Expert Systems With Applications 55 (2016) 329–340

Contents lists available at ScienceDirect

Expert Systems With Applications

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

LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues

Daniel O. Tambasco Bruno[a,1], Marcelo Z. do Nascimento[a,b,*], Rodrigo P. Ramos[c,2], Valério R. Batista[a,1], Leandro A. Neves[d,3], Alessandro S. Martins[e,4]

[a] UFABC - CMCC, av. dos Estados 5001, Bl.B, 09210–580 St. André, SP, Brazil
[b] UFU - FACOM, av. João Neves de Ávila 2121, Bl.B, 38400–902 Uberlândia, MG, Brazil
[c] UNIVASF - CENEL, av. Antônio C. Magalhães 510, 48902-300, Juazeiro, BA, Brazil
[d] UNESP - DCCE, r. Cristóvão Colombo 2265, 15054-000, S. J. Rio Preto, SP, Brazil
[e] IFTM, r. Belarmino Vilela Junqueira S/N, 38305-200, Ituiutaba, MG, Brazil

ARTICLE INFO

Keywords:
Breast cancer tissues
Texture analysis
Local binary pattern
Curvelet transform
Computer aided diagnosis
Polynomial classifier

ABSTRACT

In computer-aided diagnosis one of the crucial steps to classify suspicious lesions is the extraction of features. Texture analysis methods have been used in the analysis and interpretation of medical images. In this work we present a method based on the association among curvelet transform, local binary patterns, feature selection by statistical analysis and distinct classification methods, in order to support the development of computer aided diagnosis system. The similar features were removed by the statistical analysis of variance (ANOVA). The understanding of the features was evaluated by applying the decision tree, random forest, support vector machine and polynomial (PL) classifiers, considering the metrics accuracy (AC) and area under the ROC curve (AUC): the rates were calculated on images of breast tissues with different physical properties (commonly observed in clinical practice). The datasets were the Digital Database for Screening Mammography, Breast Cancer Digital Repository and UCSB biosegmentation benchmark. The investigated groups were normal-abnormal and benign-malignant. The association of curvelet transform, local binary pattern and ANOVA with the PL classifier achieved higher AUC and AC values for all cases: the obtained rates were among 91% and 100%. These results are relevant, specially when we consider the difficulties of clinical practice in distinguishing the studied groups. The proposed association is useful as an automated protocol for the diagnosis of breast tissues and may contribute to the diagnosis of breast tissues (mammographic and histopathological images).

1. Introduction

Routine physical breast examination should always be complemented with further analyses through clinical equipment techniques. These are screen-film mammography, ultrasonography, magnetic resonance and tomosynthesis. Regarding accessibility combined with cost-benefit for early detection of lesions, the screen-film mammography still prevails as the best one (Hussain, 2013; Nascimento et al., 2013). As a complementary protocol, histological images of breast cancers are also evaluated to confirm a preliminary diagnosis. This procedure is commonly applied in clinical practice (Cireşan, Giusti, Gambardella, & Schmidhuber, 2013; Malhotra, Zhao, Band, & Band, 2010; Tashk, Helfroush, Danyali, & Akbarzadeh-jahromi, 2015; Veta et al., 2015; Veta, Pluim, van Diest, Viergever et al., 2014).

Precision and correctness are two requirements always expected on a diagnosis, specially because studies show that cancer cases have been increasing over the years. The 2012 report of the World Health Organization's International Agency for Research on Cancer indicates that 8.2 million obits were caused by cancer in that year. This means a rise of 8% compared to the previous 2008 report. Specifically in respect to breast cancer there has been a rise of 14%, with an amount of 522,000 women in 2012 (Ferlay et al., 2012).

* Corresponding author at: UFU - FACOM, av. João Neves de Ávila 2121, Bl.B, 38400–902 Uberlândia, MG, Brazil. Tel.: +55 34 3239 4531, +55 11 4996 0077; fax: +55 34 3239 4144.

