Cirrhosis Prognostic Quantification with Ultrasound: An Approximation to Model for End-Stage Liver Disease

Ricardo Ribeiro[1,2,4,⋆], Rui Tato Marinho[3], and João Miguel Sanches[1,2,⋆⋆]

[1] Institute for Systems and Robotics
[2] Department of Bioengineering, Instituto Superior Técnico, Technical University
[3] Fauldade de Medicina da Universidade de Lisboa
[4] Escola Superior de Tecnologia da Saude de Lisboa, Portugal
ricardo.ribeiro@estesl.ipl.pt

Abstract. *Model for End-stage Liver Disease* (MELD) is a common score used in clinical practice to estimate the prognostic outcome of cirrhotic patients. This score is obtained from laboratory results.

Here a novel method is proposed to estimate the MELD score based on textural information extracted from normalized *ultrasound* (US) images of liver parenchyma. The information obtained from the co-ocorrence matrix and the monogenic decomposition of the image is linearly combined to compute the score. The application of US data for prognosis purposes is also a noteworthy novelty of this paper.

A dataset of 82 cirrhotic patients is used in this work. An optimal cut-off from a ROC analysis lead to an accuracy of 80% and an AUROC of 0.801 in the prognosis prediction. No statistical differences were found between the MELD score and the proposed US score and a strong correlation (0.65 $p < 0.01$) was attained.

Keywords: Cirrhosis prognosis, Ultrasound, model fitting, MELD.

1 Introduction

Chronic liver disease (CLD) is a major public health problem [1]. CLD final stage is cirrhosis, which in most cases evolves to *hepatocellular carcinoma* (HCC) [1].

CLD staging is based on clinical, biological and morphopathological evaluation, obtained from liver biopsy. Liver biopsy is considered a key role in the diagnosis and follow-up of CLD [2,3]. The need for biopsy reduction, due to its invasive nature and the potential for sampling errors, has led to the development of noninvasive methods [2,3]. US has gained particular interest, since it is widely available, inexpensive, non-ionizing [3]. It is already the screening choice for HCC in cirrhotic patients.

In this spectrum, the majority of the studies have been focused on the role of US in the detection of cirrhosis. However, beyond diagnosis, prognosis is an

⋆ Corresponding author.
⋆⋆ This work was supported by the FCT project [PEst-OE/EEI/LA0009/2011].

J.M. Sanches, L. Micó, and J.S. Cardoso (Eds.): IbPRIA 2013, LNCS 7887, pp. 551–558, 2013.

essential part of the baseline assessment of any disease [4]. Accurate estimation of CLD prognosis is highly important since it can be used as a guide for ordering additional tests and selecting appropriate therapies.

Liver transplantation is the proposed treatment for end-stage cirrhosis. Therefore, a reliable prognostic model for organ allocation on the liver transplantation waiting list is needed [5]. MELD score has been widely applied, replacing other models, and accurately predicts a broad spectrum of liver disease [5].

MELD correctly ranks cirrhotic patients according to their risk of death over a 3-month time period. [6] showed that the mortality increased in proportion to the increase of MELD. In [7] a strong correlation ($\rho = 0.768$) between an US quantitative scoring system and the MELD score is shown. However, some of the proposed features are based on visual inspection, thus less objective.

This paper is organized as follows: Section 2 describes the CLD patients characterisitics (2.1) and introduces the US decomposition algorithm, US feature extraction and selection of the polynomial fitting model (2.2). In section 3 the experimental results of the tested models are presented and concluding remarks are discussed in Section 4.

2 Materials and Methods

Here, we outline patients clinical characteristics, the procedure for data acquisition and the methodology used to develop the prognostic model from US images.

2.1 Data

Eighty two (82) patients with cirrhosis who had been in referred to the Department of Gastroenterology of Santa Maria Hospital affiliated to Lisbon Medicine Faculty were evaluated. Based on liver biopsy results and clinical diagnosis, 35 patients were diagnosed as compensated cirrhosis and 47 as decompensated cirrhosis, as summarized in Table 1. Patients with history of HCC were excluded.

For each patient, MELD score was calculated using the original formula [6], $\text{MELD} = 9.6 \times \ln(\text{creatinine}) + 3.8 \times \ln(\text{bilirubin}) + 11.2 \times \ln(\text{INR}) + 6.43$.

