis coordinates, R.

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), ix) sodium (Na) and x) lactate dehydrogenase (LDH).

All these features, organized in a 23 length vector, are used in a forward selection method with the criterion of 1 - Nearest Neighbor *leave-one-out cross-validation* (LOOCV) performance in order 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) *Bayesian classifier* and iii) 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 [8]. 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 [8]. In this paper, a multiclass SVM classifier was adopted, using a Gaussian radial-basis kernel function and a polynomial kernel.

In the Bayes classifier the feature set, X, is assumed multivariate normal distributed [9] with means, μ_τ and covariances matrices, Σ_τ, according to each class. The linear discriminant functions are

$$g_\tau(X) = -\frac{1}{2}(X - \mu_\tau)^T \Sigma_\tau^{-1}(X - \mu_\tau) - \frac{1}{2} log\,|\Sigma_\tau| \tag{2}$$

where $\tau \in \{N, CHC, CC, DC, HCC\}$ and the *a priori* probability, $P(\omega_\tau)$, of the classes were calculated based on their frequencies.

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

3 Experimental Results

Eighty eight data samples were obtained from 88 patients. The patients were selected from the Department of Gastroenterology and Hepatology of the Santa Maria Hospital, in Lisbon, with known diagnosis. The samples were labeled in five classes; *Normal*, ω_N, *Chronic Hepatitis without cirrhosis*, ω_{CHC}, *Compensated Cirrhosis*, ω_{CC}, *Decompensated Cirrhosis*, ω_{DC}, and *Hepatocelular Carcinoma*, ω_{HCC}. Among them, 36 belong to ω_N, 7 to ω_{CHC}, 8 to ω_{CC}, 32 to ω_{DC} and 5 patients to ω_{HCC}.

From figure 2 we can appreciate the results obtained from the *de-speckled* and contour steps. To standardize the proceedings, and as reported in the literature, we focused the study in the anterior liver surface, using a low-frequency transducer. The results showed that in the *de-speckled* US image the liver boundary was clearly defined in the cases reported (for example, normal liver (first row), cirrhotic liver without ascites (second row) and cirrhotic liver with ascites (third row)). The detected contour was plotted in the original US image, so that the physician could evaluate the liver contour.

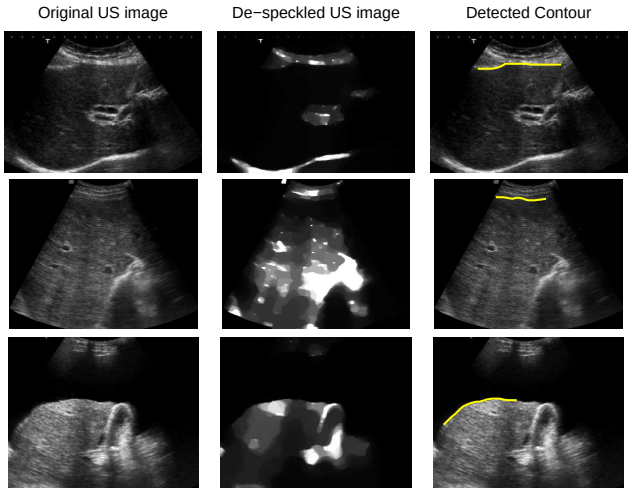


Fig. 2. Method used to detect the anterior liver surface contour. First row corresponds to a normal liver; second row to a compensated cirrhotic liver and the last row to a decompensated cirrhotic liver.

According to the criterions established for feature selection the results, reported in Table 1, showed 5 optimal features using only the contour features (optimal contour set), 8 optimal features using the clinical and laboratorial features (optimal clinical set) and 10 optimal features using the set of features (optimal feature set). In this last result, it is important to emphasize that the feature set selected is composed of five clinical features (A, $Ence$, T, $diagnose$,

GIB), four laboratory features (AST, INR, Na, $Crea$) and one US contour feature (R). Thus, the combination of features from different sources (US image, clinical, etc) integrates, in a quantitative and objective way, the information used in medical practice. This result is in accordance with [3, 12].

