COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 111 (2013) 84–92

ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Computer-aided diagnosis of breast masses using quantified BI-RADS findings

Woo Kyung Moon[a,1], *Chung-Ming Lo*[b,1], *Nariya Cho*[a], *Jung Min Chang*[a], *Chiun-Sheng Huang*[c], *Jeon-Hor Chen*[d,e], *Ruey-Feng Chang*[b,f,*]

[a] *Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea*
[b] *Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan*
[c] *Department of Surgery, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan*
[d] *Center for Functional Onco-Imaging, Department of Radiological Science, University of California, Irvine, CA, USA*
[e] *Department of Radiology, E-Da Hospital and I-Shou University, Kaohsiung, Taiwan*
[f] *Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan*

ARTICLE INFO

Article history:
Received 20 October 2012
Received in revised form
20 February 2013
Accepted 31 March 2013

Keywords:
Breast cancer
Ultrasound
Computer-assisted diagnosis
Breast Imaging Reporting and Data System

ABSTRACT

The information from radiologists was utilized in the proposed computer-aided diagnosis (CAD) for breast tumor classification. The ultrasound (US) database used in this study contained 166 benign and 78 malignant masses. For each mass, six quantitative feature sets were used to describe the radiologists' grading of six Breast Imaging Reporting and Data System (BI-RADS) categories including shape, orientation, margins, lesion boundary, echo pattern, and posterior acoustic features on breast US. The descriptive abilities were between 76% and 82% and the predicted descriptors were then used for tumor classification. Using receiver operating characteristic curve for evaluation, the area under curve (AUC) of the proposed CAD was slightly better than that of a conventional CAD based on the combination of all quantitative features (0.96 vs. 0.93, $p = 0.18$). The partial AUC over 90% sensitivity of the proposed CAD was significantly better than that of the conventional CAD (0.90 vs. 0.76, $p < 0.05$). In conclusion, the computer-aided analysis with qualitative information from radiologists showed a promising result for breast tumor classification.

1. Introduction

Breast ultrasound (US) is commonly used to distinguish benign from malignant masses. The Breast Imaging Reporting and Data System (BI-RADS) lexicon was developed by the American College of Radiology [1] to standardize the terminology and assessment of clinical examination. For radiologists to evaluate lesions, the dominant sonographic characteristics are described according to six BI-RADS descriptive categories: shape, orientation, margins, lesion boundary, echo pattern, and posterior acoustic features. The BI-RADS lexicon was confirmed so that radiologists can obtain good agreement when analyzing US images to classify masses [2]. To provide more efficient procedure for diagnosis, various computer-aided diagnosis (CAD) systems have been developed to quantize the tumor characteristics used by radiologists [3,4]. Based on the segmentation of tumor area, morphology features were proposed to describe tumor shape and texture features were suggested to describe tumor echogenicity. According

* *Corresponding author at*: Department of Computer Science and Information Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC. Tel.: +886 2 33661503; fax: +886 2 23628167.
E-mail address: rfchang@csie.ntu.edu.tw (R.-F. Chang).
[1] These authors are co-first authors based on equal contributions in separate (medicine and engineering) components of the study.
0169-2607/$ – see front matter
http://dx.doi.org/10.1016/j.cmpb.2013.03.017

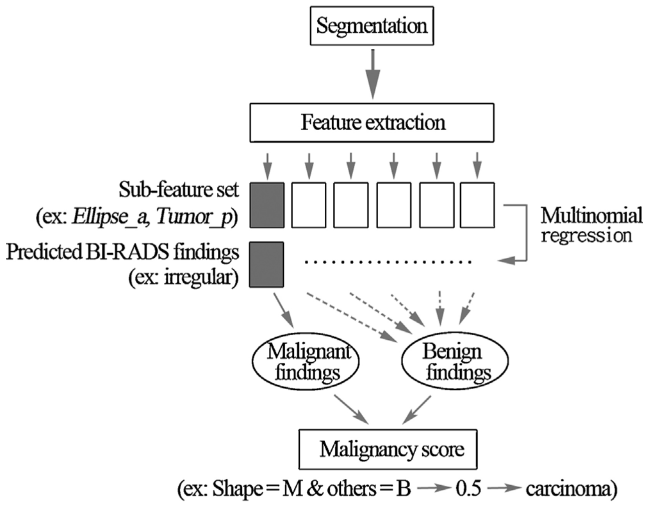

Fig. 1 – The organization of the proposed CAD system. Six quantified BI-RADS findings were used to determine the malignancy score. In our CAD algorithm, any tumor with one or more malignant findings was classified as malignant. Only the tumors that had no malignant findings and had at least one benign finding were classified as benign.

to the biopsy-proven result, the quantitative features were simply combined to classify tumors into benign or malignant in the CAD systems. However, the quantitative features were developed without radiologists' confirmations. Whether the quantitative features have successfully interpreted the malignancy of tumors is uncertain in the conventional CAD systems.

