Expert Systems With Applications 99 (2018) 71–82

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Density-wise two stage mammogram classification using texture exploiting descriptors☆

Aditya A. Shastri, Deepti Tamrakar, Kapil Ahuja*

Computer Science and Engineering, Indian Institute of Technology Indore, Simrol, Indore, MP 453552, India

ARTICLE INFO

Article history:
Received 10 May 2017
Revised 31 December 2017
Accepted 16 January 2018

Keywords:
Mammogram
Gabor filter
Histogram of gradients
Discrete Cosine Transform
Feature selection

ABSTRACT

Breast cancer is becoming pervasive with each passing day. Hence, its early detection is a big step in saving the life of any patient. Mammography is a common tool in breast cancer diagnosis. The most important step here is classification of mammogram patches as normal–abnormal and benign–malignant.

Texture of a breast in a mammogram patch plays a significant role in these classifications. We propose a variation of Histogram of Gradients (HOG) and Gabor filter combination called Histogram of Oriented Texture (HOT) that exploits this fact. We also revisit the Pass Band - Discrete Cosine Transform (PB-DCT) descriptor that captures texture information well. All features of a mammogram patch may not be useful. Hence, we apply a feature selection technique called Discrimination Potentiality (DP). Our resulting descriptors, DP-HOT and DP-PB-DCT, are compared with the standard descriptors.

Density of a mammogram patch is important for classification, and has not been studied exhaustively. The Image Retrieval in Medical Application (IRMA) database from RWTH Aachen, Germany is a standard database that provides mammogram patches, and most researchers have tested their frameworks only on a subset of patches from this database. We apply our *two new* descriptors on *all* images of the IRMA database for *density wise classification*, and compare with the standard descriptors. We achieve higher accuracy than all of the existing standard descriptors (more than 92%).

1. Introduction

Breast cancer has become the most common killer disease in the female population. Collectively India, China and US have almost one-third burden of global breast cancer (Shah, 2016). The abnormalities like the existence of a breast mass, change in shape, the dimension of the breast, differences in the color of the breast skin, breast aches, etc., are the symptoms of breast cancer. Cancer diagnosis is performed based upon non-molecular criteria like the tissue type, pathological properties and the clinical location. Cancer begins with the uncontrolled division of one cell and results in the form of a tumor.

There are several imaging techniques for examination of the breast, such as magnetic resonance imaging, ultrasound imaging, X-ray imaging, etc. Mammography is the most effective tool for early detection of breast cancer that uses a low-dose X-ray radiation. It can reveal pronounce evidence of abnormalities, such as masses and calcification, as well as subtle signs, such as bilateral asymmetry and architectural distortion. The diagnosis of breast cancer by classifying it as benign and malignant in the early stage can reduce chances of the death of the patient.

Mammographic Computer Aided Diagnosis (CAD) systems enable evaluation of abnormalities (e.g., micro-calcification, masses, and distortions) in mammography images. CAD systems are necessary to aid facilities in carrying out a more accurate diagnosis. CAD systems are designed with either fully automatic or semi-automatic tools to assist radiologists for detection and classification of mammography abnormalities (Oliver et al., 2010). In semi-automated CAD systems, enhancement techniques are first applied on a mammogram patch, radiologists then select a Region of Interest (ROI) or a patch, and finally, the patch is classified by the system.

Mammogram patch classification is often done in one stage. However, classifying a mammogram patch in multiple stages is also beneficial. Two-stage classification of mammogram patches helps in reducing the possibility of a false positive classification. In the first stage, mammogram patches are classified as normal or abnormal (mass), then in the second stage, abnormal patches are further classified into benign or malignant. This work proposes two-

☆ This material is based upon work supported by Council of Scientific and Industrial Research (India) Grant Number 25/(0220)/13/EMR-II.

