Expert Systems With Applications 42 (2015) 9499–9511

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern

Mohamed Abdel-Nasser*, Hatem A. Rashwan, Domenec Puig, Antonio Moreno

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av. Paisos Catalans 26, Tarragona 43007, Spain

ARTICLE INFO

Keywords:
Breast cancer
Breast density
Mammogram
Classification
Feature extraction

ABSTRACT

This paper proposes a computer-aided diagnosis system to analyze breast tissues in mammograms, which performs two main tasks: breast tissue classification within a region of interest (ROI; mass or normal) and breast density classification. The proposed system consists of three steps: segmentation of the ROI, feature extraction and classification. Although many feature extraction methods have been used to characterize breast tissues, the literature shows no consensus on the optimal feature set for breast tissue characterization. Specifically, mass detection on dense breast tissues is still a challenge. In the feature extraction step, we propose a simple and robust local descriptor for breast tissues in mammograms, called uniform local directional pattern (ULDP). This descriptor can discriminate between different tissues in mammograms, yielding a significant improvement in the analysis of breast cancer. Classifiers based on support vector machines show a performance comparable to the state-of-the-art methods.

1. Introduction

Breast cancer is one of the most dangerous diseases for women in their 40s worldwide (Malvezzi, Bertuccio, Levi, La Vecchia, & Negri, 2014). In 2014, breast cancer had the highest death rate among all cancer types in the European Union (DeSantis, Ma, Bryan, & Jemal, 2014). *Mammograms*, which are X-ray images of the breast, are considered the best screening tool to find breast cancer early. The common screening mammographic views are *Craniocaudal* (CC) and *Mediolateral Oblique* (MLO) (Elmore, Armstrong, Lehman, & Fletcher, 2005). The CC mammographic view is captured from the superior view of a horizontally compressed breast. In turn, the MLO view is captured from the side of a diagonally compressed breast.

The most common risk factors of breast cancer are *age, family profile, genetics* and *breast density*. Breast density represents the amount of dense tissues in the breast, and it is the strongest risk factor of breast cancer (Lokate et al., 2010). The higher is the breast density, the higher is the probability of breast cancer. In addition, there is a relation between the age of women and their breast densities, as younger women usually have denser breasts than older women. The breast masses appear brighter than the normal tissues in the mammograms, and they are defined by their *shapes* and *margins*.

In practice, mass detection is a big challenge in dense mammograms because the normal tissues also appear as bright areas, covering the places that contain masses. Some of the breast density standards classify the breast tissues into *fatty, glandular* or *dense* (Suckling et al., 1994). Fig. 1 shows examples of these tissues in mammograms. It can be noticed that it is easy to distinguish between the mass and normal regions in the *fatty* case (the red circle refers to the location of the mass region), whereas mass detection is very difficult in the case of the *dense* mammogram. The *fatty-glandular* breasts represent an intermediate situation between these two cases. Furthermore, the breast imaging reporting and data system standard (BI-RADS) (Orel, Kay, Reynolds, & Sullivan, 1999), presented by the American College of Cancer, provides the following breast density classification:

- BI-RADS I: Almost entirely fatty breast (0–25%).
- BI-RADS II: Some fibroglandular tissue (26–50%).
- BI-RADS III: Heterogeneously dense breast (51–75%).
- BI-RADS IV: Extremely dense breast (76–100%).

A breast cancer *computer-aided diagnosis* (CAD) system is a software that utilizes digitized (film-screen) or digital mammograms to assist radiologists by enhancing the quality of mammograms and detecting the early signs of breast cancer. Although radiologists try hard to estimate the breast density and to detect the masses by making a visual judgment of mammograms, they insist on requesting CAD systems to help them in this hard task.

The breast cancer CAD systems exploit various *computer vision* and *image processing* techniques. In general, a CAD system consists of three main steps: *segmentation* of the regions of interest (ROIs) from

* Corresponding author. Tel.: +34 977 55 96 77.
E-mail addresses: egnaser@gmail.com (M. Abdel-Nasser), hatem.rashwan@ieee.org (H.A. Rashwan), domenec.puig@urv.cat (D. Puig), antonio.moreno@urv.cat (A. Moreno).

http://dx.doi.org/10.1016/j.eswa.2015.07.072
0957-4174/

Table 1
Summary of breast tissue classification methods.

