Journal of Biomedical Informatics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach

J. Dheeba [a,*], N. Albert Singh [b], S. Tamil Selvi [c]

[a] Dept. of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, Kanyakumari District, Thuckalay, Tamil Nadu 629 180, India
[b] BSNL Nagercoil, India
[c] Department of Electronics and Communication Engineering, National Engineering College, Kovilpatti, India

ARTICLE INFO

Article history:
Received 26 June 2013
Accepted 17 January 2014
Available online xxxx

Keywords:
Mammograms
Breast cancer
Computer aided diagnosis
Receiver operating characteristic
Laws
Particle swarm optimization

ABSTRACT

Breast cancer is the second leading cause of cancer death in women. Accurate early detection can effectively reduce the mortality rate caused by breast cancer. Masses and microcalcification clusters are an important early signs of breast cancer. However, it is often difficult to distinguish abnormalities from normal breast tissues because of their subtle appearance and ambiguous margins. Computer aided diagnosis (CAD) helps the radiologist in detecting the abnormalities in an efficient way. This paper investigates a new classification approach for detection of breast abnormalities in digital mammograms using Particle Swarm Optimized Wavelet Neural Network (PSOWNN). The proposed abnormality detection algorithm is based on extracting Laws Texture Energy Measures from the mammograms and classifying the suspicious regions by applying a pattern classifier. The method is applied to real clinical database of 216 mammograms collected from mammogram screening centers. The detection performance of the CAD system is analyzed using Receiver Operating Characteristic (ROC) curve. This curve indicates the trade-offs between sensitivity and specificity that is available from a diagnostic system, and thus describes the inherent discrimination capacity of the proposed system. The result shows that the area under the ROC curve of the proposed algorithm is 0.96853 with a sensitivity 94.167% of and specificity of 92.105%.

1. Introduction

Breast cancer is the most frequently diagnosed cancer in women worldwide and the leading cause of cancer death among females. Breast cancer accounts for 23% of the total cancer cases and 14% of the cancer death in both developed and developing countries. It is estimated that more than 1.6 million new cases of breast cancer occurred among women worldwide in 2010 [1,2]. In 2011, nearly 1.7 million people were told to have breast cancer; statistics says that in USA 527 new cases of breast cancer were diagnosed per day and 110 people die of it per day. Early diagnosis remains important for survival, particularly in low and middle income countries where the diseases is diagnosed in late stages and resources are very limited. One proven way of reducing mortality from breast cancer is the screening of asymptotic women by mammography.

Mammography is the best screening tool that uses low dose X-rays to create an image of the breast to find breast cancer. Mammography has been proved to be effective in screening asymptomatic women to reduce mortality by as much as 30%. The American Cancer Society recommends that all women aged above 40 undergo screening mammography once in a year. Dense breast tissue can look white or light gray on a mammogram. This can make mammograms harder to interpret in younger women, who tend to have denser breasts. Many breast conditions mimic the symptoms of cancer and need tests and sometimes a biopsy for diagnosis. False positive results occur when mammogram finds something that looks like cancer, but turns out to be benign (not cancer). Depending on the density of the breasts radiologists may miss up to 30% of breast cancers [3]. Even qualified radiologists find it difficult to interpret screening mammograms in large numbers.

Two powerful indicators of cancer that are commonly used in evaluating mammograms are known as masses and microcalcifications. It is generally accepted that mass detection is a more challenging problem than the detection of micro-calcifications, not only for the large variation in size and shape in which masses can appear in a mammogram but also because masses often exhibit poor image contrast [4]. A true abnormality can usually be distinguished based on careful analysis on two different view of a

* Corresponding author.
E-mail addresses: deeps_3u4@yahoo.com (J. Dheeba), mailalbertsingh@yahoo.co.in (N. Albert Singh), tamilgopal2004@yahoo.co.in (S. Tamil Selvi).

http://dx.doi.org/10.1016/j.jbi.2014.01.010
1532-0464/© 2014 Published by Elsevier Inc.

mammogram namely Carniocaudal (CC) view and MedioLateral Obligue (MLO) view [5]. But, additional methods or imaging is need if the suspected region is not viewed on the complementary view of the mammogram or if there is a dense tissue present that could obscure an underlying abnormality.

Computer-aided detection and diagnosis (CAD) can be used on the digital images to help the radiologists analyse the overall images, and highlights potential areas of concern that needs closer study. CAD can find tumours that a radiologist might not spot. Once a CAD analysis has been done, a radiologist will do a visual check of those areas, and based on training and experience, decide how serious the lumps may actually be [6]. CAD will assist the radiologists by serving them as a "second reader". The proposed CAD system will automatically identify the areas of abnormal contrast, calling the radiologist's attention to suspicious regions. Combining mammography with CAD will improve the ability to find cancer [7]. In many cases, the microcalcifications and the cancer masses are hidden in the intense breast tissues especially in younger women, making both the diagnosis and detection more complex and intricate [8]. While mammography has been proven to be a powerful tool in the fight against breast cancer, the accurate reading of mammograms can sometimes be difficult [9]. Even the most trained radiologist can miss subtle variations in tissue that might be of concern.

