ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2011.0025

MAMMOGRAMS ANALYSIS USING SVM CLASSIFIER IN COMBINED TRANSFORMS DOMAIN

B.N. Prathibha[1] and V. Sadasivam[2]
Department of Computer Science and Engineering, Manomaniam Sundaranar University, Tamil Nadu, India
E-mail: [1]prathibhashri@yahoo.co.in, [2]vs_msu@yahoo.com

Abstract
Breast cancer is a primary cause of mortality and morbidity in women. Reports reveal that earlier the detection of abnormalities, better the improvement in survival. Digital mammograms are one of the most effective means for detecting possible breast anomalies at early stages. Digital mammograms supported with Computer Aided Diagnostic (CAD) systems help the radiologists in taking reliable decisions. The proposed CAD system extracts wavelet features and spectral features for the better classification of mammograms. The Support Vector Machines classifier is used to analyze 206 mammogram images from Mias database pertaining to the severity of abnormality, i.e., benign and malign. The proposed system gives 93.14% accuracy for discrimination between normal-malign and 87.25% accuracy for normal-benign samples and 89.22% accuracy for benign-malign samples. The study reveals that features extracted in hybrid transform domain with SVM classifier proves to be a promising tool for analysis of mammograms.

Keywords:
Mammograms, Classification, Hybrid Transforms, SVM

1. INTRODUCTION

Mammography is an X-ray examination technique that photographs the inside tissue abnormalities of the breast [1]. It is one of the best screening methods used for early detection of breast cancer for decades. Early detection of the cancer leads to significant improvements in conservative treatment. A variety of CAD techniques were developed in order to improve the accuracy of interpretation. Through the combination of digital Mammography and CAD, radiologists can give better diagnostics that will positively impact millions of women. CAD's have the potential to detect findings that might otherwise be overlooked during the review process, hence increasing the cancer detection and an effort to prevent biopsy intervention for tissue identification. The CAD systems attempt to achieve both the goals of optimality i.e., reducing the cost and increasing the accuracy of diagnostics. Thus the need for "callbacks" or "repeat mammograms" is greatly diminished, if not eliminated.

2. RELATED RESEARCH

Many computer algorithms for mammogram analysis are available using transforms. The vast use of transforms is due to the fact that they help to interpret the images at a variety of scales and translations.

The Discrete Wavelet Transform (DWT) is used by Jianhua Yao et al. [2] for breast tumor analysis on Dynamic Contrast Enhanced Magnetic Resonance images. They obtain 0.989 and 0.984 area under the curve values for different training and testing datasets. The support vector machine is used as the classifier for breast tissue classification. April Khademi and Shridhar Krishnan [3] used the statistical features from wavelet domain. The Shift Invariant DWT or the Stationary Wavelet Transform (SWT) with scale invariant representation is used, giving a generalized framework for medical image analysis. They classified small bowel images, retina images as well as mammogram images obtaining 85%, 82.2% and 69% respectively.

Ibrahim Faye et al. [4] proposed a method for classification of mammogram using wavelet features, achieving 98% accuracy. The high frequency signals of mammograms are extracted by using DWT by Weidong Xu et.al. [5] in order to find the presence of micro Calcifications. The CAD algorithm used the Adaptive network based Fuzzy Inference system (ANFIS) for more preciseness. Essam A Rashed et al. [6] used a region based multi resolution analysis for distinguishing malign and benign tumors. The biggest wavelet coefficients are used as the discriminating feature, achieving 99.5% of successful classification. The Euclidean distance measure is used to design the classifier.

An evaluation on the performance of CAD systems is done by Sheila Timp et al. [7] with double reading on radiologist's diagnostics in classifying benign and malign masses. They conclude that CAD systems with temporal analysis have the potential to help radiologists in discriminating benign and malign masses. The curvelet transform is used as a feature extraction technique by Mohamed Meselhy Eltoukhy et al. [8] for breast cancer diagnosis in digital mammograms. The Euclidean distance is then used to construct a supervised classifier. The experimental results gave 98.59% classification accuracy rate, which indicate that curvelet transformation is a promising tool for analysis and classification of digital mammograms. B. Zheng et al. [9] have proposed a scheme that uses a neural network with spatial domain features and Discrete Cosine Transform (DCT) based features as input and a spectral entropy based decision criteria. The DCT features are used by Farag A. and Mashali S. [10] in order to discriminate normal mammograms from mammograms with micro calcifications. Classification is done with a three-layer back propagation neural network.

