A computerised framework for prediction of fatty and dense breast tissue using principal component analysis and multi-resolution texture descriptors

Indrajeet Kumar* and H.S. Bhadauria

Department of Computer Science and Engineering,
GB Pant Engineering College,
Pauri Garhwal, Uttarakhand, 246194, India
Fax: 01368228062
Email: erindrajeet@gmail.com
Email: hsb76iitr@gmail.com
*Corresponding author

Jitendra Virmani

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

Abstract: The present work proposes a computerised framework for prediction of fatty and dense breast tissue using principal component analysis and multi-resolution texture descriptors. For this study, 480 MLO view digitised screen film mammograms have been taken from the DDSM dataset. A fixed ROIs size of 128 × 128 pixels are cropped from the centre location of each mammographic image. Three texture features are computed in multi-resolution transform domain, where each ROI is decomposed up to 2nd level using ten different compact support wavelet filters resulting 16 sub-band feature images. Two step feature optimisation approach (feature pruning followed by feature space dimensionality reduction using PCA) is applied. In feature pruning stage, the TFV corresponding to best basis feature is selected; result of feature pruning stage is PTFV. This PTFV is subjected to PCA for feature space dimensionality reductions. After the application, PCA accuracy increases from 92.1% to 97.9%.

Keywords: mammography; breast density classification; multi-resolution texture descriptors; principal component analysis; PCA; support vector machine classifier; SVM.

Reference to this paper should be made as follows: Kumar, I., Bhadauria, H.S. and Virmani, J. (2018) 'A computerised framework for prediction of fatty and dense breast tissue using principal component analysis and multi-resolution texture descriptors', *Int. J. Computational Systems Engineering*, Vol. 4, Nos. 2/3, pp.73–85.

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 in Classification of Breast Density from Mammographic Images from the GB Pant Engineering College, Pauri Garhwal, Uttarakhand. 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 MTech in Electronics Engineering from the Aligarh Muslim University, Aligarh in 2004. He received his PhD in 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 the Govind Ballabh Pant Engineering College, Pauri Garhwal, Uttarakhand. 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.

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 in 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 Aug 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.

1 Introduction

Breast density is considered as a dominant risk factor for indicating the growth of breast cancer (Vachon et al., 2007; Boyd et al., 2009, 2011; Colin et al., 2013; Eng et al., 2014). It is the most common life intimidating disease that is found in women (Parkin and Fernández, 2006; Ferlay et al., 2013). Even for the experienced radiologists, the early detection of breast cancer is really a tough task like therapeutic action for the diseases especially in cases when these lesions are found in the area where glandular ducts are prominent.

In studies of Wolfe (1976a, 1976b) presented the relationship between the risk of breast lesions and breast tissue density. The most frequently used imaging modalities of screening for prior detection of breast lesions is mammography (Tang et al., 2009; Kumar et al., 2015a, 2015b). In this screening methodology, dense tissue represented by bright intensity pattern while fatty tissue represented by dark intensity. The sample images of fatty and dense classes, taken randomly from the DDSM dataset (Heath
et al., 1998) are shown in Figure 1.

Developing a computer aided classification system for prediction of breast density is clinically important because of:

1 risk of breast cancer is directly related to breast density

2 for double screening of dense mammograms to look for the lesions masked behind these cases.

After the literature study, it has been observed that the breast density classification can be designed using:

1 computer aided classification systems designed using segmented tissue-based approach (ST)

2 computer aided classification systems designed using fixed size ROI-based approach (ROI).

It is well-known that segmented tissue-based approaches require additional steps viz. eliminating the background and removing the pectorial muscle. Due to these additional steps, segmented tissue-based approaches are more time consuming and complex in comparison to the fixed size ROI-based approaches.

After the extensive review of literature, it has been observed that breast density classification studies are designed using either benchmark dataset (i.e., MIAS, DDSM) or by using the dataset of mammographic images collected by authors. It is worth mentioning that mammographic images in MIAS dataset are classified into three breast density classes, i.e.:

1 fatty

2 fatty glandular

3 dense tissue.

Also, few studies have proposed 4-class breast density classification using MIAS dataset; but in these studies, the participating radiologists have labelled the images of MIAS dataset as per 4-class BIRADS density standards. The studies in literature carried out by different authors on different datasets for breast density classification are reported in Table 1.

From Table 1, few facts can be observed:

a most of the computer aided classification systems designed for breast density classification are based on ST-based approach (Miller and Astley, 1991; Karssemeijer, 1998; Bovis and Singh, 2002; Petroudi et al., 2003; Oliver et al., 2005, 2008, 2010; Bosch et al., 2006; Muhimmah and Zwiggelaar, 2006; Jamal et al., 2007; Subashini et al., 2010; Tzikopoulos et al., 2011; Qu et al., 2011; Chen et al., 2011; Masmoudi et al., 2013; He et al., 2016)

b only a few studies have been carried out on a fixed size ROI (Blot and Zwiggelaar, 2001; Liu et al, 2011; Mustra et al., 2012; Sharma and Singh, 2014, 2015; Kumar et al., 2015a; Virmani et al., 2016; Kriti and Virmani, 2015; 2016a, 2016b; Kumar et al., 2017a, 2017b) based approach.

Further, it can also be observed that most of the computer aided classification systems for 4-class breast density classification have been conducted on DDSM dataset based on segmented tissue-based approach (Bovis and Singh, 2002; Oliver et al., 2005, 2008, 2010; Bosch et al., 2006). Only few study (Kumar et al., 2015a; Kumar et al., 2017a,

2017b) that carried out 4-class breast density classification based on fixed size ROI approach using DDSM dataset and attain the accuracy of 90.8% using ensemble of neural network classifiers without any feature space dimensionality reduction approach. The maximum accuracy achieved for MIAS dataset is 79.2% using ROI approach (Mustra et al., 2012) and 86.4% on self-collected dataset for 4-class breast density classification using fixed size ROI approach (Liu et al., 2011) on 88 images only.

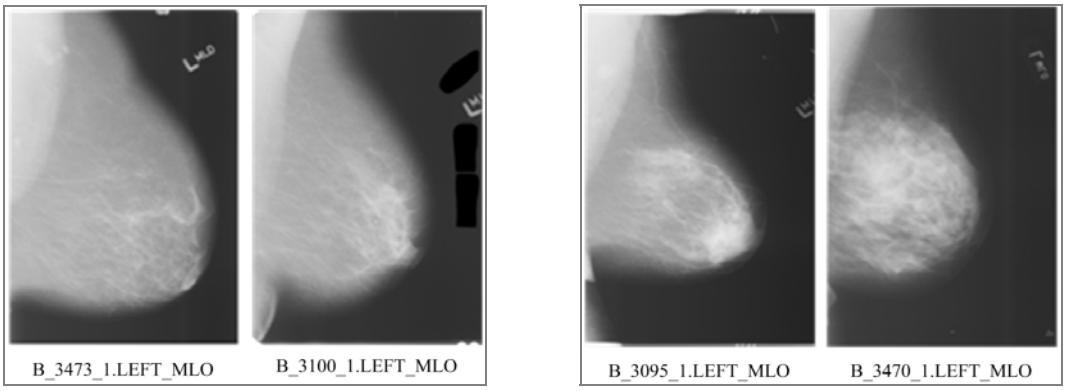
It is worth mention that the studies (Bovis and Singh, 2002; Oliver et al., 2010) have been reported as 2-class breast tissue classification systems (by considering B-I and B-II as 'fatty' and B-III and B-IV as 'dense' image class) on DDSM dataset using segmented tissue-based approach and thereby attaining the accuracy of 96.7% using ANN classifier (Bovis and Singh, 2002). The maximum accuracy achieved for MIAS dataset is 97.2% using ROI approach (Sharma and Singh, 2015) and 97.2% on self-collected dataset for 2-class breast density classification using fixed size ROI approach (Mustra et al., 2012).

Table 1 Studies in literature carried out by different authors on different datasets for breast density classification

Dataset used	*Author, year*	*ST/ROI*	*No. of images*	*Classifier used*	*Considered class*	*Accuracy (%)*
DDSM	Bovis and Singh, 2002	ST	377	ANN	4-class	71.4
	Oliver et al., 2005	ST	615	kNN	4-class	47.0
	Bosch et al., 2006	ST	500	SVM	4-class	84.7
	Oliver et al., 2008	ST	132	SFS-KNN	4-class	77.0
	Kumar et al., 2015a	ROI	480	SVM	4-class	73.7
	Kumar et al., 2017b	ROI	480	HY-HCF	4-class	84.1
	Kumar et al., 2017a	ROI	480	ENN	4-class	90.8
	Bovis and Singh, 2002	ST	377	ANN	2-class	96.7
	Oliver et al., 2010	ST	831	kNN	2-class	79.0
MIAS	Oliver et al., 2005	ST	270	Decision tree	4-class	73.0
	Bosch et al., 2006	ST	322	SVM	4-class	95.4
	Oliver et al., 2008	ST	322	SFS- kNN	4-class	66.0
	Qu et al., 2011	ST	322	*FELM	4-class	72.6
	Chen et al., 2011	ST	322	kNN	4-class	75.0
	Mustra et al., 2012	ROI	322	kNN	4-class	79.2
	Blot and Zwiggelaar, 2001	ROI	265	kNN	3-class	65.0
	Muhimmah and Zwiggelaar, 2006	ST	321	#DAG-SVM	3-class	77.5
	Subashini et al., 2010	ST	43	SVM	3-class	95.5
	Tzikopoulos et al., 2011	ST	322	SVM	3-class	85.7
	Oliver et al., 2010	ST	322	KNN	2-class	94.0
	Mustra et al., 2012	ROI	322	kNN	2-class	91.6
	Sharma and Singh, 2014	ROI	322	SMO-SVM	2-class	96.4
	Sharma and Singh, 2015	ROI	212	kNN	2-class	97.2
	Virmani and Kriti, 2016	ROI	322	kNN	2-class	96.2
	Kriti et al., 2016b	ROI	322	PCA-SVM	2-class	94.4
	Kriti and Virmani, 2015	ROI	322	PCA-NN	2-class	95.2
	Kriti et al., 2016a	ROI	322	PCA-SSVM	2-class	94.4
Self-collected datasets	Miller and Astley, 1991	ST	40	Bayesian	4-class	80.0
	Karssemeijer, 1998	ST	615	kNN	4-class	65.0
	Jamal et al., 2007	ST	100	--	4-class	78.3
	Masmoudi et al., 2013	ST	2052	kNN	4-class	79.0
	Liu et al., 2011	ROI	88	SVM	4-class	86.4
	He et al., 2016	ST	360	--	4-class	78.0

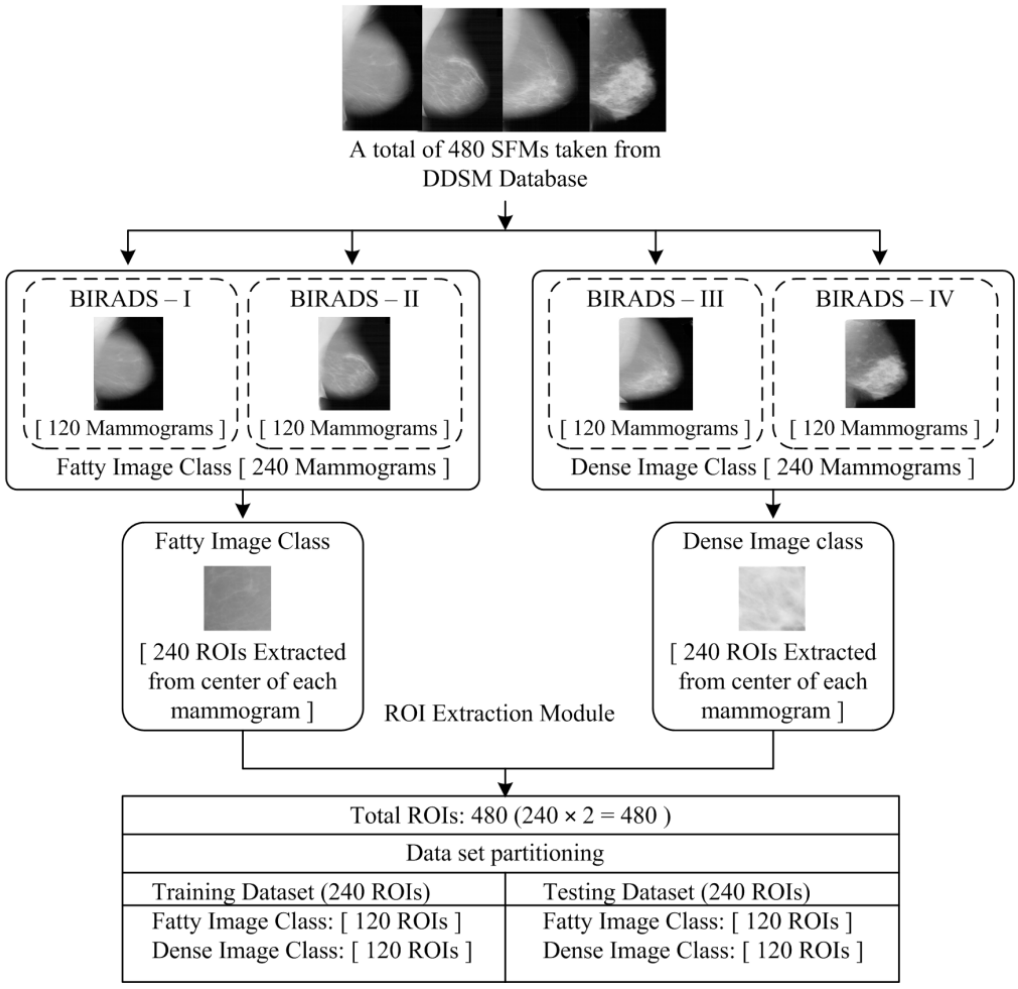
Notes: FELM*: fuzzy extreme learning machine; DAG#: directed acyclic graph; HY-HCF: hybrid hierarchical classification framework; ENN: ensemble of neural network.

Figure 1 Sample images of fatty and dense class images taken from DDSM dataset, (a) fatty image (b) dense image



(a) (b)

Figure 2 Image dataset description for 2-class breast density classification and its further bifurcation into training and testing dataset



Keeping in view the brief summary of the literature, the present study as the 2-class breast density classification has been designed using fixed size of ROI-based approach on DDSM dataset based on principal component analysis (PCA) and multi-resolution texture descriptors. It is worth observing that a very few studies have been conducted on DDSM dataset for 2-class breast density classification and maximum classification accuracy of 96.7% has been achieved using ST-based approach. However, the present study is based on fixed size ROI approach and different from the study reported in Kumar et al. (2015a). In fixed size ROI-based approach, the radiologist can easily marked square ROI at the centre location of the breast where glandular ducts are found and the computer aided classification system shall predict the density class in real-time clinical environment.

The study reported by Li et al. (2004) experimentally investigated that the density information decreases as the

location of ROI is varied from the centre location of the breast and concluded that the breast density is sufficient to account the discrimination between breast density classes. Therefore, in this study, a fixed size ROI of 128 × 128 pixels are extracted from the maximum density information region of each mammographic image. Each ROI is decomposed up to 2nd level using ten different compact support wavelet filters resulting 16 sub-band feature images. From these 16 sub-band images, three statistics (mean, standard deviation, energy) are computed. Before designing a computerised framework for prediction of breast density two step feature optimisation approach (feature pruning by exclusion of non-discriminatory features followed by feature space dimensionality reduction using PCA) is applied. The reduced feature vector set is inputted to support vector machine (SVM) classifier.

2 Materials and methods

2.1 Collection of a comprehensive and representative image dataset description

In this study, the collection of a comprehensive image dataset consisting of 480 MLO views digitised screen film mammograms (consisting of 120 mammograms belonging to each BIRADS classes) from DDSM dataset are considered. The DDSM dataset provides four images for each case, comprising of left/right MLO and left/right CC views. The overlay file of each DDSM images gives the ground truth information of breast density suggested by the expert (Heath et al., 1998). For 2-class breast density classification considering the:

a cases belonging to {BIRADS-I, BIRADS-II} classes as 'fatty' image class

b cases belonging to {BIRADS-III, BIRADS-IV} classes as 'dense' image class.

This dataset is further divided into training and testing datasets. The brief description of the comprehensive dataset and its bifurcation into training instances of dataset and testing instances of dataset is shown in Figure 2.

2.2 Proposed breast density classification system

In this study, an efficient computerised framework for the prediction of fatty and dense breast tissue using PCA and multi-resolution texture descriptors is proposed. In this proposed system, the radiologist has to just mark an ROI of size 128 × 128 pixels at the centre location of the breast where glandular ducts are found. The system will automatically compute the desired features at each stage and support vector classifier will predict the density class.

The experimental work flow diagram for the design of a computerised framework for prediction of fatty and dense breast tissue using PCA and multi-resolution texture descriptors is shown in Figure 3.

The proposed breast tissue density classification system consists of:

a ROI extraction stage

b feature extraction stage

c feature pruning stage

d feature space dimensionality reduction stage

e classification stage.

Each stage is briefly discussed here.

2.3 ROI extraction stage

Therefore, in the present study, fixed size ROI of 128 × 128 pixels are cropped from the centre location of the breast where glandular ducts are found for each mammographic image. The sample images with ROIs extracted for fatty and dense image class is shown in Figure 4.

2.4 Feature extraction stage

In this study, feature extraction is performed in transform domain using 2D WPT multi-resolution scheme. It has been mentioned in earlier studies (Wickerhauser, 1991, 1994; Meyer, 1993; Chang and Kuo, 1993; Mojsilovic et al., 2000; Rajpoot, 2003; Avci, 2008; Virmani et al., 2013) that the characterisation performance depend on the wavelet filters, i.e., bior3.1, bior3.3, bior4.4, coif1, coif2, haar (db1), db4, db6, sym3 and sym5. The comparative analysis of characteristic of these wavelet filters are given in Table 2.

Further, by using 2D WPT multi-resolution scheme each ROI is decomposed up to 2nd level and 16 sub-band feature images (as is shown in Figure 5) are obtained. The three statistics (i.e., mean (M_l), standard deviation (S_l) and energy (E_l) at level l) are computed for the 16 sub-band feature images using equations (1) to (3).

$$M_l = \left(\frac{1}{i \times j} \sum_{X=1}^{i} \sum_{Y=1}^{j} \left| SFI_l(i, j) \right| \right) \tag{1}$$

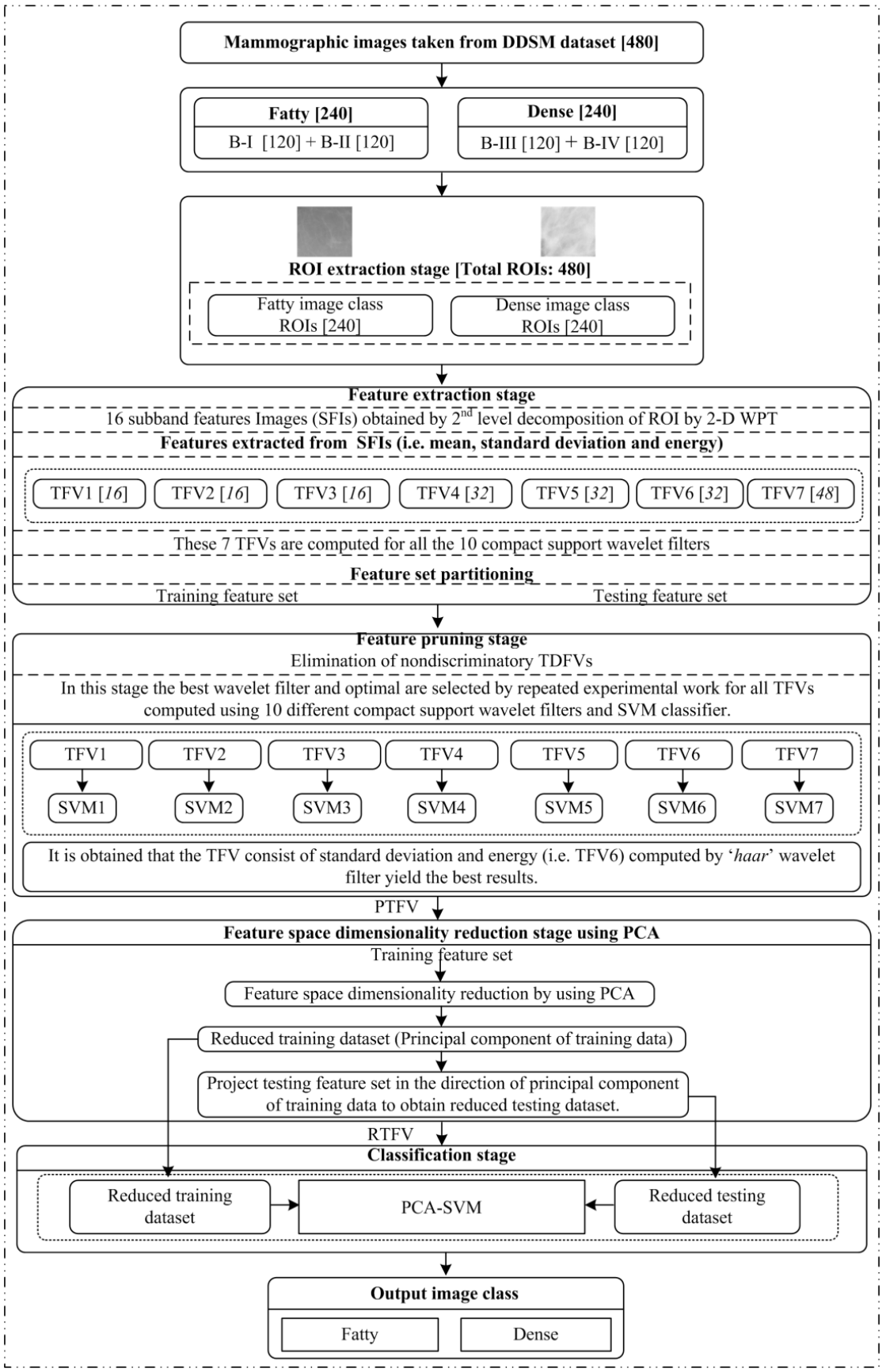
$$S_l = \left(\frac{1}{i \times j} \sum_{X=1}^{i} \sum_{Y=1}^{j} \left| SFI_l(i, j) - M_l \right|^2 \right)^{1/2} \tag{2}$$

$$E_l = \frac{1}{i \times j} \sum_{X=1}^{i} \sum_{Y=1}^{j} \left| SFI_l(i, j) \right|^2 \tag{3}$$

Here, M_l: mean computed from SFI at level l; S: standard deviation computed from SFI at level l, E: energy computed from SFI at level l and SFI_l are sub-band feature image of size $i \times j$ at level l = 1 to 7. It is well established that the maximum level of decomposition for the size of ROI $m \times n$ pixels is 2^m. Therefore, the maximum level of decomposition is seven for the ROI size 128 × 128 pixels in this study.

The extracted features (mean, standard deviation and energy) using 2D WPT multi-resolution scheme at 2nd level of decomposition from each of the 16 SFIs are shown in Figure 6.

Figure 3 Experimental work flow diagram for the design of a computerised framework for prediction of fatty and dense breast tissue using PCA and multi-resolution texture descriptors



Notes: TFVs: texture descriptor feature vectors; PTFVs: pruned TFVs; RTFVs: reduced TDVFs.

Figure 4 Sample images of ROI extraction of each BIRADS class

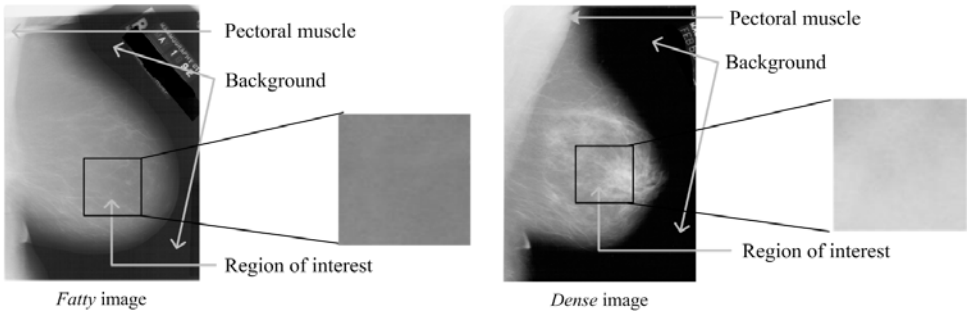
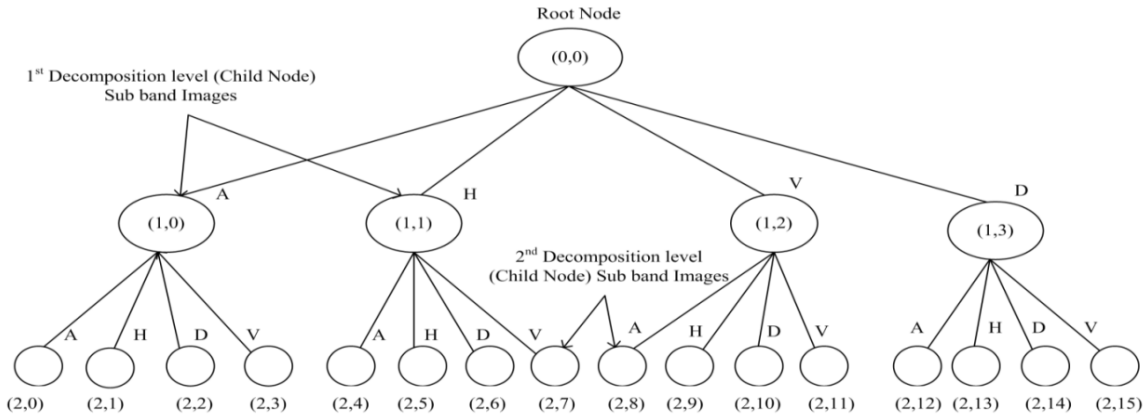


Table 2 Comparative analysis of different compact support wavelet filters

		Orthogonal	*Bi-orthogonal*	*Symmetry*	*Near symmetry*	*Asymmetry*	*Compact support*
Biorthogonal	bior3.1	No	Yes	Yes	No	No	Yes
	bior3.3	No	Yes	Yes	No	No	Yes
	bior4.4	No	Yes	Yes	No	No	Yes
Daubechies	db1#	Yes	No	Yes	No	No	Yes
	db4	Yes	No	No	No	Yes	Yes
	db6	Yes	No	No	No	Yes	Yes
Symlets	sym3	Yes	No	No	Yes	No	Yes
	sym5	Yes	No	No	Yes	No	Yes
Coiflets	coif 1	Yes	No	No	Yes	No	Yes
	coif 2	Yes	No	No	Yes	No	Yes

Note: db1# is 'haar' wavelet filter.

Figure 5 2D WPT tree up to second level decomposition from (2, 0) to (2, 15) represents 16 sub-band feature images



Notes: Here A: approximate sub-band feature image; H: horizontal sub-band feature image; D: diagonal sub-band feature image and V: Vertical sub-band feature image

Figure 6 Extracted feature (mean, standard deviation, energy) at 2nd level of decomposition using 2D-WPT

M_{AA} (2,0)	M_{AH} (2,1)	M_{HA} (2,4)	M_{HH} (2,5)
M_{AV} (2,3)	M_{AD} (2,2)	M_{HV} (2,7)	M_{HD} (2,6)
M_{VA} (2,12)	M_{VH} (2,13)	M_{DA} (2,8)	M_{DH} (2,9)
M_{VV} (2,15)	M_{VD} (2,14)	M_{DV} (2,11)	E_{DD} (2,10)

S_{AA} (2,0)	S_{AH} (2,1)	S_{HA} (2,4)	S_{HH} (2,5)
S_{AV} (2,3)	S_{AD} (2,2)	S_{HV} (2,7)	S_{HD} (2,6)
S_{VA} (2,12)	S_{VH} (2,13)	S_{DA} (2,8)	S_{DH} (2,9)
S_{VV} (2,15)	S_{VD} (2,14)	S_{DV} (2,11)	S_{DD} (2,10)

E_{AA} (2,0)	E_{AH} (2,1)	E_{HA} (2,4)	E_{HH} (2,5)
E_{AV} (2,3)	E_{AD} (2,2)	E_{HV} (2,7)	E_{HD} (2,6)
E_{VA} (2,12)	E_{VH} (2,13)	E_{DA} (2,8)	E_{DH} (2,9)
E_{VV} (2,15)	E_{VD} (2,14)	E_{DV} (2,11)	E_{DD} (2,10)

Notes: M: mean, S: standard deviation and E: energy features A: approximate sub-band feature image; H: horizontal sub-band feature image; D: diagonal sub-band feature image and V: Vertical sub-band feature image.

Table 3 The description of extracted TFVs using 2D WPT multi-resolution scheme

TFVs	*Description*	*l*
TFV1: TFV_M	Mean	16
TFV2: TFV_S	Standard deviation	16
TFV3: TFV_E	Energy	16
TFV4: TFV_{M+S}	Mean, standard deviation	32
TVF5: TFV_{M+E}	Mean, energy	32
TFV6: TFV_{S+E}	Standard deviation, energy	32
TFV7: TFV_{M+S+E}	Mean, standard deviation, energy	48

Notes: TFVs: texture descriptor feature vectors; *l*: length of TFV.

Table 4 Maximum and minimum performance obtained for TFVs with the corresponding wavelet filter

TFVs	*Max. acc. (%)*	*Wavelet filter*	*Min. acc. (%)*	*Wavelet filter*
TFV1	89.5	haar	82.9	coif2
TFV2	74.6	haar	67.5	db6
TFV3	88.3	haar	85.4	bior3.1
TFV4	90.8	haar	86.4	sym3
TVF5	88.7	haar	85.0	bior3.1
TFV6	92.1	haar	88.6	db6
TFV7	91.2	bior3.3	88.3	bior3.1

Notes: *l*: length of TFV; TFV: texture descriptor feature vector; Max. Acc.: maximum accuracy; Min. Acc.: minimum accuracy.

The brief description of extracted TFVs for 2-class breast density classification system using 2D WPT multi-resolution scheme is reported in Table 3.

To select the best wavelet filter and optimal multi-resolution features for 2-class classification tasks, all models (i.e., SVM1 to SVM7) were repeatedly trained and tested using all TFVs for ten different compact support wavelet filters. The maximum and minimum performance obtained for various TFVs using SVM classifier for 2-class breast density classification with the corresponding wavelet filter is reported in Table 4.

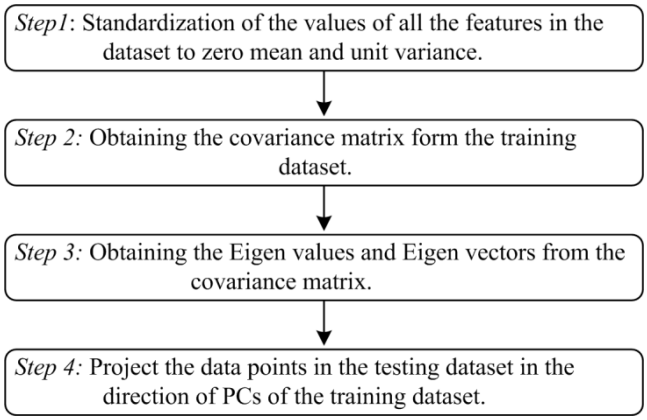
2.5 Feature pruning stage

Further, in the present work, two step feature optimisation approach (feature pruning by exclusion of non-discriminatory features followed by feature dimensionality reduction stage using PCA) is applied. In the first step of optimisation, texture descriptor feature vector pruning is performed by elimination of non-discriminatory TFVs. The classification accuracy of each TFVs is measured by the SVM classifier. The feature pruning stage yields a pruned TFV (i.e., PTFV) which gives the best classification performance for 2-class breast density classification.

2.6 Feature space dimensionality reduction stage using PCA

The pruned TFVs (i.e., PTFV1 and PTFV2) are fed to the feature space dimensionality reduction stage. It is used for reducing the dimensionality of feature space before building an efficient computerised framework (Coifman and Wickerhauser, 1992; Kadir et al., 2012). To build efficient classification module using SVM classifier for this task, optimal number of principal components, i.e., PCs is contained by exhaustive experiments carried out by first ten PCs. The four major steps concerned to PCA are shown in Figure 7.

Figure 7 Steps involved in PCA



In the present work, to retain the optimal number of PCs experimentations have been carried out by stepping through first ten PCs for 2-class breast density classification system on PTFV.

2.7 Classification stage

In this study, binary classification modules (2-class) has been implemented.

- *2-class breast density classification module*: In the present work 2-class breast density classification module has been implemented using binary SVM approach (i.e., PCA-SVM classifier). Each binary classifier is trained to separate a pair of class, and the prediction of the class for testing instance is made by majority voting mechanism. The detailed description of SVM classifier can be found in Chang and Lin (2012), Kriti and Virmani (2015), Mojsilovic et al. (2000), Burges (1998), Hassanien et al. (2013), Azar and El-Said (2014).

2.8 Performance measurement statistics

For the performance measurement overall classification accuracy (OCA), sensitivity, specificity, positive predictive value, negative predictive value are used. The brief description of these parameters are given in Figure 8.

Figure 8 Description of the performances parameters

True condition \ Predicted condition	Condition positive	Condition negative	
Test outcome positive	True positive (TP)	False positive (FP)	$PPV = \frac{TP}{TP+FP}$
Test outcome negative	False Negative (FN)	True Negative (TN)	$NPV = \frac{TN}{TN+FN}$
	$sensitivity = \frac{TP}{TP+FN}$	$specificity = \frac{TN}{TN+FP}$	$OCA = \frac{TP+TN}{TP+TN+FP+FN}$

3 Experiments, results and discussion

In this study, exhaustive experiments have been performed to design an efficient computerised framework for prediction of breast density based on PCA and multi-resolution texture descriptors. The description of these experiments is given here.

Experiment 1: 2-class breast density classification system using wavelet packet texture descriptors and SVM classifier.

In this experiment, initially compute the classification performance using SVM classifier of all TFVs (i.e., TFV1 to TFV7) for 16 sub-band feature images obtained after the 2nd level decomposition of each ROI with 2D WPT by using ten different compact support wavelet filters. From Table 4, It has been observed that the maximum accuracy yielded by TFV6 using 'haar' wavelet filter.

The classification performance for various TFVs with SVM using haar wavelet filter is given in Table 5.

From Table 5, it can be observed that maximum accuracy of 92.1% is yielded by TFV6 using SVM6 classifier for feature vector length 32. The ICA values are 92.1% and 91.6% for fatty and dense image class respectively.

Further, the performance of TFV6 using ten support wavelet filters with SVM classifier is shown in Table 6.

From Table 6, it can be observed that highest accuracy of 92.1% is yielded by TFV6 with 'haar' wavelet filter. It is also to be noted that the highest ICA values of 92.5% and 91.6% are obatained for 'fatty' image class and 'dense' image class for TFV6 with 'haar' filter.

Experiment 2: 2-class breast density classification system using PCA and multi-resolution texture descriptors with SVM classifier.

The obtained results for RTFV with PCA-SVM classifier is reported in Table 7.

From Table 7, it can be observed that the new improved accuracy for RTFV is maximum 97.9% using PCA-SVM classifier at reduced feature vector of length 7. A total of five instances (four from fatty image class and one from dense image class) are misclassified out of 240 instances for RTFV6. Therefore, the maximum ICA for fatty image class is 96.66% and 99.16% for dense image class.

Table 5 The classification performance for various TFVs with SVM using *haar* wavelet filter

TFVs	*Classifiers*	*l*	*CM*			*OCA*	*Sens.*	*Spec.*	*PPV*	*NPV*
				F	*D*					
TFV1	SVM1	16	F	109	11	89.5	88.6	90.5	90.8	88.3
			D	14	106					
TFV2	SVM2	16	F	95	25	74.5	72.5	77.0	79.1	70.0
			D	36	84					
TFV3	SVM3	16	F	108	12	88.3	87.0	89.6	90.0	86.6
			D	16	104					
TFV4	SVM4	32	F	110	10	91.6	91.6	91.6	91.6	91.6
			D	10	110					
TFV5	SVM5	32	F	106	14	88.7	89.0	88.4	88.3	89.1
			D	13	107					
TFV6	*SVM6*	*32*	F	*111*	*9*	*92.1*	*91.7*	*92.4*	*92.5*	*91.6*
			D	*10*	*110*					
TFV7	SVM7	48	F	109	11	90.4	90.0	90.7	90.8	90.0
			D	12	108					

Notes: TFVs: texture descriptor feature vectors; *l*: length of TFV; CM: confusion matrix; OCA: overall classification accuracy; F: fatty image class; D: dense image class; Sens.: sensitiviy; Spec.: specificity; PPV: positive predictive value; NPV: negative predictive value.

Table 6 Classification performance for TFV6 using 10 support wavelet filters with SVM classifier

Wavelet filter	*Confusion matrix*			*Accuracy*	*Sens.*	*Spec.*	*PPV*	*NPV*
		F	*D*					
bior3.1	F	110	10	90.8	90.1	91.5	91.6	90.0
	D	10	110					
bior3.3	F	108	12	90.0	90.0	90.0	90.0	90.0
	D	12	108					
bior4.4	F	104	16	88.7	91.8	86.0	85.0	92.5
	D	11	109					
coif1	F	108	12	90.0	90.0	90.0	90.0	90.0
	D	12	108					
coif2	F	106	14	89.5	90.5	88.6	88.3	90.8
	D	11	109					
db4	F	110	10	90.8	90.1	91.5	91.6	90.0
	D	10	110					
db6	F	104	16	88.7	90.4	87.2	86.6	90.8
	D	11	109					
haar	*F*	*111*	*9*	*92.1*	*91.7*	*92.4*	*92.5*	*91.6*
	D	*10*	*110*					
sym3	F	106	14	89.5	90.5	88.6	88.3	90.8
	D	11	109					
sym5	F	106	14	89.1	89.8	88.5	88.3	90.0
	D	12	108					

Notes: Sens.: sensitiviy; Spec.: specificity; PPV: positive predictive value; NPV: negative predictive value. All accuracies are calculated in percentage.

Table 7 The classification results for RTFV with PCA-SVM classifier

TFV	*l'*	*CM*			*Accuracy (%)*				
			F	*D*	*Sens.*	*Spec.*	*PPV*	*NPV*	*OCA*
RTFV	7	F	116	4	99.1	96.7	96.6	99.1	97.9
		D	1	119					

Notes: TFVs: texture descriptor feature vectors; *l*: length of RTFV; Max. Acc.: maximum accuracy; Min. Acc.: minimum accuracy.

Figure 9 A computerised framework for prediction of fatty and dense breast tissue using PCA and multi-resolution texture descriptors

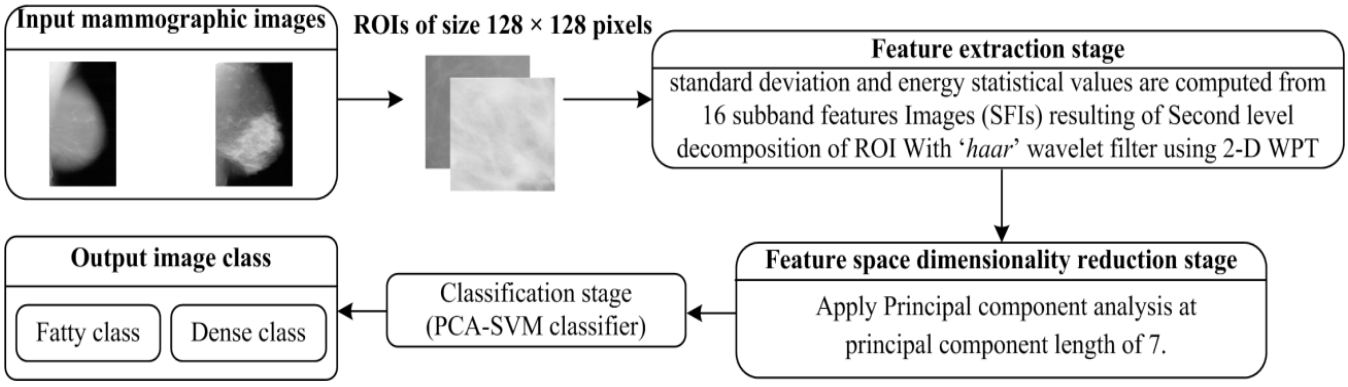


Table 8 Comparative analysis of obtained results for 4-class breast density classification and 2-class breast density classification based on multi-resolution texture descriptor and SVM classifier

Classification system	*TFVs*	*Classifier*	*CM*	*Accuracy (%)*				*Cohen's Kappa (Viera et al.) value (%)*
				F	*D*	*ICA*	*OCA*	
2-class breast density classification	TFV6	SVM6	F	111	9	92.5	92.1	84.1
			D	10	110	91.6		
	RTFV	PCA-SVM	F	116	4	96.6	97.9	95.8
			D	1	119	99.1		

Notes: TFVs: texture descriptor feature vectors; RTFVs: reduced texture descriptor feature vectors; ICS: individual class accuracy.

4 Discussion

The comparative analysis of obtained results for 4-class breast density classification and 2-class breast density classification based on multi-resolution texture descriptor and SVM classifier is shown in Table 8.

From Table 8, it can be observed that the OCA value increases from 73.7% to 77.1% in case of 4-class breast density classification and the ICA values are increases from 75.0% to 91.6%, 55.00% to 66.66% for B-I class and B-III class respectively. The obtained results show the significant improvement in OCA value and the most important improvement in ICA value of B-III class. B-III class is more confusing breast density class.

Further, it is also observed that the OCA value increases from 92.1% to 97.9% in case of 2-class breast density classification and the ICA values are increases from 92.5% to 96.6% and 91.6% to 99.1% for fatty image class and dense image class respectively.

From Table 9, it can be also observed that the kappa value (Viera and Garrett, 2005) increases from 84.1% to 95.8% for PCA-SVM with respect to SVM6 classifier. Thus, it can be concluded that the designed classification system using PCA-SVM is more stable.

After the exchaustive experiments performed for this study, the same fact as discussed in study (Kumar et al., 2015a) is observed. Thus, it has been concluded that standard deviation and energy statistics computed from sub-band feature images yielded by multi-resolution wavelet packet transform using 'haar' wavelet filter contain significant information for differential between 2-class breast density class classifications.

The designed a computerised framework for prediction of fatty and dense breast tissue using PCA and multi-resolution texture descriptors is shown in Figure 9.

5 Conclusions

In this work, a computerised framework for prediction of fatty and dense breast tissue using PCA and multi-resolution texture descriptors has been designed. In 2D WPT domain, ten compact support wavelet filters are used and after the two stage-based feature optimisation approach (i.e., feature pruning and PCA) it has been concluded that the 'haar' wavelet filter is yielding the maximum accuracy for 2-class breast density classification task.

After the exhaustive experiments performed in this work, the obtained accuracy for 2-class breast density classification is 97.9% for feature vector consisting of standard deviation and energy using 'haar' wavelet filter. Thus, it has been concluded that statistics, i.e., standard deviation and energy computed from sub-band feature images yielded by 2nd level multi-resolution decomposition using wavelet packet transform with 'haar' compact support wavelet filter contain significant information for differential between fatty and dense class of breast density. The promising results obtained by the proposed 2-class classification system indicate its usefulness to assist radiologists in routine screening mammography for differential diagnosis between different breast density classes.

References

Avci, E. (2008) 'Comparison of wavelet families for texture classification by using wavelet packet entropy adaptive network based fuzzy inference system', *Applied Soft Computing*, Vol. 8, No. 1, pp.225–231.

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

Blot, L. and Zwiggelaar, R. (2001) 'Background texture extraction for the classification of mammographic parenchymal patterns', in *Medical Image Understanding and Analysis*, July, pp.145–148.

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

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

Boyd, N.F., Martin, L., Chavez, S., Gunasekara, A., Salleh, A., Melnichouk, O., Yaffe, M., Friedenreich, C., Minkin, S. and Bronskill, M. (2009) 'Breast-tissue composition and other risk factors for breast cancer in young women: a cross-sectional study', *The Lancet Oncology*, Vol. 10, No. 6, pp.569–580.

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 Research*, Vol. 13, No. 6, p.1.

Burges, C.J. (1998) 'A tutorial on support vector machines for pattern recognition', *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp.121–167.

Chang, C.C. and Lin, C.J. (2012) *LIBSVM: A Library for Support Vector Machine, 2001*, Software [online] http://www.csie.ntu.edu.tw/~cjlin/libsvm (accessed 2016).

Chang, T. and Kuo, C.C. (1993) 'Texture analysis and classification with tree-structured wavelet transform', *IEEE Transactions on Image Processing*, Vol. 2, No. 4, pp.429–441.

Chen, Z., Denton, E. and Zwiggelaar, R. (2011) 'Local feature based mammographic tissue pattern modelling and breast density classification', in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, IEEE, October, Vol. 1, pp.351–355.

Coifman, R.R. and Wickerhauser, M.V. (1992) 'Entropy-based algorithms for best basis selection', *IEEE Transactions on Information Theory*, Vol. 38, No. 2, pp.713–718.

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

Eng, A., Gallant, Z., Shepherd, J., McCormack, V., Li, J., Dowsett, M., Vinnicombe, S., Allen, S. and dos-Santos-Silva, I. (2014) 'Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods', *Breast Cancer Research*, Vol. 16, No. 5, p.1.

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

Hassanien, A.E., El-Bendary, N., Kudělka, M. and Snášel, V. (2013) 'Breast cancer detection and classification using support vector machines and pulse coupled neural network', in *Proceedings of the Third International Conference on Intelligent Human Computer Interaction (IHCI 2011)*, Springer, Berlin, Heidelberg, Prague, Czech Republic, August, 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*, Springer International Publishing, June, pp.359–366.

Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr., P., Moore, R., Chang, K. and Munishkumaran, S. (1998) 'Current status of the digital database for screening mammography', in *Digital Mammography*, pp.457–460, Springer, Netherlands.

Jamal, N., Ng, K.H., Ranganathan, S. and Tan, L.K. (2007) 'Comparison of computerised 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*, Springer, Berlin, Heidelberg, pp.1405–1408.

Kadir, A., Nugroho, L.E., Susanto, A. and Santosa, P.I. (2012) 'Performance improvement of leaf identification system using principal component analysis', *International Journal of Advanced Science and Technology*, Vol. 44, No. 11, pp.113–124.

Karssemeijer, N. (1998) 'Automated classification of parenchymal patterns in mammograms', *Physics in Medicine and Biology*, Vol. 43, No. 2, p.365.

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

Kriti, Virmani, J. and Thakur, S. (2016a) 'Application of statistical texture features for breast tissue density classification', in *Image Feature Detectors and Descriptors*, pp.411–435, Springer International Publishing.

Kriti, Virmani, J., Dey, N. and Kumar, V. (2016b) 'PCA-PNN and PCA-SVM based CAD systems for breast density classification', in *Applications of Intelligent Optimization in Biology and Medicine*, pp.159–180, Springer International Publishing.

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 *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, March, pp.1960–1967.

Kumar, I., Bhadauria, H.S. and Virmani, J. and Thakur, S. (2017a) '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.

Kumar, I., Bhadauria, H.S., Virmani, J. and Thakur, S. (2017b) 'A hybrid hierarchical framework for classification of breast density using digitized film screen mammograms', *Multimed. Tools App*., pp.1–25, doi:10.1007/s11042-016-4340-z.

Li, H., Giger, M.L., Huo, Z., Olopade, O.I., Lan, L., Weber, B.L. and Bonta, I. (2004) 'Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location', *Medical Physics*, 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 *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, IEEE, October, Vol. 1, 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. 2013, No. 1, pp.1–9.

Meyer, Y. (1993) *Wavelets: Algorithms and Applications*, Colin Ed, 2nd ed., SIAM, Paris.

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

Mojsilovic, A., Popovic, M.V. and Rackov, D.M. (2000) 'On the selection of an optimal wavelet basis for texture characterization', *IEEE Transactions on Image Processing*, Vol. 9, No. 12, pp.2043–2050.

Muhimmah, I. and Zwiggelaar, R. (2006) 'Mammographic density classification using multiresolution histogram information', in *Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, ITAB*, October.

Mustra, M., Grgic, M. and Delac, K. (2012) 'Breast density classification using multiple feature selection', *AUTOMATIKA: Journal for Control, Measurement, Electronics, Computing and Communication*, Vol. 53, No. 4, pp.362–372.

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

Oliver, A., Freixenet, J., Marti, R., Pont, J., Pérez, E., Denton, E.R. and Zwiggelaar, R. (2008) 'A novel breast tissue density classification methodology', *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, No. 1, pp.55–65.

Oliver, A., Lladó, X., Pérez, E., Pont, J., Denton, E.R., Freixenet, J. and Martí, J. (2010) 'A statistical approach for breast density segmentation', *Journal of Digital Imaging*, Vol. 23, No. 5, pp.527–537.

Parkin, D.M. and Fernández, L.M. (2006) 'Use of statistics to assess the global burden of breast cancer', *The Breast Journal*, Vol. 12, No. s1, pp.S70–S80.

Petroudi, S., Kadir, T. and Brady, M. (2003) 'Automatic classification of mammographic parenchymal patterns: a statistical approach', *Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE*, 17 September, Vol. 1, pp.798–801, IEEE.

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

Rajpoot, N. (2003) 'Local discriminant wavelet packet basis for texture classification', in *Optical Science and Technology, SPIE's 48th Annual Meeting*, International Society for Optics and Photonics, November, pp.774–783.

Sharma, V. and Singh, S. (2014) 'CFS-SMO based classification of breast density using multiple texture models', *Medical & Biological Engineering & Computing*, Vol. 52, No. 6, pp.521–529.

Sharma, V. and Singh, S. (2015) 'Automated classification of fatty and dense mammograms', *Journal of Medical Imaging and Health Informatics*, Vol. 5, No. 3, pp.520–526.

Subashini, T.S., Ramalingam, V. and Palanivel, S. (2010) 'Automated assessment of breast tissue density in digital mammograms', *Computer Vision and Image Understanding*, Vol. 114, No. 1, pp.33–43.

Tang, J., Rangayyan, R.M., Xu, J., El Naqa, I. and Yang, Y. (2009) 'Computer-aided detection and diagnosis of breast cancer with mammography: recent advances', *IEEE Transactions on Information Technology in Biomedicine*, Vol. 13, No. 2, pp.236–251.

Tzikopoulos, S.D., Mavroforakis, M.E., Georgiou, H.V., Dimitropoulos, N. and Theodoridis, S. (2011) 'A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry', *Computer Methods and Programs in Biomedicine*, Vol. 102, No. 1, pp.47–63.

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, p.1.

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. (2016) 'Breast tissue density classification using wavelet-based texture descriptors', in *Proceedings of the Second International Conference on Computer and Communication Technologies*, Springer, India, pp.539–546.

Virmani, J. and Kriti (2016) 'Breast tissue density classification using wavelet-based texture descriptors', in *Proceedings of the Second International Conference on Computer and Communication Technologies*, pp.539–546, Springer, India.

Virmani, J., Kumar, V., Kalra, N. and Khandelwal, N. (2013) 'SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors', *Journal of Digital Imaging*, Vol. 26, No. 3, pp.530–543.

Wickerhauser, M.V. (1991) 'Lectures on wavelet packet algorithms', in *Lecture Notes*, INRIA.

Wickerhauser, M.V. (1994) *Adapted Wavelet Analysis from Theory to Software*, IEEE press.

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

Wolfe, J.N. (1976b) 'Risk for breast cancer development determined by mammographic parenchymal pattern', *Cancer*, Vol. 37, No. 5, pp.2486–2492.