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

Indrajeet Kumar [a,*], *Bhadauria H.S.* [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 27 January 2017

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

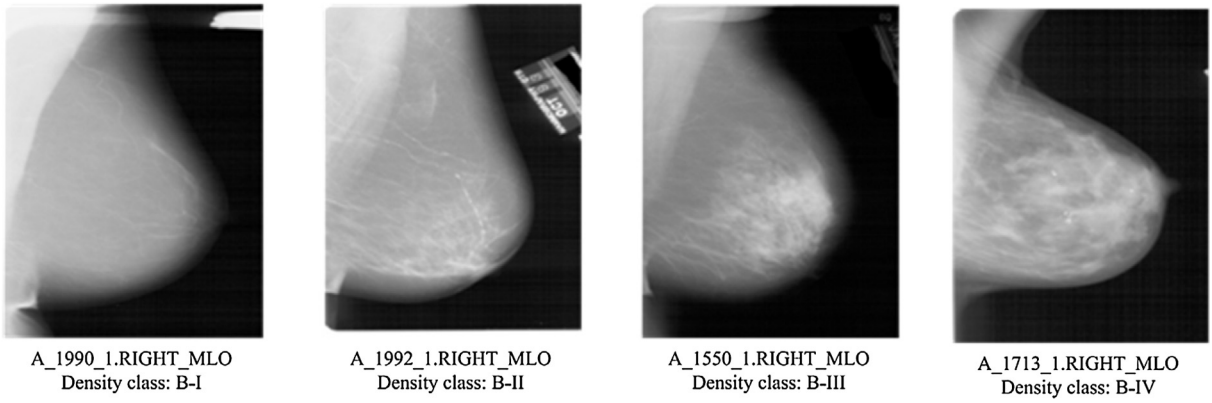

Fig. 1 – Sample images of BIRADS classes taken from DDSM dataset.

The participating radiologist opined that in clinical environment, radiologists visualize that the center location of the breast where the density information is maximum is efficient for discrimination between different breast tissue density classes. The same fact has also been experimentally investigated in study [43]. Therefore in this work fixed size ROIs of 128 × 128 pixels have been cropped from the center location of the breast where glandular ducts are prominent to obtain the accurate density class information. The texture feature vectors (TFVs) are computed from each ROI using second order statistical model. The computed TFVs are fed to 2-stage classification framework for prediction of breast density class.

2. Materials and methods

In the present work an efficient classification framework for prediction of breast density using an ensemble of neural network classifiers is proposed. In the proposed framework, the radiologist just has to mark an ROI of size 128 × 128 pixels at the center location of the breast where glandular ducts are found. The designed framework will automatically compute the desired features at each stage and an ensemble of neural network classifiers will predict the density class.

Experimental work flow for the design of an efficient classification framework for prediction of breast density

The experimental work flow followed in the present work for the design of an efficient classification framework for prediction of breast density class is shown in Fig. 2.

2.1. Dataset description

The DDSM dataset is a standard benchmark dataset which contains four digitized screen film mammographic images for each case, comprising of left/right MLO and left/right CC views. The overlay file of each image contains the expert evaluation of breast density according to BIRADS standard [44]. The image dataset consisting of 480 MLO view mammographic images was stored in a HP Z420D workstation having processor Intel® Xeon® 3.00 gigahertz with 8GB RAM. The description of dataset used in this study is shown in Fig. 3.

Table 2 – Study carried out for 4-class breast density classification.

Dataset used	Author, Year	ST/ROI	No. of images	Classifier	Accuracy (%)
DDSM	Bovis, 2002 [30]	ST	377	ANN	71.4
	Oliver, 2005 [31]	ST	615	kNN	47.0
	Bosch, 2006 [32]	ST	500	SVM	84.7
	Oliver, 2008 [33]	ST	132	SFS + kNN	77.0
	Kumar, 2015 [39]	ROI	480	SVM	73.7
MIAS	Oliver, 2005 [31]	ST	270	Decision tree	73.0
	Bosch, 2006 [32]	ST	322	SVM	95.4
	Oliver, 2008 [33]	ST	322	SFS + kNN	66.0
	Qu, 2011 [35]	ST	322	FELM*	72.6
	Chen, 2011 [36]	ST	322	kNN	75.0
	Mustra, 2012 [37]	ROI	322	kNN	79.2
Self collected dataset	Miller, 1991 [40]	ST	40	Bayesian	80.0
	Karssemeijer, 1998 [29]	ST	615	kNN	65.0
	Jamal, 2007 [41]	ST	100	–	78.3
	Liu, 2011 [34]	ROI	88	SVM	86.4
	Mustra, 2012 [37]	ROI	144	SFS + kNN	76.4
	Masmoudi, 2013 [38]	ST	2052	kNN	79.0
	He, 2016 [42]	ST	360	–	78.0

Note: ST, segmented tissue based approach; ROI, region of interest based approach; FELM*, fuzzy-extreme learning machine.