Studies show that the DWT has been widely used by classification techniques, due to its excellent spatial localization and multi resolution characteristics. But, the classical DWT suffers from lack of shift invariance. The SWT solves the problem by giving better approximation since it is redundant, linear and shift invariant. The DCT also has been in choice, as it reduces image information redundancy, it gives only a subset of the transform coefficients that is necessary to preserve the most important features. Hence, the proposed system combines both the SWT and DCT, so that combined transforms compensates the drawbacks of each other and gives more preciseness in mammograms analysis. Rest of the paper is divided into 3 sections, Section.3 discusses the methodology used, Section.4 discusses the experiments and Section.5 gives the conclusion.

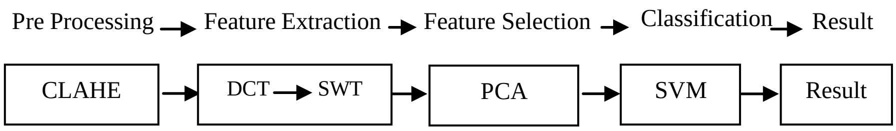

Fig.1. Flowchart of proposed method

3. MATERIALS AND METHOD

The mammography data is taken from the Mammographic Image Analysis society (MIAS) [11]. The case sample analyzed consists of 206 samples extracted from 206 individual mammograms of which 104 are normal and 102 are abnormal (benign and malign i.e., severity of abnormality). The dataset is divided into three subsets, 52 normal and 50 malign, 52 normal and 50 benign and 52 benign and 50 malign. As the size of mammograms is large a squared area of fixed size 64×64 has been cropped from each mammogram, so that the abnormality is covered. MATLAB 7.5 is extensively used through the implementation of the proposed method.

The proposed system as depicted in Fig.1 analyses the mammograms by discriminating normal-malign (cancerous, spreadable), normal-benign (non-cancerous, non spreadable), and benign-malign mammograms. The generic steps of pattern recognition are given and the specific methods, as used by the proposed system are given in the boxes under each step. During the pre processing step, the mammograms are contrast enhanced by using; Contrast Limiting Adaptive Histogram Equalization (CLAHE) algorithm and transferred to spectral domain with DCT, and then the wavelet features are extracted using SWT, which gives rise to a large pool of features. The most distinguishing features are selected using Principal component Analysis (PCA) and fed to the SVM classifier for further classification.

3.1 PRE PROCESSING

Before feature extraction the mammograms are contrast enhanced to improve the quality. A generalization of Adaptive Histogram Equalization (AHE), contrast limiting AHE is used. It has more flexibility in choosing the local histogram mapping function.

CLAHE [12] operates on small regions in the image called tiles, rather than the entire image. The contrast, especially in homogeneous areas, can be limited to avoid any noise amplification which might be present in the image, by selecting the suitable clipping level of the histogram. The clip level specifies the contrast enhancement limit, the higher value of clip level, more the contrast and vice versa.

3.2. FEATURE EXTRACTION

3.2.1. Discrete Cosine Transform

The DCT has strong capability to compress all energy of an image; hence one can often reconstruct a sequence very accurately from a few coefficients. The discrete cosine transform [13] F(u,v) of an N×N image is defined by

$$F(u,v)=C(u)C(v)\sum_{x=0}^{(N-1)} f(x,y)\cos\frac{(2x+1)u\pi}{2N}\cos\frac{(2y+1)v\pi}{2N}$$

where $C(u)=C(v)=\frac{1}{\sqrt{N}}$, for $u,v=0$,

$C(u)=C(v)=\sqrt{\frac{2}{N}}$, for $u,v=1,2,....N-1$

and f(x,y) is the intensity of the pixel in row i and column j of original image f.

The 2D-DCT compresses all the energy/information of the image and concentrates it in a few coefficients located in the upper-left corner of the resulting real-valued N X N DCT/frequency matrix. The DCT transformed image has zero or low-level pixel values except at the top left corner where the intensities are very high. These low-frequency, high intensity coefficients, are therefore, the most important coefficients in the frequency matrix and carry most of the information about the original image.