Feature Selection results for	
Contour features	m_α, R, rms_α, v_y, v_α.
Clinical/Lab. features	*AST*, *A*, *Ence*, *T*, *diagnose*, *INR*, *LDH*, *GIB*
All features	*AST*, *A*, *Ence*, *T*, *diagnose*, *INR*, *R*, *Na*, *GIB*, *Crea*.

Table 1. Feature selection results from the evaluation of the feature set, using only the contour features, only the clinical and laboratorial features and all features.

The classification technique significantly affects the final diagnosis [10]. Using the LOOCV method, the same data set was tested with different types of classifiers, namely a kNN, a *Bayes* 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 $k = 1$ for the optimal contour, clinical and feature set, which resulted in an error rate of 40.90%, 21.59% and 19.32%, respectively.

Considering the SVM classifiers, using the proposed sets, the best result of the SVM_P corresponds to a degree of 1, attaining an error rate of 45.45% for the optimal contour set, 25.0% for the optimal clinical set and 23.86% for the feature set. With the SVM_R the best performance for the optimal contour set was obtained with a radius of 1 showing an error of 48.86%, an error rate of 21.59% for the optimal clinical set with a radius of 1.9, and a radius of 1.8, for the case of the optimal feature set, with an error of 27.27%.

In the case of the *Bayesian* classifier, for each class, the mean and covariance were estimated using the LOOCV method.

Optimal Contour Set						
	ω_N	ω_{CHC}	ω_{CC}	ω_{DC}	ω_{HCC}	Overall
Bayes	27.78	28.57	0.00	37.50	80.00	31.82
kNN (k=1)	88.89	28.57	0.00	56.25	0.00	59.10
SVM_P	83.33	0.00	0.00	56.25	0.00	54.55
SVM_R	69.44	0.00	0.00	62.50	0.00	51.14
Optimal Clinical Set						
Bayes	0.00	28.57	0.00	0.00	100.00	7.95
kNN (k=1)	94.44	14.29	37.50	87.50	60.00	78.41
SVM_P	91.67	0.00	0.00	90.63	40.00	72.73
SVM_R	94.44	0.00	37.50	93.75	40.00	78.41

Table 2. Overall and individual class accuracies (%) obtained with different classifiers, using the optimal contour and clinical set.

Table 2 resumes the classification results obtained using the optimal contour and clinical set. The best overall accuracy, using the optimal contour set, of 59.10% was achieved with the kNN classifier, followed by the SVM_P, 54.55%, the SVM_R, 51.14% and the *Bayesian* classifier, with the worst result of the tested classifiers, achieving an overall accuracy of 31.82%. The diagnostic yield was improved from 59.10% to 78.41% when using the optimal clinical set, for kNN and SVM_R classifier. The accuracy results were greatly improved with this feature set for the individual class classification. By means of SVM_R it was obtained an accuracy of 94.44%, 93.75% and 37.50% for ω_N, ω_{DC} and ω_{CC}, respectively. For ω_{HCC} and ω_{CHC}, the best results were 100.0% and 28.27%, respectively, with the Bayes classifier.

	ω_N	ω_{CHC}	ω_{CC}	ω_{DC}	ω_{HCC}	Overall
Bayes	0.00	28.57	12.50	0.0	100.00	9.10
kNN (k=1)	91.67	71.43	25.00	87.50	60.00	80.68
SVM_P	91.67	28.57	0.00	87.50	80.00	76.14
SVM_R	88.88	0.00	0.00	93.75	40.00	72.73

Table 3. Overall and individual class accuracies (%) obtained with different classifiers, using the optimal feature set.

Combining features further improves the classifiers performance, as summarized in Table 3. With the optimal feature set, the best overall result was obtained with the kNN classifier. This result outperformed the best result obtained with the previous feature sets, which reinforce the idea of feature combination from different sources.

