Ultrasound in Med. & Biol., Vol. ■, No. ■, pp. 1–14, 2017

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2017.05.002

● *Original Contribution*

A MACHINE-LEARNING ALGORITHM TOWARD COLOR ANALYSIS FOR CHRONIC LIVER DISEASE CLASSIFICATION, EMPLOYING ULTRASOUND SHEAR WAVE ELASTOGRAPHY

Ilias Gatos,* Stavros Tsantis,* Stavros Spiliopoulos,† Dimitris Karnabatidis,‡ Ioannis Theotokas,§ Pavlos Zoumpoulis,§ Thanasis Loupas,¶ John D. Hazle,‖ and George C. Kagadis*‖

*Department of Medical Physics, School of Medicine, University of Patras, Rion, Greece; †2nd Department of Radiology, School of Medicine, University of Athens, Athens, Greece; ‡Department of Radiology, School of Medicine, University of Patras, Patras, Greece; §Diagnostic Echotomography SA, Kifissia, Greece; ¶SuperSonic Imagine SA, Aix-en-Provence, France; and ‖Department of Imaging Physics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

(*Received* 5 *December* 2016; *revised* 26 *April* 2017; *in final form* 1 *May* 2017)

Abstract—The purpose of the present study was to employ a computer-aided diagnosis system that classifies chronic liver disease (CLD) using ultrasound shear wave elastography (SWE) imaging, with a stiffness value-clustering and machine-learning algorithm. A clinical data set of 126 patients (56 healthy controls, 70 with CLD) was analyzed. First, an RGB-to-stiffness inverse mapping technique was employed. A five-cluster segmentation was then performed associating corresponding different-color regions with certain stiffness value ranges acquired from the SWE manufacturer-provided color bar. Subsequently, 35 features (7 for each cluster), indicative of physical characteristics existing within the SWE image, were extracted. A stepwise regression analysis toward feature reduction was used to derive a reduced feature subset that was fed into the support vector machine classification algorithm to classify CLD from healthy cases. The highest accuracy in classification of healthy to CLD subject discrimination from the support vector machine model was 87.3% with sensitivity and specificity values of 93.5% and 81.2%, respectively. Receiver operating characteristic curve analysis gave an area under the curve value of 0.87 (confidence interval: 0.77–0.92). A machine-learning algorithm that quantifies color information in terms of stiffness values from SWE images and discriminates CLD from healthy cases is introduced. New objective parameters and criteria for CLD diagnosis employing SWE images provided by the present study can be considered an important step toward color-based interpretation, and could assist radiologists' diagnostic performance on a daily basis after being installed in a PC and employed retrospectively, immediately after the examination. (E-mail: gkagad@gmail.com) **© 2017 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Fibrosis, Shear wave elastography, Computer-aided diagnosis, Classifier design, Ultrasonics.**

INTRODUCTION

Chronic liver disease (CLD) is considered the 12th leading cause of death in the United States, with a continuously increasing rate in recent years. Especially at the final stage, cirrhosis, mortality rates are significant for those 45–85 y of age (Murphy et al. 2013). In 2013, cirrhosis was responsible for more than 1.2 million deaths worldwide, which represented a 50% increase from the early 1990s (GBD Mortality and Causes of Death Collaborators 2015). In addition, hepatocellular carcinoma (HCC) caused by cirrhosis is the third leading cause of cancer mortality worldwide (Altekruse et al. 2009).

Chronic liver disease is caused by hepatitis viruses (HAV, HBV, HCV, HDV and HEV), extensive alcohol consumption, unhealthy dietary patterns, which may lead to non-alcoholic fatty liver disease (NAFLD) or non-alcoholic steatohepatitis (NASH), autoimmune hepatitis (AIH) and genetic metabolic disorders such as Wilson's disease and haemochromatosis. It can be also caused by primary biliary cirrhosis (PBC), secondary biliary cirrhosis (SBC) or primary sclerosing cholangitis (PSC). CLD causes liver tissue inflammation, which in turn leads to repetitive injury and fibrosis. The disease, if not treated, leads in several years to cirrhosis and partial

Address correspondence to: George C. Kagadis, Associate Professor of Medical Physics–Medical Informatics, Department of Medical Physics, School of Medicine, University of Patras, Rion, GR 265 04, Greece. E-mail: gkagad@gmail.com

or total liver malfunction. A major challenge for clinicians today is to monitor the disease progression and estimate the patient's fibrosis stage to select the best treatment. In the case of cirrhosis and HCC, liver transplantation and surgical removal of the malignant lesion are to be decided.

Fibrosis staging is based on the Metavir classification system, according to liver biopsy (LB). The Metavir scale consists of five stages ranging from 0 to 4 (F0 = no fibrosis, F1 = mild fibrosis, F2 = significant fibrosis, F3 = severe fibrosis, F4 = cirrhosis) (Goodman 2007). Although LB is still considered the gold standard for estimation of fibrosis stage, it carries serious limitations. It is invasive and costly, and nearly 30% of patients have post-surgical effects such as substantial pain, pneumothorax, bleeding, infection, septicemia, biloma, haemobilia, accidental injury to adjacent structures, biliary peritonitis and even death (Gilmore et al. 1995).