To decrease intra-patient variability, all patients underwent US and biochemical tests in the same day. A Philips CX© 50 US scanner was used, with a broadband curved array transducer C5-1©. Images were captured in a 1024x1024 pixel matrix, DICOM format, with a grey level resolution of 8 bits/pixel.

An acquisition preset was set using the following parameters: fundamental frequency of 3.5 MHz, without the use of harmonics or compounding technologies, depth of 15cm and two focal zones were used and set at the central portion of the image (7.5 cm). The dynamic range was set in 70 dB and the gain was variable, according to the patient biotype. Time gain compensation was set to its central position and kept constant. US images were acquired in the right liver lobe and according to patient biotype different transducer orientation angles were performed, using as protocol the same liver anatomical landmarks. Patients were positioned in supine, comfortable and asked to breath gently, avoiding major patient motion.

Table 1. Demographic, clinical and biochemical features of the studied population

	All patients (n=82)	**Compensated Cirrhosis** (n=35)	**Decompensated Cirrhosis** (n=47)
Demographic (mean±SD)			
Age (yr)	61±12	60±10	61±13
Aetiology (%)			
Alcohol	40.24	17.14	57.45
HBV	7.32	11.43	4.26
HCV	23.17	40.00	10.64
Alcohol + HBV	1.22	-	2.13
Alcohol + HCV	10.98	2.86	17.02
Other causes	13.41	22.86	6.38
Biochemical (mean±SD)			
Albumin (g/dL)	2.82±1.23	2.66 ± 1.77	2.9±0.54
Serum bilirubin (mg/dL)	2.87±4.61	1.53 ± 1.94	23.85±5.69
Serum creatinine (mg/dL)	1.29±1.81	0.97± 0.48	1.52±1.50
INR	1.29±0.47	1.18 ± 0.61	1.37±0.31
MELD (mean±SD)			
	12.42±7.33	8.72 ± 5.50	14.98±7.39

HBV - Hepatitis B virus; HCV - Hepatitis C virus

2.2 Prognostic Model

To assemble an accurate, objective and realistic model from US images, the following steps are used:

US Image Pre-processing

To eliminate the influence of the US scanner and operator, US images are normalized and decomposed. The procedure described in [8] to separate the textural and intensity information within US images is here adopted. In this, 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 US scanner parameters tuned by the clinician during the US exam.

The estimated RF image is decomposed in *de-speckled* and *speckle* fields according to the following model [8]

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

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, $x(i,j)$. Figure 1 illustrates an example of the decomposition methodology in a US liver image of a patient with decompensated cirrhosis.

US Feature Extraction and Selection

From each *speckle* field, a ROI, of 128x128 pixels, is manually selected by an expert operator along medial axis with the criteria: i) representative of liver

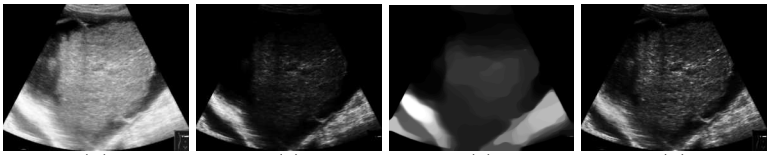


Fig. 1. Decomposition procedure of US liver parenchyma. a) Observed *B-mode* US image. Estimated b) envelope RF image, c) *de-speckle* and d) *speckle* fields.

parenchyma; ii) avoid major vessels and ligaments; and iii) as superficial as possible, to avoid US beam distortions. The following features are then extracted.

The elements of the *Co-ocurrence* matrix, $\mathbf{Co} = \{c_{i,j}(\Delta_l, \Delta_c)\}$, describe image gray level spatial inter-relationship [9]. More precisely, element $c_{i,j}(\Delta_l, \Delta_c)$ represents the joint probability of the pixel intensities i and j in relative spatial position of (Δ_l, Δ_c) [9] and can be computed as follows

$$c_{i,j}(\Delta_l, \Delta_c) = \sum_{l=1}^{N} \sum_{c=1}^{M} \begin{cases} 1 \; if(\eta_{l,c} = i) \wedge (\eta_{l+\Delta l, c+\Delta c} = j) \\ 0 \; otherwise \end{cases} \tag{2}$$

For a pixel distance of 1, we have four angular $[0^o, 45^o, 90^o, 135^o]$ *Co-occurrence* tensors for $(\Delta_l, \Delta_c) \in \{(0,1), (-1,1), (-1,0), (-1,-1)\}$, where the following features are calculated: *Contrast*, *Correlation*, *Energy* and *Homogeneity*.