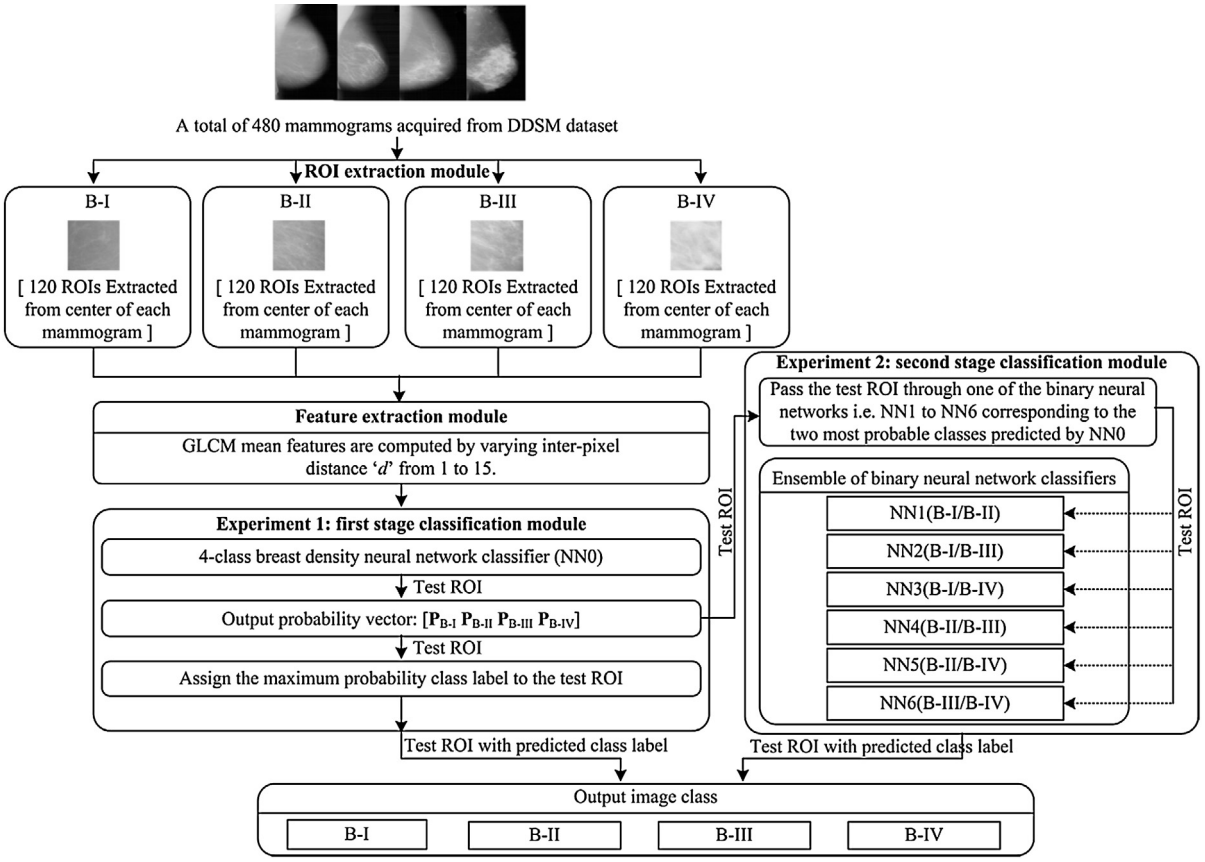

Fig. 2 – Experimental work flow followed in the present work for the design of an efficient classification framework for prediction of breast density class. [*Note*: $P_{B\text{-}I}$: probability value for B-I class; $P_{B\text{-}II}$: probability value for B-II class; $P_{B\text{-}III}$: probability value for B-III class; $P_{B\text{-}IV}$: probability value for B-IV class].

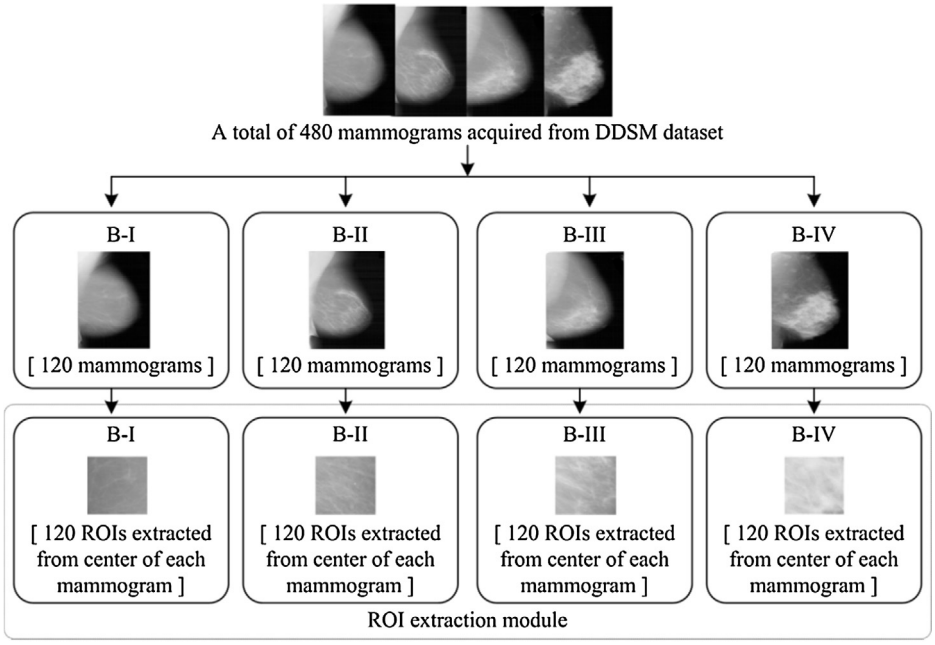

Fig. 3 – Dataset description.

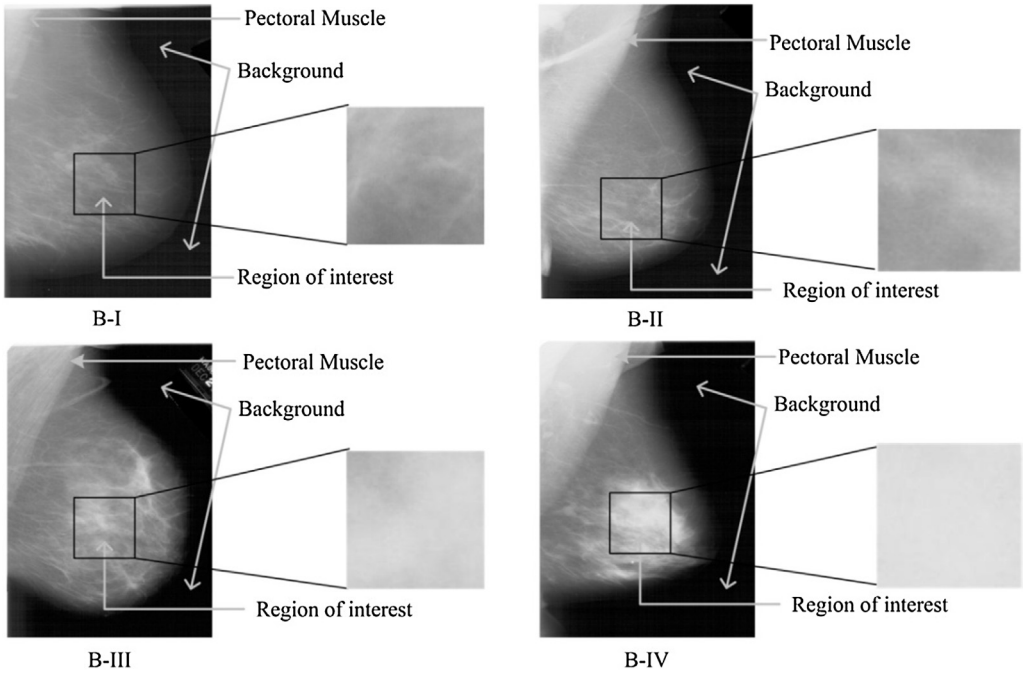

Fig. 4 – Sample of extracted ROI for each BIRADS breast density class.

2.2. *ROI extraction module*

The fact that the textural variations exhibited by the central region of the breast tissue are significant to account for discrimination between different breast density classes has been experimentally verified in the study carried by Li et al. [43]. Also the participating radiologist opined that the central area of breast tissue where glandular ducts are prominent accounts the significant information for discrimination between different breast density classes. In this study, ROIs of fixed size 128 × 128 pixels have been cropped from the central region of the breast where glandular ducts are found for each mammographic image. The sample of extracted ROI for each BIRADS breast density class is shown in Fig. 4.

