International Journal of Soft Computing 3 (3): 212-219, 2008
ISSN: 1816-9503

Mammogram Classification Using Support Vector Machines

S. Thamarai Selvi and R. Malmathanraj
Department of Information Technology, Madras Institute of Technology, Anna University, India

Abstract: The clustered microcalcifications on X-ray mammogram provides an important cue for early detection of breast cancer. Texture analysis methods can be applied to detect clustered micro calcifications in digitized mammograms. The clustered microcalcifications on X-ray mammogram provides an important cue for early detection of breast cancer. Texture analysis methods can be applied to detect clustered micro calcifications in digitized mammograms. In this study a novel 2 stage method for mammogram segmentation is implemented to facilitate automatic segmentation of micro calcification. The first stage is the Modified combined morphological spectral unsupervised Image segmentation. The first stage includes watershed transform, anisotrophic filtering technique, band pass filtering scheme, gradient synthesisation and Complex Wavelet Transform (CWT) subband extraction. The second stage of the segmentation scheme is the Random walkers segmentation technique. Finally, features are derived from the Ridgelet subbands of the segmented image. The cooccurrence matrix features are also used for classification. This study also implements the Support Vector Machines (SVM) for effective classification of Mammogram into Benign or malignant mammogram. The validation of the classification scheme was performed by using the Receiver Operating Curve (ROC) analysis, the overall sensitivity of the technique measured by the value of A_z which was found to be ranging from 0.8-0.928.

Key words: CWT, SVN, ROC, mammogram, classification, X-ray

INTRODUCTION

Cancerous tumor mass is one type of the leading causes of cancer related mortality among American Women. Although, the breast cancer is fatal, people have the highest chance of survival if physicians can detect the cancer at early stages. Thus early diagnosis and treatment play the critical roles in increasing the chance of survival. Screening mammography has been recommended as the most effective method for early detection of breast cancer. The digital mammography is one of the available screening methodologies. Digital mammography is a technique for recording x-ray images in computer code instead of on x-ray film, as with conventional mammography. The images are displayed on a computer monitor and can be enhanced (lightened or darkened) before they are printed on film. Images can also be manipulated; the radiologist can magnify or zoom in on an area. The images can be stored and retrieved electronically, which makes long-distance consultations with other mammography specialists easier. The algorithm is used to serve the mankind by serving as a second opinion for the radiologists. The focus is on the detection of cancerous masses from mammograms, which is one of the major types of breast cancer.

Currently, there are several image processing methods proposed for the detection of tumors in mammograms. Various technologies such as fractal analysis (Rama and Shankar, 1985), multiresolution based image processing (Tuecyran and Jain, 1998; Argenti *et al*., 1990) and Markov Random Field (MRF) (Patrizio *et al*., 2002) have been used. In Argenti *et al*. (1990), the wavelet analysis is used for image segmentation and micro calcification detection. In Tuecyran and Jain (1998), Sonka described an algorithm for tumor detection from mammogram that is based on fuzzy pyramid linking and multiresolution segmentation. In Cross and Jain (1983), Canny reported a two stage adaptive Density-Weighted Contrast Enhancement (DWCE) algorithm for tumor detection from mammograms. Although there are various tumor detection algorithms in the literature, the detection rate is still not high. For example, Patrizio *et al*. (2002) reports a detection rate of 90% and Smith and Chang (1994) also reports a detection rate around 90% with over

Corresponding Author: S. Thamarai Selvi, Department of Information Technology, Madras Institute of Technology, Anna University, India

ten FPs per image. This is due to the high variance in size and shape of the tumors and also due to the disturbance (noise) from the fatty tissues, veins and glands.