US features are also extracted from the monogenic decomposition (see Figure 2). Given a two dimensional signal $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, [10] define the three-component monogenic signal as

$$f_m(\mathbf{x}) = (f(\mathbf{x}), Re(\mathcal{R}f(\mathbf{x})), Im(\mathcal{R}f(\mathbf{x}))) = (f, f_1, f_2), \tag{3}$$

where $\mathcal{R}f(\mathbf{x})$ is the Reisz transform, the local amplitude of the signal is given by $A(\mathbf{x}) = \|f_m(\mathbf{x})\| = \sqrt{f^2 + f_1^2 + f_2^2}$ while its local orientation θ and local phase ψ are specified by the following relations

$$f = A\cos\psi, \quad f_1 = A\sin\psi\cos\theta, \quad f_2 = A\sin\psi\sin\theta. \tag{4}$$

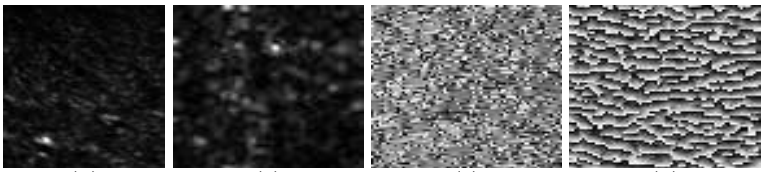


Fig. 2. Monogenic decomposition of an US image. (a) *speckle* field, (b) local amplitude (A), (c) local orientation (θ) and (d) local phase (ψ).

Here, a 3 level monogenic decomposition is used with the algorithm proposed by [11]. From each component (A, θ, ψ), at each level, the *autoregressive* (AR) coefficients of a first order 2D model $\{a_{1,1}, a_{1,0}, a_{0,1}\}$, energy and mean, are extracted.

A total of 61 features were extracted from each US image. To avoid overfitting, feature selection is performed with the stepwise regression analysis method ($p < 0.05$ to add; $p > 0.1$ to remove) [12]. An optimal subset of 2 features was produced: F_1: *Contrast* (-1,-1) and F_2: $a_{1,1}\psi_1$.

Model Selection and Accuracy

The polynomial model is used to fit MELD (M) score data and it is defined as follows,

$$\hat{M}(F_1, F_2) = \sum_{i=0}^{D} \sum_{j=0}^{D} w_{i,j} F_1^i F_2^j. \tag{5}$$

In this study, we tested the polynomial model raging the degree, D, from $D = 1, ..., 4$. The coefficients of each tested model are calculated in a least squares sense. The following *goodness of fit* tests are used to assess the adequacy of a model and by comparing it with order models under consideration.

***Sum of squares due to error* (SSE).** Measures the total deviation of the response values from the fit to the response values.

$$SSE = \sum (M_i - \hat{M}_i)^2 = \sum e_i^2. \tag{6}$$

***Root mean square error* (RMSE).** Estimates of the deviation of the random component in the data.

$$RMSE = \sqrt{\frac{\sum (M_i - \hat{M}_i)^2}{n - 2}} \tag{7}$$

***R-square*(r^2).** Measures how successful the fit is in explaining the variation of the data.

$$r^2 = \frac{\sum (M_i - \bar{M}_i)^2 - \sum (M_i - \hat{M}_i)^2}{\sum (M_i - \bar{M}_i)^2} \tag{8}$$

***Adjusted R-square*(r_a^2).**

$$r_a^2 = r^2 - \frac{1 - r^2}{n - 2}. \tag{9}$$

Assessing the Prognostic Outcome

MELD is an objective score, measured by widely available laboratory tests [5]. US_{score} $(\hat{M})$ does not intend to replace MELD score, but to use it as a reference.

Even at low MELD scores (< 15) differentiation in prognosis survival can be made, leading to different clinical approaches and surveillance. Patients with a

MELD score < 9 experienced a 1.9% mortality and with MELD ≥ 10 a mortality rate superior to 6% at 3 months [6]. Two classes are defined: ω_{RRD} - reduced risk of death (MELD$<$ 9) and ω_{IRD} increased risk of death (MELD$\geq$ 10).