2.3. *Feature extraction module*

From the exhaustive review of the literature it is observed that the texture features computed using second order statistical texture feature model (GLCM) contain significant information to account for variations in texture patterns exhibited by different breast density classes [45–55]. To verify this fact initially four directional GLCMs were obtained for a sample ROI belonging to each of B-I, B-II, B-III and B-IV breast density classes. The $GLCM_{mean}$ for a ROI belonging to a particular breast density class is obtained by using Eq. (1).

$$GLCM_{mean,B\text{-}I(d=i)} = \frac{GLCM_{B\text{-}I(0^\circ,d=i)} + GLCM_{B\text{-}I(45^\circ,d=i)} + GLCM_{B\text{-}I(90^\circ,d=i)} + GLCM_{B\text{-}I(135^\circ,d=i)}}{4} \quad (1)$$

In the similar manner $GLCM_{mean,B\text{-}II(d=i)}$, $GLCM_{mean,B\text{-}III(d=i)}$ and $GLCM_{mean,B\text{-}IV(d=i)}$ are computed by varying the inter-pixel distance 'd' = 'i' from 1 to 15.

The $GLCM_{mean}$ obtained at $d = 10$ for sample ROI images belonging to B-I, B-II, B-III and B-IV breast density classes have been plotted as images. These $GLCM_{mean}$ images are shown in Fig. 5(a)–(d) respectively.

From Fig. 5, it can be observed that the elements of $GLCM_{mean}$ are less dispersed in case of ROI belonging to BIRADS-I class and this dispersion is increasing for ROIs belonging to B-II, B-III and B-IV classes. Looking at the $GLCM_{mean}$ images for the sample ROI images belonging to each BIRADS density class, it can be observed that $GLCM_{mean}$ textural features may contain significant information to account for variations in texture patterns exhibited by different breast density classes.

The steps involved to compute $GLCM_{mean}$ are given here. One of the $GLCM_{mean}$ feature, i.e. inverse difference moment (IDM_{mean}) is computed at inter-pixel distance 'd' = 10 by using Eq. (2).

$$IDM_{mean(d=10)} = \left(\frac{W + X + Y + Z}{4}\right) \quad (2)$$

Here, $W = IDM_{(\theta=0^\circ,d=10)}$ $X = IDM_{IDM(\theta=045^\circ,d=10)}$ $Y = IDM_{(\theta=90^\circ,d=10)}$ $Z = IDM_{(\theta=135^\circ,d=10)}$

In the same manner, remaining 12 $GLCM_{mean}$ texture features angular second moment, sum average, sum variance, variance, sum entropy, entropy, difference variance, correlation, difference entropy, contrast, information measures of correlation-1, information measures of correlation-2 [45,49] have been computed by varying the inter-pixel distance 'd' from 1 to 15.

2.4. *Classification module*

The proposed classification framework consists of two stages, (a) *first stage classification module*: this module consisting of a single 4-class neural network classifier and (b) *second stage*

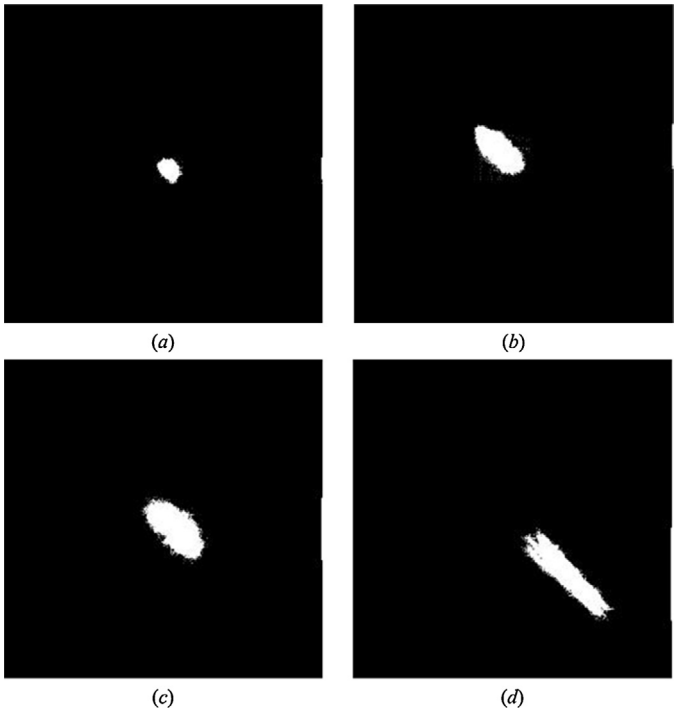

Fig. 5 – Mean of four directions (for $\theta = 0^\circ$, 45°, 90° and 135°) GLCM elements [*Note*: (a) B-I ROI, (b) B-II ROI, (c) B-III ROI and (d) B-IV ROI].

classification module: this module consisting of an ensemble of six binary neural network classifiers.

2.4.1. *First stage classification module*

This module consists of a single 4-class neural network classifier, i.e. 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 class. In this stage the predicted maximum class label is assigned to the test ROI.

The study was conducted by collecting a comprehensive image database of 480 MLO view digitized screen film mammograms comprises of (1) 120 mammograms of BIRADS-I, (2) 120 mammograms of BIRADS-II, (3) 120 mammograms of BIRADS-III and (4) 120 mammograms of BIRADS-IV taken from DDSM dataset. The instances of training data and testing data for the design of 4-class neural network classifier (NN0) is consisting of (1) 60 ROIs of BIRADS-I, (2) 60ROIs of BIRADS-II, (3) 60ROIs of BIRADS-III and (4) 60 ROIs of BIRADS-IV.

For obtaining the optimal training model for 4-class neural network classifier (NN0) the model was repeatedly trained and tested using $\text{GLCM}_{\text{mean}}$ TFVs computed by varying inter-pixel distance 'd' from 1 to 15. It was observed that the maximum classification accuracy of 79.5% is achieved by using $\text{GLCM}_{\text{mean}}$ features computed at inter-pixel distance $d = 10$.

In the present study extensive experiments have been conducted to analyze the performance of CAD systems for breast density classification using various statistical texture feature models. The obtained accuracy for various texture models are given in Table 3.

From Table 3, it has been observed that $\text{GLCM}_{\text{mean}}$ texture features yields the maximum classification accuracy, i.e. 79.5 at interpixels distance $d = 10$ for classification module stage one. Thus $\text{GLCM}_{\text{mean}}$ texture features are considered for further analysis, i.e. for the second stage classification module.

2.4.2. *Second stage classification module*

This module consists of an ensemble of six binary neural network classifiers, i.e. 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 classification module, i.e. output probability vector $[\mathbf{P}_{\text{B-I}}\ \mathbf{P}_{\text{B-II}}\ \mathbf{P}_{\text{B-III}}\ \mathbf{P}_{\text{B-IV}}]$ predicted by NN0 is used to obtain the two most probable classes for a test ROI. In the second stage classification module 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.

The bifurcation of instances of training data and testing data for the design of six binary neural network classifiers (NN1 to NN6) is shown in Fig. 6.

It is not necessary that the TFV ($\text{GLCM}_{\text{mean}}$ features computed at inter-pixel distance $d = 10$) which yield maximum

Table 3 – Obtained accuracy for various texture models.

