Available online at www.sciencedirect.com
ScienceDirect
Nuclear Instruments and Methods in Physics Research A 580 (2007) 1071–1074
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A
www.elsevier.com/locate/nima

Can texture of tissue surrounding microcalcifications in mammography be used for breast cancer diagnosis?

A. Karahaliou, I. Boniatis, P. Sakellaropoulos, S. Skiadopoulos, G. Panayiotakis, L. Costaridou*

Department of Medical Physics, School of Medicine, University of Patras, Patras 265 00, Greece

Available online 30 June 2007

Abstract

This study investigates whether texture properties of the tissue surrounding microcalcifications (MCs) can contribute to breast cancer diagnosis. A case sample of 100 MC clusters (46 benign, 54 malignant) from 85 dense mammographic images included in the Digital Database for Screening Mammography, is analyzed. Regions of interest containing clusters are processed using wavelet-based enhancement and individual MCs are segmented by local thresholding. The segmented MCs are removed from original image data and the surrounding tissue area is subjected to texture analysis. The feasibility of four texture feature sets (first-order statistics, gray level co-occurrence matrices, gray level run length matrices and Laws' texture energy measures) in discriminating malignant from benign tissue was investigated using a *k*-nearest neighbor classifier. Laws' texture energy measures achieved the best classification accuracy 89% (sensitivity 90.74% and specificity 86.96%).

PACS: 87.57.Ra; 87.57.Nk; 87.59.Ek

Keywords: Mammography; Microcalcifications; Surrounding tissue; Texture analysis; Classification

1. Introduction

Microcalcification (MC) clusters are strong indicators of malignancy, and they appear in 30–50% of the mammographically diagnosed cases [1]. While mammography is the most sensitive imaging modality for breast cancer screening, its specificity in discriminating malignant from benign lesions is low, resulting in increased number of benign biopsies [2]. Computer-aided diagnosis is one of the most promising approaches in improving the efficacy of mammography. Automated discrimination between malignant and benign MC clusters remains a challenging task, due to high variation of individual MC size and morphology, as well as low contrast especially in case of dense breast parenchyma. A variety of computer-aided diagnosis algorithms have been proposed for MC clusters, based either on human-extracted features, or on computer-extracted image features (morphological or textural) [3,4]. The performance of the morphology-based schemes depends strongly on the robustness of MC segmentation algorithms, mostly challenged by the presence of dense breast parenchyma. Texture-based schemes overcome this limitation since no segmentation stage is required, including MCs in the regions to be further analyzed. However, this rationale introduces bias since the MC, a tiny deposit of calcium in breast tissue, can neither be malignant nor benign. This characterization corresponds to the tissue surrounding and underlying the MC cluster, which is also subjected to pathoanatomical analysis.

To the best of the authors' knowledge, there is only one study focused on texture analysis of the tissue surrounding MCs for breast cancer diagnosis [5], performed on digitally acquired views during stereotactic biopsy. The current study investigates whether texture properties of the tissue surrounding MC clusters, as depicted on screening mammograms, can be used as a computer-aided diagnosis module.

*Corresponding author. Tel.: +30 2610 969111; fax: +30 2610 996113.
E-mail address: costarid@upatras.gr (L. Costaridou).

0168-9002/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2007.06.064

2. Materials and methods

2.1. Case sample

The case sample consists of 100 biopsy proven MC clusters (46 benign, 54 malignant) from 85 dense mammographic images (heterogeneously dense and extremely dense breast parenchyma), included in the Digital Database for Screening Mammography (DDSM) [6]. Mammograms are digitized with the LUMISYS 200 scanner (12-bit pixel depth, 50 μm spatial resolution).

2.2. Preprocessing

Images are processed using a wavelet spatially adaptive enhancement method [7], based on local modification of multiscale gradient magnitude values provided by the redundant dyadic wavelet transform. Fig. 1 presents (a) original and (b) enhanced Region Of Interest (ROI) containing a MC cluster.

2.3. Removing MCs from original image data

A radiologist delineated MC cluster areas on enhanced ROIs (Fig. 1b) and selected experimental local thresholds to segment individual MCs. False positive MCs corresponding to normal dense tissue were excluded using a size criterion. Segmented MCs were removed from original ROIs, providing the surrounding tissue ROIs (ST-ROIs), shown in Fig 1c.

