A computerised framework for characterisation of breast tissue using mammographic images

Indrajeet Kumar*

Graphic Era hill University,
Dehradun, India
Fax: 01368228062
Email: erindrajeet@gmail.com
*Corresponding author

Jitendra Virmani

Council of Scientific & Industrial Research –
Central Scientific Instruments Organization (CSIR-CSIO),
Ministry of Science & Technology,
Government of India,
Sector 30-C, Chandigarh-160030, India
Email: jitendra.virmani@gmail.com

H.S. Bhadauria

Department of Computer Science and Engineering,
GB Institute of Engineering and College,
Pauri Garhwal, Uttarakhand, 246164, India
Email: hsb76iitr@gmail.com

Abstract: In this study various experiments have been conducted for four-class and two-class breast tissue pattern characterisation using various texture feature models in transform domain. These experiments have been conducted on 480 mammographic images taken from the DDSM dataset. From each image, ROI of predefined size 128 × 128 pixels has been extracted from the core location of the breast tissue where remarkable amount of glandular tissues are found. Various statistical parameters have been computed from each ROI using transform domain texture feature models. The computed texture feature vectors are passed to the classification phase. The classification phase is consisted of SVM classifier. From the exhaustive experiments conducted in this study, it can be concluded that the maximum accuracy of 86.2% has been obtained for four-class breast tissue characterisation and the maximum accuracy of 91.2% has been obtained for two-class breast tissue characterisation.

Keywords: mammography; breast tissue characterisation; BIRADS classification; texture feature models; support vector machine classifier.

Reference to this paper should be made as follows: Kumar, I., Virmani, J. and Bhadauria, H.S. (2019) 'A computerised framework for characterisation of breast tissue using mammographic images', *Int. J. Computational Systems Engineering*, Vol. 5, No. 4, pp.193–210.

Biographical notes: Indrajeet Kumar received his BTech in Computer Science and Engineering from the Sant Longowal Institute of Engineering and Technology, Punjab in 2009 and MTech in Computer Engineering with specialisation in Networking from the YMCA University of Science and Technology, Faridabad, Haryana, India in 2013. Presently, he is pursuing his PhD on Classification of Breast Density from mammographic images from the GB Institute of Engineering and College, Pauri Garhwal, Uttarakhand, India. His research interests include application of machine learning and soft computing techniques for analysis of medical images.

Jitendra Virmani received his BTech (Hons.) in Instrumentation Engineering from the Sant Longowal Institute of Engineering and Technology, Punjab in 1999 and MTech in Electrical Engineering with specialisation in Measurement and Instrumentation Engineering from the Indian Institute of Technology, Roorkee in 2006. He received his PhD on Analysis and Classification of B-Mode Liver Ultrasound Images from the Biomedical Signal and Image Processing Laboratory, Indian Institute of Technology – Roorkee in 2014. He served in academia in various reputed organisations like Jaypee University of Information Technology – Solan, H.P,

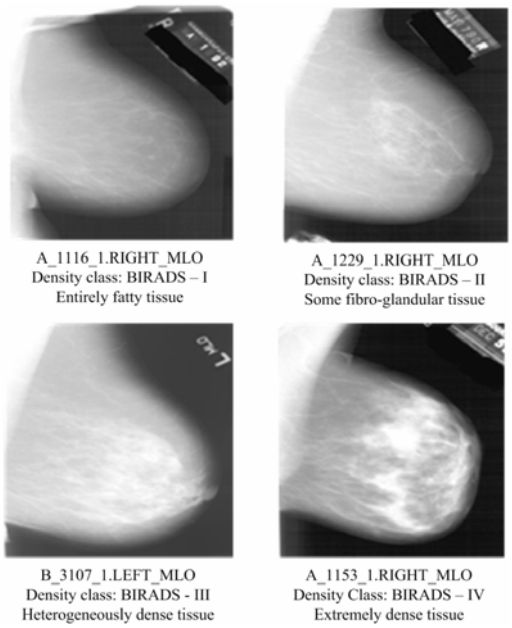
India and Thapar University – Patiala, Punjab, India for 13 years before joining the CSIR-CSIO, Chandigarh, India during August 2016. He is a Life Member of Institute of Engineers (IEI), India. His research interests include application of machine learning and soft computing techniques for analysis of medical images.

H.S. Bhadauria received his BTech in Computer Science and Engineering from the Aligarh Muslim University, Aligarh in 1999, and M.Tech. in Electronics Engineering from the Aligarh Muslim University, Aligarh in 2004. He received his PhD on Detection and Segmentation of Brain Hemorrhage using CT images from the Biomedical Signal and Image Processing Laboratory, Indian Institute of Technology – Roorkee in 2013. During his PhD, he worked on enhancing the detection and segmentation of brain hemorrhage using CT imaging modality. He served in academia for 12 years. He is presently serving as an Assistant Professor at G B Institute of Engineering and College, Pauri Garhwal, Uttarakhand, India. He is a Life Member of Institute of Engineers (IEI), India. He has published more than 60 research papers in international and national journals and conferences. His areas of interest are digital image and digital signal processing.

1 Introduction

Mammography is frequently used modality for the prediction of breast lesions. Among the various breast lesions, breast tumour is the major cause for death of women fertility in developed country as well as in developing country across the world (Ferlay et al., 2013; Parkin and Fernandez, 2006). Breast tissue density is considered as a well-built and an important indicator for the development of breast cancer (Wolfe, 1976, 1977; Vachon et al., 2007; Boyd et al., 2011; Colin et al., 2013; Eng and Gallant, 2014). Fundamentally, variations in breast density patterns are reflected by variations in intensity values, therefore the problem of breast tissue pattern characterisation can be treated as the problem of textural representation, description and analysis.

Figure 1 Sample image of different BIRADS breast tissue density classes



The breast tissue pattern characterisation is important for daily clinical environment because in many cases lesions are shrouded at the back the dense tissue so it is not visible and also same facts have been observed by the radiologist. Thus it is suggested that if the detection of testing case is dense (BIRADS-III or BIRADS-IV) then such type of suspicious cases have to be passed through double screened for the detection of lesions.

The BIRADS breast tissue density definitions (Huo et al., 2001; Yaghjyan et al., 2011; Papaevangelou et al., 2011; Al Mousa et al., 2014; Kumar et al., 2015a) and the sample of mammographic images of each class, taken from the digital database for screening mammography (DDSM) is given in Figure 1.

Studies in literature indicate that there has been remarkable curiosity amongst the research community to design computer assisted characterisation system for prediction of breast tissue pattern distribution. These computer aided characterisation systems have been designed by using

a segmented tissue-based approaches (STBAs)

b fixed size ROI-based approaches (RBAs).

It is worth mentioning that more studies have been carried out on computer aided characterisation system designs using STBAs and the related studies on computer aided characterisation system designs based on RBAs are few. It may be noted that computer aided characterisation system designs using STBAs require automatic segmentation of breast tissue which involves extra steps so STBAs are time overwhelming and difficult with respect to the RBAs.

After the literature study of studies carried out in past, the development of breast tissue pattern characterisation (four-class and two-class) systems have been summarised as given in Figure 2.

Table 1(a) Related studies for breast tissue pattern characterisation using STBA

Used dataset	*Author(s), year*	*No. of images*	*Classifier used*	*Considered class*	*Individual class accuracy (%)*	
DDSM	Bovis and Singh (2002)	377	ANN	Four-class	B1	(--)
					B2	(--)
					B3	(--)
					B4	(--)
			ANN	Two-class	Fatty	(--)
					Dense	(--)
	Oliver et al. (2005)	300	kNN	Four-class	B1	(46.0)
					B2	(48.0)
					B3	(50.0)
					B4	(36.0)
	Bosch et al. (2006)	500	SVM	Four-class	B1	(--)
					B2	(--)
					B3	(--)
					B4	(--)
	Oliver et al. (2008)	831	SFS-kNN	Four-class	B1	(54.7)
					B2	(87.7)
					B3	(76.8)
					B4	(69.4)
MIAS	Bosch et al. (2006)	322	SVM	Four-class	B1	(96.0)
					B2	(93.5)
					B3	(94.2)
					B4	(97.7)
	Oliver et al. (2008)	322	Bayesian	Four-class	B1	(91.0)
					B2	(84.0)
					B3	(89.0)
					B4	(73.0)
	Qu et al. (2011)	322	FELM#	Four-class	B1	(85.0)
					B2	(68.0)
					B3	(74.7)
					B4	(51.3)
	Chen et al. (2011)	301	kNN	Four-class	B1	(89.0)
					B2	(87.0)
					B3	(96.0)
					B4	(100)
			kNN	Two-class	Fatty	(88.0)
					Dense	(98.0)
Self collected datasets	Miller and Astley (1991)	40	Bayesian	Four-class	B1	(--)
					B2	(--)
					B3	(--)
					B4	(--)
	Karssemeijer (1998)	615	kNN	Four-class	B1	(76.1)
					B2	(90.1)
					B3	(74.4)
					B4	(60.0)

Note: FELM#: fuzzy extreme learning machine, B1: BIRADS-I; B2: BIRADS-II; B3: BIRADS-III; B4: BIRADS-IV.