Texture model	Accuracy (%)	Texture model	Accuracy (%)
FOS	69.1	NGTDM at $d = 10$	69.6
$\text{GLCM}_{\text{mean}}$ at $d = 10$	**79.5**	SFM	51.6
GLDS	47.1	GLRLM	66.6

The bold values in the table signify the highest accuracy achieved among the various texture features for the present work.

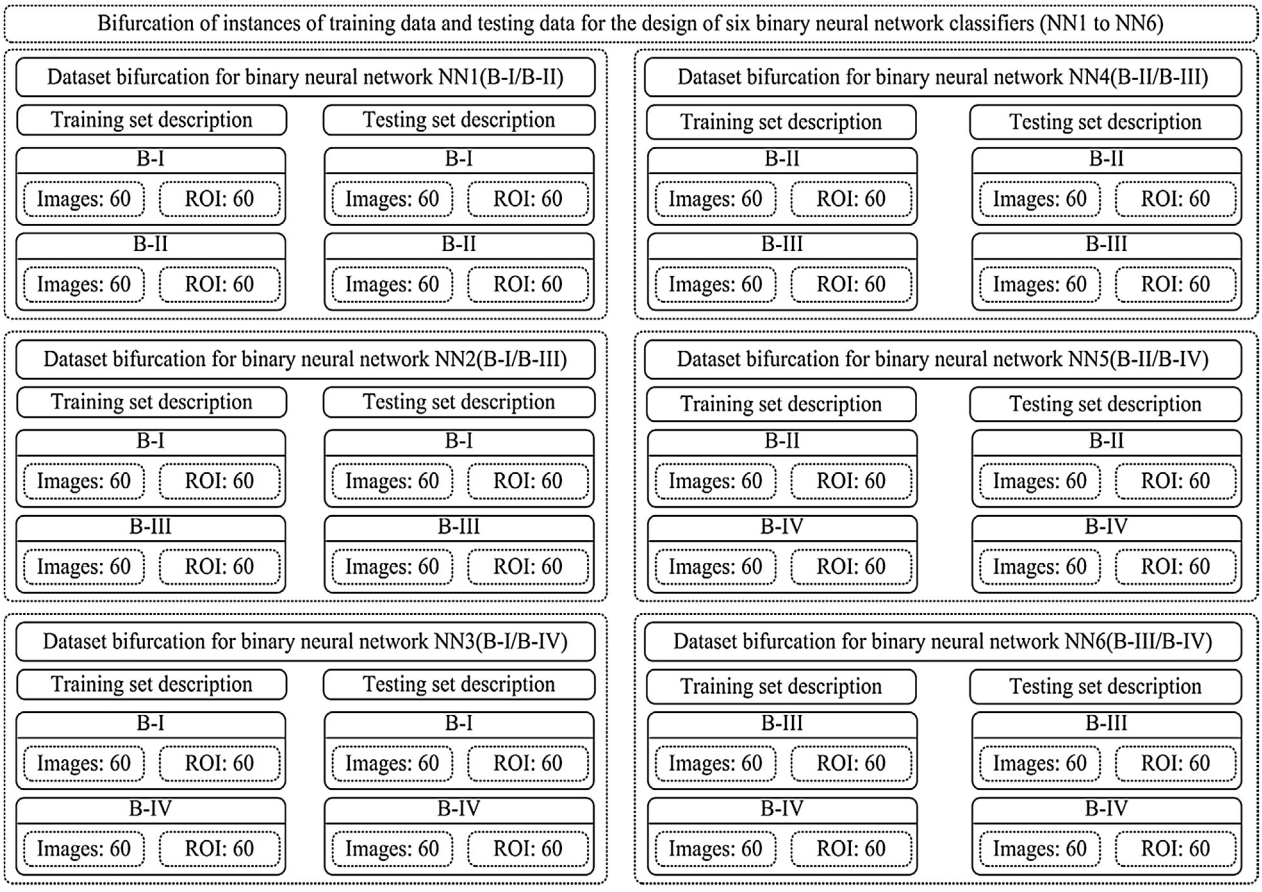

Fig. 6 – Bifurcation of instances of training data and testing data for the design of NN1 to NN6 classifiers.

classification accuracy for NN0 will also yield maximum classification accuracy for binary neural network classifiers (NN1 to NN6). Therefore for obtaining the optimal training model for each binary neural network classifier, all models were repeatedly trained and tested using $GLCM_{mean}$ TFVs computed by varying inter-pixel distance '*d*' from 1 to 15. The optimal TFV with which maximum classification accuracy was achieved for each neural network is given in Table 4.

For designing the 4-class neural network classifier (NN0) and for designing the six binary neural network classifiers (NN1 to NN6) the trial-and-error procedure was used for the optimization of hidden layer neurons. After exhaustive experimentation with different numbers of hidden layer neurons, it was concluded that with 9 neurons in hidden layer of NN0 and with eight neurons in hidden layers of 'NN1 to NN6 a reasonable tradeoff between convergence and accuracy was obtained. The back propagation algorithm with adaptive learning rate and momentum has been carried out for learning of these supervised networks [56–62].

The work performed in first stage classification module (i.e. passing the test ROI through 4-class neural network classifier NN0) is equivalent to showing the test image to an experienced radiologist who has expertise of making differential diagnosis between the 4-class BIRADS breast density

Table 4 – Optimal TFV with which maximum classification accuracy was achieved for each neural network.

Classification stage	TFV	*l*	Neural network [(classes) (I:H:O)]	Accuracy (%)
Stage 1	TFV1: $GLCM_{mean}$ (d = 10)	13	NN0 [(B-I/B-II/B-III/B-IV) (13:9:4)]	79.5
Stage 2	TFV1: $GLCM_{mean}$ (d = 10)	13	NN1 [(B-I/B-II) (13:8:2)]	95.8
	TFV1: $GLCM_{mean}$ (d = 10)	13	NN2 [(B-I/B-III) (13:8:2)]	95.8
	TFV2: $GLCM_{mean}$ (d = 5)	13	NN3 [(B-I/B-IV) (13:8:2)]	100
	TFV3: $GLCM_{mean}$ (d = 9)	13	NN4 [(B-II/B-III) (13:8:2)]	85.0
	TFV4: $GLCM_{mean}$ (d = 7)	13	NN5 [(B-II/B-IV) (13:8:2)]	98.3
	TFV5: $GLCM_{mean}$ (d = 6)	13	NN6 [(B-III/B-IV) (13:8:2)]	88.3

Note: TFV: texture feature vector; l: length of TFV; I: input layer neurons; H: hidden layer neurons; O: output layer neurons.

Table 5 – Results obtained by the first stage of classification module (stage 1).

