Applied Soft Computing xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

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

Localization of tumor and its stage using intelligent techniques

G. Wiselin Jiji*, Jini R. Marsilin

Department of Computer Science and Engineering, Dr. Sivanthi Aditanar College of Engineering, Tiruchendur 628215, India

ARTICLE INFO

Article history:
Received 30 June 2012
Received in revised form
15 September 2012
Accepted 12 October 2014
Available online xxx

Keywords:
Back propagation neural network
Breast cancer staging
Early detection
Euclidean distance
Gabor feature
Mahalanobis distance
Pattern

ABSTRACT

Breast cancer is the most common disease, which is the leading cause of cancer deaths among women. This deadly disease is curable, if identified at the initial stage. In this paper, we propose a scheme for predicting the real stage of breast cancer by retrieving the mammogram images from the past cases that are similar to the query image. First, Gabor and energy features are extracted to differentiate abnormal tissue textures from normal tissue texture. In the second step, back propagation neural network (BPNN) algorithm is used to detect the tumor. Next, pattern is extracted from the query image and the real stage of breast cancer is identified using the depth of the tumor. In this paper, Euclidean distance metric and Mahalanobis distance metric are used to compute the pattern similarity between the images for retrieval. In the same way, pattern base images also retrieved. The tumor found in these retrieved images shows the same stage of breast cancer related to the query image. Performance of Euclidean and Mahalanobis distance metric is compared using precision recall measures. The proposed approach achieved 87% classification rate.

1. Introduction

Breast cancer is a malignant tumor that forms in the tissues of breast. It is a group of cancer cells with an uncontrolled growth of breast cells and has ability to grow into surrounding tissues or spread to other parts of the body. The primary risk factors of breast cancer are age, later age at child birth, lack of childbearing or breastfeeding, fear of self examination, fear of chemotherapy, higher hormone levels and also the economic status. These factors increase a person's chance of getting a breast cancer.

National Cancer Institute (NCI) surveyed that one in eight women in the United States (approximately 12.5%) were affected by this cancer during their lifetime [4]. The survey reported that women of 40–60 years have 16% breast-related problems. Hence women may have the risk of breast cancer in above 40 years in their breast lumps. Another survey on the year 2004 alone estimated that 5,19,000 deaths happened because of breast cancer worldwide [2]. Another one survey reported in 2008 that 4,58,503 deaths worldwide [3]. It enforces the need for improved screening techniques and increases the awareness of women about the potential risk of breast cancer for early detection. However, death rates have remained relatively steady over the past 20 years because of early detection and proper checkup with better treatments [1].

The stage of the breast cancer is categorized based on the tumor size and whether the cancer has spread to other parts of the body beyond the breast. Stage 0 describes noninvasive breast cancers and there is no evidence of cancer cells breaking out of the part of the breast, in which they started. Stage I describes invasive breast cancer of tumor size up to 2 cm. Stage II is also the invasive breast cancer with tumor size of 2–5 cm. Stage III is the advanced stage (i.e.) the cancer is any size and has spread to the chest wall and/or skin of the breast. Stage IV cancer has spread to other parts of the body, most often the bones, lungs, liver, or brain. Stages I, II and III are the "early-stage" breast cancer (Table 1).

Mammography is the effective method to detect breast cancer at initial stage [5–9]. Early detection using mammography screening aims to increase treatment options and decrease death rate. Women having higher risk of breast cancer must go through mammography screening for early detection. The present breast cancer detection techniques and methods concerned only to determine the stage of the breast cancer and help the doctor to make decision about the proper treatment. In this paper, the real stage of breast cancer is predicted and similar images that are relevant to the query image are retrieved using similarity measure.

Many techniques are available for detecting tumor from mammogram images. Early detection technique using SVM classifier [10] segments the tumor from mammogram images. Features extracted from mammogram images play an important role to distinguish the abnormal tissue. Shape based features in [11] uses the extracted shape content for retrieving mammographic masses.

* Corresponding author. Tel.: +91 9443087064; fax: +91 4639242482.
E-mail address: jijivevin@yahoo.co.in (G.W. Jiji).

http://dx.doi.org/10.1016/j.asoc.2014.10.012
1568-4946/

Table 1
Detection results of the tumor in Stage III.

	Patient case		True positive	False positive	True negative	False negative	Sensitivity	Specificity	Positive prediction value
	N_p	N_n							
Proposed work	50	50	48	7	43	2	96	86	87.27
Work [28]	50	50	43	7	43	7	86	86	86

Gabor features used in [12] provide the best pattern retrieval accuracy. Similarly, CBIR approach used in [13] is effectively applied for radiographic images. The approach used in [14] offers high retrieval effectiveness.

The major objective of this project is to identify the correct stage of breast cancer by retrieving images and get the proper treatment in appropriate time to reduce the high death rate. Early stage detection is the efficient way to cure the breast cancer.

