BIOCYBERNETICS AND BIOMEDICAL ENGINEERING XXX (2017) XXX–XXX

Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Article

A classification framework for prediction of breast density using an ensemble of neural network classifiers

Q1 *Indrajeet Kumar* [a,*], *H.S. Bhadauria* [a], *Jitendra Virmani* [b], *Shruti Thakur* [c]

[a] *Department of Computer Science and Engineering, G B Pant Engineering College, Pauri Garhwal, Uttarakhand 246194, India*
[b] *CSIR-Central Scientific Instruments Organization, Chandigarh, India*
[c] *Department of Radiology, IGMC, Shimla, Himachal Pradesh, India*

ARTICLE INFO

Article history:
Received 11 September 2016
Received in revised form
10 January 2017
Accepted 15 January 2017
Available online xxx

Keywords:
Mammography
Breast density classification
Gray level co-occurrence matrix
Neural network classifier
Ensemble classifier

ABSTRACT

The present work proposes a classification framework for the prediction of breast density using an ensemble of neural network classifiers. Expert radiologists, visualize the textural characteristics of center region of a breast to distinguish between different breast density classes. Accordingly, ROIs of fixed size are cropped from the center location of the breast tissue and GLCM mean features are computed for each ROI by varying inter-pixel distance '*d*' from 1 to 15. The proposed classification framework consists of two stages, (*a*) *first stage*: this stage consists of a single 4-class neural network classifier NN0 (B-I/B-II/B-III/B-IV) which yields the output probability vector [$\mathbf{P}_{\text{B-I}}$ $\mathbf{P}_{\text{B-II}}$ $\mathbf{P}_{\text{B-III}}$ $\mathbf{P}_{\text{B-IV}}$] indicating the probability values with which a test ROI belongs to a particular breast density class. (*b*) *second stage*: this stage consists of an ensemble of six binary neural network classifiers NN1 (B-I/B-II), NN2 (B-I/B-III), NN3 (B-I/B-IV), NN4 (B-II/B-III), NN5 (B-II/B-IV) and NN6 (B-III/B-IV).

The output of the first stage of the classification framework, i.e. output on NN0 is used to obtain the two most probable classes for a test ROI. In the second stage this test ROI is passed through one of the binary neural networks, i.e. NN1 to NN6 corresponding to the two most probable classes predicted by NN0. After passing the entire test

* *Corresponding author at*: Department of Computer Science and Engineering, G B Pant Engineering College, Pauri Garhwal, Uttarakhand 246194, India.
E-mail addresses: erindrajeet@gmail.com (I. Kumar), hsb76iitr@gmail.com (H.S. Bhadauria), jitendra.virmani@gmail.com (J. Virmani), tshruti878@yahoo.in (S. Thakur).

Abbreviations: ROI, region of interest; FELM, fuzzy-extreme learning machine; NN, neural network classifier; BIRADS, breast imaging-reporting and data system; DDSM, digital database for screening mammography; MIAS, Mammographic Image Analysis Society; ST, segmented tissue; ANN, artificial neural network classifier; *k*-NN, *k*-nearest neighbors; MI, misclassified instances; SVM, support vector machine classifier; SFS, sequential forward search; MLO/CC, mediolateral oblique/cranial-caudal; TFV, texture feature vector; CM, confusion matrix; OCA, overall classification accuracy; NGTDM, neighborhood gray tone difference matrix; TI, testing instance; ICA, individual class accuracy; FOS, first order statistics; GLCM, gray level co-occurrence matrix; GLDS, gray-level difference statistics; SFM, statistical feature matrix; GLRLM, gray level run length matrix.

http://dx.doi.org/10.1016/j.bbe.2017.01.001

ROIs through the second stage, the overall accuracy increases from 79.5% to 90.8%. The promising results achieved by the proposed classification framework indicate that it can be used in clinical environment for differentiation between breast density patterns.

1. Introduction

Breast density is considered as a prominent indicator for the growth of breast lesions [1–16]. Fundamentally, variations in breast density patterns are associated with amounts of fatty tissue and glandular tissues. Women with fatty breast have lower percentage of fibrous tissue and higher percentage of fatty tissue and women with dense breast have vice versa. The distribution of tissue density according to BIRADS specification [17–22] is given in Table 1.