	CM				Accuracy (%)	
	B-I	B-II	B-III	B-IV	ICA	OCA
B-I	56	3	1	0	93.3	79.5
B-II	3	47	9	1	78.5	
B-III	1	14	32	13	53.3	
B-IV	0	0	4	56	93.3	

classes. This radiologist is asked to provide the probability value with which a test ROI belongs to a particular class. Let's assume that the radiologist predicts an output probability vector [$\mathbf{P}_{B\text{-}I}$ (0.3012) $P_{B\text{-}II}$ (0.7875) $\mathbf{P}_{B\text{-}III}$ (0.7242) $\mathbf{P}_{B\text{-}IV}$ (0.4125)] for a particular test ROI. In the first stage of classification module depending upon the prediction of this radiologist the predicted maximum probability class label, i.e. $P_{B\text{-}II}$ (0.7875) = *max* ([$\mathbf{P}_{B\text{-}I}$ (0.3012) $P_{B\text{-}II}$ (0.7875) $\mathbf{P}_{B\text{-}III}$ (0.7242) $\mathbf{P}_{B\text{-}IV}$ (0.4125)]) is assigned to the test ROI.

It can be witnessed from the probability vector that the radiologist had confusion between $P_{B\text{-}II}$ (0.7875) and $\mathbf{P}_{B\text{-}III}$ (0.7242) classes.

The work performed in second stage classification module (i.e. passing the test ROI through the binary neural network classifier corresponding to the two most probable classes, i.e. NN4 in this case) is equivalent to taking the second opinion from an another experienced radiologist who has expertise of making differential diagnosis between B-II and B-III breast density classes only. Let's assume that this radiologist predicts an output probability vector [$P_{B\text{-}II}$ (0.7925) $\mathbf{P}_{B\text{-}III}$ (0.8228)] for the test ROI. In the second stage classification module depending upon the prediction of this radiologist the predicted maximum probability class label, i.e. $P_{B\text{-}III}$ (0.8228) = *max* ([$P_{B\text{-}II}$ (0.7925) $\mathbf{P}_{B\text{-}III}$ (0.8228)]) is assigned to the test ROI.

3. Experiments and results

In this work extensive experiments have been performed for the design of an efficient classification framework for prediction of breast density using an ensemble of neural network classifiers. The description of these experiments is given here.

3.1. *Experiment 1*

In this experiment a single 4-class neural network classifier, i.e. NN0 (B-I/B-II/B-III/B-IV) has been designed. The NN0 classifier yields the output probability vector [$\mathbf{P}_{B\text{-}I}$ $P_{B\text{-}II}$ $\mathbf{P}_{B\text{-}III}$ $\mathbf{P}_{B\text{-}IV}$] indicating the probability values with which a test ROI belongs to a particular class. In this experiment the maximum probability class labels, as predicted by NN0, are assigned to the test ROIs.

The performance of TFV ($GLCM_{mean}$ features computed at $d = 10$) for classification between four BIRADS density classes has been tested using NN0 classifier. The results obtained (i.e. results obtained by the first stage of classification module) are reported in Table 5.

It can be observed from Table 5 that 4-class neural network classifier (NN0) yields overall classification accuracy (OCA) value of 79.5% (i.e. from total 240 testing instances 191 have been correctly classified). The 191(191/240) correctly classified instances consist of 56 (56/60) B-I, 47 (47/60) B-II, 32 (32/60) B-III and 56 (56/60) B-IV cases. Thus the individual class classification accuracy (ICA) values of 93.3% (56/60), 78.3% (47/60), 53.3% (32/60) and 93.3% (56/60) are obtained for B-I, B-II, B-III and B-IV classes respectively.

From the confusion matrix of Table 5, it has been observed that the 49 (49/240) instances are misclassified. The description of 240 testing instances according to the output of first stage classification module is reported in Table 6.

It is worth mentioning that from total 240 testing instances 49 (49/240) have been misclassified. These 49(49/240) misclassified instances consist of 4(4/60) B-I, 13(13/60) B-II, 28 (28/60) B-III and 4 (4/60) B-IV cases. Out of these 49 (49/240) misclassified instances, 42 (42/49) instances having the 2nd highest probability value for belonging to correct class, 5 (5/49) instances having the 3rd maximum probability value for belonging to correct class and the remaining 2 (2/49) instances having the 4th highest probability value for belonging to correct class.

3.2. *Experiment 2*

In this experiment an ensemble of six binary neural network classifiers, i.e. 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) has been designed. All the 240 testing instances are passed through one of the binary neural networks, i.e. NN1 to NN6 corresponding to the two most probable classes predicted by NN0.

On the basis of the output probability vector [$\mathbf{P}_{B\text{-}I}$ $P_{B\text{-}II}$ $\mathbf{P}_{B\text{-}III}$ $\mathbf{P}_{B\text{-}IV}$] predicted by NN0 the number of testing instances that pass through the each binary neural network classifier is shown in Table 7.

Table 6 – Description of 240 testing instances according to the output of first stage of classification module.

Total testing instances	Correctly classified instances#	Misclassified instances	Instances having 2nd highest probability for correct class	Instances having 3rd highest probability for correct class	Instances having 4th highest probability for correct class
{60 ∈ B-I}	56	4	3	1	–
{60 ∈ B-II}	47	13	10	2	1
{60 ∈ B-III}	32	28	25	2	1
{60 ∈ B-IV}	56	4	4	–	–
Total: 240	191	49	42	5	2

Note: Correctly classified instances#: instances having 1st highest probability for correct class.

Table 7 – Distribution of 240 testing instances according to the two most probable classes predicted by first stage of classification module.

Neural network	Testing instances	Neural network	Testing instances
NN1(B-I/B-II)	61	NN4(B-II/B-III)	64
NN2(B-I/B-III)	17	NN5(B-II/B-IV)	14
NN3(B-I/B-IV)	8	NN6(B-III/B-IV)	76

The performance of TFVs ($GLCM_{mean}$ feature vectors) for classification between two most probable classes (as evaluated using binary neural networks, i.e. NN1 to NN6) for 240 testing instances is depicted in Table 8.

From Table 8 it can be visualized that after the second stage of the classification module 218 (218/240) instances are correctly classified and the remaining 22 (22/240) instances are misclassified. The total number of misclassified instances is computed using Eq. (3).

$$TotalMIs = \sum(\{MIs \in NN1\}, \{MIs \in NN2\}, \{MIs \in NN3\}, \{MIs \in NN4\}, \{MIs \in NN5\}, \{MIs \in NN6\}) \quad (3)$$

Here MIs stand for misclassified instances and NN1 to NN6 are binary neural network classifiers.

Table 9 – Results obtained by the second stage of classification module (stage 2).

	CM				Accuracy (%)	
	B-I	B-II	B-III	B-IV	ICA	OCA
B-I	59	1	0	0	98.3	90.8
B-II	1	55	4	0	91.6	
B-III	0	7	48	5	80.0	
B-IV	0	0	4	56	93.3	

The classification performance of TFVs $GLCM_{mean}$ features for classification between 4 BIRADS density classes has been tested using binary neural network classifiers (NN1 to NN6) and the results obtained (i.e. the results obtained by second stage of classification module) are reported in Table 9.

It can be observed from Table 9 that the second stage of classification module (NN1–NN6) yields overall classification accuracy (OCA) value of 90.8% (i.e. from total 240 testing instances 218 have been correctly classified). The 218 (218/240) correctly classified instances consist of 59 (59/60) B-I cases, 55 (55/60) B-II cases, 48 (48/60) B-III cases and 56 (56/60) B-IV cases. Thus the individual class classification accuracy (ICA) values of 98.3% (59/60), 91.6% (55/60), 80.0% (48/60) and 93.3% (56/60) are obtained for B-I class, B-II class, B-III class and B-IV class cases respectively.