Table 1(a) Related studies for breast tissue pattern characterisation using STBA (continued)

Used dataset	*Author(s), year*	*No. of images*	*Classifier used*	*Considered class*	*Individual class accuracy (%)*	
Self collected datasets	Petroudi et al. (2003)	132	kNN	Four-class	B1	(91.0)
					B2	(64.0)
					B3	(70.0)
					B4	(78.0)
				Two-class	Fatty	(91.0)
					Dense	(94.0)
	Jamal et al (2007)	100	---	Four-class	B1	(--)
					B2	(--)
					B3	(--)
					B4	(--)
	Masmoudi et al. (2013)	2,052	kNN	Four-class	B1	(72.0)
					B2	(65.0)
					B3	(63.0)
					B4	(78.0)
	He et al. (2016)	360	---	Four-class	B1	(69.3)
					B2	(90.5)
					B3	(76.0)
					B4	(45.4)

Note: FELM#: fuzzy extreme learning machine, B1: BIRADS-I; B2: BIRADS-II; B3: BIRADS-III; B4: BIRADS-IV.

Table 1(b) Related studies for breast tissue pattern characterisation using RBA

Used dataset	*Author(s), year*	*No. of images*	*Classifier used*	*Considered class*	*Individual class accuracy (%)*	
DDSM	Kumar et al. (2015a)	480	SVM	Four-class	B1	(75.0)
					B2	(68.3)
					B3	(55.0)
					B4	(96.6)
	Kumar et al. (2017a)	480	HHC*	Four-class	B1	(--)
					B2	(--)
					B3	(--)
					B4	(--)
	Kumar et al. (2017b)	480	ENN**	Four-class	B1	(98.3)
					B2	(91.6)
					B3	(80.0)
					B4	(93.3)
MIAS	Sharma and Singh (2015)	212	kNN	Two-class	Fatty	(95.1)
					Dense	(100)
	Sharma and Singh (2014)	322	kNN	Two-class	Fatty	(94.4)
					Dense	(100)
	Virmani et al. (2015)	322	kNN	Two-class	Fatty	(--)
					Dense	(--)
	Kriti et al. (2016a)	322	PCA-PNN	Two-class	Fatty	(84.9)
					Dense	(99.0)
	Kriti (2015a)	322	kNN	Two-class	Fatty	(86.7)
					Dense	(100)

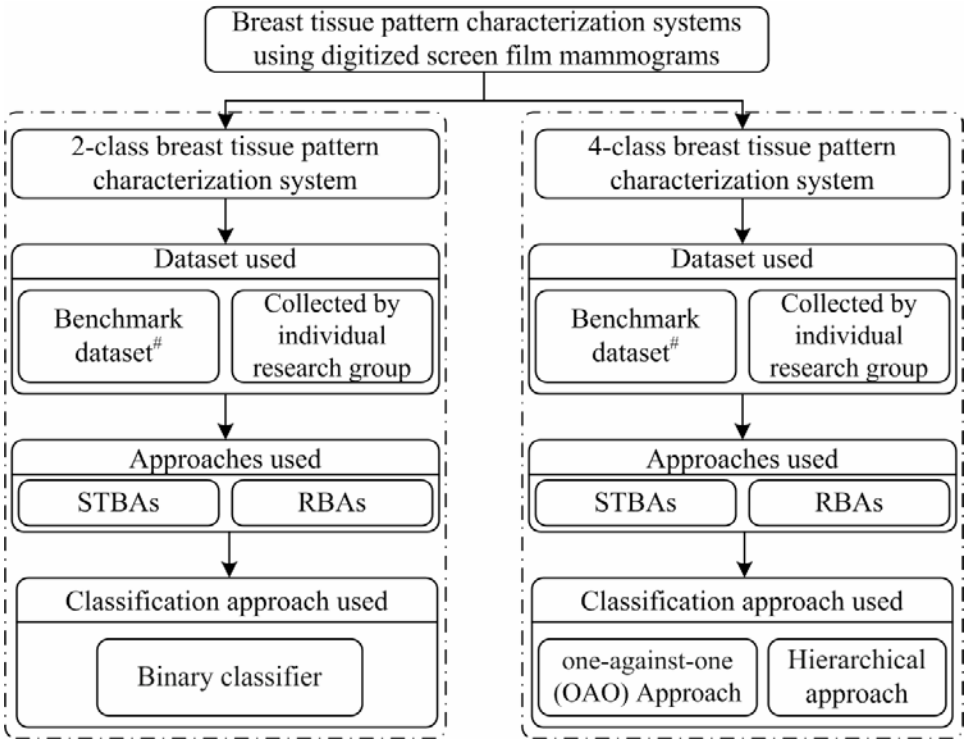
Note: HHC*: hybrid hierarchical classifier; ENN**: ensemble of neural network.

Table 1(b) Related studies for breast tissue pattern characterisation using RBA (continued)

Used dataset	*Author(s), year*	*No. of images*	*Classifier used*	*Considered class*	*Individual class accuracy (%)*	
MIAS	Kriti et al. (2016b)	322	PCA-SSVM	Two-class	Fatty	(88.6)
					Dense	(97.2)
	Mustra et al. (2012)	322	kNN	Four-class	B1	(--)
					B2	(--)
					B3	(--)
					B4	(--)
			kNN	Two-class	Fatty	(--)
					Dense	(--)
Self collected datasets	Liu et al. (2011)	88	SVM	Four-class	B1	(100)
					B2	(82.4)
					B3	(75.0)
					B4	(100)
	Mustra et al. (2012)	144	SFS-kNN	Four-class	B1	(--)
					B2	(--)
					B3	(--)
					B4	(--)
			ANN	Two-class	Fatty	(--)
					Dense	(--)

Note: HHC*: hybrid hierarchical classifier; ENN**: ensemble of neural network.

Figure 2 Development of breast tissue pattern characterisation (four-class and two-class) systems



The brief details of the related studies conducted in past for breast tissue pattern characterisation using STBA and RBA are given in Table 1(a) and Table 1(b), respectively.

From Table 1(a), it has been observed that in study (Oliver et al., 2008) the authors have attempted four-class breast tissue characterisation using STBA on 831 images of DDSM dataset. In this study author perform the feature

selection SFS and selected features are inputted to kNN classifier and the accuracy of 54.7% for B1, 87.7% for B2, 76.8% for B3 and 69.4% for B4 with overall classification accuracy of (54.7 + 87.7 + 76.8 + 69.4) / 4 = 72.1% is obtained.

It can be also noticed that the classification accuracy of 95.3% ((96.0 + 93.5 + 94.2 + 97.7) / 4)) has been reported in (Bosch, et al., 2006) on 322 mammographic images of MIAS dataset using STBA with 96.0% for B1, 93.5% for B2, 94.2% for B3 and 97.7% B4 respectively. It is worth to mention that the allots of research work has been carried out on self collected mammographic images and obtained accuracy of 75.7% in study (Petroudi et al., 2003) along with 91.0% accuracy for B1, 64.0 accuracy for B2, 70.0% accuracy for B3 and 78.0% for B4 class respectively.

From Table 1(b) following key points are drawn

a It has been observed that in study (Kumar et al., 2017b) the authors had proposed four-class breast tissue pattern characterisation using RBA on 480 images of DDSM dataset. In this study author extracted the $GLCM_{mean}$ features. The assembly of neural network classifiers paradigm is used for the classification purpose. The obtained highest classification accuracy of 90.8% is achieved with 98.3% for B1, 91.6% for B2, 80.0% for B3 and 93.3% for B4 respectively.

b It is worth to mention that few study (Mustra et al., 2012) proposed a BIRADS breast tissue classification as well as two-class breast tissue characterisation on MIAS using RBAs. The obtained maximum accuracy for two-class characterisation using RBA is 97.5% reported in study (Sharma et al., 2014).

c Allots of computer assisted BIRADS breast tissue characterisation system has been designed on self collected mammographic images and the accuracy of 89.3% reported in study (Liu et al., 2011) is maximum obtained accuracy.

Keeping in view, the brief summary of the literature it has been concluded that extensive studies performed for the characterisation of breast tissue by different research community using texture analysis in spatial domain. In recent research region transform domain-based feature analysis is playing very important role to develop a computer assisted characterisation system. So in this study BIRADS breast tissue characterisation has been performed using transform domain based statistical analysis.

The present study reports the breast tissue characterisation using RBA-based approach. ROI of fixed sizes 128 × 128 pixels have been manually extracted from the core location of the breast tissue where remarkable amount of glandular tissues are found which the significant information regarding the tissue characterisation. The various texture features models in transform domain

a Fourier power spectrum (FPS)

b 2D discrete wavelet transform (2D DWT)

c 2D Gabor wavelet transform (2D GWT).

The extracted feature set is inputted to classification module.

2 Materials and methods

2.1 Description of image dataset

In this work, 480 mammographic images consist of 120 ∈ B1, 120 ∈ B2, 120 ∈ B3 and 120 ∈ B4 were taken from DDSM dataset. The DDSM dataset consists of a total of 2620 cases having four image of each case, i.e., 2 MLO and 2 CLO views image. The more details about the DDSM dataset were also found in study (Heath et al., 1998). For each class B1 to B4, a total of 120 mammograms are taken and a total of 480 ROIs are cropped from these 480 mammograms. In the similar manner for two-class classification, a total of 240 mammograms, i.e., (120 ∈ B1, 120 ∈ B2) are combined to form 'fatty' image class and a total of 240 mammograms, i.e., (120 ∈ B3, 120 ∈ B4) are combined to form 'dense' image class. The depiction of used dataset for this work is shown in Figure 3.

2.2 Computerised frameworks for characterisation of breast tissue using mammographic images

In this study a computerised frameworks for

a Four-class breast tissue characterisation using mammographic images (experiment 1).

b Two-class (fatty, dense) breast tissue characterisation using mammographic images (experiment 2) have been designed.

The proposed computerised framework for characterisation of breast tissue using mammographic images is shown in Figure 4.

The proposed computerised framework for characterisation of breast tissue using mammographic images is consisting of

1 ROI extraction phase

2 feature extraction phase

3 classification phase.

The brief description of each section is given here one by one.

Figure 3 Image dataset description, (a) four-class breast tissue characterisation (b) two-class breast tissue characterisation

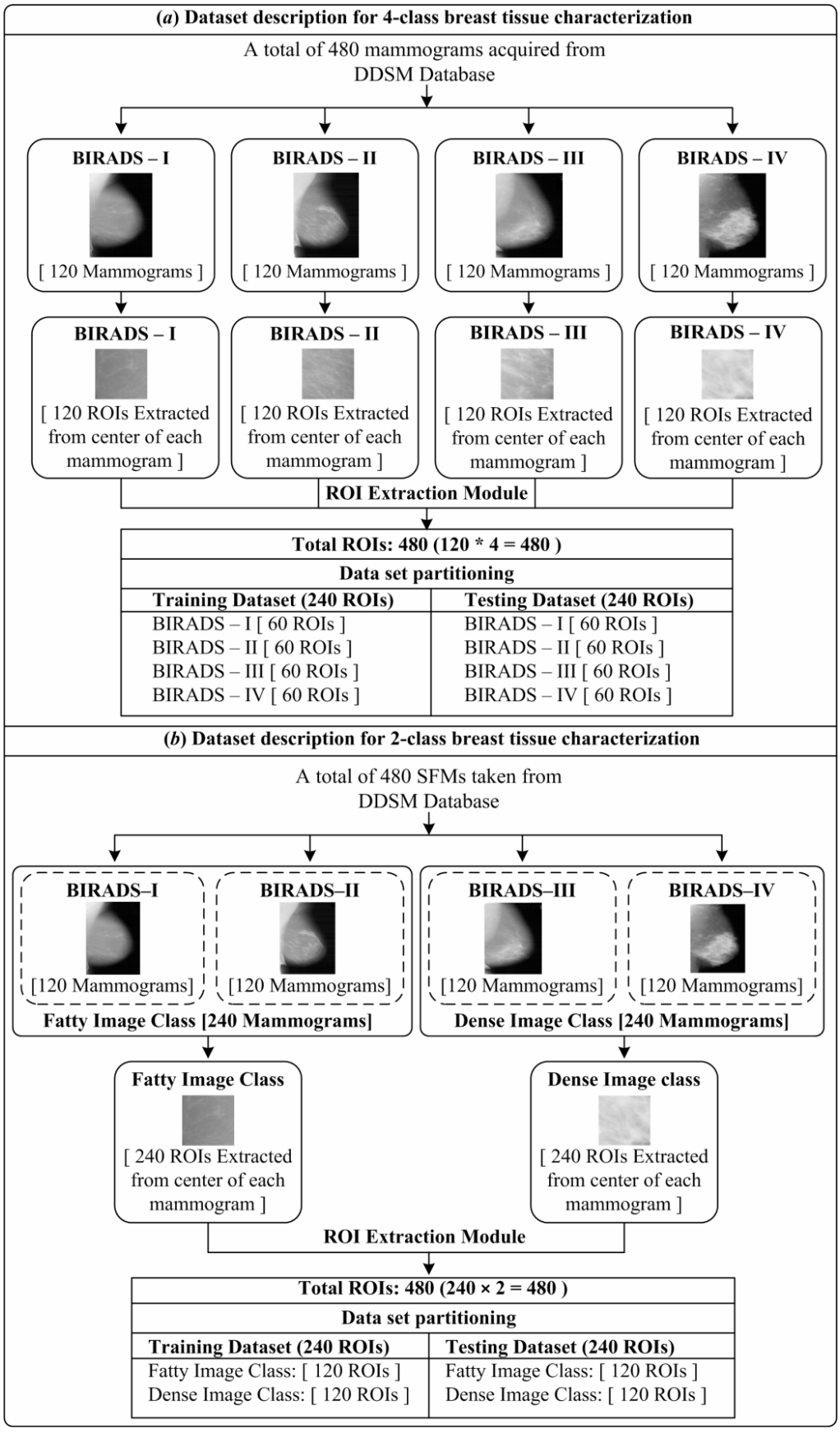
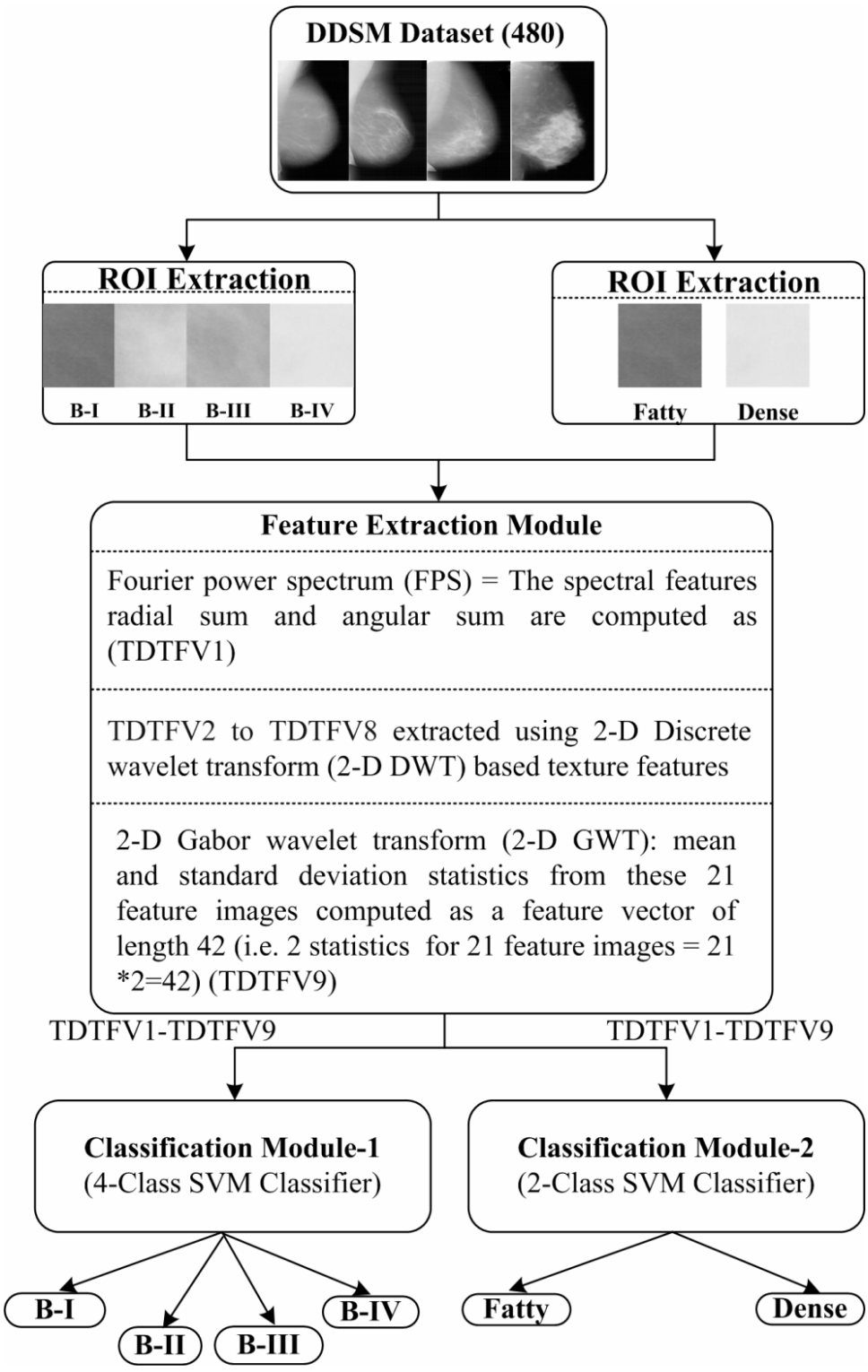


Figure 4 Proposed computerised framework for characterisation of breast tissue using mammographic images

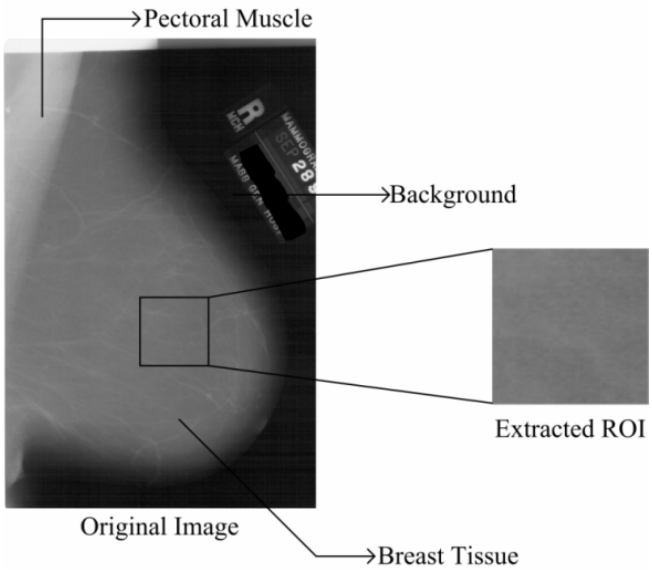


2.2.1 ROI extraction phase

The visual appearance in changes in textural behaviours reflected by the core location of the mammograms are accounts the sufficient information for characterisation between different breast tissue pattern. The similar observations have been also observed in the study carried by (Li et al., 2004) after extensive experimentations. Also the participating radiologist suggested that the participating radiologist visualise the core area of mammograms plays the major contribution for the discriminating between different breast density/tissue pattern classes. Accordingly, in the present work, ROIs of fixed size, i.e., 128 ×128 pixels have been are extracted from the central location of the breast (behind the nipple) from each mammogram. The

sample images of BIRADS class mammograms with marked ROI are shown in Figure 5.

Figure 5 Sample of BIRADS class mammograms with marked ROI



2.2.2 Feature extraction phase

Recently, texture descriptor computed in transform domain using various multi-resolution schemes such as FPS, 2D DWT) and 2D GWT gains the attention for discrimination between different breast tissues (Virmani et al., 2014, 2015; Sharma et al., 2001; Lee et al., 2006; Kumar et al., 2015a). The brief description of various transform domain texture feature models used in the present work is given here.

FPS-based texture feature model

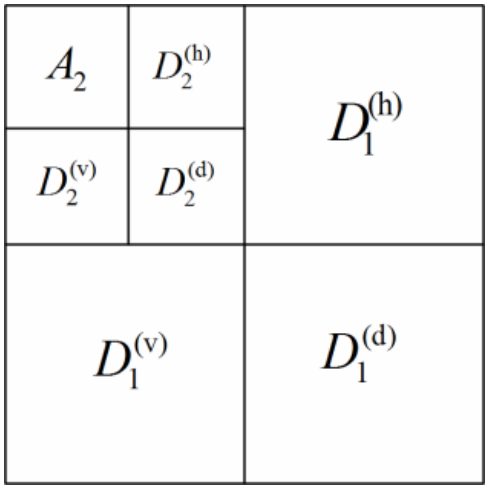
The TDTFV1 (as shown in first row of Table 2, computed using FPS-based texture feature model) consists of spectral features, i.e., radial sum and angular sum (Virmani et al., 2014).

2D Discrete wavelet transform-based texture feature model

In the present work each ROI image has been decomposed up to second level by using 2D DWT multi-resolution scheme with different compact supporting filters. After the exhaustive experiments it has been found that the compact support filter 'haar' yields the maximum discrimination information between different density patterns. Thus 'haar' wavelet filter is considered for this study. After decomposition seven subband images, i.e., one approximate subband image and six orientation selective detail subband images are generated for each ROI image as shown in Figure 6. Normalised energy values are computed from each of these subband images.

The brief description of different TFVs (TDTFV2 to TDTFV8) extracted using 2D discrete wavelet transform-based texture feature model is reported in Table 2.

Figure 6 2D DWT tree up to second level decomposition



Notes: A_2: approximate subband feature image at second level of decomposition; $D_2^{(h)}$: detail coefficient of horizontal subband feature image at second level of decomposition; $D_2^{(d)}$: detail coefficient of diagonal subband feature image at second level of decomposition; $D_2^{(v)}$: detail coefficient of vertical subband feature image at second level of decomposition; $D_1^{(h)}$: detail coefficient of horizontal subband feature image at second level of decomposition; $D_1^{(d)}$: detail coefficient of diagonal subband feature image at first level of decomposition; $D_1^{(v)}$: detail coefficient of vertical subband feature image at first level of decomposition; A_i or D_i: i^{th} level of decomposition.

2D GWT-based texture feature model

In the present work, 2D Gabor wavelet multi-scale decompositions have been performed using magnitude values (0, 1 and 2) and seven orientations (22.5°, 45°, 67.5°, 90°, 112.5°, 135° and 157.5°) resulting in a group of 21 (7 × 3) wavelets of Gabor filter bank. It is worth mentioning that these 21 wavelets capture energy at a particular scale and orientations. The real parts of Gabor filter bank is shown in Figure 7.

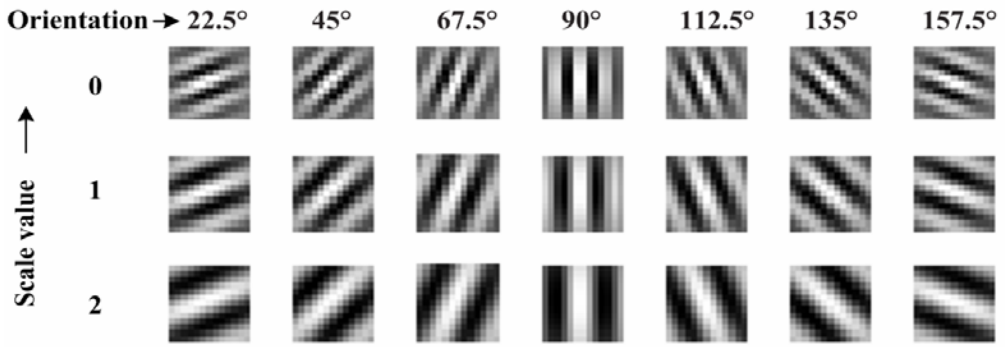
A group of total 21 Gabor filters is convolved with an input ROI image to getting filtered images called features images. Each filtered image depicts the texture information at a certain scale and orientation value. After that compute two statistics (mean, standard deviation) from obtained 21 Gabor outputs or filtered images that provides a length of feature vector 42 (i.e., 2 statistics for 21 feature images). The brief description of different TFV (i.e., TDTFV9) extracted using 2D Gabor wavelet transform-based texture feature models used in this work is reported in Table 2.

Table 2 Different TFVs extracted using transform domain texture feature models

TFM	*TFV*	*Extracted features*	*l*
FPS	TDTFV1	Radial sum, angular sum.	2
2-D DWT	TDTFV2	$\frac{\|\| A_2 \|\|_F^2}{area(A_2)}, \frac{\|\| D_2^{(h)} \|\|_F^2}{area(D_2^{(h)})}, \frac{\|\| D_2^{(v)} \|\|_F^2}{area(D_2^{(v)})}, \frac{\|\| D_2^{(d)} \|\|_F^2}{area(D_2^{(d)})}, \frac{\|\| D_1^{(h)} \|\|_F^2}{area(D_1^{(h)})}, \frac{\|\| D_1^{(v)} \|\|_F^2}{area(D_1^{(v)})}, \frac{\|\| D_1^{(d)} \|\|_F^2}{area(D_1^{(d)})}$	7
	TDTFV3	$\frac{\|\| D_1^{(h)} \|\|_F^2}{area(D_1^{(h)})}, \frac{\|\| D_1^{(v)} \|\|_F^2}{area(D_1^{(v)})}, \frac{\|\| D_1^{(d)} \|\|_F^2}{area(D_1^{(d)})}$	3
	TDTFV4	$\frac{\|\| D_1^{(h)} \|\|_F^2}{area(D_1^{(h)})}, \frac{\|\| D_1^{(v)} \|\|_F^2}{area(D_1^{(v)})}, \frac{\|\| D_1^{(d)} \|\|_F^2}{area(D_1^{(d)})}, \frac{\|\| D_2^{(h)} \|\|_F^2}{area(D_2^{(h)})}, \frac{\|\| D_2^{(v)} \|\|_F^2}{area(D_2^{(v)})}, \frac{\|\| D_2^{(d)} \|\|_F^2}{area(D_2^{(d)})}$	6
	TDTFV5	$\frac{\|\| D_1^{(h)} \|\|_F^2}{area(D_1^{(h)})}, \frac{\|\| D_1^{(v)} \|\|_F^2}{area(D_1^{(v)})}, \frac{\|\| D_1^{(d)} \|\|_F^2}{area(D_1^{(d)})}, \frac{\|\| D_2^{(d)} \|\|_F^2}{area(D_2^{(d)})}$	4
	TDTFV6	$\frac{\|\| A_2 \|\|_F^2}{area(A_2)}, \frac{\|\| D_1^{(h)} \|\|_F^2}{area(D_1^{(h)})}, \frac{\|\| D_1^{(v)} \|\|_F^2}{area(D_1^{(v)})}, \frac{\|\| D_1^{(d)} \|\|_F^2}{area(D_1^{(d)})}$	4
	TDTFV7	$\frac{\|\| A_2 \|\|_F^2}{area(A_2)}, \frac{\|\| D_2^{(h)} \|\|_F^2}{area(D_2^{(h)})}, \frac{\|\| D_2^{(v)} \|\|_F^2}{area(D_2^{(v)})}, \frac{\|\| D_2^{(d)} \|\|_F^2}{area(D_2^{(d)})}$	4
	TDTFV8	$\frac{\|\| D_2^{(h)} \|\|_F^2}{area(D_2^{(h)})}, \frac{\|\| D_2^{(v)} \|\|_F^2}{area(D_2^{(v)})}, \frac{\|\| D_2^{(d)} \|\|_F^2}{area(D_2^{(d)})}$	3
2-D GWT	TDTFV9	mean and standard deviation are calculated for seven orientation 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5° and three magnitude value 0, 1 and 2 ($3 \times 7 \times 2 = 42$).	42

Notes: TFM: texture feature model; TDTFV: transform domain texture feature vector; l: length of TFV; A: approximate subband image; D: detail subband image; h: horizontal, v: vertical; d: diagonal; A_i or D_i: i represent i^{th} level of decomposition; area: total number of pixels to the respective subband image.

Figure 7 Real parts of Gabor filter bank for seven orientations and three magnitude values



2.2.3 Classification phase

In this study, classification module is consisting of two different modules as

a four-class breast tissue pattern characterisation module

b two-class breast tissue pattern characterisation module.

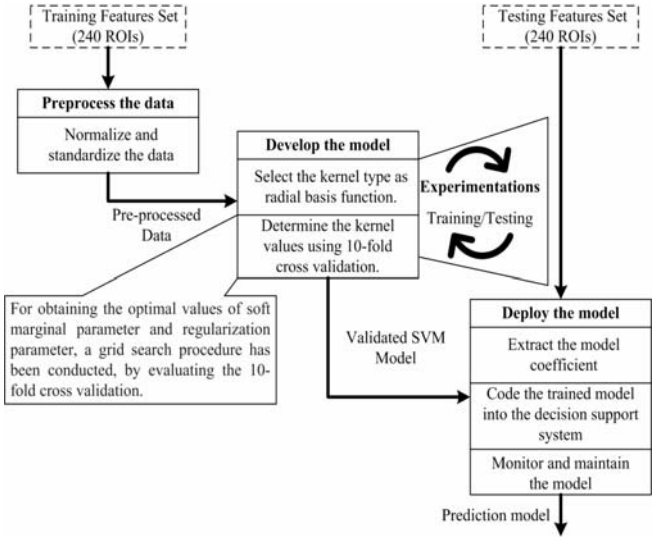
The brief description of each module is given here.

a Four-class breast tissue pattern characterisation module: This classification module uses one-against-one (OAO) approach [available in LibSVM library (Chang and Lin, 2012)] by constructing X(X−1)/2 binary sub-classifiers (here, X is the number of classes).

It is necessary to train each binary sub-classifier separately for the differentiation between a pair of classes, and majority voting technique is used for the making prediction of an input test ROI. In 4-class breast tissue pattern characterisation, the prediction of the class of the test ROI is made by widely used majority voting scheme on the predictions of six binary sub-classifiers as SVM (BIRADS-I versus BIRADS-II), SVM (BIRADS-I versus BIRADS-III), SVM (BIRADS-I versus BIRADS-IV), SVM (BIRADS-II versus BIRADS-III), SVM (BIRADS-II versus BIRADS-IV) and SVM (BIRADS-III versus BIRADS-IV).

b Two-class breast tissue pattern characterisation module: It has been implemented using binary SVM approach available in LibSVM library [67].

Figure 8 Development of SVM model



2.2.3.1 SVM classifier

In this study SVM classifier has been used for the development of computerised framework for characterisation of breast tissue (Azar and El-Said, 2014; Virmani et al., 2014; Kumar et al., 2015a; Hassanien et al., 2011;, Chang and Lin, 2012). The library functions for SVM are imported from LibSVM library. From the earlier study it has been observed that Gaussian radial basis kernel (RBF) function yields better performance for medical image processing. So that Gaussian RBF kernel-based SVM is investigated. The development of SVM model for this work is shown in Figure 8. For obtaining the optimal values of regularisation parameter C and kernel parameter γ a grid search procedure has been conducted, by evaluating the ten-fold cross validation training accuracy for each combination of (C, γ), such that, $C \in \{2^{-4}, 2^{-3}, \ldots, 2^{15}\}$ and $\gamma \in \{2^{-12}, 2^{-11}, \ldots, 2^{4}\}$.

2.3 Performance analysis

The performance of proposed computerised framework is expressed in overall characterisation accuracy (OC) and individual class characterisation accuracy (IC). The values of the OC and IC are computed using equation given in (1) and (2), respectively.

$$\text{OC} = \frac{\text{CPS}}{\text{TTS}} * 100 \tag{1}$$

where CPS is used to define the number of correctly predicted samples and TTS is used for total number of test samples.

$$\text{IC} = \frac{\text{ICPS}}{\text{NTIC}} * 100 \tag{2}$$

where ICPS is used to define the number of correctly predicted samples belonging to a particular class and NTIC is used for total number of test samples for a particular class.

The statistical analysis is computed by the most trustable Cohen's kappa method defined in study (Viera and Garrett, 2005). It is used for testing the stability of the developed computerised framework for breast tissue pattern characterisation. For the computation of Cohen's kappa coefficient (k) a confusion matrix as shown in Figure 9 is used.

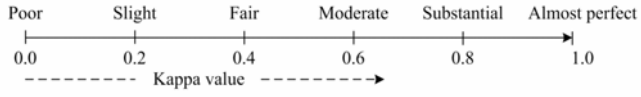
Figure 9 Sample representation of a confusion matrix

Confusion matrix

		Observer2	
		C1	C2
Observer1	C1	*N11*	*N12*
	C2	*N21*	*N22*

Note: C1: class1; C2: class2.

Figure 10 Importance of the obtained Cohen's kappa value



Kappa value	Significance
<0	Less significant
0.01-0.20	Slightly significant
0.21-0.40	Fairly significant
0.41-0.60	Moderately significant
0.61-0.80	Substantially significant
0.81-1.0	Almost perfect

The value of kappa coefficient k is computed by using equation (3).

$$k = \frac{p_o - p_e}{1 - p_e} \tag{3}$$

Here k is kappa coefficient, p_o is known as observed accuracy and p_e is known as expected accuracy. The value of observed accuracy p_o and expected p_e accuracy are computed using equation (4) and equation (5) respectively.

$$p_o = \frac{N11 + N22}{N11 + N12 + N21 + N22} \tag{4}$$

$$p_e = \frac{A + B}{N11 + N12 + N21 + N22} \tag{5}$$

where A and B is calculated using equation (6) and equation (7), respectively.

$$A: \text{class/observer } 1 = (N11 + N12) * (\text{N}11 + N21) \quad (6)$$

$$B: \text{class/observer } 2 = (N21 + N22) * (\text{N}12 + N22) \quad (7)$$

The interpretations of the obtained kappa coefficient value are shown in Figure 10.

3 Experiments and results

The experimental work flow for the design of computerised framework for characterisation of breast tissue using mammographic images is shown in Figure 11.

3.1 Experiment 1

Experiment carried out for the design of computerised framework for characterisation of four-class breast tissue pattern using mammographic images

In this experiment performance of various TDTFVs has been computed for four-class breast tissue pattern characterisation task using SVM1 classifier. The obtained performances are reported in Table 3.

From Table 3 it can be inferred that for four-class characterisation of breast tissue pattern, the highest OC achieved is 86.2% ((54 + 51 + 44 + 56) / 240) with TDTFV2 of length 7. The individual class classification accuracy for class B1, the IC value is 90.0% (54/60), for class B2, the IC value is 85.0% (51/60), for class B3, the IC value is 76.6% (44/60) and for class B4 the IC value is 93.3% (56/60).

From Table 3, it has been observed that the TDTFV2 yields maximum kappa coefficient value 0.8167 using SVM1 classifier. It demonstrates that the developed computerised framework for four-class breast tissue pattern characterisation using SVM1 for TDTFV2 is more reliable for clinical environment than systems designed using others TDTFVs. The results of the experiment demonstrate that the ROI size of 128 × 128 pixels extracted from the core region of the breast where glandular tissues are found that provides sufficient information for characterisation between different breast tissue classes.

3.2 Experiment 2

Experiment carried out for the design of computerised framework for characterisation of two-class breast tissue using mammographic images

In this experiment the performance of various TDTFVs has been computed for two-class breast tissue pattern characterisation task using SVM2 classifier. The obtained performances of the system are given in Table 4.

From Table 4, the performances of the examination carried out for this work inferred that for two-class breast tissue pattern characterisation, TDTFV2 of length 7 yields the OC value of 91.6%. The IC values of 92.5% (111/120) and 90.8% (109/120) have been observed for the 'fatty' and 'dense' classes using TDTFV2.

It can be also found in Table 4 that the TDTFV2 of length seven yields maximum kappa coefficient value of 0.8250 using SVM2 classifier. It confirms that the designed computerised framework for binary (two-class) classification task using SVM2 classifier for TDTFV2 is more significant for daily clinical environment than systems designed using other TDTFVs.

3.3 Statistical analysis

The statistical analysis of the proposed system is performed by using Cohen's kappa coefficient as per discussed in study (Viera and Garrett, 2005).

a For four-class breast tissue pattern characterisation

Observed accuracy (p_o) = 0.8625	Expected accuracy (p_e) = 0.2500
$p_o - p_e$ = 0.6125	$1 - p_e$ = 0.7500
Cohen's kappa value = 0.8167	kappa error = 0.0296

The Cohen's kappa value for the computerised framework for four-class breast tissue pattern characterisation is 0.8167. The obtained kappa value shows that the proposed system has considerable importance for the radiologists to classify between different BIRADS breast density classes.

b For two-class breast tissue pattern characterisation

Observed accuracy (p_o) = 0.9125	Expected accuracy (p_e) = 0.5000
$p_o - p_e$ = 0.4125	$1 - p_e$ = 0.5000
Cohen's kappa value = 0.8250	kappa error = 0.0365

The Cohen's kappa value for the computerised framework for two-class breast tissue characterisation is 0.8250. The obtained kappa value shows that the proposed system is acceptable and suitable for the radiologists to classify between fatty and dense breast density classes.

3.4 Comparative analysis

The comparative analysis of various experiments carried out in this study for breast tissue characterisation is reported in Table 5. The comparative study is analysed in terms of misclassified instances and performance of four-class breast tissue characterisation and two-class breast tissue characterisation.

From Table 5, it is observed that the designed computerised framework by experiment 2 is better that the proposed computerised framework by experiment 1 with respect to the number of TMIs, OC value as well as kappa value. Out of 240 testing instances only 20 instances are misclassified at experiment 2 while 33 instances are misclassified at experiment 1. It can also be noticed that the obtained OC value for experiment 1 is 86.25% and 91.6% for experiment 2. The Cohen's kappa value of experiment 2

is 0.8250 while 0.8167 is for experiment 1, which shows the classification framework designed using experiment 2, is more reliable and suitable for the real time breast tissue characterisation problem.

Table 3 Classification performance of various TDTFVs for four-class breast tissue pattern classification using SVM1 classifier

Classification performance of transform domain texture feature vectors (TDTFV1-TDTFV9) for different ROI sizes								
TDTFVs (l)	*Confusion matrix*					*IC (%)*	*OC (%)*	*Kappa value*
		B1	*B2*	*B3*	*B4*			
TDTFV1 (2)	B1	52	6	2	0	86.6	82.1	0.7611
	B2	7	50	3	0	83.3		
	B3	1	8	42	9	70.0		
	B4	0	2	5	53	88.3		
TDTFV2 (7)	B1	54	4	2	0	90.0	86.2	0.8167
	B2	3	51	5	1	85.0		
	B3	2	5	46	7	76.6		
	B4	0	1	3	56	93.3		
TDTFV3 (3)	B1	49	8	3	0	81.6	68.3	0.5778
	B2	14	37	6	3	61.6		
	B3	2	13	31	14	51.6		
	B4	0	2	11	47	78.3		
TDTFV4 (6)	B1	46	9	4	1	76.6	69.6	0.5944
	B2	7	37	11	5	61.6		
	B3	2	15	33	10	55.0		
	B4	1	1	7	51	85.0		
TDTFV5 (4)	B1	48	7	4	1	80.0	72.9	0.6389
	B2	8	41	6	5	68.3		
	B3	2	12	36	10	60.0		
	B4	0	1	9	50	83.3		
TDTFV6 (4)	B1	50	7	2	1	83.3	77.1	0.6944
	B2	5	47	8	0	78.3		
	B3	2	8	38	12	63.3		
	B4	0	2	8	50	83.3		
TDTFV7 (4)	B1	48	10	2	0	80.0	75.8	0.6778
	B2	7	46	6	1	76.6		
	B3	2	11	37	10	61.6		
	B4	0	0	9	51	85.0		
TDTFV8 (2)	B1	51	7	2	0	85.0	77.1	0.6944
	B2	6	44	7	3	73.3		
	B3	2	9	39	10	65.0		
	B4	0	2	7	51	85.0		
TDTFV9 (42)	B1	54	4	1	1	60.0	79.2	0.7222
	B2	6	46	7	1	76.6		
	B3	4	8	38	10	63.3		
	B4	0	2	6	52	83.3		

Note: *l*: length of TFV; TDTFV: transform domain texture feature vector.

Figure 11 Experimental work flow for the design of computerised framework for characterisation of breast tissue using mammographic images

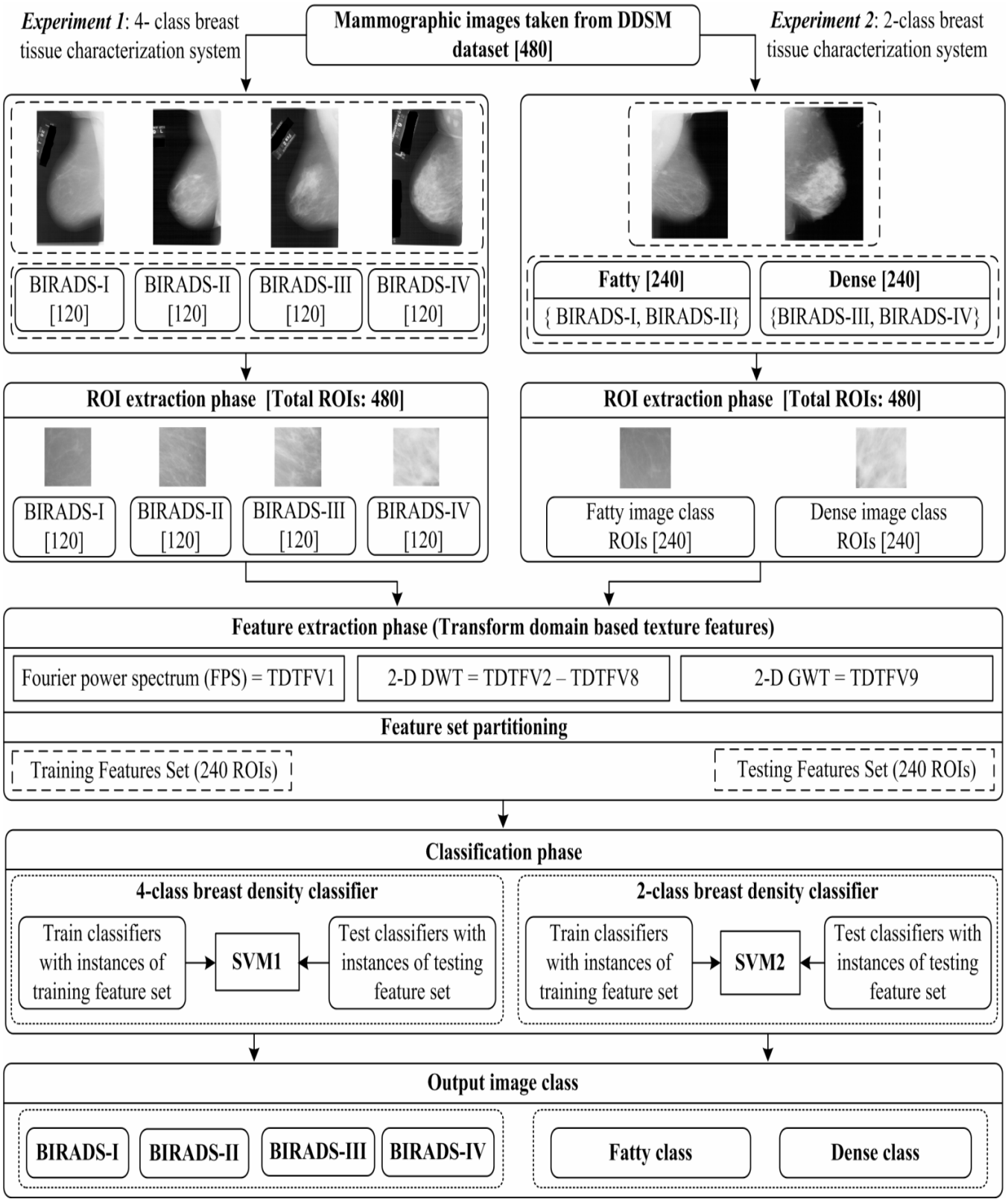


Figure 12 Comparison between earlier studies for four-class and two-class breast tissue pattern characterisation using RBA and proposed work for breast tissue pattern characterisation

Comparison between earlier study carried out by Kumar, et al. 2017b and proposed work for 4-class breast tissue density characterization using RBA				
Author(s), Year	**No. of images**	**Classifier used**	**Individual class accuracy (%)**	
Kumar, et al. 2017b	480	Ensemble of neural networks	B1	98.3
			B2	91.6
			B3	80.0
			B4	93.3
Proposed	480	Support vector machine (SVM)	B1	90.0
			B2	85.0
			B3	76.6
			B4	93.3
Comparison between earlier study carried out by Sharma, et al. 2015 and proposed work for 2-class breast tissue density characterization using RBA				
Author(s), Year	**No. of images**	**Classifier used**	**Individual class accuracy (%)**	
Sharma, et al. 2015	212 (from MIAS)	*k*-nearest neighbor classifier	Fatty	98.3
			Dense	91.6
Proposed	480 (from DDSM)	Support vector machine (SVM)	Fatty	92.5
			Dense	90.8

Table 4 Classification performance yielded by SPTFV1 for two-class breast tissue pattern classification.

Classification performance of transform domain texture feature vectors (TDTFV1-TDTFV9) using SVM classifier						
TDTFV (l)		*CM*		*Accuracy (%)*		*Kappa value*
		F	*D*	*IC*	*OC*	
TDTFV1 (2)	F	102	18	85.0	82.9	0.6583
	D	23	97	80.8		
TDTFV2 (7)	F	111	9	92.5	91.6	0.8250
	D	11	109	90.8		
TDTFV3 (3)	F	109	11	90.8	89.6	0.7917
	D	14	106	88.3		
TDTFV4 (6)	F	102	18	85.0	82.1	0.6417
	D	25	95	79.1		
TDTFV5 (4)	F	101	19	84.1	83.7	0.6750
	D	20	100	83.3		
TDTFV6 (4)	F	108	12	90.0	88.3	0.7667
	D	16	104	86.6		
TDTFV7 (4)	F	104	16	86.6	87.1	0.7417
	D	15	105	92.1		
TDTFV8 (2)	F	103	17	85.8	83.7	0.6750
	D	22	98	81.6		
TDTFV9 (42)	F	103	17	85.8	87.1	0.7417
	D	14	106	88.3		

Table 5 Comparative analysis of various experiments carried out in this study for breast tissue characterisation

	Considered class	*No. of TIs*	*No. of MIs*	*IC (%)*	*TMIs*	*OC (%)*	*Kappa value*
Experiment 1	BIRADS-I	60	6	90.0	33	86.2	0.8167
	BIRADS-II	60	9	85.0			
	BIRADS-III	60	14	76.6			
	BIRADS-IV	60	4	93.3			
Experiment 2	Fatty	120	9	92.5	20	91.6	0.8250
	Dense	120	11	90.8			

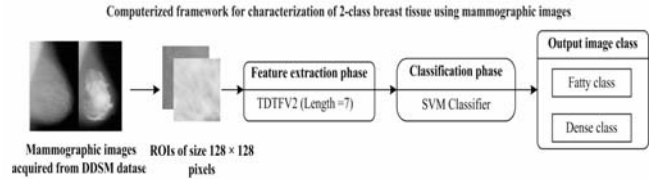
Notes: TIs: testing instances; MIs: misclassified instances; TMIs: total misclassified instances. [A total of 20 (20/240) testing samples are misclassified by experiment 2].

The comparison between earlier studies for four-class and two-class breast tissue pattern characterisation using RBA and proposed work for breast tissue pattern characterisation is shown in Figure 12.

4 Conclusions

From the extensive experimentation carried out in this work it can be concluded that ROI size of 128 × 128 pixels using texture models in transform domain for four-class breast tissue characterisation as well as for two-class breast tissue characterisation accounts significant information to design computerised framework for breast tissue characterisation. The maximum accuracy of 86.2% has been achieved for four-class breast tissue characterisation and the maximum accuracy of 91.6% has been achieved for two-class breast tissue characterisation using TDTFV2 of length seven with SVM classifier.

Figure 13 Proposed computerised framework for characterisation of two-class breast tissue using mammographic images



It may be also noticed that higher accuracy of 91.6% is achieved by two-class breast tissue characterisation module in comparison to 86.2% as achieved by four-class breast tissue characterisation framework. Therefore, the present study recommends the use of the proposed computerised framework for breast tissue pattern characterisation as shown in Figure 13.

The promising results obtained by the current study indicate the usefulness of the proposed system for characterisation of breast tissue pattern using mammographic images during real time clinical practice.

References

Al Mousa, D.S., Brennan, P.C., Ryan, E.A., Lee, W.B., Tan, J. and Mello-Thomas, C. (2014) 'How mammographic breast density affects radiologists' visual search patterns', *Acad. Radiol*, Vol. 21, No. 11, pp.1386–1393.

Azar, A.T. and El-Said, S.A. (2014) 'Performance analysis of support vector machine classifiers in breast cancer mammography recognition', *Neural Comput Appl*., Vol. 24, No. 5, pp.1163–1177.

Bosch, A., Munoz, X., Oliver, A. and Marti, J. (2006) 'Modeling and classifying breast tissue density in mammograms', in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 'CVPR'06*, New York, Vol. 2, pp.1552–1558.

Bovis, K. and Singh, S. (2002) 'Classification of mammographic breast density using a combined classifier paradigm', in *Proceeding of Medical Image Understanding and Analysis 'MIUA' Conference*, Portsmouth, pp.177–180.

Boyd, N.F., Martin, L.J., Yaffe, M.J. and Minkin, S. (2011) 'Mammographic density and breast cancer risk: current understanding and future prospects', *Breast Cancer Res Tr*., Vol. 13, No. 6, pp.223–234.

Chang, C.C. and Lin, C.J. (2012) *LIBSVM, a Library of Support Vector Machines* [online] http://www.csie.ntu.edu.tw /~cjlin/libsvm.5 (accessed July 2016).

Chen, Z., Denton, E. and Zwiggelaar, R. (2011) 'Local feature-based mammographic tissue pattern modeling and breast density classification', in Proceedings *of 4th International Conference on Biomedical engineering and Informatics*, Shanghai, pp.351–355.

Colin, C., Prince, V. and Valette, P.J. (2013) 'Can mammographic assessments lead to consider density as a risk factor for breast cancer', *Eur J Radiol*., Vol. 82, No. 3, pp.404–411.

Eng, A. and Gallant, Z. (2014) 'Digital mammographic density and breast cancer risk a case-control study of six alternative density assessment methods', *Breast Cancer Res Tr*., Vol. 16, No. 5, pp.439–451.

Ferlay, J., S-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W., Comber, H., Forman, D. and Bray, F. (2013) 'Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012', *Eur J Cancer*, Vol. 49, No. 6, pp.1374–1403.

Hassanien, A.E., Bendary, N.E., Kudelka, M. and Snasel, V. (2011) 'Breast cancer detection and classification using support vector machines and pulse coupled neural network', in *Proceedings of 3rd International Conference on Intelligent Human Computer Interaction 'IHCI 2011'*, pp.269–279.

He, W., Harvey, S., Juette, A., Denton, E.R. and Zwiggelaar, R. (2016) 'Mammographic segmentation and density classification: a fractal inspired approach', in *International Workshop on Digital Mammography*, pp.359–366.

Heath, M., Bowyer, K., Kopans, D., Moore, R. and Kegelmeyer, P.J. (1998) 'The digital database for screening mammography', in *Proc. Int. Workshop Digital Mammography*, pp.212–218.

Huo, Z., Giger, M.L. and Vyborny, C.J. (2001) 'Computerized analysis of multiple-mammographic views: potential usefulness of special view mammograms in computer-aided diagnosis', *IEEE T Med Imaging*, Vol. 20, No. 12, pp.1285–1292.

Jamal, N., Ng, K.H., Ranganathan, S. and Tan, L.K. (2007) 'Comparison of computerized assessment of breast density with subjective BI-RADS classification and Tabar's Pattern from two-view CR mammography', in *World Congress on Medical Physics and Biomedical Engineering 2006*, pp.1405–1408.

Karssemeijer, N. (1998) 'Automated classification of parenchymal patterns in mammograms', *Phys Med Biol*., Vol. 43, No. 2, pp.365–389.

Kriti, V.J. (2015a) 'Breast density classification using Laws' mask texture features', *International Journal of Biomedical Engineering and Technology*, Vol. 19, No. 3, pp.279–302.

Kriti, V.J. (2015b) 'Breast tissue density classification using wavelet-based texture descriptors', in *Proceedings of the Second International Conference on Computer and Communication Technologies, Advances in Intelligent Systems and Computing*, Vol. 381, pp.539–546.

Kriti, V.J. and Thakur, S. (2016) 'Application of statistical texture features for breast tissue density classification', *Image Feature Detectors and Descriptors, Studies in Computational Intelligence*, Vol. 630, pp.411–435.

Kriti, V.J., Dey, N. and Kumar, V. (2016) 'PCA-PNN and PCA-SVM-based CAD Systems for breast density classification', *Applications of Intelligent Optimization in Biology and Medicine, Intelligent Systems Reference Library*, Vol. 96, pp.159–180.

Kumar, I., Bhadauria, H.S. and Virmani, J. (2015a) 'Wavelet packet texture descriptors-based four-class BIRADS breast tissue density classification', *Procedia Computer Science*, Vol. 70, pp.76–84.

Kumar, I., Virmani, J. and Bhadauria, H.S. (2015b) 'A review of breast density classification methods', in *Proceeding of 2nd International Conference on Computing for Sustainable Global Development 'INDIACom – 2015'*, pp.1960–1967.

Kumar, I., Bhadauria, H.S. and Virmani, J. and Thakur, S. (2017a) 'A hybrid hierarchical framework for classification of breast density using digitized film screen mammograms', *Multimed Tools App*, pp.1–25.

Kumar, I., Bhadauria, H.S. and Virmani, J. Thakur, S. (2017b) 'A classification framework for the prediction of breast density using an ensemble of neural network classifiers', *Biocybernetics and Biomedical Engineering*, Vol. 37, No. 1, pp.217–228.

Lee, C. and Chen, S.H. (2006) 'Gabor wavelets and SVM classifier for liver diseases classification from CT images', in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp.548–552.

Li, H., Giger, M.L., Huo, Z., Olopade, O.I., Lan, L., Weber, W.L. and Bonta, I. (2004) 'Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location', *Med Phys*., Vol. 31, No. 3, pp.549–555.

Liu, Q., Liu, L., Tan, Y., Wang, J., Ma, X. and Ni, H. (2011) 'Mammogram density estimation using sub-region classification', in *Proceedings of 4th International Conference on Biomedical Engineering and Informatics 'BMEI'*, Shanghai, pp.356–359.

Masmoudi, A.D., Ayed, N.G.B., Masmoudi, D.S. and Abid, R. (2013) 'LBPV descriptors-based automatic ACR/BIRADS classification approach', *EURASIP Journal on Image and Video Processing*, Vol. 1, No. 1, pp.1–9.

Miller, P. and Astley, S. (1991) 'Classification of breast tissue by texture analysis', in *BMVC91*, pp.258–265.

Mustra, M., Grgic, M. and Delac, K. (2012) 'Breast density classification using multiple feature selection', *Automatika*, Vol. 53, No. 4, pp.362–372.

Oliver, A., Freixenet, J. and Zwiggelaar, R. (2005) 'Automatic classification of breast density', in *Proceedings of the IEEE International Conference on Image Processing*, Vol. 2, pp.1258–1261.

Oliver, A., Freixenet, J., Marti, R., Pont, J., Perez, E., Denton, E.R.E. and Zwiggelaar, R. (2008) 'A novel breast tissue density classification methodology', *IEEE T Inf Technol B*, Vol. 12, No. 1, pp.55–65.

Papaevangelou, A., Chatzistergos, S., Nikita, K.S. and Zografos, G. (2011) 'Breast density computerized analysis on digitized mammograms', *J Hellenic Stud*., Vol. 83, No. 3, pp.133–138.

Parkin, D.M. and Fernandez, L.M. (2006) 'Use of statistics to assess the global burden of breast cancer', *Breast*, Vol. 12, No. 1, pp.70–80.

Petroudi, S., Kadir, T. and Brady, M. (2003) 'Automatic classification of mammographic parenchymal patterns: a statistical approach', in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Cancun, Mexico, Vol. 1, pp.798–801.

Qu, Y., Shang, C. and Shen, Q. (2011) 'Evolutionary fuzzy extreme learning machine for mammographic risk analysis', *I Journal of Fuzzy Systems*, Vol. 13, No. 4, pp.282–291.

Sharma, M., Markou, M. and Singh, S. (2001) 'Evaluation of texture methods for image analysis', in *Proceedings of the Seventh Australian and New Zealand Intelligent Information Systems Conference*, Vol. 4, pp.117–121.

Sharma, V. and Singh, S. (2014) 'CFS-SMO-based classification of breast density using multiple texture models', *Med Biol Eng Comput*., Vol. 52, No. 6, pp.521–529.

Sharma, V. and Singh, S. (2015) 'Automated classification of fatty and dense mammogram', *Journal of medical imaging and health informatics*, Vol. 5, No. 7, pp.520–526.

Vachon, C.M., Van Gils, C.H., Sellers, T.A., Ghosh, K., Pruthi, S., Brandt, K.R. and Pankratz, V.S. (2007) 'Mammographic density, breast cancer risk and risk prediction', *Breast Cancer Research*, Vol. 9, No. 6, pp.1–9.

Viera, A.J. and Garrett, J.M. (2005) 'Understanding interobserver agreement: the kappa statistic', *Fam Med*., Vol. 37, No. 5, pp.360–363.

Virmani, J., Kumar, V., Kalra, N. and Khandelwal, N. (2014) 'Neural network ensemble-based CAD system for focal liver lesions from B-mode ultrasound', *J Digit Imaging*, Vol. 27, No. 4, pp.520–537.

Wolfe, J.N. (1976) 'Breast patterns as an index of risk for developing breast cancer', *Am J Roentgenol*., Vol. 126, No. 6, pp.1130–1137.

Wolfe, J.N. (1977) 'Risk for breast cancer development determined by mammographic parenchymal pattern', *Cancer*, Vol. 37, No. 5, pp.2486–2492.

Yaghjyan, L., Pinney, S.M., Mahoney, M.C., Morton, A.R. and Buckholz, J. (2011) 'Mammographic breast density assessment: a methods study', *Atlas J. Med. Biol. Sci*., Vol. 1, No. 1, pp.8–14.