* Corresponding author.
E-mail addresses: phd1501201001@iiti.ac.in (A.A. Shastri), deeptitm@iiti.ac.in (D. Tamrakar), kahuja@iiti.ac.in (K. Ahuja).
URL: http://iiti.ac.in/people/~kahuja/ (K. Ahuja)

https://doi.org/10.1016/j.eswa.2018.01.024
0957-4174/

stage mammogram patch classification. The system is trained with normal, benign and malignant mammogram patches separately.

Generally, CAD systems consist of basic modules as follows: mammogram patch pre-processing, breast segmentation, enhancement, feature extraction and classification (Rangayyan, Ayres, & Desautels, 2007). Pre-processing step helps in removal of irrelevant regions present in a mammogram patch such as pectoral muscles and digit information. Breast region is segmented using a threshold. Enhancement techniques such as adaptive histogram equalization, non-linear filtering are applied on the breast region to improve visualization of tissues or a tumor in a mammogram patch (Anand & Gayathri, 2015; Jenifer, Parasuraman, & Kadirvelu, 2016; Sundaram, Ramar, Arumugam, & Prabin, 2011). In most works, shape features of a mammogram patch have only been considered. The shape of a mammogram patch plays an important role for benign and malignant classification. While benign masses have round or oval shapes with clear margins, malignant masses with spicule have jagged edges (Mudigonda, Rangayyan, & Desautels, 2000). Appropriate features of mammogram patches help in accurate classification.

Mammogram patches can be better classified by using their texture properties.[1] This work proposes a descriptor that captures the textural features of a mammogram patch, i.e. Histogram of Oriented Texture (HOT), which is a variation of Histogram of Gradients (HOG) and Gabor filter combination. We also apply the existing Pass Band - Discrete Cosine Transform based descriptor (PB-DCT) here because of its advantage in helping filter textural features. These descriptors have not been used yet for mammogram patch classification. We use Discrimination Potentiality (DP) to select appropriate features of mammogram patches in these two descriptors, resulting in two new descriptors. The proposed descriptors are compared with the six standard descriptors for mammogram patch classification; Zernike moments (Tahmasbi, Saki, & Shokouhi, 2011), MLPQ (Nanni, Brahnam, & Lumini, 2012), GRsca (Nanni, Brahnam, Ghidoni, Menegatti, & Barrier, 2013), Wavelet Gray Level Co-occurrence Matrix (WGLCM) (Beura, Majhi, & Dash, 2015), Local Configure Pattern (LCP) and Histogram of Gradients (HOG) (Ergin & Kilinc, 2015). SVM is the most suitable classifier for two-class classification and is widely used in this field. Hence, we use this.

Breasts with high density have a higher chance of cancer. However, high dense tissues and masses appear as mostly white in a gray scale of a mammogram patch. Hence, it is very difficult to detect a tumor in high dense tissues. Especially, the difference between benign and malignant tumors is hard to determine (De Oliveira, De Albuquerque Araújo, & Deserno, 2011; Oliver et al., 2010; 2008; 2012; Petroudi & Brady, 2011). Generally, breasts are classified based upon density in three different ways by the Breast Imaging Reporting And Database Systems (BIRADS); two classes (fatty and dense), three classes (fatty, glandular, and dense) or four classes (mostly fatty, scattered density, consistent density and extremely dense) (De Oliveira et al., 2011; Oliver et al., 2008). Most researchers in this area have not considered the density of a breast for mammogram patch classification (normal–abnormal and benign–malignant). Hence, in this work, using a two-stage mammogram patch classification system, we test our two proposed descriptors for each BIRADS class separately.

CAD systems are usually tested on the MIAS and DDSM mammogram patch datasets of the IRMA database (Deserno, 2012). The MIAS dataset consists of a small set of images, while DDSM includes few thousand images. Several descriptors and methodologies have been proposed for mammogram patch classification, but their performances have been investigated only for a small set of images. Moreover, these systems have not achieved desired accuracy (De Oliveira et al., 2011; Petroudi & Brady, 2011). The performance of our system is tested on all mammogram patches of the MIAS and DDSM datasets. The experimental results show the effectiveness of our approach; we achieve near to 92% accuracy.