A digital image is usually characterized by 2 main aspects: tone and texture. The image tone consists of background greylevel variations of the pixels throughout the entire image. The image texture represents the intrinsic spatial variability of neighboring pixel values for each pixel within the image. It follows that methods of image analysis can be broadly divided into 2 categories: the spectral one and the textural one. In spectral analysis the interest is in studying broad variations of the grey levels of the pixels (image tone) in mono-or multispectral bands. The aim of textural analysis is to characterize spatially the grey-level relationships between the pixels of a neighborhood. Tone and texture are usually not independent in an image, so that when processing it we observe that one influences the other. The image tone characteristics can be separated from the texture and these 2 features can be processed independently. This situation promises to be very useful in enhancing the texture perception on the one hand and in improving the result of conventional texture analysis on the other.

The solution to the mammogram mass calculations is by texture analysis. Textures are replications, symmetries and combinations of various basic patterns, usually with some random variation one of the gray-level statistics. Texture analysis is an important task in many computer applications of Computer image analysis for classification, detection or segmentation of images based on local spatial patterns of intensity. Textures are replications, symmetries and combinations of various basic patterns, usually with some random variation (Cross and Jain, 1983). The major task in texture analysis is the texture segmentation of an image, that is, to partition the image space into a set of sub regions, each of which is homogeneously textured. In texture segmentation the goal is to assign an unknown sample image to one of a set of known texture classes.

Texture segmentation process involves 2 phases: the learning phase and the recognition phase. In the learning phase, the target is to build a model for the texture content of each. Texture class present in the training data generally comprises of images with known class labels. The texture content of the training images is captured with the chosen texture analysis method, which yields a set of textural features for each image. These features, which can be scalar numbers or discrete histograms (Kaplan, 1999) or empirical distributions, characterize given textural properties of the images, such as spatial structure, contrast, roughness, orientation, etc. In the recognition phase the texture content of the unknown sample is first described with the same texture analysis method. Then the textural features of the sample are compared to those of the training images with a classification algorithm and the sample is assigned to the category with the best match. Optionally, if the best match is not sufficiently good. According to some predefined criteria, the unknown sample can be rejected instead. The basic assumption in this approach to texture classification is based on a discrimination function using several texture characteristics. Texture is one of the fundamental features for describing image characteristics. The ability to effectively and efficiently classify and segment images based on textural features is of key importance in scene analysis, medical image analysis, remote ensing as well as automation of industrial applications. The 2 general approaches performing texture segmentation are analogous to the methods for image segmentation: region-based approaches that are based on the similarity of some texture property and try to identify regions of the image having a uniform texture and boundary-based approaches that detect boundaries where there are differences in texture and are based upon the detection of differences in texture in adjacent regions.

In this study the segmentation scheme involves 2 stages to ensure an accurate segmentation of micro calcification. The Modified combined spectral morphological unsupervised Image segmentation provides an approximate segmentation of micro calcifications. The random walker Image segmentation scheme provides an accurate segmentation. This study effectively utilizes the Ridgelet transform feature extraction technique, the sub band features of the complex wavelet Transform and the co occurrence features of the Ridgelet subband for mammogram Tumor classification. The Support Vector Machines (SVM) is introduced for classification of the Ridgelet decomposed medical Image features.

The Ridgelet transform features such as mean and standard deviation from subbands of Ridgelet decomposed image and the concurrence features from the subbands offers better classifications using mathematical classifiers (Kaizer, 1955) for natural texture images. The efficiency of kernel methods in the texture classifications in explained by Tuceryan and jain (1990) and Cross and Jain (1983).

A Receiver Operating Characteristics (ROC) analysis (Chang and Kuo, 1992, 1993) was used to evaluate the classification performances of the textural features. ROC analysis is based on statistical decision theory and has been applied extensively to the evaluation of clinical diagnosis. The ROC curve represents the relation ship between True Positive Fraction (TPF) and False Positive

Fraction (FPF) for variations of decision threshold (Laws, 1980). Ideally the area under the curve (A_z) should be as close to 1 as possible for an accurate classification technique.

MAMMOGRAM SEGMENTATION

The segmentation of micro calcification is performed in 2 steps:

- Approximate tumor segmentation by Modified combined morphological spectral unsupervised segmentation.
- Accurate Tumor segmentation using Random walker's segmentation scheme.

