Neurocomputing 154 (2015) 1–14

Contents lists available at ScienceDirect

Neurocomputing

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

Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer

Shradhananda Beura*, Banshidhar Majhi, Ratnakar Dash

Pattern Recognition Research Laboratory, Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

ARTICLE INFO

Article history:
Received 22 June 2014
Received in revised form
23 September 2014
Accepted 8 December 2014
Communicated by Huiyu Zhou
Available online 24 December 2014

Keywords:
Mammogram
ROI
GLCM
DWT
Null hypothesis
Confusion matrix

ABSTRACT

In this paper, we propose a mammogram classification scheme to classify the breast tissues as normal, benign or malignant. Feature matrix is generated using GLCM to all the detailed coefficients from 2D-DWT of the region of interest (ROI) of a mammogram. To derive the relevant features from the feature matrix, we take the help of *t-test* and *F-test* separately. The relevant features are used in a BPNN classifier for classification. Two standard databases MIAS and DDSM are used for the validation of the proposed scheme. It is observed that *t-test* based relevant features outperforms to that of *F-test* with respect to accuracy. In addition to the suggested scheme, the competent schemes are also simulated for comparative analysis. It is observed that the proposed scheme has a better say with respect to accuracy and area under curve (*AUC*) of receiver operating characteristic (ROC). The accuracy measures are computed with respect to normal vs. abnormal and benign vs. malignant. For MIAS database these accuracy measures are 98.0% and 94.2% respectively, whereas for DDSM database they are 98.8% and 97.4%.

1. Introduction

Breast cancer is still the most common cancer throughout the world and a frequent cause of cancer death among women. According to Globocan project, it has been estimated for the year 2012 that 1.67 million new cancer cases were diagnosed worldwide, which is 25% of all types of cancers. In India, the breast cancer is considered as the most common cancer and in the year 2012, 144,937 women were newly detected with this cancer and 70,218 patients died among them. So, it can be noticed that one patient is dying out of every two newly diagnosed women [1]. It has been studied that the recovery of the breast cancer as well as survival rate can be improved by the early detection through periodic screening. Regarding this context, mammography is the most effective and reliable method for an accurate detection of breast cancer in recent years [2]. Mammograms are X-ray images of breasts. Reading of mammograms is a very important task for radiologists as they suggest patients for biopsy. However, human interpretation of mammograms varies as it depends on training and experience. This leads to different judgments by different radiologists. Mammogram interpretation is a repetitive task which requires maximum attention for avoidance of misinterpretation. Therefore, computer-aided diagnosis (CAD) system is currently a very popular and efficient method which analyzes the digital mammograms with the use of image processing. CAD system helps radiologists in accurate interpretation of mammograms for detection of suspicious lesions and classification. It has been observed that 60–90% of the biopsies of cancers predicted by radiologists found benign later [3]. So, it is very important to develop a CAD system, which can distinguish normal–abnormal as well as benign–malignant mammograms. The main objective of CAD system is to increase diagnosis accuracy and enhancing the mammogram interpretation. Thus, CAD system can reduce the variability in judgments among radiologists by providing an accurate diagnosis of digital mammograms. Regarding this responsibility, one important step is to find out a set of significant features from the mammographic images that can distinguish the normal mammograms from abnormal as well as the benign lesions from malignant ones. Different techniques and methods have been studied for this purpose.

For mammogram feature extraction and classification, several researches have been carried out over the year. One of the effective methods is the multiresolution analysis in which the original mammographic image is decomposed into several sub-images that preserve informations about both high and low frequencies. Wavelet transform is one of the important methods for the texture analysis of the image. Many researchers worked on multiresolution analysis of mammograms based on wavelets by using different types of feature spaces. Dhawan et al. used wavelet decomposition and gray level image structure features for classification of mammograms and obtained an area under curve (*AUC*) of 0.81 in a receiver operating

* Corresponding author.
E-mail addresses: shradhananda.beura@gmail.com, beura.shradhananda@gmail.com (S. Beura).

http://dx.doi.org/10.1016/j.neucom.2014.12.032
0925-2312/

characteristic (ROC) curve [4]. Wei et al. achieved *AUC* of 0.96 through ROC analysis in the classification of abnormal–normal mammographic tissue classification by using multiresolution texture features [5]. In their method, wavelet transform has been used to decompose the mammographic region of interest (ROI) to collect different detail coefficients and consequently, texture features were extracted from these coefficients. Liu et al. used a set of statistical features based on wavelets and found 84.2% accuracy rate by using binary tree as classifier in mammogram classification [6]. Rashed et al. obtained 87.06% of classification accuracy by using different types of Daubechies wavelets in the classification of mammograms [7]. Prathibha et al. used multiscale wavelet transformation for extraction of texture features from the mammographic images. They obtained the classification performance as *AUC* of 0.95 in ROC to classify normal and abnormal mammograms by using the nearest neighbor classifier [8]. Buciu et al. achieved *AUC* values as 0.79 and 0.78 for classification of normal–abnormal and benign–malignant mammogram classes respectively [9]. They have used Gabor wavelets with principal component analysis for reduction in dimension of directional features with the help of support vector machine as classifier. Görgel et al. used wavelet based support vector machine (SVM) in their proposed method for mammographic mass classification and achieved an accuracy of 84.8% [10]. In another proposed method, Görgel et al. found 96.0% and 93.59% classification accuracy rates for normal–abnormal and benign–malignant mammogram classification using spherical wavelet transform (SWT) for extraction of features and SVM as the classifier [11]. In their proposed method, a local seed region growing algorithm has been used to detect ROIs of mammograms.