To optimize the cut-off (c) value of the US$_{score}$ for the defined mortality classes *receiver operator characteristic* (ROC) analysis is performed. ROC curve provide a graphical representation of the tradeoff between *sensitivity* (sens) and *specificity* (spec). The ROC curve can be defined as a parametric curve, $s(c) : R \rightarrow R^2$, where $s(c) = (1 - spec(c), sens(c))$. The optimal cut-off value is based on the c value that maximize the *area under the ROC curve* (AUROC).

3 Experimental Results

Table 2 summarizes the performance of each model, computed in a leave-one-out cross-validation basis. Linear model is the best choice overall, achieving the lowest SSE and $RMSE$ of 267.4 and 2.044, respectively, a r^2 of 0.919 and a r_a^2 of 0.917. In the linear model, described as

$$\hat{M} = w_1 \times F_1 + w_2 \times F_2 + w_3, \tag{10}$$

the estimated coefficients (with 95% confidence bounds) were found to be $w_1 = 3.99(3.07, 4.92)$, $w_2 = -42.43(-48.1, -36.75)$ and $w_3 = 29.58(26.84, 32.32)$. A Pearson correlation coefficient of 0.65 was achieved and no statistical differences were observed between the MELD (M) value and the proposed US$_{score}$ ($\hat{M}$) ($p < 0.01$). Figure 3 projects the fitting surface of $\hat{M}$.

To define the best cut-off value, a ROC analysis was performed with three coefficients combinations: model fit 1 - best fitting result ($29.58 + 3.99 \times F_1 - 42.43 \times F_2$), model fit 2 - coefficients of the lower confidence bounds ($26.84 + 3.072 \times F_1 - 48.1 \times F_2$) and model fit 3 - coefficients of the higher confidence bounds ($32.32 + 4.923 \times F_1 - 36.7 \times F_2$). The results are resumed in Table 3.

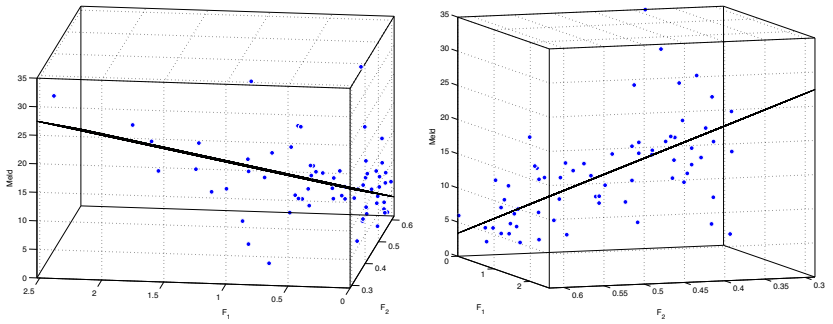
Table 2. Goodness of fit results of the tested models

Model	SSE	R-square	Adjusted R-square	RMSE
Linear (d=1)	267.4	0.919	0.917	2.044
Quadratic (d=2)	2115	0.363	0.311	5.89
d=3	1992	0.400	0.305	5.91
d=4	1244	0.625	0.524	4.89

Table 3. US$_{score}$ model performance with three different set of estimated coefficients

Model	AUROC(95% CI)	cut-off	OA (%)	Sens(%)	Spec(%)	PPV (%)	NPV(%)
fit 1	0.80 (0.70 - 0.91)	10.8	80	74.4	85.9	87.9	70.6
fit 2	0.79 (0.69 - 0.90)	11.6	79.1	76.9	82.1	85.7	71.9
fit 3	0.70 (0.69 - 0.90)	17.4	79.1	76.9	82.1	85.7	71.9

AUROC, area under the ROC curve; CI, confidence interval; OA, overall accuracy; PPV ad NPV, positive and negative predictive values



(*a*) (*b*)

Fig. 3. The linear model describing MELD score as a function of the US features: (a) F_1 and (b) F_2 view

Linear model fit 1 achieved the best performance with an AUROC of 0.801, a PPV of 87.9% and a NPV of 70.6%. The cost-effective cut-off was set at 10.8, leading to a detection rate of 74.4% for ω_{IRD} and 85.9% for ω_{RRD}. US_{score} accurately predicted, based on the MELD values, 66 of the 82 patients with cirrhosis, yielding an overall accuracy of 80.59%.

4 Discussion and Conclusions

The aim of this study was to develop an US *computer-aided diagnosis* (CAD) tool to predict the outcome of cirrhotic patients. The underlying hypothesis is that the morphological changes detected by US are correlated with the changes in liver function and metabolism. The proposed approach was based on the prediction of MELD, which as acquired particular interest due to its objective scoring system. It is calculated from three biochemical variables and accurately predicts 3-month mortality in liver cirrhosis.