E-mail addresses: danieltbruno@gmail.com (D.O. Tambasco Bruno), marcelo.zanchetta@gmail.com (M.Z. do Nascimento), rodrigo.ramos@univasf.edu.br (R.P. Ramos), vramos1970@gmail.com (V.R. Batista), neves.leandro@gmail.com (L.A. Neves), alessandrosm2006@yahoo.com.br (A.S. Martins).

[1] Tel.: +55-11-4996-0077.
[2] Tel.: +55-74-2102-7630.
[3] Tel.: +55-17-3221-2211.
[4] Tel.: +55-34-3271-4000.

http://dx.doi.org/10.1016/j.eswa.2016.02.019
0957-4174/

As we have mentioned, clinical equipment techniques help hinder these rates. The cutting-edge one is tomosynthesis but very few women can afford it in developing countries. As an example, in Brazil the women population is approximately 100 million, where the cases of breast cancer for 2014 were predicted to be 57,120 with a fatal rate of 23.4% (INCA, 2014).

In spite of being the most accessible technique, mammographic images are difficult to analyse. This occurs because the breast can have some adjacent tissues similar to lesions. Such false positive lesions must be ruled out by the analysis of radiologists (Thurfjell, Lernevall, & Taube, 1994; Warren & Duffy, 1995). However, nowadays there is a high demand for diagnoses, which increases both the operational costs and the workload of these professionals. This leads to more frequent human error resulting in incorrect diagnoses.

Since early 1990, computer aided diagnosis (CAD) has been an alternative tool to supplementary analysis that helps radiologists achieve more precise diagnoses. The CAD system can be divided into two parts: computer-aided detection (CADe) and computer-aided diagnosis (CADx). CADe schemes are systems that locate suspicious lesions in mammograms automatically. In Oliver et al. (2010) there is a good survey on the state of the art of CADe systems.

CADx systems are devoted to classify lesions. They complement the CADe analysis with the characterisation of regions and the computation of probabilities of lesion malignancy (Elter & Horsch, 2009). In CADx schemes, one of the crucial steps to classify suspicious lesions is the feature extraction. Typically, two classes of characteristics are extracted from mammograms, namely morphological and non-morphological features.

Morphological features give information about size and shape of the lesion. Non-morphological features work on grey level properties. For instance, a consistent variation of grey levels in the image is detected as a pattern. Patterns are then grouped to form a set called *image texture*.

Texture is a general concept that also applies to colour images. In this case patterns are identified by colour gradients along the image (Gonzalez & Richard, 2002). Texture involves basic grey levels or colour texture primitives that form elements called *textons* (Zhu, Guo, Wang, & Xu, 2005) or *texels* (Haralick, 1979). Several methods have been developed for texture analysis, being classified as statistical, model-based and frequency-based (Tang & He, 2013).

Statistical methods are based on the distribution of the grey scale. Some of them use the grey scale histogram as a measure of texture (Gupta & Markey, 2005; Souza Jacomini, Nascimento, Dantas, & Ramos, 2012). The grey scale distribution can be described by either first or second order statistics and summarised as a co-occurrence matrix (Nanni, Brahnam, Ghidoni, & Menegatti, 2014; Nanni, Brahnam, Ghidoni, Menegatti, & Barrier, 2013). The metric considers only pixels individually, but this causes measures to be more sensitive to image pixel changes. Finally, in Schwartz, Roberti de Siqueira, and Pedrini (2012) the authors use a statistical method that works on transitions between grey levels of pixels.

Model-based approaches aim at establishing stochastic properties that can define texture. Some techniques used for this purpose are Markov random fields (Suliga, Deklerck, & Nyssen, 2008; Yu & Huang, 2010), fractal features (Neves, do Nascimento, & de Godoy, 2014) and autoregressive model (Mayerhoefer et al., 2010). They can describe texture through some few parameters of microtextures, which however give macrotexture information in the end. Therefore, either little is known about the texture or there exist more than one possible texture.