3.2.2 Stationary Wavelet Transform

The SWT, also called un-decimated wavelet transform is very similar to DWT. In DWT, the translated version of a signal is not the same as the original signal, which is due to lack of shift invariance, but the SWT achieves shift-invariance by removing the down samples thus obtaining a redundant decomposition. The digital implementation of the SWT [14] up to 2 levels of decomposition is as shown in Fig.2 where n is the original image and g[n], h[n] are the low pass and high pass filters. The filters at each level of decomposition are up sampled versions of previous level.

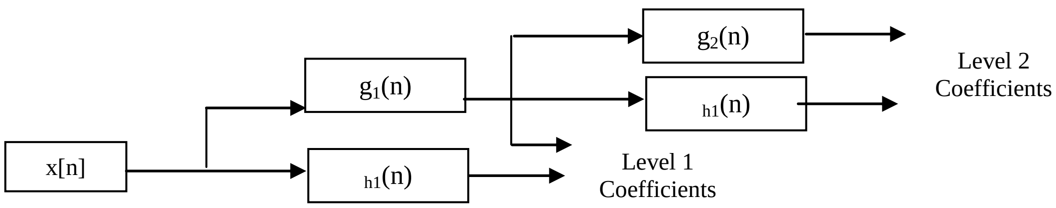

Fig.2. Stationary Wavelet Transform

3.3. FEATURE SELECTION

Through the transformations a high dimension feature vectors set is obtained. Only the features which have most discerning ability are fed to the classifier. PCA is the commonly used statistical technique for dimensionality reduction [15], in the sense that it replaces a large set of observed variables with a smaller set of new variables in such a way that they highlight their similarities and differences.

PCA involves the calculation of the eigen value decomposition of a data covariance matrix or singular value decomposition (SVD) of a data matrix, usually after mean centering the data for each attribute. The results of a PCA are usually discussed in terms of the principal component coefficients, also known as loadings and the component scores.

3.4. CLASSIFICATION

The SVM are non-parametric, supervised classifiers, [16] i.e., they use labeled examples to build models for classification of new data. Given a training set, (x_1, y_1), (x_2, y_2)………(x_3, y_3) such that $x_i \in R^P$ are feature vectors of dimensionality p and $y_i \in \{+1, -1\}$ are labels for a binary classifier. SVM require the solution of the following optimization problem.

$$\min\left(\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i\right)$$

subject to $y_i(w^T\Phi(x_i)+b) \geq 1 - \xi$

where w and b are hyper plane parameters, $C > 0$ is the penalty parameter and $\Phi(x)$ is a function to map vector x into a higher dimensional space. $K(x_i, x_j) = \Phi(x_i)^T\Phi(x_j)$ is called the kernel function. . Three types of kernels are used and given in following equations.

1. Linear: $K(x_i, x_j) = x_i^T x_j$ (1)
2. Polynomial: $K(x_i, x_j) = (x_i, x_j)^d$ (2)
3. RBF: $K(x_i, x_j) = \exp\left(\frac{-\| x_i - x_j \|^2}{2\sigma^2}\right)$ (3)

4. EXPERIMENTS AND RESULTS

The mammograms are contrast enhanced to improve the quality and then the discrete Cosine transformed. The DCT modulus contains the low frequency information of mammogram image, as long as this information does not lose the principal features can be retained. The resultant image is then decomposed using SWT with Daubechies filter at different lengths. The principal features are selected from each filter through PCA, normalized [17] and fed to the SVM classifier. The overall performance of the filters is combined for precise classification. The performance of the classifier is cross validated by using leave-M-out method. In leave-M-out method M samples from the database are kept for testing and remaining is used for training. The process is repeated until all samples of database have been tested.

Fitness criteria

The degree of reliability of the classification is found with the confusion matrix and some of the figures of merit associated with it. First, by defining arbitrarily which of the two sets is positive (P) or negative (N), the following four quantities were defined: true positives (TP) is the number of positive ROI that are correctly classified to class positive; true negatives (TN) is the number of negative ROI that are correctly classified to class negative; false positives (FP) is the number of negative ROI that are incorrectly classified to be class positive; and false negative (FN) is the number of ROI that are classified to be negative despite they are of class positive.

The Sensitivity, Specificity, Accuracy and Matthews's correlation coefficient (MCC) are considered as fitness criteria and calculated as follows,

$$Sensitivity = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(FP + TN)}$$

$$Accuracy = \frac{(TP + TN)}{(P + N)} \text{ and}$$

$$MCC = \frac{(TP.TN - FP.FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

The MCC is considered along with accuracy since it considers failure classification rate along with successful classification rate where as accuracy considers only successful classification rate.