Another drawback of LB is the lack of fibrosis uniformity across the liver tissue. Biopsy needles provide a small volume of tissue (1/50,000th of the total mass of the liver), resulting in staging variability between samples derived from different liver areas (Carey and Carey 2010; Goodman 2007).

New and non-invasive approaches to fibrosis stage assessment have been developed in the last few years. These include biochemical serum markers (BSMs) and imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US). BSMs have exhibited adequate performance in estimating significant fibrosis (F ≥ F2), and cirrhosis (F4), but failed in early-stage estimation (Carey and Carey 2010; Frulio and Trillaud 2013; Martinez et al. 2011). With respect to CT imaging, Romero-Gómez et al. (2008) reported area under the curve (AUC) values of 0.83 for F ≥ F2 and 0.86 for F ≥ F3 in patients with chronic hepatitis C. Magnetic resonance elastography (MRE) is a contemporary imaging technique exhibiting great accuracy at all stages. Although very accurate in the few studies in which it has been tested (Carey and Carey 2010; Huwart et al. 2008), MRE is expensive and requires extensive validation. B-Mode ultrasound (US) only predicts diffuse abnormalities (Sanford et al. 1985) and confirms the absence of cirrhosis (Giorgio et al. 1986; Harbin et al. 1980).

Over the last decade, US elastography (USE) has been introduced to correlate liver stiffness to fibrosis stages. Several elastographic approaches have been introduced for CLD evaluation, such as transient elastography (TE), acoustic radiation force impulse (ARFI), real-time elastography (RTE) and shear wave elastography (SWE) imaging. All clinical studies dealing with USE are based on stiffness cutoff values that correspond to fibrosis stages. TE carried out with the Fibroscan US system (Echosens, Paris, France) is the oldest and most validated elastographic method employed for liver fibrosis evaluation. The mean diagnostic accuracy of TE has an AUC value of 0.85. with average sensitivity and specificity values of 0.78 and 0.79 for F ≥ F2, F ≥ F3 and F ≥ F4 (Bota et al. 2013; Chung et al. 2013; Frulio and Trillaud 2013; Tsochatzis et al. 2011). TE suffers from optimal stiffness cutoff value validation because of the overlap between different fibrosis stages in several studies. RTE, developed by Hitachi, also has a mean AUC value of 0.85 and mean sensitivity and specificity values of 0.83 and 0.77 for all stages above F2 (Chung et al. 2013; Hong et al. 2014; Kobayashi et al. 2015). Meta-analyses have indicated that the qualitative nature of RTE provides only useful visualization of stiffness contrasts and not quantification of fibrosis stages (Hong et al. 2014; Kobayashi et al. 2015). ARFI elastography, recently introduced by Siemens, is considered a useful tool for fibrosis evaluation, with results similar to those for TE (0.85 mean AUC value) (Bota et al. 2013; Chung et al. 2013; Frulio and Trillaud 2013; Nierhoff et al. 2013). Nevertheless, ARFI imaging carries several limitations; elasticity measurements are not given in real time, and only one acquisition can be made each time in a small pre-determined and constant area. Until today, only few ARFI-based studies have been published, and the validity and utility of the method have not been established. SWE is yet another recently introduced technique that offers real-time elasticity imaging as well as stiffness quantification (in units of kPa) over a 2-D region of interest; it is commercially available on the Aixplorer US scanner (SuperSonic Imagine, Aix-en-Provence, France). SWE offers good classification results (mean AUC values of 0.81 for F ≥ F1 and 0.86 for F ≥ F2), but, like the previously mentioned techniques, also has limitations such as stiffness cutoff threshold variability between different studies, uncertainties in the selection of the optimum quantification region and a lack of clear guidelines on how to avoid artifacts and areas of poor image quality (Bota et al. 2015; Deffieux et al. 2015; Ferraioli et al. 2012a, 2015; Gerber et al. 2015; Sporea et al. 2014). Only few elastographic studies (TE, RTE, ARFI, and SWE) provide mild fibrosis (F ≥ F1) classification results, compared with significant fibrosis (F ≥ F2) and cirrhosis (F = F4). For SWE clinical studies, only Sporea et al. (2014) provided liver stiffness cutoff values for predicting mild fibrosis using TE as the gold standard. The cutoff value for F ≥ F1 was set at 7.1 kPa, achieving an AUC value of 0.825 and total accuracy of 76% with sensitivity and specificity values of 74.5%, and 78%, respectively (Sporea et al. 2014).

Several attempts that quantify sonographic findings by means of pattern recognition algorithms to classify CLD have been reported so far. Regarding B-mode US