In this study, the sonographic interpretation of six BI-RADS descriptive categories from expert breast radiologists were quantified using six relevant feature sets for tumor classification. In our CAD algorithm, any tumor with one or more malignant findings was classified as malignant. Only the tumors that had no malignant findings and had at least one benign finding were classified as benign. The organization of the proposed CAD system is shown in Fig. 1. The performance of the proposed CAD system was compared to that from a conventional CAD system using only quantitative features.

2. Materials and methods

2.1. *Patients and data acquisition*

This study was approved by our institution review board, and informed consent was waived for this retrospective study. From January 2003 to July 2004, a total of 244 patients with suspicious findings on US images underwent core needle biopsy or fine-needle aspiration cytology. These US images were acquired using an ATL HDI 5000 scanner (Philips, Bothell, WA) with linear probe with a frequency of 5–12 MHz. The biopsy-proven tumors used in this study included 166 (68%) benign and 78 (32%) malignant tumors. The benign tumors included 103 cases of fibroadenoma and 63 cases of fibrocystic changes. The malignant tumors included 76 cases of invasive ductal carcinoma and 2 cases of invasive papillary carcinoma. The mean age of patients with benign tumors was 44 (range 21–65), and the patients with malignant tumors had a mean age of 49 (range 33–70).

Two breast radiologists who were blinded to the pathologic report classified the tumors of all patients into BI-RADS assessment categories according to the US findings in consensus [2]. The radiologists had 5 and 15 years of experience with breast US. There were 52 (21%) tumors in BI-RADS 3 (probably benign), 148 (61%) tumors in BI-RADS 4 (suspicious abnormality), and 44 (18%) tumors in BI-RADS 5 (highly suggestive of malignancy). The sonographic characteristics of the tumors including shape, orientation, margins, lesion boundary, echo pattern, and posterior acoustic features are summarized in Table 1.

2.2. *Tumor segmentation*

To quantify the tumor characteristics, the tumors had to be segmented first. The level set method [5] was employed as a segmentation tool to separate the tumors from background tissues. In the beginning, the image contrast was improved using a sigmoid filter [6]. Fig. 2(b) shows the result of applying the sigmoid filter in Fig. 2(a), the original sonographic image. Next, the gradient image, which presents the intensity variations in the horizontal and vertical directions, was calculated with the gradient magnitude filter [7] on the contrast-enhanced image. Fig. 2(c) shows the gradient magnitude image of Fig. 2(b). Then, the sigmoid filter was applied again to the gradient magnitude image for contrast enhancement, as shown in Fig. 2(d). After preprocessing, the level set method, proposed to model a complex shape with changing topology, was applied to the enhanced gradient image for outlining the contour of tumors. Fig. 2(e) shows the result of applying the level set method in Fig. 2(d).

2.3. *Quantitative features*

The quantitative features extracted from the segmented tumors can be classified into two groups: morphology and texture features. Morphological features were used to describe the geometric characteristics of the tumors, such as shape, orientation, and margins. To extract the morphological features, the best-fit ellipse was utilized for approximating the size and position of each tumor. For example, the tumor orientation can be computed by the angle of the major axis of the ellipse. Other studies developed quantitative features based directly on the original properties of tumors. Rangayyan et al. [8] estimated the compactness of tumors according to the perimeter and area. Nie et al. [9] proposed the use of the normalized radial length (NRL) to describe the roundness of tumors. The NRL was defined as the Euclidean distance between the tumor center and the pixels on the tumor boundary normalized by the maximum distance.

The other group of quantitative features is texture features. Texture features were used to describe the tissue composition inside the tumor [9]. Different tissues have different echogenicity patterns that result in various distributions of gray level values. Gray level co-occurrence matrices (GLCM)