In terms of individual class accuracy, the best result for ω_N was obtained using the kNN and SVM_P classifiers, both with an accuracy of 91.67%. The best outcome in differentiating chronic hepatitis without cirrhosis samples (ω_{CHC}) and compensated cirrhosis (ω_{CC}), from the other classes was achieved by means of the kNN classifier with an accuracy of 71.43% and 25.00%, respectively. Regarding the classification of ω_{DC}, the best individual accuracy result was reached with the SVM_R classifier, yielding 93.75%. The detection of ω_{HCC} was 100.0% using the *Bayes* classifier.

4 Conclusions

In this work a semi-automatic detection of liver surface is proposed to help in the diagnosis of diffuse liver disease. The results shown in this paper suggest the usefulness of combining US liver contour features with laboratorial and clinical parameters for accurately identifying different stages of diffuse liver disease.

The pre-classification steps showed good results, since the *de-speckled* image field aided the detection of liver surface contour.

The optimal feature set outperformed the optimal contour and clinical set. In the classification procedure, using the optimal feature set, the kNN outperformed the rest of the classifiers in terms of overall accuracy. The low accuracy results in ω_{CC}, maybe due to the small sample size of the class. Another interesting result was the classification accuracy improvement in ω_N using the optimal clinical set.

This finding demonstrated the problem of the semi-automatic detection of the contour, since it has an operator-dependent component, the initialization step.

Promising results were obtained, which showed the discriminant power of the features as well as of the classifier, specially in terms of individual class accuracy. These results promote the development of more robust classification techniques, particularly classification combiners.

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, (iii) perform a more exhaustive analysis in terms of classifiers, such as using a combination of classifiers and (iv) use state-of-the-art automatic snakes, in order to create a fully automatic detection method.

References

1. S. Sherlock and J.Dooley. *Diseases of the liver and Biliary System*. Blackwell Science Ltd, 11 edition, 2002.
2. Stefano Gaiani, Laura Gramantieri, Nicola Venturoli, Fabio Piscaglia, Sebastiano Siringo, Antonia D'Errico, Gianni Zironi, Walter Grigioni, and Luigi Bolondi. What is the criterion for differentiating chronic hepatitis from compensated cirrhosis? a prospective study comparing ultrasonography and percutaneous liver biopsy. *Journal of Hepatology*, 27(6):979 – 985, 1997.
3. V Simonovsky. The diagnosis of cirrhosis by high resolution ultrasound of the liver surface. *Br J Radiol*, 72(853):29–34, 1999.
4. J A Ladenheim, D G Luba, F Yao, P B Gregory, R B Jeffrey, and G Garcia. Limitations of liver surface US in the diagnosis of cirrhosis. *Radiology*, 185(1):21–23, 1992.
5. 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.
6. 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*, ISBI'10, pages 284–287. IEEE Press, 2010.
7. Chris Bregler and Malcolm Slaney. *Snakes-A MatLab MEX file to demonstrate snake contour-following*, 1995.
8. W. Yeh, Y. Jeng, C. Li, P. Lee, and P. Li. Liver fibrosis grade classification with b-mode ultrasound. *Ultrasound in Medicine & Biology*, 29:1229–1235, 2003.
9. A. Mojsilovic, S. Markovic, and M. Popovic. Characterization of visually similar diffuse diseases from b-scan liver images with the nonseparable wavelet transform. *Image Processing, International Conference on*, 3:547, 1997.
10. 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*, 15:466–478, 1996.
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. Annalisa Berzigotti, Juan G. Abraldes, Puneeta Tandon, Eva Erice, Rosa Gilabert, Juan Carlos Garca-Pagan, and Jaime Bosch. Ultrasonographic evaluation of liver surface and transient elastography in clinically doubtful cirrhosis. *Journal of Hepatology*, 52(6):846 – 853, 2010.