Table 8 – Classification performance of TFV ($GLCM_{mean}$ features) tested by binary neural networks classifiers (NN1 to NN6).

TFV	Neural network	No. of TIs	Testing class labels	Class label predicted by stage 2	
				Correctly classified	Misclassified
TFV1: $GLCM_{mean\ (d=10)}$	NN1(B-I/B-II)	61	{37 ∈ B-I}	{37 ∈ B-I} ✓	–
			{23 ∈ B-II}	{23 ∈ B-II} ✓	–
			{1 ∈ B-III}	-	{1 ∈ B-III} ×
				Description of misclassified instances	
				1 instance belonging to B-III class is misclassified as BI-II class.	
TFV1: $GLCM_{mean\ (d=10)}$	NN2(B-I/B-III)	17	{17 ∈ B-I}	{17 ∈ B-I} ✓	–
TFV2: $GLCM_{mean\ (d=5)}$	NN3(B-I/B-IV)	8	{5 ∈ B-I}	{5 ∈ B-I} ✓	–
			{1 ∈ B-II}	–	{1 ∈ B-II} ×
			{2 ∈ B-IV}	{2 ∈ B-IV} ✓	–
				1 instance belonging to B-II class is misclassified as B-I class.	
TFV3: $GLCM_{mean\ (d=9)}$	NN4(B-II/B-III)	64	{1 ∈ B-I}	–	{1 ∈ B-I} ×
			{33 ∈ B-II}	{31 ∈ B-II} ✓	{2 ∈ B-II} ×
			{30 ∈ B-III}	{24 ∈ B-III}✓	{6 ∈ B-III} ×
				1 instance belonging to B-I class is misclassified as B-II class, 2 instances belonging to B-II class are misclassified as B-III class and 6 instances belonging to B-III class are misclassified as B-II class.	
TFV4: $GLCM_{mean\ (d=7)}$	NN5(B-II/B-IV)	14	{1 ∈ B-II}	{1 ∈ B-II} ✓	–
			{2 ∈ B-III}	–	{2 ∈ B-III} ×
			{11 ∈ B-IV}	{11 ∈ B-IV}✓	–
				2 instances belonging to B-III class are misclassified as B-IV class.	
TFV5: $GLCM_{mean\ (d=6)}$	NN6(B-III/B-IV)	76	{2 ∈ B-II} {27 ∈ B-III}	–	{2 ∈ B-II} ×
			{47 ∈ B-IV}	{24 ∈ B-III}✓	{3 ∈ B-III} ×
				{43 ∈ B-IV}✓	{4 ∈ B-IV} ×
				2 instances belonging to B-II class and 4 instances belonging to B-IV class are misclassified as B-III class and 3 instances belonging to B-III are misclassified as B-IV class.	
		Total: 240		Total: 218	Total: 22

Note: TIs: testing instances.

Table 10 – Comparative analysis of misclassified instances at different stages of classification module.

Testing instances	Misclassified at stage 1	Misclassified at stage 2
{60 ∈ B-I}	{4 ∈ B-I}	{1 ∈ B-I}
{60 ∈ B-II}	{13 ∈ B-II}	{5 ∈ B-II}
{60 ∈ B-III}	{28 ∈ B-III}	{12 ∈ B-III}
{60 ∈ B-IV}	{4 ∈ B-IV}	{4 ∈ B-IV}
Total: 240	Total: 49	Total: 22

Note: Out of total 240 testing instances 49 are misclassified by first stage.
Out of total 240 testing instances 22 are misclassified by second stage.
Total 27 (49 – 22) testing instances misclassified by first stage are correctly classified by second stage.

3.3. *Comparative analysis of misclassified instances*

It is worth mentioning that from total 240 testing instances 22 have been misclassified after the second stage of classification module. These 22(22/240) misclassified instances consist of 1 (1/60) belonging to B-I, 5(5/60) belonging to B-II, 12 (12/60) belonging to B-III and 4 (4/60) belonging to B-IV classes respectively.

The comparative analysis of misclassified instance by different stages of classification module is reported in Table 10.

From Table 10, it has been observed that most of the cases in misclassified instances belong to the B-II or B-III breast density classes. However 16 misclassified instances belonging to B-III class, 8 misclassified instances belonging to B-II class and 3 misclassified instances belonging to B-I class have been correctly classified after the inclusion of second stage of classification module. Thus total 27 (16 + 8 + 3) instances misclassified by first stage of classification module have been correctly classified after the inclusion of second stage of classification module and the remaining 22 instances remain misclassified.

From Table 10, it can be seen that the number of misclassified instances are decreased from 49 (49/240) to 22 (22/240) after the inclusion of the second stage of classification module; thus, the overall classification accuracy has improved from 79.5% to 90.8%. The comparison of the results obtained after stage 1 and stage 2 of classification module is given in Table 11.

From Table 11, it is observed that after the inclusion of the second stage of classification module the OCA value has increased by 11.3% and the ICA values for B-I, B-II and B-III have increased by 5%, 13% and 26.7% respectively. However no change has been observed for the ICA value of B-IV class.

The statistical performance analysis is performed by the calculating Cohen's kappa coefficient value. The kappa value