Approximate tumor segmentation by Modified combined morphological spectral unsupervised segmentation.

The watershed algorithm has the advantages of a region growing algorithm (the regions are spatially consistent, with boundaries forming a closed, connected set) while still making use of edge information, as captured by the gradient surface. However, noise, which tends to be amplified by the derivative operator, manifests itself as over segmentation. Clustering or thresholding approaches, on the other hand, which can exploit the bulk properties of feature distributions, often do not naturally satisfy spatial consistency. These methods also pose the problem of how to choose seeds for what essentially reduces to a classification task. In this research a combined, two-stage approach for image segmentation is proposed. The first stage is a refinement of the texture-gradient watershed algorithm described by Hill *et al.* (2003).

Approximate segmentation: The watershed Transform (Haralick *et al.*, 1973) is a morphological segmentation tool. When applied to a gradient image, the resulting image corresponds to local troughs in the gradient, while the boundary or "watershed" correspond to peaks. The important disadvantage as watershed transform is that it over segments the input Image. To over come this the Image is transformed into frequency domain using DT-CWT and the sub bands are filtered by using the anisotropic filtering scheme and band pass filtering scheme.

The anisotropic filtering scheme is used to punish the normalized sub band features $D_{i,è}$ (x,y) with values all about the same level, while amplifying those with a small number of relatively large peaks. The finer peaks

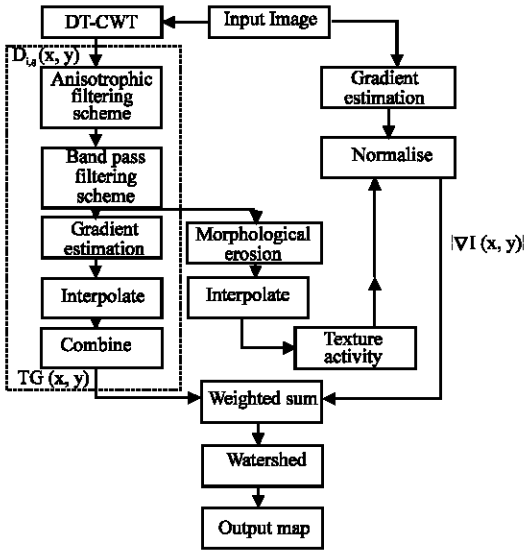

Fig. 1: Modified combined morphological spectral unsupervised segmentation scheme. Dotted line indicate that the procedures are followed for multi subbands of DT-CWT. (Discrete Time Complex wavelet transform)

correspond to the thin edges such as veins in mammograms and are filtered using Band pass filtering scheme.

As a whole the procedure may be summarized as:

- Compute a texture representation that characterizes a local area surround each pixel.
- Post process the texture features to make them suitable for meaningful gradient estimation.
- Generate Gradient images for each of the texture features as well as for gray scale intensity.
- Normalize/weight the contribution of each gradient image.
- Combine the various gradient images to from the single valued gradient surface.
- Segment by applying watershed transform to this surface.

The final single valued gradient surface is computed as the combination of the texture gradient and modulated Intensity gradient as illustrated in Fig. 1.

$$Gs(x,y) = (|\nabla\_I (x,y)|/ (Activity (x,y) * W_I)) + (TG (x,y))/ W_T$$

W_T, denotes median value of texture gradient while W_I is defined to be four times the median Intensity gradient. Activity (x,y) is calculated from the energy function (Fig. 2) (Challaghan and Bull, 2005).