Table.1 gives the figures of TP, FN, TN, FP of the data sets used and the classification power of the proposed system in terms of sensitivity, specificity, accuracy and MCC, for leave-1-out cross validation.

Table.1. Performance analysis on different datasets

Samples	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	MCC
Normal- Malign	46	49	3	4	92.00	94.23	93.14	0.86
Normal-Benign	42	47	5	8	84.00	90.38	87.25	0.75
Benign-Malign	44	47	5	6	88.00	90.38	89.22	0.78

The performance of the classifier in discriminating normal to malign, normal to benign and from benign to malign are experimented and the results are tabulated. It is clear that the proposed method works well in classifying normal from malign samples (93.33) than classification of normal from benign samples (87.25) and benign from malign samples (89.22).

Table.2. Performance of Daubechies filter

Filters	Sensitivity	Specificity	Accuracy	MCC
Db2	82.00	86.54	84.31	0.69
Db4	84.00	88.46	86.27	0.73
Db8	90.00	88.46	89.22	0.79

Table.2 shows the results of Daubecheies filter at different lengths for normal and malign samples. It is clear that the Db8 performs well compared to Db2 and Db4 filters. However the efficiency of three filters is combined for the better classification.

Table.3. Cross Validation Results

M	Sensitivity	Specificity	Accuracy	MCC
M=2	86.54	94.00	90.20	0.81
M=3	86.54	92.00	89.22	0.79
M=4	84.62	92.00	88.24	0.77
M=5	86.54	92.00	89.22	0.79

Table.3 gives the successful classification rate in classifying the normal and cancerous mammograms using leave-M-out cross validation technique, for different values of M. From the Table.3 it is clear that the results are quite consistent for different values of M. The Fig.3 shows the effect of clip limit used in CLAHE on classification accuracy, as discussed in section 3.1.

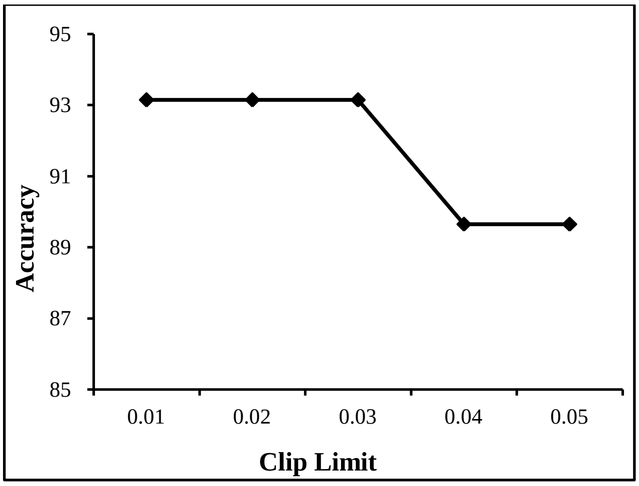

Fig.3. Effect of clip limit on accuracy

The value of the clip limit is varied from 0.01 to 0.05 and is observed that from 0.01 to 0.03 (lower contrast level) the accuracy is high and stable (93.33), from 0.04 (high contrast level) the accuracy drops (90.65).

The performance of the SVM classifier is evaluated by using different kernels (Eq.1, 2 and 3). The Fig.4 illustrates the ability of RBF kernel with polynomial kernel and linear kernel. The RBF kernel performs well by giving an accuracy (93.14) comparing with polynomial (83.33) and linear (73.53).

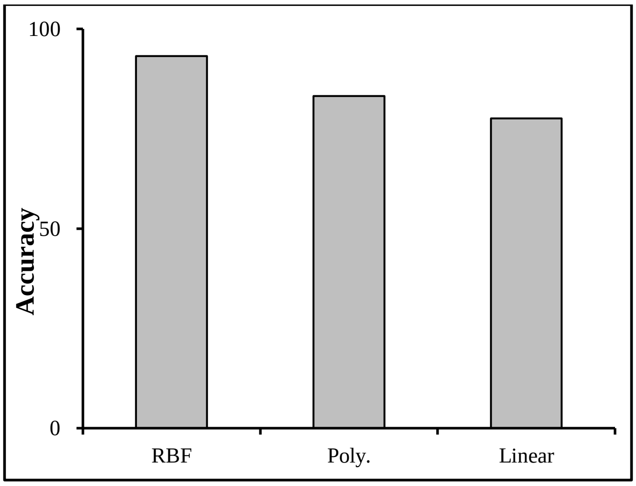

Fig.4. Performance of SVM kernels

The choice of sigma in the RBF kernel has a fair impact on the classification accuracy. The value of sigma (Eq. 3) is varied from 0.1 to 1, and the results are shown in Fig.5 in terms of accuracy. It is clear that the proposed system achieves best classification rate at 0.4, then there is a deteriotion in accuracy.