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 experimental results. Finally, Section 5 gives conclusion and discusses future work.

2. Literature review

Researchers have reviewed existing techniques for detection and analysis of abnormalities in mammogram patches as discussed earlier (calcification, masses, tumors, bilateral asymmetry, and architectural distortion) (Rangayyan et al., 2007). Some people have also reviewed the contribution of texture to risk assessment for each density separately (Oliver et al., 2010). Mammogram patches consist of directionally oriented, texture image due to its fibroglandular tissues, ligaments, blood vessels and ducts. These texture features for mammogram patches can be categorized into four groups; statistical (Beura et al., 2015; Oliver et al., 2012; Peng, Mayorga, & Hussein, 2016; Rabottino, Mencattini, Salmeri, Caselli, & Lojacono, 2008; Shanthi & Murali Bhaskaran, 2012), local pattern histogram (Abdel-Nasser, Rashwan, Puig, & Moreno, 2015; De Oliveira et al., 2011; Ergin & Kilinc, 2015; Petroudi & Brady, 2011), directional (Buciu & Gacsadi, 2011; Gedik, 2016; Leena Jasmine, Govardhan, & Baskaran, 2009; Mudigonda et al., 2000; 2001), and transform based (Dabbaghchian, Ghaemmaghami, & Aghagolzadeh, 2010; Laadjel, Al-Maadeed, & Bouridane, 2015).

Statistical features such as mean, variance, energy, entropy, skewness, and kurtosis are mostly utilized as a descriptor for classification (Beura et al., 2015; Oliver et al., 2012; Peng et al., 2016; Rabottino et al., 2008; Shanthi & Murali Bhaskaran, 2012). Gray Level Co-occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRLM) provide the relationship between neighboring pixels of a mammogram patch. Statistical properties of these matrices have also been exploited for mammogram patch classification (Beura et al., 2015). These features are extracted by directly using spatial data from images.

Some works have exploited local distribution of textural properties of mammogram patches for classification. Histogram of Gradients (HOG) (Ergin & Kilinc, 2015), Local Configure Pattern (LCP) (Ergin & Kilinc, 2015), Uniform Directional Pattern (UDP) (Abdel-Nasser et al., 2015), Local Ternary Pattern (LTP) (Muramatsu, Hara, Endo, & Fujita, 2016), Local Phase Quantization (LPQ) (Ojansivu & Heikkilä, 2008) and Local Binary Pattern (LBP) (Oliver, Lladó, Freixenet, & Martí, 2007) are some such examples. There are three different variants of LBP, which are usually used for exploiting local textural properties; Uniform Local Binary Pattern (LBP-u), Rotation Invariant Local Binary Pattern (LBP-ri) and Rotation Invariant Uniform Local Binary Pattern (LBP-riu) (Nanni et al., 2012; Ojala, Pietikainen, & Maenpaa, 2002). Block-wise feature extraction gives better performance as compared to global feature vectors. Statistical properties of local histogram have also been used as a mammogram patch descriptor for classification (Wajid & Hussain, 2015).

Coming to directional features, wavelet, dual-tree complex wavelet Gabor, Contourlet, finite Shearlet, etc. have been exploited for multi-resolution and multi-orientation texture or tissue analysis of a mammogram patch (Gedik, 2016; Leena Jasmine et al., 2009; Mudigonda et al., 2000; 2001). Gabor based feature extraction schemes are widely used for mass classification as benign–malignant. Gabor features can be extracted from mammogram patches in different ways (Buciu & Gacsadi, 2011; Khan et al., 2016). Recently, some people have proposed directional features of

[1] It is relatively hard to classify a mammogram patch as benign–malignant (as compared to normal–abnormal) due to their lack of distinctive properties (Mudigonda, Rangayyan, & Desautels, 2001).