The different breast tissue density pattern reflects different texture properties, therefore the problem of 4-class (i.e. BIRADS breast density) classification can be considered as the problem of texture description and representation. Breast density classification is clinically important because most often the lesions masked behind the dense tissue are missed during screening mammography.

Mammography [23–28] is the first examination carried out for detection of breast diseases. It is based on different levels of X-ray absorption for the various breast tissues. The sensitivity and specificity of mammography is high, so small tumors and micro calcifications can be detected easily. In many cases it is observed that the breast lesions masked behind the dense tissue are missed during screening mammography. Thus the accurate prediction of density class is a prerequisite for detection of these lesions in the dense mammograms.

In the present work, a 2-stage classification framework for prediction of BIRADS breast density classes has been designed using an ensemble of neural network classifiers. The image dataset used for this work comprises of 480 mammograms taken from DDSM dataset such that (1) 120 mammograms belong to B-I class, (2) 120 mammograms belong to B-II class, (3) 120 mammograms belong to B-III class and (4) 120 mammograms belong to B-IV class. The sample images belonging to B-I, B-II, B-III and B-IV classes, taken from the DDSM dataset are shown in Fig. 1.

The development of a computer aided classification system for prediction of breast density plays an important role in clinical environment because of (1) the detection of lesions in dense mammograms is difficult in such cases if lesions are masked behind the dense tissue and are often missed during screening mammography and (2) increment in breast tissue density is considered as risk factor of breast cancer. The computer aided classification systems designed for characterization of breast density can be classified as (1) systems designed using segmented tissue (ST) based approach [29–33,35,36,38,40–42] and (2) systems designed using fixed size ROI based approach [34,37,39]. It is well known that segmented tissue based approach requires additional tasks viz. eliminating the background and removing the pectoral muscle. Due to these additional steps segmented tissue based approaches are time consuming and complex in comparison to fixed size ROI based approaches.

After the extensive review of literature it has been observed that breast density classification systems have been designed using either benchmark datasets (i.e. MIAS, DDSM) or datasets of mammographic images are collected by the authors. It is worth mentioning that the DDSM dataset contains images which are already labeled according to BIRADS density standard by the experts, however in case of MIAS dataset as well as in case of datasets collected by authors the images have been labeled according to BIRADS standard by the participating radiologists. The studies conceded in past for 4-class breast density classification is reported in Table 2.

It can be visualized from Table 2 that the studies on 4-class breast density classification have been conducted on (*a*) benchmark DDSM dataset [30–33,39], (*b*) benchmark MIAS dataset [31–33,35–37] and (*c*) dataset of mammographic images collected by authors [29,34,37,38,40–42]. Further from Table 2, it can be observed that most of the related researches in literature for 4-class breast density classification have been conducted on using segmented tissue based approach [29–33,35,36,38,40–42]. It may also be noted that 4-class breast density classification based on fixed size ROI has been carried out in studies [34,37,39]. It is worth observing that the maximum accuracy of 84.7% has been attained using segmented tissue based approach [32] and accuracy of 73.7% has been attained using ROI based approach [39] on DDSM dataset. The maximum accuracy of 79.2% has been attained on MIAS dataset [37] and the maximum accuracy of 86.4% has been attained on dataset of mammographic images collected by authors [34] using ROI based approach.

In study [39] the authors have attempted 4-class breast density classification using wavelet packet texture descriptors with fixed ROI size on 480 mammograms of DDSM dataset. The study reports the accuracy of 73.7% using SVM classifier. In study [34] the authors have attempted 4-class breast density classification on the self-collected dataset consisting of only 88 mammograms using SVM classifier yielding the accuracy of 86.4%. It may be noted that the study [39] can be only directly related to present work as it has been carried out on DDSM dataset using fixed size ROI based approach.

Table 1 – Distribution of tissue density according to BIRADS specification.

BIRADS class	Density (%)	Breast density
BIRADS-I (B-I)	00–25	Entirely fatty tissue
BIRADS-II (B-II)	26–50	Some-fibroglandular tissue
BIRADS-III (B-III)	51–75	Hetero-geneously dense tissue
BIRADS-I V (B-IV)	76–100	Extremely dense tissue