



Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, LBP and SVM



Wener Borges de Sampaio^{a,*}, Aristófanés Corrêa Silva^a, Anselmo Cardoso de Paiva^a, Marcelo Gattass^b

^a Federal University of Maranhão - UFMA, Applied Computing Group - NCA Av. dos Portugueses, SN, Campus do Bacanga, Bacanga 65085-580 São Luís, MA, Brazil

^b Pontifical Catholic University of Rio de Janeiro - PUC-Rio R. São Vicente, 225, Gávea, 22453-900 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Keywords:

Breast cancer
Computer-aided detection
Micro-genetic algorithm
Phylogenetic trees
Local binary patterns
Support vector machine

ABSTRACT

Breast cancer is the second commonest type of cancer in the world, and the commonest among women, corresponding to 22% of the new cases every year. This work presents a new computational methodology, which helps the specialists in the detection of breast masses based on the breast density. The proposed methodology is divided into stages with the objective of overcoming several difficulties associated with the detection of masses. In many of these stages, we brought contributions to the areas. The first stage is intended to detect the type of density of the breast, which can be either dense or non-dense. We proposed an adaptive algorithm capable of analyzing an image and telling if it is dense or non-dense. The first stage consists in the segmentation of the regions that look like masses. We propose a novel use of the micro-genetic algorithm to create a texture proximity mask and select the regions suspect of containing lesions. The next stage is the reduction of false positives, which were generated in the previous stage. To this end, we proposed two new approaches. The first reduction of false positives used DBSCAN and a proximity ranking of the textures extracted from the ROIs. In the second reduction of false positives, the resulting regions have their textures analyzed by the combination of Phylogenetic Trees, Local Binary Patterns and Support Vector Machines (SVM). A micro-genetic algorithm was used to choose the suspect regions that generate the best training models and maximize the classification of masses and non-masses used in the SVM. The best result produced a sensitivity of 92.99%, a rate of 0.15 false positives per image and an area under the FROC curve of 0.96 in the analysis of the non-dense breasts; and a sensitivity of 83.70%, a rate of 0.19 false positives per image and an area under the FROC curve of 0.85, in the analysis of the dense breasts.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Breast cancer is the most frequent type of cancer among the female population. It is also the type of cancer which most kills women (Parkin, Bray, Ferlay, & Pisani, 2005). The early diagnosis of this disease is the main form of fighting it.

A mammogram is a radiography of the breast, which allows the early detection of cancer, since it is capable of displaying lesions at their initial stage, having sizes in the range of millimeters. It is made through an appropriate X-ray device, called mammographer. The precision of the mammogram depends on several factors, such as size and location of the lesion, the density of the breast tissue and on the quality of the technical resources used. Besides, the task of carefully

interpreting a large number of cases demands time and an elevated degree of attention from the specialist physician.

According to Norman, et al. (Boyd et al., 2007), one of the factors that hinder the detection of masses by the specialists is the type of density of the breast, which can be dense (fibrous) or non-dense (fat). The fat tissue appears as a dark region in a mammogram. On the other hand, fibrous structures (including masses) appear as clearer regions of the mammogram. Due to these characteristics, it is more difficult for a radiologist to find lesions in dense breasts.

All these factors have motivated much research over the last decades, aiming at the development of computational systems to help the specialist physician in the task of interpreting radiological images. These Computer-Aided Detection (CAD)/Diagnosis (CADx) systems have gained more and more space in modern medicine, serving as a second information source for specialists and increasing the rates of correct detections in the identification of serious diseases, such as breast cancer (Fenton et al., 2007). However, most studies seen in the literature use the same techniques and configurations

* Corresponding author. Tel.: +559881607911.

E-mail addresses: wenersampaio@gmail.com, wenersampaio@yahoo.com.br (W.B. de Sampaio), ari@dee.ufma.br (A.C. Silva), paiva@deinf.ufma.br (A.C. de Paiva), mgattass@tegraf.puc-rio.br (M. Gattass).

both for dense and non-dense masses, whereas those techniques could be more appropriate for a specific kind of density.

The efficiency of CAD systems depend on the image processing techniques. The available literature brings acknowledged works that deal with the same problem approached by the methodology proposed herein, that is, the development of computational methods for aiding the specialist in the detection of lesions in mammograms.

The present bibliography contains a wide variety and combination of techniques intended to detect masses in mammograms. They normally employ techniques for describing the geometry and the texture of suspect regions.

For the geometric analysis, descriptors are used to find a shape pattern capable of representing and differentiating between masses and non-masses. Some examples of geometric descriptors are: area, perimeter, circularity, shape factor (Dong et al., 2015), eccentricity, circular density, circular disproportion and density (Sampaio, Moraes Diniz, Corrêa Silva, Cardoso de Paiva, & Gattass, 2011) active contour (Liu, Xu, Liu, & Feng, 2011), template matching (Nunes, Silva, & de Paiva, 2009), contourlet (Moayed, Azimifar, Boostani, & Katebi, 2010), generalized moment patterns (Deepak, Medathati, & Sivaswamy, 2012), density-weighted contrast enhancement (Petrick, Chan, Sahiner, & Wei, 1996), etc.