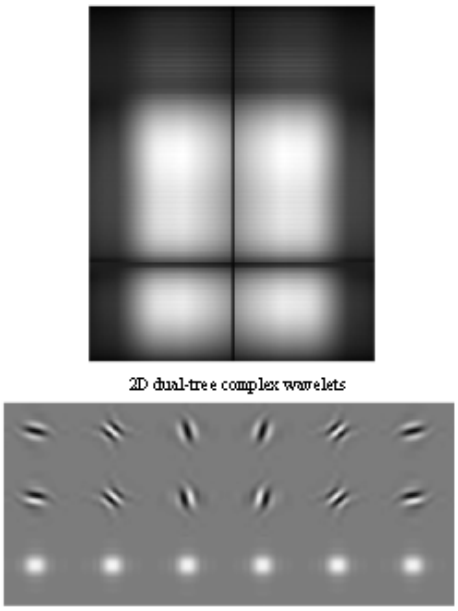

Fig. 2: DT- CWT sub bands used in Modified combined morphological spectral unsupervised segmentation scheme

The technique described by Callaghan and Bull (2005) median filters the image and the results of our modified combined morphological spectral unsupervised segmentation scheme and the combined morphological spectral unsupervised segmentation scheme is given in Fig. 3 for comparison.

Accurate segmentation of micro calcifications: The accurate tumor segmentation is performed using the random walks for Image segmentation technique (Bajcsy and Lieberman, 1976). The random walks for image segmentation is proposed for performing multi label, interactive image segmentation. Given a small number of pixels with user defined labels, one can analytically and quickly determine the probability that a random walker starting at each unlabeled pixel will first reach one of the prelabeled pixels. The output is shown in Fig. 4.

Specifically, it was shown in Bajcsy and Lieberman (1976) that the random walker algorithm has the following properties:

- The solution for the probabilities is unique.
- The expected value of the probabilities for an image of pure noise, given by identically distributed (not necessarily independent) random variables, is equal to those obtained on a uniform image.

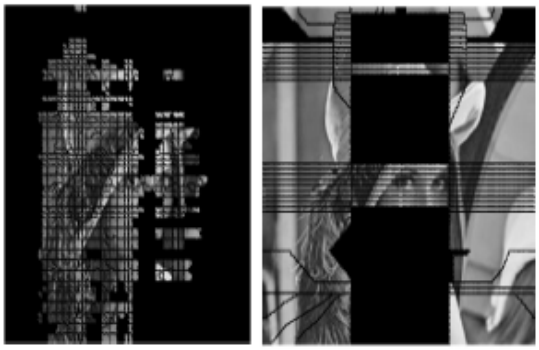

Fig. 3: Output of the Modified combined morphological spectral unsupervised segmentation scheme using Lena image

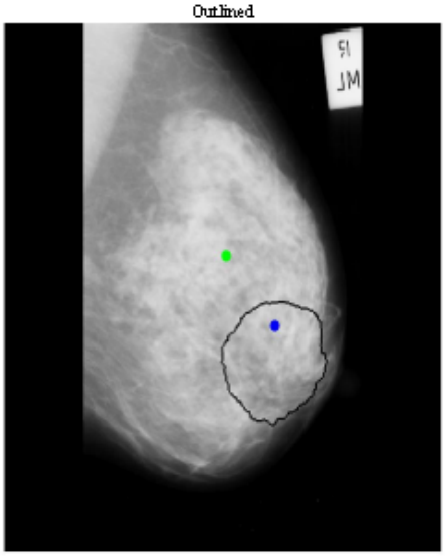

Fig. 4: Output image showing the segmented tumor region

- The expected value of the probabilities in the presence of random, uncorrelated weights is equal to the probabilities obtained by using weights equal to the mean of each random variable.

The ridgelet transform allows representing edges and other singularities in a more efficient way.

Features such as mean and standard deviation are calculated from each of the ridgelet sub bands of the segmented micro calcifications and stored in the database for training. In order to improve the value of A_z, the cooccurence matrix is formed for every subband of DRT and features such as energy, entropy, contrast, shade and prominence are calculated Arivazhagan and Ganesan (2003) and stored in the database. The Ridgelet decomposed image is shown in Fig. 5. Classification is performed by using the support vector Machine.