2. Methods

We propose a technique to retrieve the mammogram images that gives the real stage of breast cancer using pattern similarity. This is shown in Fig. 1.

2.1. Feature extraction

An image feature is a distinguishing primitive characteristic or attribute of an image. Feature extraction uses the function of raw pixel values for determining relevant content from the image. The efficiency and accuracy of any image retrieval system depends on the feature extraction step. Each of the images is scanned with the window size of 3×3 and obtained feature matrix. Most of the image analysis applications use the texture features efficiently. Texture is the regular repetition of pattern on a surface. To differentiate abnormal breast tissue from normal tissue, texture features played an important role and proved to be efficient [15]. Gabor features [16,17] are used with many applications like texture classification and image recognition. This paper focuses on energy and Gabor features.

2.1.1. Energy

Energy is the second-order feature which is extracted using gray level co-occurrence matrix (GLCM) [18] feature. It is the sum of squared elements in GLCM. It is also known as angular second moment or uniformity. The GLCM feature energy is calculated by,

$$\text{Energy} = \sum_i \sum_j C_d(i,j)^2 \tag{1}$$

Here each entry (i,j) in GLCM refers to the number of occurrences of the pair of gray levels i and j which are a distance d apart in original image.

2.1.2. Gabor feature

The Gabor filter (Gabor wavelet) [19,20] represents a band-pass linear filter. It was first introduced by David Gabor in 1946 [21]. A Gabor filter is obtained by modulating a sinusoidal wave of particular frequency and orientation with a Gaussian. It achieves an optimal resolution in both spatial and frequency domains. A two dimensional Gabor function $g(x, y)$ is given as,

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \tag{2}$$

Here σ_x, σ_y are the standard deviations of the Gaussian envelope along the x- and y-dimensions. Its Fourier transform $G(u, v)$ can be written as,

$$G(u,v) = \exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\} \tag{3}$$

where $\sigma_u = 1/2\pi\sigma_x$, $\sigma_v = 1/2\pi\sigma_y$. Gabor functions form a complete but nonorthogonal basis set. Let $g(x, y)$ be the mother Gabor wavelet, then this self-similar filter dictionary can be obtained by appropriate dilations and rotations of $g(x, y)$ through the generating function

$$g_{mn}(x,y) = a^{-m}G(x', y'), \quad a > 1 \quad m, n = \text{interger}$$

$$x' = a^{-m}(x\cos\theta + y\sin\theta), \quad \text{and} \quad y' = a^{-m}(-x\cos\theta + y\sin\theta) \tag{4}$$

where $\theta = n\pi/K$ and K is the total number of orientations. The scale factor a^{-m} ensures energy is independent of m. The redundant information of the filtered images is reduced by the following strategy. Let U_l and U_h denote the lower and upper center frequencies of interest. Let K be the number of orientations. S is the number of scaled in the multiresolution decomposition. The filter parameters σ_u and σ_v are computed as,

$$a = (U_h/U_l)^{1/S-1}, \quad \sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\ \ln 2}},$$
$$\sigma_v = \tan\left(\frac{\pi}{2k}\right)\left[U_h - 2\ \ln\left(\frac{\sigma_u^2}{U_h}\right)\right], \tag{5}$$
$$\left[2\ \ln 2 - \frac{(2\ \ln 2)^2\sigma_u^2}{{U_h}^2}\right]^{-\frac{1}{2}}$$

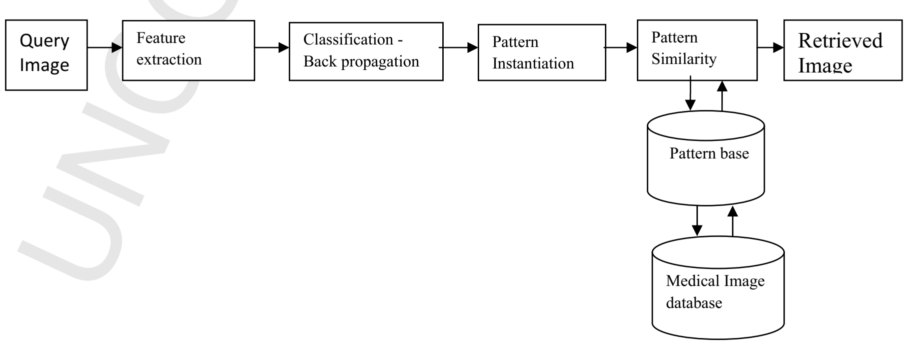

Fig. 1. Block diagram of proposed image retrieval system.

Please cite this article in press as: G.W. Jiji, J.R. Marsilin, Localization of tumor and its stage using intelligent techniques, Appl. Soft Comput. J. (2014), http://dx.doi.org/10.1016/j.asoc.2014.10.012