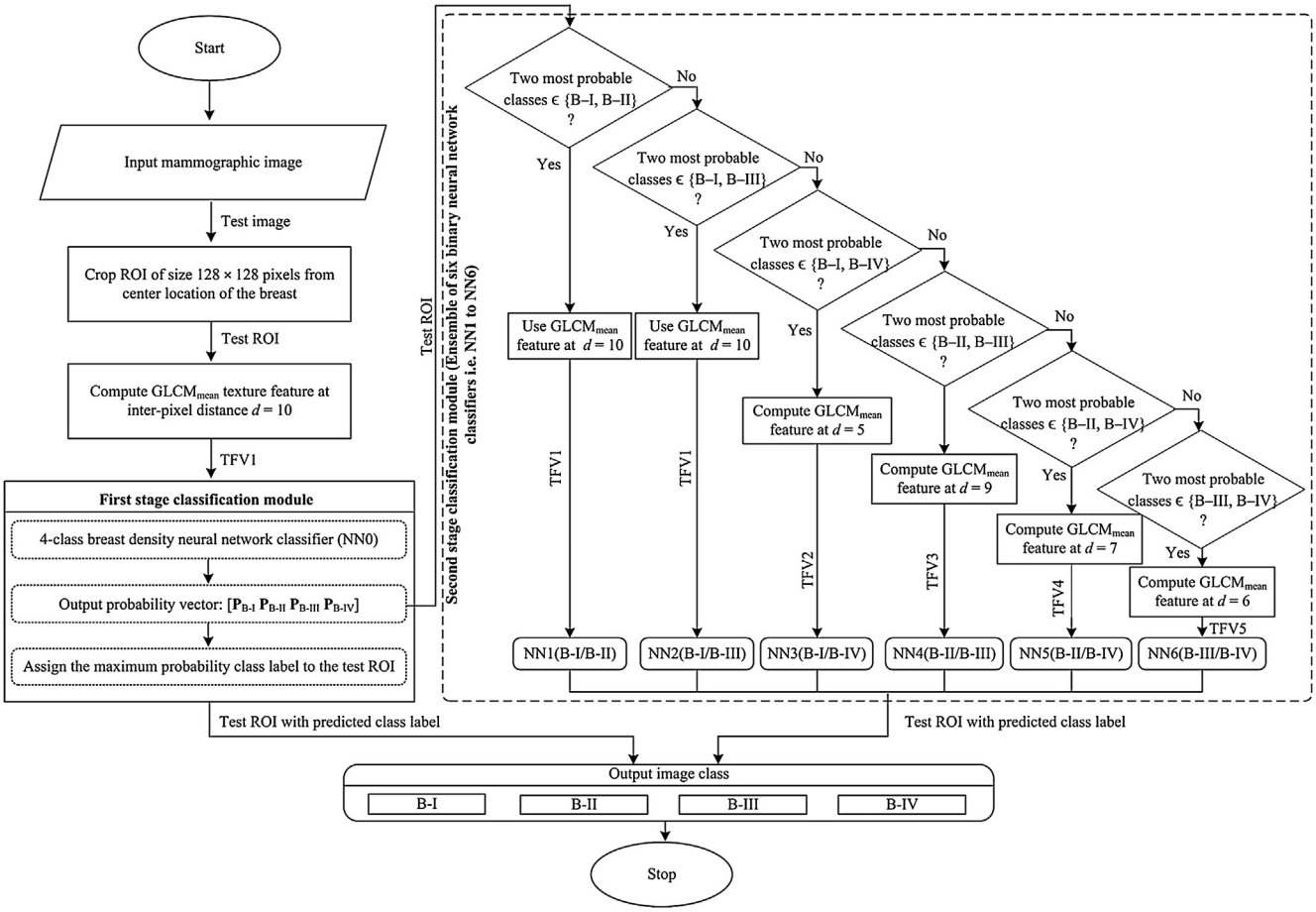

Fig. 7 – Flowchart for classification framework for prediction of breast density using an ensemble of neural network classifiers.

Table 11 – Comparison of the results obtained after stage 1 and stage 2 of classification module.

Classification stage	Accuracy (%)					Kappa value
	OCA	ICA-I	ICA-II	ICA-III	ICA-IV	
After stage 1	79.5	93.3	78.3	53.3	93.3	0.7278
After stage 2	90.8	98.3	91.6	80.0	93.3	0.8778

Note: OCA: overall classification accuracy; ICA-I: individual class accuracy for B-I; ICA-II: individual class accuracy for B-II; ICA-III: individual class accuracy for B-III; ICA-IV: individual class accuracy for B-IV.

of stage 1 is 0.7228 which shows the designed system performance is average. The kappa value for stage 2 is 0.8778 which shows the designed system is acceptable and recommended for the clinical practice.

The flow chart of the proposed framework for prediction of breast density using an ensemble of neural network classifiers is shown in Fig. 7.

4. Conclusion

From the extensive experiments performed in this work, it has been observed that the second order statistical texture features, i.e. $GLCM_{mean}$ texture features are significant to account for design of breast density classification framework. Further, it has been noticed that there is considerable improvement in classification accuracy of the proposed classification framework after the inclusion of an ensemble of binary neural network classifiers in the second stage.

The results obtained by the proposed classification framework for 4-class breast density classification using ensemble of neural network classifiers indicate that the framework can be routinely used in clinical environment for differential diagnosis between different breast density patterns.

REFERENCES

[1] Parkin DM, Fernandez LM. Use of statistics to assess the global burden of breast cancer. Breast 2006;12:70–80.

[2] Boyd NF, Martin L, Chavez S, Gunasekara A, Salleh A, Bronskill M. Breast-tissue composition and other risk factors for breast cancer in young women: a cross-sectional study. Lancet Oncol 2009;10:569–80.

[3] Colin C, Prince V, Valette PJ. Can mammographic assessments lead to consider density as a risk factor for breast cancer? Eur J Radiol 2013;82:404–11.

[4] Ekpo EU, Ujong UP, Mello-Thoms C, McEntee MF. Assessment of interradiologist agreement regarding mammographic breast density classification using the fifth edition of the BI-RADS atlas. Am J Roentgenol 2016;206:1119–23.

[5] Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, et al. Digital mammographic density and breast cancer risk: a case–control study of six alternative density assessment methods. Breast Cancer Res 2014;16:439–51.

[6] Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. Breast Cancer Res 2011;13:223–34.

[7] Vachon CM, Van-Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR, et al. Mammographic density, breast cancer risk and risk prediction. Breast Cancer Res 2007;9:1–9.

[8] Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007;356:227–36.

[9] Maskarinec M, Pagano I, Lurie G, Wilkens LR, Kolonel L. Mammographic density and breast cancer risk the multiethnic cohort study. Am J Epidemiol 2005;162:743–52.

[10] Pinsky RW, Helvie MA. Mammographic breast density: effect on imaging and breast cancer risk. J Natl Compr Canc Netw 2010;8:1157–65.

[11] McCormack VA, Silva ID. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidem Biomar 2006;15:1159–69.

[12] Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer 1976;37:2486–92.

[13] Wolfe JN. Breast patterns as an index of risk for developing breast cancer. Am J Roentgenol 1976;126:1130–7.

[14] Boyd NF, Rommens JM, Vogt K, Lee v, Hopper J, Yaffe MJ, et al. Mammographic breast density as an intermediate phenotype for breast cancer. Lancet Oncol 2005;6(10):798–808.

[15] Warren R. Hormones and mammographic breast density. Maturitas 2004;49:67–78.

[16] Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. Cancer Epidem Biomar 2008;7:1133–44.

[17] Ganesan K, Acharya U, Chua CK, Min LC, Abraham KT, Ng KB. Computer-aided breast cancer detection using mammograms: a review. IEEE Rev Biomed Eng 2013;6:77–98.

[18] Kumar I, Virmani J, Bhadauria HS. A review of breast density classification methods. Proceeding of 2nd International Conference on Computing for Sustainable Global Development 'INDIACom – 2015; 2015.

[19] Zhang G, Wang W, Moon J, Pack JK, Jean S. A review of breast tissue classification in mammograms. Proceedings of ACM Symposium on Research in Applied Computation; 2011.

[20] Abdel-Gawad EA, Khalil OA, Ragaee SM. Assessment of breast lesions using BI-RADS US lexicon in mammographically dense breasts (ACR categories 3 and 4) with histopathological correlation. Egypt J Radiol Nucl Med 2014;45:1301–7.

[21] Tagliafico A, Tagliafico G, Tosto S, Chiesa F, Martinoli C, Derchi LE, et al. Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one. Breast 2009;18:35–40.