A lot of researches in the area of CAD systems for breast cancer and developing intelligent techniques for improving classification accuracy have been conducted in last few decades [10–12]. Different studies have demonstrated that Computer Aided Detection (CAD) of breast cancer can improve the detection rate from 4.7% to 19.5% compared to radiologists. Regarding classification of abnormalities in mammogram, a number of techniques have been presented using machine learning approaches to classify samples as normal and abnormal.

Karahaliou et al. [13] investigated multi-scale texture properties of the tissue surrounding microcalcifications (MCs) for breast cancer diagnosis using probabilistic neural network. Kupinski and Giger [14] presented a radial gradient index based algorithm and a probabilistic algorithm for detecting lesions in digital mammograms. Sahiner et al. [15] used a Convolution Neural Network (CNN) classifier to classify the masses and the normal breast tissue. Eltonsy et al. [16] presented a method based on the presence of concentric layers surrounding a focal area with suspicious morphological characteristics and low relative incidence in the breast region. Zheng et al. [17] presented a mixed feature based neural network for detection of microcalcification clusters in digital mammograms. Features are computed in both the spatial and spectral domain and uses spectral entropy as a decision parameter. Backpropagation with Kalman filtering (KF) is employed to allow more computationally efficient training as required for evaluation of different features and input images.

Among existing CAD techniques, the main problem of developing an acceptable CAD system is inconsistent and low classification accuracy. In order to improve the training process and accuracy, in this paper a novel intelligent classifiers that use texture information as input to classify the normal and abnormal tissues in mammograms is investigated. Moreover, the intelligent machine learning classifiers are optimized using heuristic algorithms for finding appropriate hidden neurons, learning rate and momentum constant during the training process.

This paper concentrates on developing a CAD system as an artificial second radiologist. Texture helps to understand image content based on textural properties in images. Texture is the most important visual cue in identifying different types of homogeneous regions and gives information about the surface property, depth and orientation [18–22]. This texture information helps to extract specific characteristics from a data. Mammographic images possess textural information that could bear discriminant features. The Laws texture features were extracted from the mammogram to differentiate between abnormal and normal pixels. New artificial intelligent techniques such as neural network have been used in medical applications for discriminating the normal and abnormal tissues in mammograms. The thriving of artificial intelligence which utilizes the human experience in a more relaxed form than the conventional mathematical approach has recently attracted more attention.

Designing optimal neural network architecture is made by a human expert and it requires a tedious trial and error process. Especially automatic determination of artificial neural network parameters is the most critical task. This paper focuses mainly on designing a CAD system based on the optimized wavelet neural network evaluated using Particle Swarm Optimization approach (PSOWNN). Optimization of WNN is carried out to improve the classification accuracy in breast cancer detection thereby reducing the misclassification rate.

2. Materials

The proposed CAD system is based on a pattern recognition system which intelligently identifies the abnormal regions. CAD schemes using digital image processing techniques have the goal of improving the detection performance. Typically CAD systems are designed to provide a "second opinion" to aid rather than replacing the radiologist. Fig. 1 shows the proposed approach for detection of abnormality in mammograms.

2.1. Mammogram database

Clinical mammogram database consisting of 216 images of 54 patients were taken from mammogram screening centers. The real time database includes a wide spectrum of cases that are difficult to classify by radiologists. All clinical mammograms that were collected from screening clinics were positive for presence of abnormalities. Mammograms were collected from 54 patients and all these patients have agreed to have their mammograms to be used in research studies. For each patient 4 mammograms were taken in two different views, one is the Craniocaudal (CC) and the other is the Mediolateral Oblique (MLO) view. The two projections of each breast (right and left) were taken for every case. The suspicious regions were identified by the automated system based on various machine learning algorithms and was reviewed by experienced radiologists. For this study a total of 216 mammograms were taken, all the mammograms were digitized to a resolution of 290×290 Dots per Inch (DPI) which produces 24 bits/pixel. Each digitized mammograms was incorporated into a 2020×2708 pixel image (5.47 Mpixels). Screening mammography is taken on asymptotic women to detect clinically occult cancers. Table 1 shows the summary of the different types of abnormalities in the database. The mammograms obtained are from women with an age group of 20–69 years old.

2.2. Mammogram image preprocessing

The goal of pre-processing the image is to simplify recognition of cancers (abnormalities) without throwing away any important information. Mammograms has breast region and is superimposed over background structures to which analysis is not necessary. One way would be to restrict the analysis to Region of Interest (ROI) that does not contain any background. The initial preprocessing is done in the digital mammogram to separate the region of interest (breast) and the dark background. The separation of ROI from the dark background is done using a global thresholding technique.