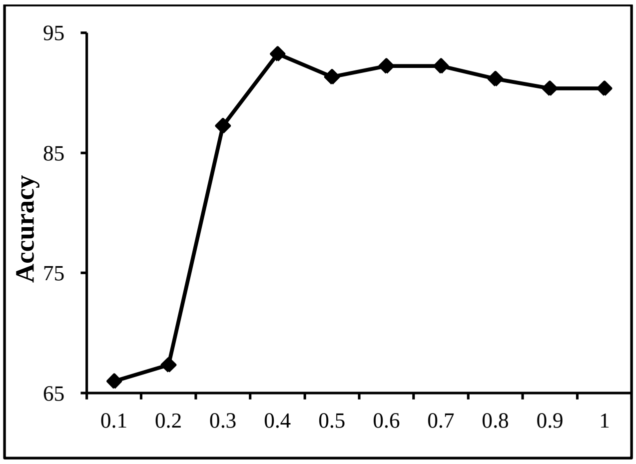

Fig.5. Effect of Sigma value on accuracy

The number of features used for classification, also has great impact on classification accuracy. The effect of number of features on accuracy is depicted in Fig.6. The number of features used is experimented from 5 to 35. Initially, the accuracy of the classifier increases as number of features increases reaches a peak and then begins to decrease. The classifier gives its best accuracy for a total of 20 features.

The classification rate of the proposed system is compared with the Kernel Discriminate Analysis (KDA) and the k Nearest Neighbor (kNN) classifier.

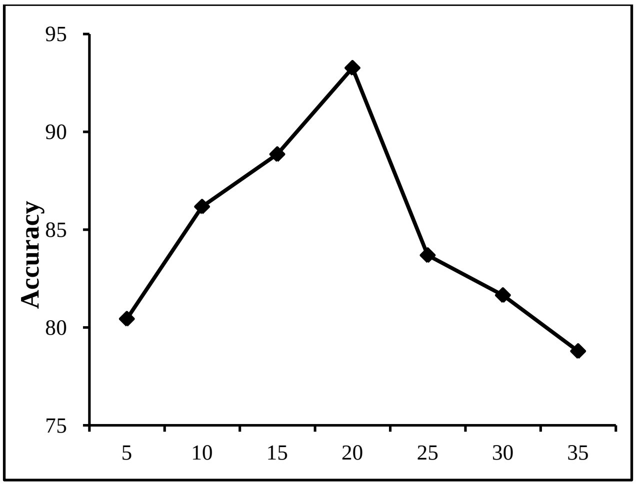

Fig.6. Effect of number of features

The KDA is a binary classifier, [18] it uses Parzen window density estimation to obtain an unknown probability density function directly from the data and overcomes the crisis of knowing prior probability density of the dataset thus skipping the learning phase. The kNN [19] is the well known intuitive statistical classifier, it can be defined as, for a given unlabeled sample x, the k "closest" labeled samples in the reference data set are found and x is assigned to the class that appears most frequently within the k subset. The value of k is varied from 3 to 11 and is observed that best performance is achieved for value of 3. From the Table.4 it is clear the proposed system performs better than the above said classifiers.

Table.4. Performance evaluation with other classifiers

Classifiers		**Sensitivity**	**Specificity**	**Accuracy**	**MCC**
KDA		92.73	74.55	83.64	0.69
KNN	**K=3**	94.55	74.55	84.55	0.71
	K=5	94.55	69.09	81.82	0.66
	K=7	94.55	70.91	82.73	0.67
SVM		92.00	94.23	93.14	0.86

5. CONCLUSION

The proposed work analyses the mammograms into normal, benign and malign using cascade of transforms; SWT and DCT joined with SVM classifier**.** The DCT gives a subset of the transform coefficients that is sufficient to preserve the most important features. The SWT gives a better approximation since it is redundant, linear and shift invariant. Both the transforms have been in use by many classification algorithms individually, the proposed system extracts SWT features in DCT domain, in order that the combined transforms compensates the drawbacks of each other, giving more precise classification. Through the experiments it is found that the SVM classifier with RBF kernel works excellently in the hybrid transform domain comparing to kNN and KDA classification techniques. Experiments shows that the level of the contrast enhancement, the number of features used and the value of sigma in RBF kernel are the optimal smoothing parameters for the proposed classification system. Future work concentrates on combining DCT with projection based transforms like Fanbeam and Radon.