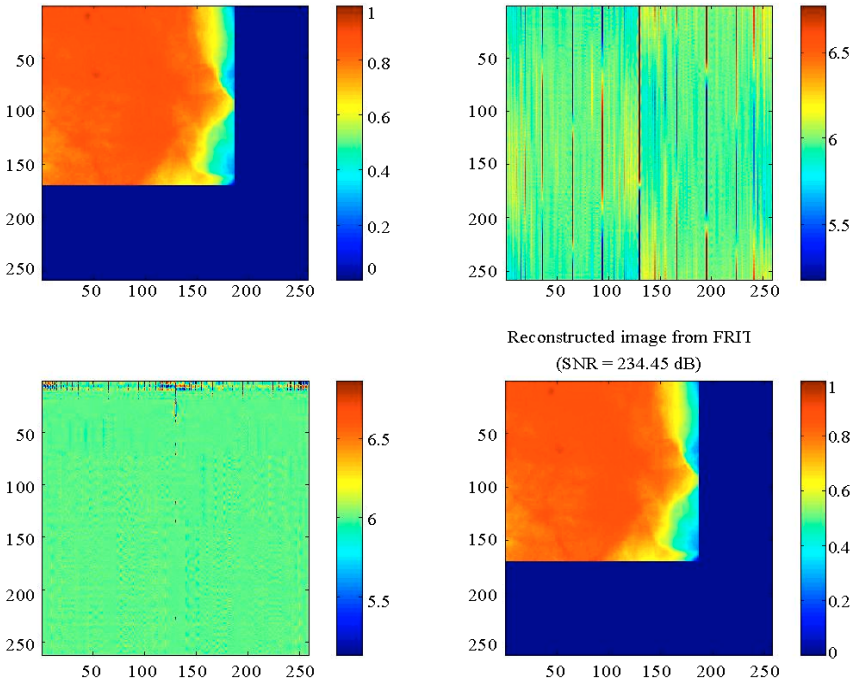

Fig. 5: Ridgelet transformed segmented image

Traditionally neural networks had been used for classification which is based on Empirical Risk Minimization (ERM). SVM Leo (2006) was developed by Vapnik and had become popular tools for data mining. The formulation embodies the Structural Risk Minimization (SRM), which is superior to empirical risk minimization. SRM minimizes the upper bound on expected risk as supposed to ERM that minimizes the error on training data (Fig. 6). So, SVM generalizes much better. There are many linear classifier that can separate data but SVM only maximizes the margin i.e. the distance between it and the nearest data point in each class.

We have N training data $\{(x_1,y_1), (x_2,y_2),\ldots.. (x_N,y_N)\}$ Where $x_i \in R^d$ and $y_i \in \{+1,-1\}$. It needs to be classified using a linear hyper plane classifier

$$f(x) = \text{sgn}(w.\ x - b)$$

This hyper plane will have maximum distance between each class (Fig. 7). This hyper plane

$$H : y = w.\ x - b = 0$$ and two hyper planes parallel to it

$$H_1 : y = w.\ x - b = +1$$
$$H_2 : y = w.\ x - b = -1$$

With no data points between H_1 and H_2 and distance between H_1 and H_2 maximized. Some training point will lie on the hyper plane H_1 and H_2, they are called support vector machines because they define the separating plane and the other training points can be removed or moved provided they don't cross the planes H_1 and H_2. The distance between hyper plane H_1 and H_2 is $2/\|w\|$. To maximize the distance between the two data sets or between H_1 and H_2 we need to minimize $\|w\|$ with the condition that there are no data points between H_1 and H_2:

$$w.\ x - b \geq +1 \text{ for } y_i = +1$$
$$w.\ x - b \geq -1 \text{ for } y_i = -1$$

Combining the above two equations,

$$y_i\ (w.\ x - b) >= 1$$

So the problem of maximizing the distance between hyper plane H1 and H2 is formulated as

$$\min \tfrac{1}{2}\, w^T w \text{ subject to } y_i\ (w.\ x - b) \geq 1$$

This is a convex quadratic problem in w, b in a convex set. The solution is found by solving using lagrangian method by introducing lagrangian multipliers. It is easier to solve using lagrangian dual equation given by

$$L_D = \Sigma_i\, \alpha_i - \Sigma_i\, \Sigma_j\, \alpha_i\, \alpha_j\, y_i\, y_j\, x_i\, .x_j$$

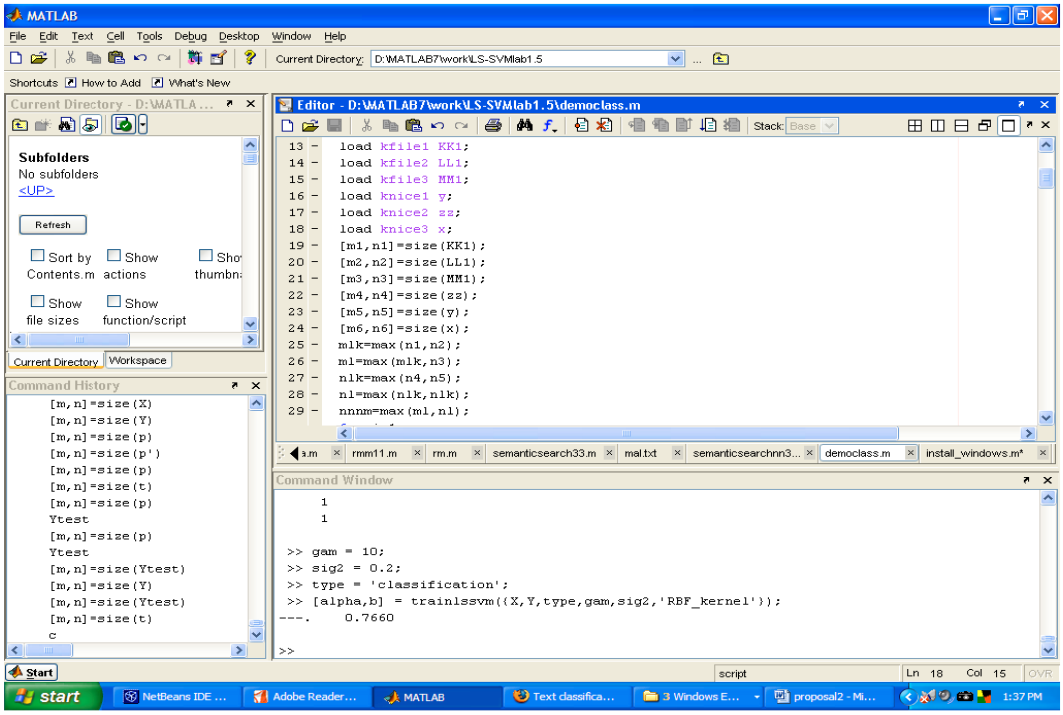

Fig. 6: Plot to show the training session in SVM classification module and the time for execution

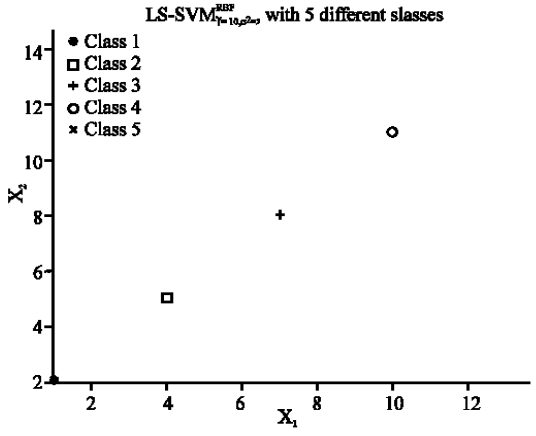

Fig. 7: Plot to show the multi class classification ability of LS SVM

The significance of the above equation is that the training input vectors appear only as dot product. So when the data is not linearly separable it is required to transform the data into a higher dimensional. This causes complex calculations is neural networks but in SVM as data appear only as an dot product all calculation can be carried explicitly in low dimension if an kernel function exists for

$$L_D = \Sigma_i \alpha_i - \Sigma_i \Sigma_j \alpha_i \alpha_j y_i y_j \Phi(x_i).\Phi(x_j)$$

$$\text{as } \Phi(x_i).\Phi(x_j) = K(x_i, x_j)$$

Where, K is the kernel function. This is equivalent as the dot product in high dimension is equal to kernel function in input space. The common kernel function used is Gaussian kernel,

$$K(xi, xj) = e^{-\| xi - xj \| 2/\sigma 2}$$

Mercers condition determines whether a function g(x) can be used as a kernel or not,
$\int g(x)^2 dx$ should be finite.

RESULTS AND DISCUSSION

In this experiment a data set containing 75 X-ray mammogram is used for analysis. In the modified combined morphological spectral segmentation scheme, the discrete time-complex wave let transform contains imaginary terms and the absolute value calculation is performed to have real coefficient subbands.

The feature vector for each image is calculated from the sub bands of the DRT decomposition. The Ridgelet transform was performed by using the FRIT Tool box. The support vector machine learning scheme involves training and testing procedure. In the testing procedure a segmented mammogram is decomposed using FRIT and its features are calculated to form the feature vector similar to that of the learning phase. The support vectors

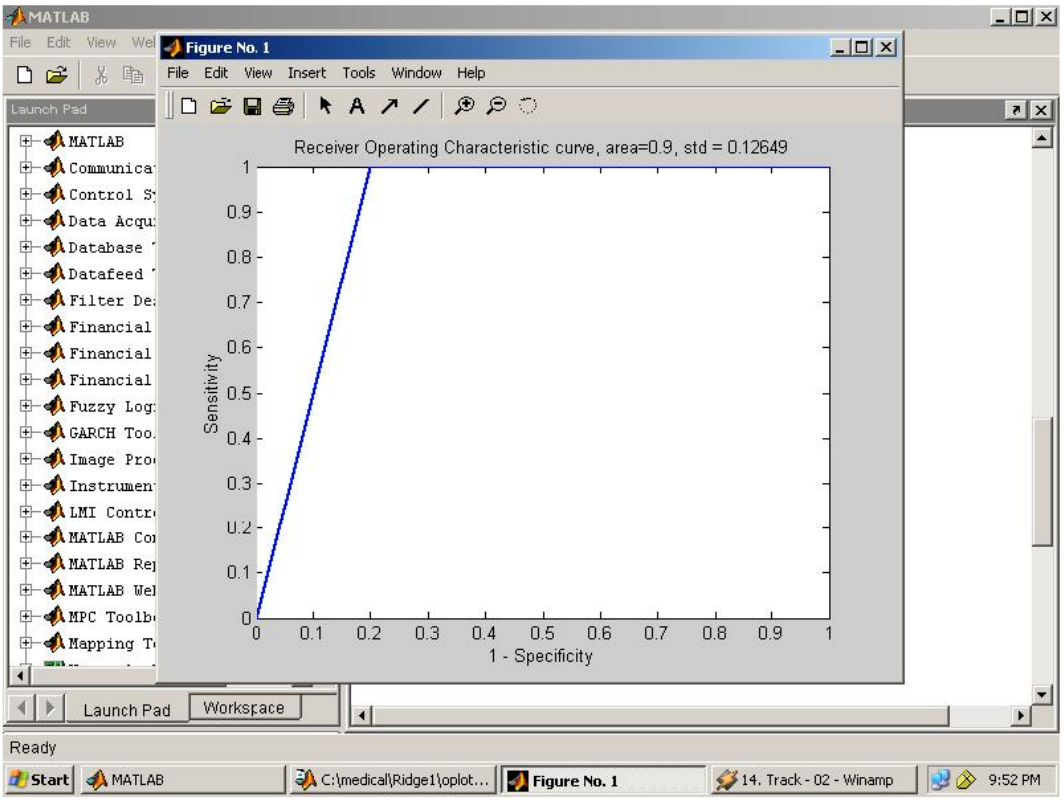

Fig. 8: Plot to show the receiver operating curve characteristics

machine was implemented by using quadp function of MATLAB Software. The classification performance of the SVM classifier is evaluated using ROC analysis. To form an efficient feature vector pattern the FAST ICA analysis is used together with the Mahalanobis distance discrimination.