[22] John M, Karen K, Carol S, Seibert J. A breast density index for digital mammograms based on radiologists ranking. J Digit Imaging 1998;2:101–15.

[23] Pereira R, Marques PM, Honda OM, Kinoshita SK, Engelmann R, Muramatsu C, et al. Usefulness of texture analysis for computerized classification of breast lesions on mammograms. J Digit Imaging 2007;20:248–55.

[24] Oliver A, Tortajada M, Llado X, Freixenet J, Ganau S, Tortajada L, et al. Breast density analysis using an automatic density segmentation algorithm. J Digit Imaging 2015;28:604–12.

[25] Papaevangelou A, Chatzistergos S, Nikita KS, Zografos G. Breast density: computerized analysis on digitized mammograms. Hellenic J Surg 2011;83:133–8.
[26] Heine J, Carton MJ, Scott CG. An automated approach for estimation of breast density. Cancer Epidem Biomar 2008;17:3090–7.
[27] Huo Z, Giger ML, Vyborny CJ. Computerized analysis of multiple-mammographic views: potential usefulness of special view mammograms in computer-aided diagnosis. IEEE T Med Imaging 2001;20(12):1285–92.
[28] Yaghjyan L, Pinney S, Mahoney M, Morton A, Buckholz J. Mammographic breast density assessment: a methods study. Atlas J Med Biol Sci 2011;1:8–14.
[29] Karssemeijer N. Automated classification of parenchymal patterns in mammograms. Phys Med Biol 1998;43:365–9.
[30] Bovis K, Singh S. Classification of mammographic breast density using a combined classifier paradigm. 4th International Workshop on Digital Mammography. 2002. pp. 177–80.
[31] Oliver A, Freixenet J, Zwiggelaar R. Automatic classification of breast density. Proceedings of the IEEE International Conference on Image Processing 'ICIP 2005' Genova 2005; 2005.
[32] Bosch A, Munoz X, Oliver A, Marti J. Modeling and classifying breast tissue density in mammograms. Proceeding of Computer Vision and Pattern Recognition; 2006.
[33] Oliver A, Freixenet J, Marti R, Pont j, Perez E, Denton ER, et al. A novel breast tissue density classification methodology. IEEE T Inf Technol B 2008;12:55–65.
[34] Liu Q, Liu L, Tan Y, Wang J, Ma X, Ni X. Mammogram density estimation using sub-region classification. Proceeding of 4th International Conference on Biomedical Engineering and Informatics; 2011.
[35] Qu Y, Shang C, Wu W, Shen Q. Evolutionary fuzzy extreme learning machine for mammographic risk analysis. Int J Fuzzy Systems 2011;13:282–91.
[36] Chen Z, Denton E, Zwiggelaar R. Local feature based mammographic tissue pattern modeling and breast density classification. Proceeding of 4th International Conference on Biomedical Engineering and Informatics; 2011.
[37] Mustra M, Grgic M, Delac K. Breast density classification using multiple feature selection. Automatika: J Contr Measurement Electron Comput Commun 2012;53:362–72.
[38] Masmoudi AD, Ayed NGB, Masmoudi DS, Abid A. LBPV descriptors-based automatic ACR/BIRADS classification approach. J Image Video Process 2013;1(1):1–9.
[39] Kumar I, Virmani J, Bhadauria HS. Wavelet packet texture descriptors based four-class BIRADS breast tissue density classification. Procedia Comput Sci 2015;70:76–84.
[40] Miller P, Astley S. Classification of breast tissue by texture analysis. Proceeding of BMVC91; 1991.
[41] Jamal N, Ng NH, Ranganathan S, Tan LK. Comparison of computerized assessment of breast density with subjective BI-RADS classification and Tabar's pattern from two-view CR mammography. Proceeding of World Congress on Medical Physics and Biomedical Engineering; 2006.
[42] He W, Harvey S, Juette A, Denton A, Zwiggelaar R. Mammographic segmentation and density classification: a fractal inspired approach. Proceeding of International Workshop on Digital Mammography; 2016.
[43] Li H, Giger ML, Huo Z, Olopade OI, Lan L, Weber BL, et al. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location. Med Phys 2004;31:549–55.
[44] Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP. The digital database for screening mammography. Proceedings of the 5th International Workshop on Digital Mammography; 2000.
[45] Haralick RM, Shanmugam K, Dinstein JH. Textural features for image classification. IEEE T Syst Man Cyb 1973;6:610–21.
[46] Li H, Giger ML, Olopade OI, Margolis A, Lan L, Chinander MR. Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. Acad Radiol 2005;12:863–73.
[47] Tourassi GD. Journey toward computer-aided diagnosis: role of image texture analysis. Radiology 1999;213:317–20.
[48] Mudigonda NR, Rangayyan RM, Desautels JL. Gradient and texture analysis for the classification of mammographic masses. IEEE T Med Imaging 2000;19:1032–43.
[49] Mudigonda NR, Rangayyan RM, Desautels JL. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE T Med Imaging 2001;20:1215–27.
[50] Vasantha M, Bharathi VS, Dhamodharan R. Medical image feature extraction, selection and classification. Int J Eng Sci Tech 2010;2:2071–6.
[51] Mohanaiah P, Sathyanarayana P, Kumar LG. Image texture feature extraction using GLCM approach. IJSR 2013;3:862–6.
[52] Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. Clin Radiol 2004;5:1061–9.
[53] Amadasun M, King R. Textural features corresponding to textural properties. IEEE T Syst Man Cyb 1989;19:1264–74.
[54] Weszka JS, Dyer CR, Rosenfeld A. A comparative study of texture measures for terrain classification. IEEE T Syst Man Cyb 1976;4:269–85.
[55] Kim JK, Park HW. Statistical textural features for detection of microcalcifications in digitized mammograms. IEEE T Med Imaging 1999;18:231–8.
[56] Mougiakakou SG, Valavanis IK, Nikita A, Nikita KS. Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. Artif Intell Med 2007;41(1):25–37.
[57] Virmani J, Kumar V, Kalra N, Khandelwal N. Prediction of cirrhosis based on singular value decomposition of gray level co-occurrence matrix and an neural network classifier. Proceedings of the IEEE International Conference on Developments in E-systems Engineering, Dubai (DeSe); 2011.
[58] Lee WL, Hsieh KS, Chen YC. A study of ultrasonic liver images classification with artificial neural networks based on fractal geometry and multiresolution analysis. Biomed Eng-App Bas 2004;16:59–67.
[59] Andre TC, Rangayyan RM. Classification of breast masses in mammograms using neural networks with shape, edge sharpness, and texture features. J Electron Imaging 2006;15:013019–29.
[60] Wu y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology 1993;187:81–7.
[61] Virmani J, Kumar V, Kalra N, Khandelwal N. Neural network ensemble based CAD system for focal liver lesions from B-mode ultrasound. J Digit Imaging 2014;27(4):520–37.
[62] Sujana H, Swarnamani S, Suresh S. Application of artificial neural networks for the classification of liver lesions by image texture parameters. Ultrasound Med Biol 1996;22:1177–81.