Ultrasound in Med. & Biol., Vol. ■, No. ■, pp. 1–11, 2015

0301-5629/$ - see front matter

ELSEVIER

http://dx.doi.org/10.1016/j.ultrasmedbio.2015.11.014

● *Original Contribution*

ULTRASOUND IMAGE DISCRIMINATION BETWEEN BENIGN AND MALIGNANT ADNEXAL MASSES BASED ON A NEURAL NETWORK APPROACH

Verónica Aramendía-Vidaurreta,* Rafael Cabeza,* Arantxa Villanueva,* Javier Navallas,* and Juan Luis Alcázar†

*Department of Electric and Electronic Engineering, Public University of Navarra, Pamplona, Spain; and †Department of Obstetrics and Gynecology, Clinical University of Navarra, Pamplona, Spain

(*Received* 17 *July* 2015; *revised* 2 *November* 2015; *in final form* 13 *November* 2015)

Abstract—The discrimination between benign and malignant adnexal masses in ultrasound images represents one of the most challenging problems in gynecologic practice. In the study described here, a new method for automatic discrimination of adnexal masses based on a neural networks approach was tested. The proposed method first calculates seven different types of characteristics (local binary pattern, fractal dimension, entropy, invariant moments, gray level co-occurrence matrix, law texture energy and Gabor wavelet) from ultrasound images of the ovary, from which several features are extracted and collected together with the clinical patient age. The proposed technique was validated using 106 benign and 39 malignant images obtained from 145 patients, corresponding to its probability of appearance in general population. On evaluation of the classifier, an accuracy of 98.78%, sensitivity of 98.50%, specificity of 98.90% and area under the curve of 0.997 were calculated. (E-mail: veronica.aramendia@gmail.com) **© 2015 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Adnexal mass, Texture feature, Classification, Neural network.**

INTRODUCTION

Adnexal masses are a common problem in clinical practice. Correct differential diagnosis is essential to provide optimal management. Benign masses can be treated conservatively (Alcazar et al. 2013a) or by surgical removal thorough minimally invasive surgery (Canis et al. 2002), whereas masses suspected of being malignant should be referred to a tertiary care center with a high volume of ovarian cancer cases (Engelen et al. 2006). Transvaginal ultrasound is currently considered the first-line imaging technique for this differential diagnosis (American College of Obstetricians and Gynecologists [ACOG] 2007). The main problem with this technique is that its diagnostic performance is based mainly on the subjective impression of the sonographer or sinologist (Valentin 2004), and therefore, it is highly dependent on the examiner's experience (Van Holsbeke et al. 2010). In fact, the accuracy and level of inter-observer agreement are both correlated with experience and the accuracy varies between 82% and 92% for the least and most and less experienced observers, respectively (Timmerman et al. 1999a). For this reason, some scoring systems (Alcazar et al. 2003), simple rules (Timmerman et al. 2005) and logistic models (Timmerman et al. 2008) have been proposed for use by less experienced examiners. Although good results have been reported (Alcazar et al. 2008, 2013b), certainly these approaches are not superior to the subjective impression of an experienced examiner (Valentin et al. 2011). A new alternative has emerged based on computer-aided diagnostic (CAD) technique using ultrasound images and image mining algorithms for classifying accurately adnexal masses (Acharya et al. 2014).

Research on this field is varied. A decision tree (DT) classifier was used successfully, with an average classification accuracy of 95.1% (Acharya et al. 2012b) and accuracy of 97% (Acharya et al. 2012c). However, only 20 patients were included (10 benign and 10 malignant), and 100 images were obtained for each patient volume. Therefore, the classifier benefited from an abundant supply of images, 1000 images of benign tumors and

Address correspondence to: Verónica Aramendía-Vidaurreta, Universidad Pública de Navarra DIEE, Department of Electrical and Electronical Engineering, Campus de Arrosadía s/n, Pamplona 31006, Spain. E-mail: veronica.aramendia@gmail.com