Stepwise regression model selected two US features ($a_{1,1}$ ψ_1 and contrast Co(-1,-1)), that best describes the heterogeneous pattern of cirrhotic livers. The monogenic decomposition enhanced different US patterns, most often imperceptible to the human eye, since the local amplitude and phase includes intensity and structural information, respectively. Also, the use of the *speckle* field proved to be an appropriate approach, since it highlight the textural pattern and eliminate the influence of US scanner parameters.

Several polynomial models were then tested. The linear model achieved the best performance with a low $RMSE$ and high R-square. A strong positive correlation coefficient with MELD (0.65, $p < 0.01$) was also found, similar to correlation report in [7] of 0.76 ($p < 0.05$).

To work as a prognostic tool in a first line assessment of cirrhosis outcome, two classes were set: patients with reduced risk of death (ω_{RRD}) and patients with increased risk of death (ω_{IRD}) in a 3 months window. By means of the ROC analysis an optimal cut-off was selected and an accuracy of 80% was attained.

In future studies the influence of US scanner parameters and ROI location will be assessed. Also an integrated support system for clinicians will be developed and a wider cirrhotic population will be used.

In conclusion, a new and objective algorithm as been proposed for the assessment of cirrhotic patients outcomes based on US liver images. This algorithm is built upon monogenic signal and co-occurrence US features, allowing a rapid analysis of liver function without the need for further laboratorial tests apart from an common US scanner.

References

1. Marinho, R., Giria, J., Moura, M.: Rising costs and hospital admissions for hepatocellular carcinoma in portugal (1993-2005). World J. Gastroenterol. 13(10), 1522–1527 (2007)
2. Aube, C., Oberti, F., Korali, N., Namour, M.-A., Loisel, D., Tanguy, J.-Y., Valsesia, E., Pilette, C., Rousselet, M.C., Bedossa, P., Rifflet, H., Maiga, M.Y., Penneau-Fontbonne, D., Caron, C., Cales, P.: Ultrasonographic diagnosis of hepatic fibrosis or cirrhosis. Journal of Hepatology 30(3), 472–478 (1999)
3. Allan, R., Thoirs, K., Phillips, M.: Accuracy of ultrasound to identify chronic liver disease. World J. Gastroenterol. 28(16), 3510–3520 (2010)
4. D'Amico, G., Garcia-Tsao, G., Pagliaro, L.: Natural history and prognostic indicators of survival in cirrhosis: Systematic review of 118 studies. Journal of Hepatology 44(1), 217–231 (2006)
5. Cholongitas, E., Papatheodoridis, G.V., Vangeli, M., Terreni, N., Patch, D., Burroughs, A.K.: Systematic review: the model for end-stage liver disease: should it replace child-pugh's classification for assessing prognosis in cirrhosis? Alimentary Pharmacology & Therapeutics 22(11-12), 1079–1089 (2005)
6. Wiesner, R., Edwards, E., Freeman, R., Harper, A., Kim, R., Kamath, P., Kremers, W., Lake, J., Howard, T., Merion, R.M., Wolfe, R.A., Krom, R.: Model for end-stage liver disease (meld) and allocation of donor livers. Gastroenterology 124, 91–96 (2003)
7. Yan, G., Duan, Y., Ruan, L., Chao, T., Yang, Y.: A study on the relationship between ultrasonographic score and clinical score (meld, cpt) in cirrhosis. Hepatogastroenterology 52, 1329–1333 (2005)
8. Seabra, J.C., Sanches, J.A.M.: 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 2010, pp. 284–287. IEEE Press (2010)
9. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics SMC-3(6), 610–621 (1973)
10. Felsberg, M., Sommer, G.: The monogenic signal. IEEE Transactions on Signal Processing 49(12), 3136–3144 (2001)
11. Kovesi, P.: Phase congruency: A low-level image invariant. Psychological Research, 136–148 (2000)
12. Rawlings, J.O., Pantula, S.G., Dickey, D.A.: Applied regression analysis: a research tool. Springer (1998)
13. Zheng, R., Wang, Q., Lu, M., Xie, S., Ren, J., Su, Z., Cai, Y., Yao, J.: Liver fibrosis in chronic viral hepatitis: An ultrasonographic study. World J. Gastroenterol. 9(11), 2484–2489 (2003)