REFERENCES

[1] American Cancer Society, Inc. Cancer Facts and Figures 2003, Atalnta: ACS, 2003

[2] Jianhua Yao, Jeremy Chen and Catherine Chow, 2009, "Breast tumor analysis in DCE MRI using texture features and wavelet transforms," IEEE Transactions of Selected Topics in Signal Processing, Vol. 3, No.1, pp.94–100.

[3] April Khademi, Shridhar Krishnan, 2008, "Medical image texture analysis: A case study with small bowel, retinal and mammogram images," In Canadian conference on Electrical and Computer Engineering.

[4] Ibrahim Faye, Brahim Belhaouri and Mohamed Meselhy Eltoukhy, 2009, "Digital Mammograms Classification Using a Wavelet Based Feature Extraction Method", In Proceedings of the Second International Conference on Computer and Electrical Engineering, Vol.2.

[5] Weidong Xu, Lihua Li Shaofang Zou, 2007, "Detection and classification of Micro calcifications based on DWT and ANFIS", First international conference on Bioinformatics and Biomedical Engineering, pp.547 – 550.

[6] Essam A Rashed, Ismail A Ismail and Sherif I. Zaki, 2007, "Multiresolution mammogram analysis in multilevel decomposition", Pattern Recognition Letters, Vol.28, No.2, pp.286–292.

[7] Sheila Timp, Celia Varela, and Nico Karssemeijer, 2010, "Computer-Aided Diagnosis With Temporal Analysis to Improve Radiologists' Interpretation of Mammographic Mass Lesions", IEEE transactions on information technology in biomedicine, Vol.14, No.3, pp.803-808.

[8] Mohamed Meselhy Eltoukhy , Ibrahima Faye , Brahim Belhaouari Samir, 2010, "Breast cancer diagnosis in digital mammogram using multiscale curvelet Transform", Computerized Medical Imaging and Graphics, Vol.34, No.4, pp.269–276.

[9] B. Zheng, Wei Quan and Laurence P. Clark, 1996, "Digital Mammography: Mixed Feature Neural Network with Spectral Entropy Decision for Detection of Micro calcifications", IEEE Transactions on Medical Imaging**,** Vol.15, No.5, pp.589-597.

[10] Farag A, Mashali, 2003, "DCT based features for the detection of micro calcifications in digital mammograms", IEEE International Symposium on Micro-NanoMechatronics and Human Science, Vol.1, pp.352-355.

[11] Mammographic Image Analysis Society, http://peipa.essex.ac.uk/ipa/pix/mias/

[12] S.M.Pizer and E.P.Amburn, 1987, "Adaptive histogram equalization and its variations", Computer Vision, Graphics, and Image processing, Vol.39, pp.355-368.

[13] Gonzalez, Woods, and Eddins, 2004, Digital Image Processing Using MATLAB , Prentice Hall, 3rd edition.

[14] S Mallat, 1989, "A theory for multiresolution signal decomposition: The wavelet representation", IEEE Transaction Pattern Analysis Machine Intelligence, Vol.11, No.7, pp.674–693.

[15] Weilong Chen, Meng Joo Er, and Shiqian Wu, 2005, "PCA and LDA in DCT domain," Pattern Recognition Letters, Vol.26, pp.2474-2482.

[16] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. , 2003, "A practical guide to support vector classification,"

Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei. http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[17] B.N. Prathibha and V.Sadasivam, 2010, "Multi Resolution Texture Analysis of Mammograms Using Nearest Neighbor Classification Techniques", International journal of Information Acquisition, Vol.7, No.2, pp. 109-118.

[18] Earl Gose, Richard Johns Baugh and Steve fast, 2007, "Pattern recognition and image analysis", pp.159-175, PHI Pvt. Ltd., New York.

[19] Donald F. Specht, 1990, "Probabilistic neural networks", Neural Networks, Vol.3, No.1, pp. 109– 118.