1000 images of malignant tumors, but less image heterogeneity. A support vector machine (SVM) classifier was used together with a radial basis function (RBF) to automatically classify benign and malignant ovarian tumor images, with a high accuracy of 99.9% (Acharya et al. 2012a). However, the study was also restricted to 20 patients. In addition, with the help of a gynecologist and radiologist, a region of interest (ROI) was delineated to enclose the suspicious portion of the image. Therefore, it is not a completely automatic method and depends on the operator's experience. A SVM classifier was also implemented on 187 ultrasound images from 177 patients (Khazendar et al. 2014, 2015). The first, based on classification decisions of high, medium and low confidence, respectively, provides accuracies of 90%, 81% and 69%. The second provides an average accuracy of 77% with a confidence interval of 95%. However, although a large number of patients are taken into account, a wide range of suspicious images are rejected in the high-confidence range. A probabilistic neural network (PNN) classifier was used to classify benign and malignant ovarian masses with an accuracy of 99.8% on a database of 2600 images on 20 patients (Acharya et al. 2014). However, in addition to the lower heterogeneity mentioned earlier, the work was performed using 3-D color Doppler images rather than commonly available B-mode ultrasound images. Also, a ROI was manually extracted. In 173 patients, an artificial neural network (ANN) model using patient age, menopausal status and serum CA-125 level as extracted features was employed, with a sensitivity of 95.9% and specificity of 81% (Timmerman et al. 1999b). However, the work was also evaluated with help of 3-D color Doppler images. A Bayesian least-squares support vector machine (LS-SVM) achieved an accuracy of 84.38% on 425 patients (Lu et al. 2003). However, among the 27 variables extracted, color Doppler images and blood flow indexes were acquired before operation.

The objective of the study described here was to develop a CAD technique for ultrasound images able to discriminate between benign and malignant adnexal masses and based on neural networks capable of overcoming the limitations of the available approaches. Moreover, the system must be able to deal with a heterogeneous group of patients and have a higher sensitivity–specificity balance with the future goal of being used as a practical tool during diagnostic decision making, while assisting doctors in the interpretation of suspicious ultrasound images.

METHODS

A method for automatic discrimination of adnexal masses is proposed. Figure 1 is a block diagram of the proposed CAD technique. It consists of an online classification system that predicts the class label (benign or malignant) of a test image based on a neural network classifier determined by an offline learning system. The offline learning classifier produces the trained networks using a combination of the offline training features and the respective offline ground truth (gold standard) of the training class labels (0/1 for benign/malignant).

Data

The risk that an ovarian neoplasm is malignant is 13% in pre-menopausal women and 45% in post-menopausal women (Koonings et al. 1989). In this study, 145 women (age: 35–65 years, mean = 43) with adnexal masses were selected. At the Clinical University of Navarra (CUN), all women gave informed written consent to use and analyze their ultrasound and clinical data, and institutional review board approval was obtained. Biopsies indicated that 39 had malignancy in their ovaries and 106 had benign conditions. All patients were evaluated by 3-D transvaginal ultrasonography using a Voluson 730 Expert or E8 ultrasound systems (GE Healthcare, Irvine, CA, USA) according to a pre-defined scanning protocol using a RIC 5- to 9-MHz endovaginal probe by a single examiner (J.L.A.) with more than 12 y of experience in 3-D ultrasound. Once the adnexal mass was identified on real-time scanning, the 3-D box was activated and superimposed over the mass. One to four volumes, depending on the size and characteristics of the mass, were acquired and stored for further evaluation. A single image was manually chosen from each of the volumes on which the adnexal mass could be clearly observed. Thus, the evaluated database consists of 145 images, each from a different patient. MATLAB software (R2014a, The MathWorks, Natick, MA, USA) was used with the gray-scale information contained in the image. The hardware platform used MSI GP60 2PE (Processor Intel Core i5-4210H CPU 2.90GHz). Histology results corresponding to the images used were available for all adnexal masses (Table 1).

Three-dimensional data acquisition scans were essential to this study, although 2-D images were used, because these ensure that the adnexal mass is accurately retained (manual selection of a single image from the whole volume). Simplifying the process to 2-D images lets us compare the results obtained with this discrimination algorithm with the results from existing research in this field, which are based mainly on 2-D images.

Feature extraction

Originally, clinical discrimination of the data in terms of image texture was attempted. Malignant images are supposed to have an assorted gray-level distribution, whereas benign images are supposed to have a more homogeneous distribution. Mathematically, a blind set of