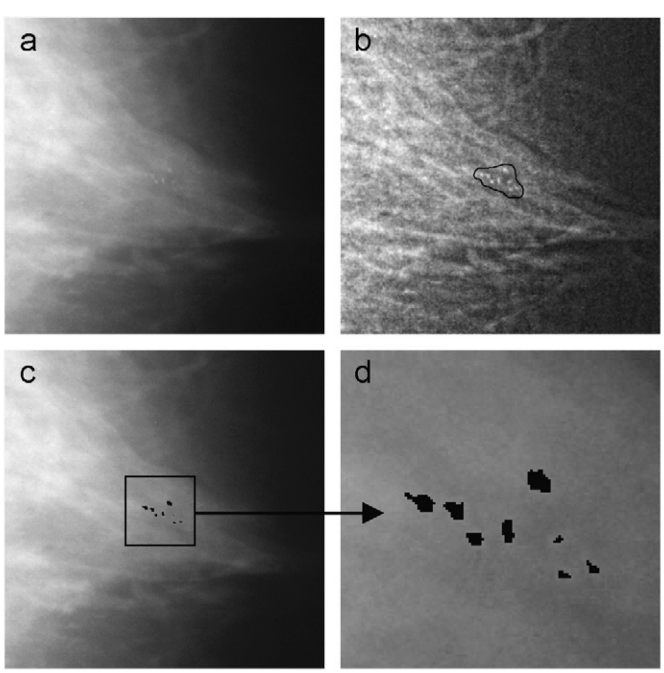

Fig. 1. (a) ROI containing a MC cluster in original mammogram, (b) enhanced ROI, with delineated MC cluster area, (c) ST-ROI, (d) magnified 128×128 pixels subregion of ST-ROI.

2.4. Texture analysis of tissue surrounding MCs

Texture analysis is performed in a 128×128 pixels subregion of each ST-ROI (Fig. 1d), positioned to contain the cluster at its center. For clusters larger than a single ROI, multiple ROIs (with minimum overlap) were used to cover the entire cluster area. Four categories of textural features were extracted: First-Order Statistics (FOS), Gray Level Co-occurrence Matrices (GLCMs) features, Gray Level Run Length Matrices (GLRLMs) features and Laws' Texture Energy Measures (LTEMs).

FOS describe texture-based on the statistical moments of the intensity histogram of an image. Four FOS features were calculated: Mean, Standard Deviation, Kurtosis and Skewness.

The GLCM characterizes the spatial distribution of gray levels in an image [8]. Thirteen textural features were calculated from four GLCMs (corresponding to distance = 1 pixel, and four angles: 0°, 45°, 90°,135°): energy, entropy, contrast, local homogeneity, correlation, shade, promenance, sum of squares, sum average, sum entropy, difference entropy, sum variance and difference variance. The mean and range of each feature over the four GLCMs were calculated, giving in total 26 GLCMs features.

The GLRLM characterizes the coarseness of image texture in specified directions [9]. Five run-length features were calculated from four GLRLMs corresponding to four angles (0°, 45°, 90°, 135°): short runs emphasis, long runs emphasis, grey level non-uniformity, run length non-uniformity and run percentage. The mean and range of each feature over the four GLRLMs were calculated, giving in total 10 GLRLMs features.

LTEMs were extracted from images that had been previously filtered by each one of the 25 Laws' masks [10]. These filtered images are defined as Texture Energy (TE) images. Averaging the TE images corresponding to symmetrical kernels, 15 TR images were produced (R stands for "Rotational invariance"). From each one of the 15 TR images, 4 first-order statistics (mean, standard deviation, skewness and kurtosis) were computed, giving in total 60 Laws' textural features.

2.5. Classification of tissue surrounding MCs

A k-nearest neighbor (kNN) classifier was used for classification of the tissue surrounding MCs, employing the inverse distance-weighted voting rule [11]. In this approach, the contribution of each one of the k neighbors is weighed according to its Euclidean distance (d_k) from the test sample, giving greater weight to closer neighbors. Specifically, the vote of the kth neighbor was defined as

$$\text{vote}(k) = 1/(d_k + 1), \quad k = 1, 2, ..9.$$

The votes of each class are summed and the test sample is assigned to the class with the highest sum of votes. For each textural feature category a best feature set was selected with respect to overall accuracy achieved, employing an