Ultrasound in Med. & Biol., Vol. ■, No. ■, pp. 1–12, 2017

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2016.12.016

● *Original Contribution*

SONOELASTOMICS FOR BREAST TUMOR CLASSIFICATION: A RADIOMICS APPROACH WITH CLUSTERING-BASED FEATURE SELECTION ON SONOELASTOGRAPHY

Qi Zhang,* Yang Xiao,† Jingfeng Suo,* Jun Shi,* Jinhua Yu,‡ Yi Guo,‡ Yuanyuan Wang,‡ and Hairong Zheng†

*Institute of Biomedical Engineering, Shanghai University, Shanghai, China; †Paul C. Lauterbur Research Center for Biomedical Imaging, Institute of Biomedical and Health Engineering, Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; and ‡Department of Electronic Engineering, Fudan University, Shanghai, China

(*Received* 19 *June* 2016; *revised* 9 *December* 2016; *in final form* 24 *December* 2016)

Abstract—A radiomics approach to sonoelastography, called "sonoelastomics," is proposed for classification of benign and malignant breast tumors. From sonoelastograms of breast tumors, a high-throughput 364-dimensional feature set was calculated consisting of shape features, intensity statistics, gray-level co-occurrence matrix texture features and contourlet texture features, which quantified the shape, hardness and hardness heterogeneity of a tumor. The high-throughput features were then selected for feature reduction using hierarchical clustering and three-feature selection metrics. For a data set containing 42 malignant and 75 benign tumors from 117 patients, seven selected sonoelastomic features achieved an area under the receiver operating characteristic curve of 0.917, an accuracy of 88.0%, a sensitivity of 85.7% and a specificity of 89.3% in a validation set via the leave-one-out cross-validation, revealing superiority over the principal component analysis, deep polynomial networks and manually selected features. The sonoelastomic features are valuable in breast tumor differentiation. (E-mail: zhangq@shu.edu.cn or zhangq@t.shu.edu.cn) **© 2017 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Radiomics, Sonoelastography, Breast tumor, Classification, Feature selection, Hierarchical clustering.**

INTRODUCTION

Ultrasound elastography, or sonoelastography, has emerged as a valuable tool for breast tumor characterization because of its depiction of tissue hardness in color images (Barr et al. 2015; Zhang et al. 2014b, 2016b). Malignant and benign tumors have different color patterns on sonoelastography because of their different hardness distributions. There are two main categories of sonoelastography, strain elastography (Kadour and Noble 2009; Ophir et al. 1991) and shear wave elastography (Bercoff et al. 2004; Nightingale et al. 2003). Strain elastography is easy to use and provides elasticity images in a manner similar to that of palpation (Shiina et al. 2015). Many manufacturers produce medical ultrasound devices with a strain elastography function (Shiina et al. 2015). Considering its increasing availability, the present study is focused on strain elastography.

Address correspondence to: Qi Zhang, Institute of Biomedical Engineering, Shanghai University, Xiangying Building, Room 803, No. 333, Nanchen Road, Shanghai, 200444, China. E-mail: zhangq@shu.edu.cn or zhangq@t.shu.edu.cn

In clinical practice of strain elastography, the Tsukuba score is usually used for qualitative assessment of breast tumors; this 5-point scale is used to visually grade the hardness of a mass (Itoh et al. 2006). Ten-point grading (Zhi et al. 2013), three-point grading (Kim et al. 2015) and another five-point grading (Alhabshi et al. 2013) are also employed. However, these grading methods suffer from considerable inter-observer variability because of their subjective and qualitative description of lesion hardness (Yoon et al. 2011).

Quantitative assessment has been proposed to provide less subjective and less operator-dependent descriptions. It usually measures the ratio of the strain in fat or gland to the strain in a tumor, that is, fat-to-lesion strain ratio or gland-to-lesion strain ratio (Cho et al. 2010; Fausto et al. 2015; Zhao et al. 2012; Zhou et al. 2014), or the ratio of the hard area within a tumor to the area of the entire tumor (*i.e.*, area ratio) (Zhang et al. 2014b). These ratios were proposed based on the fact that malignant breast tumors are usually harder than benign tumors. A feature related to tumor shape was also derived as the ratio of the lesion size on elastography to the B-mode size (*i.e.*, size ratio)

(Alhabshi et al. 2013; Barr et al. 2015). However, these few descriptors have attained limited diagnostic performance, probably because they focus only on a certain aspect of the tumor hardness or shape while neglecting other useful information such as tumor heterogeneity. Breast tumor is a heterogeneous tissue with intra-tumoral regional variations in proliferation, cell death, metabolic activity, vascular structure and other factors (Asselin et al. 2012; Zhang et al. 2015). The heterogeneity is also a pattern trait of malignancy (Chaddad et al. 2015; Zhang et al. 2015). Thus, tumor shape, hardness and heterogeneity should all be taken into consideration in breast tumor classification.