The value of A_z is found to be 0.9 for the ridgelet sub band features and 0.928 for the concurrence features of the ridgelet subband. The ROC diagram is shown in Fig. 8.

CONCLUSION

From the mammogram analysis it is inferred that the better classification can be achieved due to Ridgelet features and better segmentation scheme. The highest value of A_z calculated in this analysis is found to the 0.928. This analysis effectively implements the novel combination of Ridgelet features and support vector machine for mammogram classification. The texture analysis scheme is effectively adapted to the mammogram classification to improve the value is A_z. It can also be verified that the Ridgelet transform has good directional selectivity and also able to locally and sparsely represent the signal. It can also be inferred that the support vector machines offer better classification gain. In future, this work can be extended for an efficient classification system design with minimum features having better discrimination ability.

REFERENCES

Argenti, F., L. Alparone and G. Benelli, 1990. Fast algorithms for texture analysis using co-occurrence matrices. IEEE Proceedings-F, 137: 443-448.

Arivazhagan, S. and L. Ganesan, 2003. Texture classification using wavelet transform. Pattern Recognition Letter, pp: 1513-1521.

Bajcsy, R. and L. Lieberman, 1976. Texture gradient as a depth cue. Computer Graphics and Image Processing, 5: 52-67.

Chang, T. and C. Kuo, 1992. Texture classification with tree-structured wavelet transform. Pattern Recognition Methodology and Systems, Proceedings of the 11th IAPR International Conference.

Chang, T. and C. Kuo, 1993. Texture analysis and classification with tree-structured wavelet transform. IEEE Transactions Image Processing, 2: 429-441.

Cross, G.R. and A.K. Jain, 1983. Markov random field texture models. IEEE Transactions on Pattern Analysis Machine Intelligence, PAMI-5, pp: 25-39.

Haralick, R.M., K. Shanmugam and I. Dinstein, 1973. Texture features for image classification. IEEE Transactions on System Man Cybernat, 8: 610-621.

Hill, P., C. Canagarajah and D. Bull, 2003. Image segmentation using a texture gradient-based watershed transform. IEEE Transactions on Image Processing, 12: 1618-1633.

Kaizer, H., 1955. A quantification of textures on aerial photographs. Technical Note 121, AD 69484, Boston University Research Laboratory.

Kaplan, L.M., 1999. Extended fractal analysis for texture classification and segmentation. IEEE Transactions on Image Processing, 8: 1572-1585.

Laws, K.L., 1980. Rapid texture identification. Proc. SPIE 238: 376-380.

Leo Grady, 2006. Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis Machine Intelligence, 28: 1768-1783.

Patrizio Campisi, Alessandro Neri and Gaetano Scarano, 2002. Model based rotation invariant texture classification. IEEE International Conference on Image Processing, pp: 117-120.

Rama Chellappa and Shankar Chatterjee, 1985. Classification of textures using Gaussian Markov random fields. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP, 33: 959-963.

Robert, J. O'Callaghan and David R. Bull, 2005. Combined Morphological-Spectral Unsupervised Image Segmentation. IEEE Transactions on Image Processing, Vol. 14.

Smith, J. and S.F. Chang, 1994. Transform features for texture classification and discrimination in large image databases. IEEE International Conference Image Processing.

Tuceryan, M. and A.K. Jain, 1990. Texture Segmentation Using Voronoi Polygons. IEEE Transactions on Pattern Analysis Machine Intelligence, PAMI, 12: 211-216.

Tuceyran, M. and A.K. Jain, 1998. Texture Analysis, in Handbook of Pattern Recognition and Computer Vision. 2nd Edn. World Scientific Publishing Co., Chapter 2.1, pp: 207-248.