Method	Feature extraction method	Utilized classifiers
(Oliver et al., 2007)	Local binary pattern (LBP)	Support vector machines (SVM)
(Liu & Zeng, 2015)	Gray level co-occurrence matrix (GLCM) and completed local binary pattern	SVM
(Beura et al., 2015)	GLCM	Back propagation neural network
(Chu et al., 2015)	Shape features and texture features	SVM
(Jen & Yu, 2015)	Mean and standard deviation	Principal component analysis
(Wang, Yu, Kang, Zhao, & Qu, 2014)	GLCM and morphological features	Extreme learning machine
(de Oliveira et al., 2015)	Taxonomic diversity index and taxonomic distinctness	SVM
(Dheeba, Singh, & Selvi, 2014)	Laws texture energy measures	Particle swarm optimized wavelet neural network
(Zheng, 2010)	Gabor filters	Threshold-based approach
(Pomponiu et al., 2014)	Histogram of oriented gradients (HoG)	SVM
(Abdel-Nasser, Moreno, & Puig, 2015)	LBP, robust LBP, center symmetric LBP, fuzzy LBP, local greylevel appearance, local directional number (LDN), GLCM, HoG, Gabor filters	Linear SVM, non-linear SVM, random forest, *K*-nearest neighbor Fisher linear discriminant

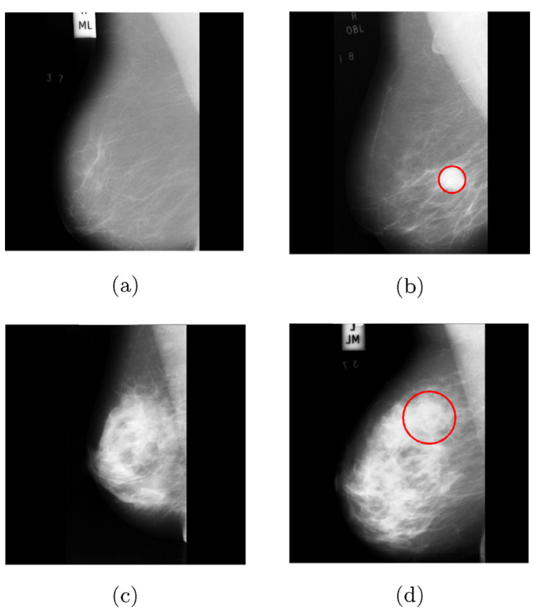

Fig. 1. Examples of mammograms from the mini-MIAS breast cancer database (Suckling et al., 1994). A *fatty* mammogram containing: (a) normal tissue and (b) mass tissue. A *dense* mammogram containing: (c) normal tissue and (d) mass tissue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the images, *feature extraction* from the ROIs and a final *classification* step. Both *breast density* classification and *breast mass* detection play an important role on the improvement of the treatment of breast cancer. Unfortunately, the performance of CAD systems decreases in the case of dense mammograms (see Fig. 1). Noise and artifacts that exist in mammograms may also degrade the performance of feature extraction methods. The common artifacts that may exist in mammograms are: detector-based, machine-based, patient-related, processing and storage artifacts. The literature shows no consensus on the optimal feature set for breast tissue characterization. A poor description of breast tissues leads to a high number of false positives (mammograms interpreted by a CAD system as abnormal cases when they are actually normal).

The work reported in this paper proposes a CAD system to analyze breast tissues. This CAD system performs two tasks: breast tissue classification within a region of interest (mass or normal) and breast density classification. The main contributions of this paper are:

- We propose the *uniform local directional pattern* (ULDP) as a texture descriptor for breast tissues in mammograms. ULDP codes a local neighborhood in the breast region based on its edge responses. As will be shown in this paper, the proposed descriptor can discriminate between different masses in mammograms regardless of their size, shape or margin, yielding a significant improvement in the analysis of breast cancer.
- ULDP is used to classify breast tissues into mass or normal, and to estimate breast density. Two publicly available mammographic databases are used: mini-MIAS (screen-film database) and INbreast (full field digital database).
- We studied the effect of breast density on the performance of ULDP when classifying breast tissues into mass or normal.

The rest of this paper is organized as follows. Section 2 provides a summary of the related work. Section 3 explains the proposed CAD system. Section 4 presents the experimental results, which are discussed in Section 5. Section 6 summarizes our work and describes some lines of future work.

2. Related work

This section summarizes some of the related works that have been proposed in mass/normal breast tissue classification and breast density classification. We discuss the feature extraction methods used in each work and highlight the superior advantages of the proposed descriptor.

2.1. Breast tissue classification

Fig. 2 shows the common shapes and margins of breast masses. The shape of a certain breast mass can be *round, oval, lobular* or *irregular*. The circumscribed oval and round shaped masses are usually *benign*, whereas *malignant* masses usually have irregular shapes. The margins of breast masses can be *circumscribed, microlobulated, obscured, indistinct* or *spiculated* (García-Manso, García-Orellana, González-Velasco, Gallardo-Caballero, & Macías-Macías, 2013).

Several features extraction methods have been proposed for breast tissue classification. Here we present several works related to breast tissue classification, and discuss the descriptors utilized in them. Table 1 summarizes some of this previous work.

Oliver, Lladó, Freixenet, and Martí have used the LBP to reduce the number of the false positives in breast mass detection. LBP may assign the same pattern to a pixel in a *tumorous region* and to another pixel in a *normal dense* tissue, which leads to a noticeable number of false detections, as illustrated in the next section. This happens when the values of all the neighbors are *higher/smaller* than the value of the central pixel. This problem of LBP is called the *saturation problem*.