Texture of a mammographic image is the quantitative statistical measurements of pixel intensities in a region. The textural information of mammographic images is very important for distinguishing the abnormal pattern from the normal. A popular method for texture analysis can be taken into notice is the gray-level co-occurrence matrix (GLCM) which estimates the second order-statistical properties of images [12–14]. Chan et al. achieved performance index value of $AUC = 0.89$ by using texture morphology features based on GLCM in the classification of mammograms [15]. In their proposed method, a feature selection technique based on genetic algorithm (GA) has been used to select effective features from multidimensional feature spaces. Mutaz et al. developed a method in which the textural features were extracted from ROI using GLCM [16]. Utilizing these features, they discriminated the benign and malignant mammograms with the help of neural network and achieved an accuracy of 91.67% sensitivity. Jona et al. used GLCM to extract the features from the mammographic images [17]. They optimized the feature set by using a hybrid particle swarm optimization and genetic algorithm, and obtained 94% of classification accuracy by using SVM to classify the normal and abnormal mammograms.

The literature survey reveals about the existing classification schemes for digital mammogram images. However, most of them are not able to provide a good accuracy. It has been seen that the dimension of extracted feature space is so high due to large varieties of normal and abnormal tissues present in the breast. The use of high dimensional feature space may degrade the performance of the classification scheme. From a large feature space, only some of the features are effective and significant for the mammogram classification. Therefore, in addition to feature extraction, feature selection is also the key step in mammogram classification, which selects only the significant features from available feature space. So there is a need to develop some new feature extraction and selection algorithms to increase the accuracy of classification rate. In this paper, we have proposed an effective feature extraction algorithm using two dimensional discrete wavelet transform (2D-DWT) based multiresolution analysis along with gray-level co-occurrence matrix (GLCM) to compute texture features for mammographic images. A feature selection algorithm has been applied using two statistical feature selection methods such as *two-sample t* and *F-test* to select significant features from extracted features. Utilizing these significant features, a back propagation neural network (BPNN) has been used as classifier to predict the mammogram, whether it is a normal or abnormal. In addition, the severity with respect to malignant or benign is also estimated in abnormal cases. The rest of this paper is organized as follows: Section 2 deals with the proposed scheme, where extraction and selection of features and classification is discussed in detail. Section 3 describes the results obtained on standard databases. Section 4 gives the concluding remarks.

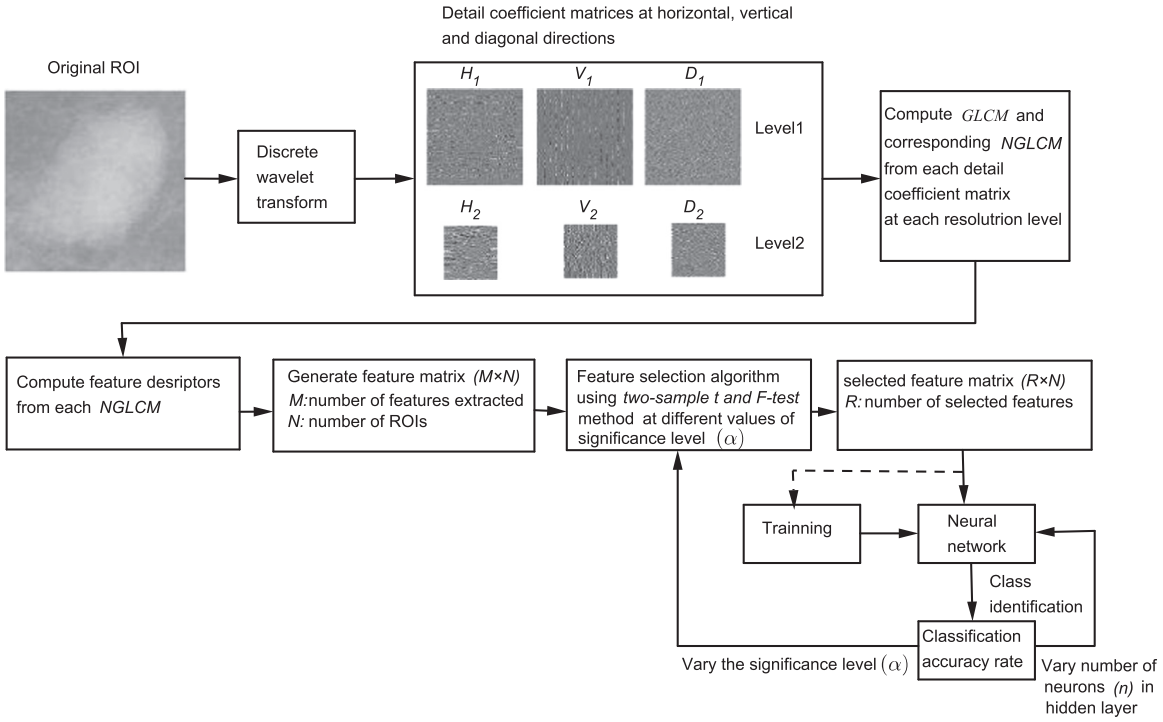

Fig. 1. Block diagram of the proposed scheme for classification of mammograms using back propagation neural network (BPNN).