Recent advances in machine learning algorithms allow for more objective and precise quantitative imaging descriptors, which could be used to comprehensively evaluate breast tumor intensity, shape and texture and could potentially be used as non-invasive biomarkers for discrimination between malignant and benign tumors (Venkatesh et al. 2015). *Radiomics* refers to the extraction and analysis of a large number of quantitative features with high throughput from medical images (Aerts et al. 2014; Kumar et al. 2012; Lambin et al. 2012). Radiomics has been increasingly used in computer tomography, magnetic resonance imaging and positron emission tomography (Gillies et al. 2015; Huang et al. 2016; Vallières et al. 2015), but it is seldom employed in ultrasonography. In this article, we propose using a radiomics approach on sonoelastography for breast tumor classification, and thus we name the approach *sonoelastomics*. The high-throughput features are then selected for feature reduction using hierarchical clustering (HC). We hypothesize that the sonoelastomic features capture distinct differences in breast tumors and may have discriminative ability for tumor classification.

METHODS

Image acquisition, hardness retrieval and image segmentation

Ethical approval was obtained and the informed consent requirement was waived for this retrospective study. A sonoelastography data set containing 117 patients with 117 breast tumors (42 malignant and 75 benign) was used in the study. The elastograms were acquired before tumor biopsy using the HI VISION Preirus system (Hitachi Medical System, Tokyo, Japan) equipped with a 5- to 13-MHz linear array probe. All tumors were subjected to core biopsy or fine-needle aspiration cytology for histopathologic diagnosis as the gold standard. To examine the repeatability of elastography, we acquired two images from each of 110 tumors at two scanning planes or at an interval of around 10 s, and only one image was acquired for each of the remaining 7 tumors.

The Hitachi Preirus elastography system provides dual-modality visualization on a full screen (Fig. 1a), where the right part is a gray-scale B-mode image, and the left part is a composite color RGB image displayed as a translucent color elastographic image superimposed on the gray-scale B-mode image. Therefore, a pure color elastogram was obtained by subtracting the B-mode image from the composite image, but still in RGB format (Fig. 1b) (Zhang et al. 2014b, 2015, 2016a). The hardness distribution was then retrieved by computing the hue (H) values from the pure elastogram (Zhang et al. 2014b) as

$$H = \begin{cases} H0, \text{ if } B \leq G \\ 1-H0, \text{ if } B > G \end{cases} \quad (1)$$

$$H0 = \frac{1}{2\pi}\cos^{-1}\left\{\frac{2R-G-B}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right\} \quad (2)$$

where R, G and B are three color values of a pixel in the pure elastogram. The Hitachi elastography system only uses five-sixths of the full hue scale, namely, from red to blue (color bar in Fig. 1a), but without colors such as purple and purplish red that are covered in the remaining one-sixth. Therefore, the H value calculated from eqn (1) quantifies tissue hardness and ranges from 0 (red, softest) to 5/6 (blue, hardest), depicted as the gray-scale image in Figure 1e. There are missing areas without hardness information in the elastograms, which appear as black holes or shades (Fig. 1a,b). The pixels in these areas have invalid hue values and were automatically detected and excluded from further analysis (Fig. 1e).

An automated image segmentation method using the Chan-Vese level sets was applied to B-mode images to detect tumor boundaries, followed by a morphologic closing operation (Zhang et al. 2014b, 2015). The tumor boundaries detected on B-mode images (Fig. 1c) were then mapped to the retrieved elastograms (Fig. 1e) to specify the regions of interest.

Feature generation

Four categories of features were calculated: shape features, intensity statistics, gray-level co-occurrence matrix (GLCM) texture features and contourlet texture features.

The shape features quantified the morphology of tumors. They included area, convex area, perimeter, equivalent diameter, long- and short-axis lengths, orientation, solidity and eccentricity, as well as the mean, median and maximal thicknesses and the mean, median and maximal widths.