Frequency-based models produce information derived from local operators and statistical attributes of images in the frequency domain. Some examples of these models are the wavelet transform (Dheeba & Tamil Selvi, 2012; Jacomini, Nascimento, Dantas, & Ramos, 2013; Nascimento et al., 2013), the ridgelet transform (Ramos, Nascimento, & Pereira, 2012) and the curvelet transform (Eltoukhy, Faye, & Belhaouari Samir, 2010a). They decompose the original image into subbands that preserve high and low frequency information. These are the most adequate techniques to extract texture, for they enable the image to be represented by multiple scales in a way that is quite close to what is done by the human eyes (Tang & He, 2013).

Several recent works have been developed for the diagnosis of suspicious regions in mammograms, mainly focused on the understanding of feature extraction from images of breast lesions. Thus, a relevant problem in CAD is the study of normal, benign and malignant lesions in images with suspicion of breast cancer. For instance, in Liu and Tang (2014) the authors extract texture features based on a grey level co-occurrence matrix (GLCM). In Rouhi, Jafari, Kasaei, and Keshavarzian (2015) the authors present approaches based on artificial neural network, cellular neural network and genetic algorithm. In Liu and Zeng (2015) the authors develop a method considering an adaptive region growing method to locate suspicious regions. Afterwards, they extract geometrical and texture features (GLCM and completed local binary pattern (CLBP)) that are then applied to classify regions of interest (ROIs). The ROIs are classified by means of support vector machine (SVM), with supervision provided by the diagnosis of a radiologist. In Abdel-Nasser, Rashwan, Puig, and Moreno (2015) the authors present a system with segmentation of the ROIs, feature extraction based on uniform local directional pattern and classifications using the SVM. In Rouhi and Jafari (2016) the authors consider texture feature computed from GLCM and CLBP, with classification obtained via SVM and supervised by radiologists. In these works the authors conclude that these approaches can describe breast tissues properly and with promising results. However, as mentioned in Abdel-Nasser et al. (2015), it is not possible to elect just one of these feature descriptors as optimal to quantify breast tissues.

Also, there are many techniques to identify patterns, such as local and mid-level features (Kong, Jiang, & Yang, 2015). For instance, local binary patterns (LBP) have been relevant to texture analysis in several applications (Guo, Zhao, & Pietikäinen, 2014; Nanni, Lumini, & Brahnam, 2010; Song, Yan, Chen, & Zhang, 2013). The LBP is an effective texture description operator with many significant advantages, for it generates histograms that are very useful to represent texture features that are invariant by rotations and brightness levels. LBP labels can be regarded as local primitives such as curved edges, spots and flat areas, among others. They have been largely applied to classify abnormalities detected in images (Choi, Kim, & Ro, 2012; Llad, Oliver, Freixenet, Mart, & Mart, 2009).

In fact, there is not a universal texture descriptor that always gives the best quantification for all different kinds of images. However, curvelet transform and LBP appear to be more effective for texture analysis. Eltoukhy, Faye, and Belhaouari Samir (2010a) used the 100 greatest coefficients from each decomposition level of the curvelet transform in order to identify lesions in mammograms. Zhang and Pham (2011) applied curvelet transform with Haralick to describe microscopy images. Zhou Li-Jian (2012) also applied curvelet transform and LBP to solve problems in face recognition when illumination is variable. In their approach the LBP technique was applied only over subband level 1. Nagaraja, Prabhakar, and Praveen Kumar (2013) proposed a technique for facial expression representation based on a combination that uses curvelet transform to obtain the several approximation subbands, of which the ones with highest energy are selected and then submitted to CLBP. Al-Hammadi, Muhammad, Hussain, and Bebis (2013) used the LBP technique over the curvelet transform coefficients and SVM classifier to describe information from chrominance components for image forgery detection. Dhahbi, Barhoumi, and Zagrouba (2015) investigated the moment theory to characterise the distribution