The vast majority of studies use some kind of texture analysis. These analysis usually employs statistical and geostatistical descriptors, diversity and richness-of-species indexes, with the objective of finding relations between the distribution of pixels belonging to masses and non-masses. Some examples of texture descriptors include: Local Binary Patterns (LBP) (Berbar, Reyad, & Hussain, 2012), Complete Local Binary Patterns (Liu et al., 2011), statistical fusion (Bajger, Ma, Williams, & Bottema, 2010), Gray Level Co-occurrence Matrix (GLCM) (Abdalla, Dress, & Zaki, 2011; Anitha & Peter, 2015; Tai, Chen, & Tsai, 2014), first-order and second-order statistical functions (Berbar et al., 2012; Jen & Yu, 2015), Gabor filters, image phase analysis, angular analysis of energy propagation, fractal analysis, Laws texture and Haralick descriptors (Banik, Rangayyan, & Desautels, 2011), gray-scale invariant ranklet texture (Masotti, Lanconelli, & Campanini, 2009), optical density transformation (Tai et al., 2014), pyramid decomposition (Lin, Chang, Yeh, Liu, & Yeh, 2014). Other descriptors are obtained by the Fourier and Wavelet transforms (Agrawal, Vatsa, & Singh, 2014; Kuo, Lin, Hsu, & Cheng, 2014; Lin et al., 2014), Phylogenetic Trees (Oliveira, Carvalho Filho, Silva, de Paiva, & Gattass, 2015), Vector Field Convolution (Dong et al., 2015).

The texture and geometry descriptors are, in general, used together with some machine learning technique, which will tell if the Region of Interest (ROI) analyzed belongs to a mass or to a non-mass. Some examples include the Support Vector Machine (SVM) (Agrawal et al., 2014; Berbar et al., 2012; Dong et al., 2015; Liu et al., 2011; Oliveira et al., 2015; Sampaio et al., 2011), Linear Discriminant Analysis (LDA) (Abdalla et al., 2011; Bajger et al., 2010), Artificial Neural Networks (ANN) (Abdalla et al., 2011; Banik et al., 2011; Lin et al., 2014) and k-Nearest Neighbors (KNN) (Berbar et al., 2012), Stepwise Discriminant Analysis (Tai et al., 2014), Particle Swarm Optimization (PSO) (Kuo et al., 2014).

Great efforts were joined to create image databases for open use by the scientific community. Among these databases, we highlight the Digital Database for Screening Mammography DDSM (Abdalla et al., 2011; Bajger et al., 2010; Berbar et al., 2012; Dong et al., 2015; Jen & Yu, 2015; Liu et al., 2011; Oliveira et al., 2015; Sampaio et al., 2011; Tai et al., 2014), and the Mammographic Image Analysis Society (MIAS) (Agrawal et al., 2014; Anitha & Peter, 2015; Deepak et al., 2012; Dong et al., 2015; Kuo et al., 2014; Moayed et al., 2010). Some studies use private databases with the same purpose (Bajger et al., 2010; Banik et al., 2011; Jen & Yu, 2015; Lin et al., 2014).

Analyzing the studies numbered above, we notice that most of them use the MIAS and DDSM databases. Nevertheless, none of them tries to use a methodology adapted to the density of the breast.

The cited works, in general, use supervised learning techniques to perform the classification of the ROIs. The SVM classifier, which got much attention, was also used in this work.

This work used two stages (segmentation and reduction of false positives using DBSCAN), which use information of the from the image under analysis, dismissing the use of a specific knowledge base for each stage.

The work by Oliveira et al. (2015) also extracts features using phylogenetic trees. However, it only takes into account the frequency of the gray levels of the ROIs under analysis. In our work, the phylogenetic trees use the spatial relation among gray levels, as well as their occurrences, thus performing both a local and a global analysis. We must stress that the work by Oliveira et al. (2015) did not apply the stages of segmentation and reduction of false positives, and for this reason had a superior performance, since classification errors in these stages were not computed.

Opposite the present work, none of the cited work use a methodology in an attempt to optimize their training models.

Table 17 summarizes the related works presented in this section.

This work presents a CAD methodology for helping the specialist physician in the task of detecting masses in mammographic images. The technique is adapted to the density of the breast. The first stage of the methodology removes artifacts outside the breast and makes the reduction of noise. The next stage classifies the density of the breast as dense or non-dense. The segmentation chooses the regions of the image that probably contain masses by means of a micro-genetic algorithm (μ GA). A first reduction of false positives (RFP) uses the Density Based Spatial Clustering of Applications with Noise (DBSCAN) and a proximity ranking of the textures extracted from the regions of interest (ROI). In the second RFP, the ROIs have their textures analyzed by the combination of Phylogenetic Trees, Local Binary Patterns and Support Vector Machines (SVM).

The next sections can be briefly described as follows. Section 2 presents the bibliographic review, containing the theoretical basis necessary for understanding the methodology. In Section 3 we present the five stages that compose the present work: image acquisition, density detection, pre-processing, segmentation of the regions of interest and reduction of false positives. In Section 4, we present and discuss the results. We also present case studies of the application of the proposed methodology. Section 5 concludes the work, presenting the main contributions and the efficiency of the methods used.

2. Theoretical basis

This section brings a quick review on Phylogenetic Trees, Local Binary Patterns (LBP), taxonomic diversity and distinction indexes and Micro-genetic Algorithm.

2.1. Phylogenetic trees and taxonomic indexes

Phylogenetic trees are used in Biology to describe the evolutionary relation among species. In these trees, the leaves represent the species and the internal nodes represent the common ancestors between the species. So, it is possible to relate the species under study from an evolutionary standpoint. The inclined cladogram is a graphic representation used to describe the phylogenetic relation between ancestor species (Hall & Greenstreet, 1998). These trees allow the extraction of indexes that relate the diversity, richness and parenthood between species.

Fig. 1 presents an example of ape phylogenetic tree, represented by an inclined cladogram, where one may notice that a chimpanzee has higher phylogenetic proximity to humans than a siamang. In this tree, the leaf nodes are the species analyzed, the internal nodes correspond to a common ancestor and the edges indicate the phylogenetic distance between both species.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات