

Research Article

Content Based Mammogram Retrieval based on Breast Tissue Characterization using Statistical Features

K. Vaidehi and T.S. Subashini

Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, 608002, India

Abstract: The aim of the study is to retrieve the similar mammographic images based on the type of breast tissue density of the given query image. Statistical descriptors were extracted from the candidate blocks of the breast parenchyma. The mean of extracted features are fed into the SVM classifier for classification of the tissue density into any of the three classes namely dense, glandular and fatty and the classification accuracy obtained is 91.54%. After classification the mammogram images along with its feature vector are stored into three separate databases based on tissue type. Then K-means clustering algorithm is used to divide each database into 2 clusters. For content based retrieval of the mammograms based on the given query image, first the query image is classified into any of the three tissue class. Then the feature vector of the query image is compared with the two cluster centroids of the corresponding class, so as to confine the search within the closest cluster. Top 5 similar images are retrieved from its corresponding class database. Euclidean distance based k-NN is used for mammogram retrieval and this study obtained the highest precision rate ranging between 98 and 99%.

Keywords: Breast tissue density, CBIR, mammography, statistical features, SVM

INTRODUCTION

Globally, breast cancer is the second leading lethal disease among women. In India, breast cancer cases are expected to be double by 2025. The Indian Cancer Society has declared 2013 as a breast cancer awareness year and is taking various initiatives to create awareness in people. Achieving benign detection and adequate treatment will lead to better long term survival as well as a better quality of life.

Mammography is the highly used imaging modality for detecting and diagnosing breast cancer. Digital mammogram is a mammography system which takes an electronic image of the breast and stores it directly in a computer. Computer-aided detection/diagnosis can be applied easily to the digital mammogram. In the medical domain, computers have become a vital role for image acquisition to image analysis. There are two major processes involved in the mammogram image interpretation: Computer Aided Detection (CADE) system and Computer Aided Diagnosis (CADi) collectively referred to as CAD. CADE systems are usually marking the abnormal patterns in the medical imaging. CADi systems evaluate the abnormal tissue patterns in the medical image analysis. CAD is fundamentally based on the image processing and pattern classification techniques which have been used to support the physician/radiologist to

make important medical decisions through physician-computer interaction.

Content Based Image Retrieval (CBIR) is a technique for retrieving similar images based on the content of the query image that may aid radiologist to interpret the visual content of the medical images. The term content may describe color, texture, shape or any other useful information of the image. Day to day medical image collections are increased rapidly, CBIR is the only way to manage the large medical image collections and also helps the radiologist in decision making. CBIR coupled CAD has been an active research area in medical imaging. The proposed CBIR system based on breast density is useful for identifying and retrieving similar mammogram images from huge mammogram databases and archives. Here content means some property extracted from the image such as textures, shapes and interest points which helps to identify the breast tissue type. These contents are extracted using image processing techniques and are called feature vectors. Similarity between the two images (the query image and the image from the database) is computed by first indexing the query image based on tissue type and then measuring the distance between the feature vectors corresponding to the mammogram images having similar tissue type of that of the query image in question.

Corresponding Author: K. Vaidehi, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, 608002, India

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

The objective of this study is carried out in two steps viz.:

- Classifying the mammograms based on the breast tissue characterization
- Retrieving similar mammogram images from the predicted class relevant to the query image

In the first step, candidate blocks are selected from the artifact less and pectoral muscle removed breast parenchyma. From these candidate blocks statistical features are extracted. The extracted features are fed into SVM classifier for classifying the breast tissue into dense, glandular and fatty. In the second step, classified mammograms are stored into 3 different databases. To speed up the retrieval, three class databases are clustered into 2 groups using K-means algorithm. Then k-NN algorithm is used to retrieve the similar mammographic density images from the clustered databases.

LITERATURE REVIEW

CBIR is a demanding technique in the medical field since it can retrieve the relevant images from the large database based on the image content and provide the decisions to the physician. The first study on CAD system in mammography was developed by Winsberg *et al.* (1967). Many researchers developed automatic breast density characterization with different CAD approaches (Wang *et al.*, 2003; Sheshadri and Kandaswamy, 2007; Oliver *et al.*, 2008; Tagliafico *et al.*, 2009; Oliver *et al.*, 2010; Subashini *et al.*, 2010; Tzikopoulos *et al.*, 2011; Liasis *et al.*, 2012; Muštra *et al.*, 2012). Even though many papers on automatic classification of mammograms are present in the literature, CBIR for mammograms are in the benign stage of development.

Optimal approaches of CBIR based CAD schemes in digital mammography are reviewed and their performance are assessed in Zheng (2009). The authors in Kinoshita *et al.* (2007) developed a mammogram image retrieval system, where visual features such as shape, histogram, texture, moments, granulometric and radon features are used to measure the similarity of breast density patterns. The Kohonen self-organizing map was used to perform the retrieval of mammograms. Since many features were used in this study, it takes a lot of computational time and it may not be suitable to medical images. The work in Deserno *et al.* (2012) applied 2DPCA features for retrieving mammogram images based on breast density and lesions. These features were evaluated using Gaussian kernel and polynomial kernel of SVM for classification. They obtained average precision rates are in the range of 72.14 to 80.64%. Self-Organizing Map (SOM) network was used to cluster the images based on the breast tissue patterns and Genetic Algorithm (GA) was applied for retrieving the mammogram images in Jose and Mythili (2009).

Commonly used image processing algorithms for detection of mass and calcification are surveyed in Bozek *et al.* (2009). The author in Rangayyan *et al.* (2007) reviewed the detection of subtle signs of breast cancer using CAD. The work in Wei *et al.* (2012) proposed a mammogram retrieval system, where similarity of mass lesions is based on the shape and margin features. The author El-Naqa *et al.* (2004) demonstrated in his experiment, the two stage learning machine (MSVM-SVM and MSVM-GRNN) based framework for modeling human perceptual similarity for CBIR. It was developed and evaluated for retrieval of clinical mammograms containing micro calcification clusters. Retrieval driven classification approach with an adaptive SVM is proposed in Wei *et al.* (2009a). In this study, mammogram image database containing micro calcifications clusters were used. Genetic algorithm is used for feature selection to improve the ranking quality of medical image retrieval and it is experimented with three image datasets, comprising breast images and lung nodules (Da Silva *et al.*, 2011). Similarity measure models such as SVM, DANN and Ncut were used for retrieving the mammograms in Wei *et al.* (2009b). This study concluded that supervised learning approach achieves significant improvement than unsupervised learning methods for content based mammogram retrieval.

PROPOSED METHODOLOGY

The main goal of this study is to retrieve the mammogram images relevant to the breast tissue density of the query mammogram. This study helps the radiologist ineffective diagnosis and decision making. The various modules of the proposed method are shown in Fig. 1. In this proposed CBIR system, a two-step approach is used to retrieve the similar mammogram images. The first step classifies the mammogram images based on the mammographic density into three classes of breast tissue namely fatty, glandular and dense using statistical features. The second step then retrieves the most similar images within the predicted class.

Step 1: Classification of mammograms: The initial step is preprocessing, which includes artifact removal, image enhancement and pectoral muscle removal. In this study, the breast region is segmented into blocks and features were extracted from these blocks. The region of the whole breast parenchyma is splitted into 9×9 blocks and the interleaved 9×9 blocks which contains less than 60% of black pixels is only considered as candidate blocks. From these blocks statistical features are extracted. The extracted features are fed into SVM classifier for classifying the breast tissue density into dense, glandular and fatty.

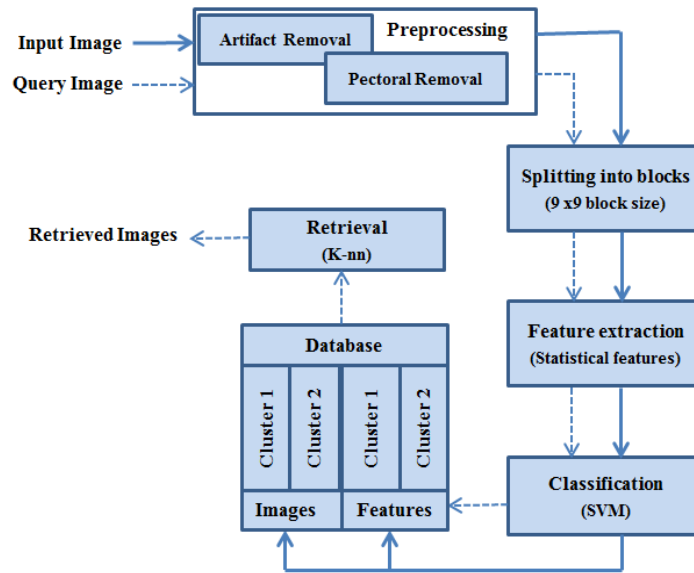


Fig. 1: Various modules of the proposed method

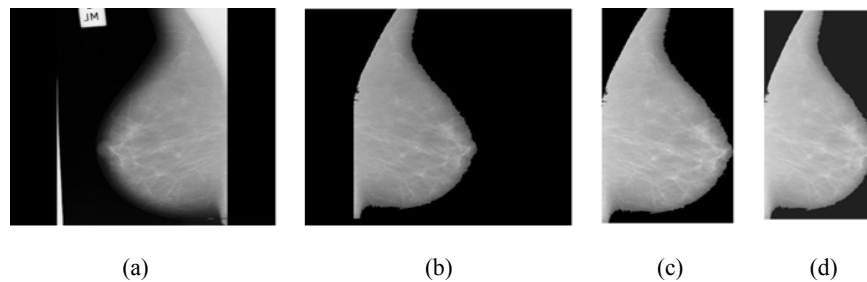


Fig. 2: (a) Original image of size 1024×1024 (mini-mias), (b) preprocessed image of size 1024×1024, (c) bounded breast region alone of size 987×476, (d) resized image of size 981×468

Step 2: Mammogram retrieval: In the second step, the classified mammograms are stored in a database along with the features of the mammograms. Before applying k-NN, Each class database is grouped into 2 clusters by using K-means clustering algorithm. The mammogram images similar to the query mammogram image are retrieved using k-NN algorithm. This method has been tested on the Mammographic Image Analysis Society digital mammogram database (Mini-MIAS) (Suckling *et al.*, 1994).

Preprocessing: To reduce false positive, false negative and the misclassification rates, the mammogram images are first preprocessed. High intensity radio opaque artifacts such as labels, wedges present in the mammogram image are removed in this stage since it may lead to misclassification. Further the mammogram image quality is improved by noise reduction. In Mediolateral Oblique views (MLO) of mammograms, pectoral muscles are present and which affects the CAD results in the detection of breast densities (Kwok *et al.*,

2000). So, the pectoral muscle removal is inevitable before the mammogram is further processed for classification.

A mammogram contains two distinctive regions: breast region and non-uniform background region as shown in Fig. 2a. Most of the background region contains high intensity artifacts. For artifact removal, first Otsu's global thresholding method is used to binarize the image. Otsu's global thresholding method maximizes the between class variance i.e., well thresholded classes should be distinct with respect to the intensity values of their pixels (Otsu, 1979). Next, a connected component labeling algorithm is utilized to recover the biggest region that is breast region along with pectoral muscle. Connected component labeling scans an image, pixel by pixel (from top to bottom and left to right) in order to locate the connected pixel regions, i.e., region of pixels that share the same intensity values. After artifact removal, median filtering is applied for removing noise and morphological techniques are used for enhancing the image. The pectoral muscle is removed using our previous study (Vaidehi and Subashini, 2013) which uses a straight

line method. Pectoral muscles locate on the right or left top corner depending on the view of the image. To make processing easy the right MLO mammogram is first flipped to the left side before removing the pectoral region. To remove the pectoral muscle a straight line method proposed in our previous study (Subashini *et al.*, 2010; Vaidehi and Subashini, 2013) is applied to the top left quadrant which contains the pectoral region. Artifact and pectoral muscle removed image is shown in Fig. 2b.

Splitting into blocks: The preprocessed image is of size 1024×1024 . But the image contains both background black pixels and the region of interest namely the breast parenchyma. Since the statistical features here to be extracted only for the breast region it is first cropped using boundary box. However the size of the bounding box varies from one mammogram to another mammogram depending upon the size of the breast, so resizing the images to the same size would not be a correct option for extracting features since that will not result in correct values. To overcome this problem the image is resized to the nearest multiple of 9 since it is proposed to extract features from 9×9 blocks.

Assuming that the bounded region is of size $m \times n$ and as we need to divide the bounded region into non-overlapping sub blocks of 9×9 , we take mod 9 with m and n values. The resized image is obtained as follows (Fig. 2c and d):

$$\begin{aligned} mr &= m \bmod 9 \\ nr &= n \bmod 9 \\ \text{new } m &= m - mr \\ \text{new } n &= n - nr \end{aligned}$$

Example: Assuming the bounded region is of size 987×476 .

Here,

$$\begin{aligned} m &= 987 \text{ and } n = 476 \\ mr &= 987 \bmod 9 = 6 \\ nr &= 476 \bmod 9 = 8 \\ \text{new } m &= m - mr = 987 - 6 = 981 \\ \text{new } n &= n - nr = 476 - 8 = 468 \\ \text{Resized image is of size} &= \text{new } m \times \text{new } n \\ &= 981 \times 468 \end{aligned}$$

In our previous study (Subashini *et al.*, 2010) the segmented breast region is used to extract features for classification. But the results show the dense and glandular tissues were misclassified as there is only a minor intensity variation between the glandular and dense breast. Since the statistical features represent the entire breast region, the misclassification rate was high with respect to glandular and dense tissues. So to overcome this, in this study, it is proposed to take features from interleaved 9×9 blocks, which are able to

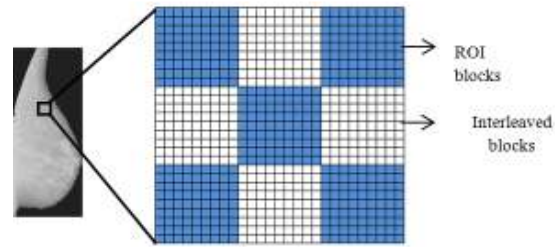


Fig. 3: Sample 27×27 image portion is splitted into 9×9 interleaved blocks

capture the sufficient texture information of the mammogram. Then the mean of all the candidate block features is taken as the feature vector of a particular mammogram.

The sample 27×27 image portion of original image splitted into 9×9 blocks with interleaved blocks are shown in the Fig. 3. The interleaved blocks which contains less than 60% of black pixels is only considered for further processing and it is denoted as candidate blocks. The remaining blocks containing more than 60% black pixels are discarded because black pixels represent the background pixels which need not be processed.

Feature extraction: Features are the most useful information of the particular image which plays a major role in medical image analysis. Feature extraction is an important phase for achieving good classification and retrieval results. Extracting numerical features from the region of interest which represent the particular image is called feature vectors. CAD coupled CBIR systems mainly depends on the feature vectors for perfect classification and retrieval of images.

Many number of feature extraction techniques are used for breast density classification and CAD coupled CBIR in digital mammograms (Subashini *et al.*, 2010; Tzikopoulos *et al.*, 2011; Liasis *et al.*, 2012; Muštra *et al.*, 2012; Chandy *et al.*, 2014; Vález *et al.*, 2011). Statistical features give more significant information in pattern recognition area. It characterize texture by the statistical distribution of the image gray level intensity (Chandy *et al.*, 2014; Srinivasan and Shobha, 2008). Statistical features such as mean, standard deviation, entropy, skewness and kurtosis are extracted from the candidate blocks. Then mean is calculated from all the candidate block features for reducing the dimension of feature vector. Summary of statistical descriptors are given in the Table 1. These feature vectors are fed into the SVM classifier for classifying the breast tissues namely dense, fatty and glandular.

Classification: The classification phase determines the class of a given mammogram based on the features extracted using statistical descriptors. In this study,

Table 1: Statistical features used

Feature	Expression
Mean	$\mu = \frac{\sum_{ij} X_{ij}}{N}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{ij} (X_{ij} - \mu)^2}{N}}$
Skewness	$\frac{\sum_{ij} (X_{ij} - \mu)^3}{N\sigma^3}$
Kurtosis	$\frac{\sum_{ij} (X_{ij} - \mu)^4}{(N-1)\sigma^4}$
Entropy	$S = -\sum_{i=1}^{\Omega} X_i \log_2(X_i)$

retrieval mainly depends on classification part because accurate classification predicts the better retrieval performance. Based on the literature (Subashini *et al.*, 2010; Tzikopoulos *et al.*, 2011; Liasis *et al.*, 2012; Muštra *et al.*, 2012; Vallez, 2011) SVM has chosen as a classifier for classifying the mammograms into three classes namely dense, glandular and fatty.

SVM classifier: In the 1990s, support vector machine learning algorithm was developed by Vapnik (1998). SVM can be used for pattern classification and nonlinear regression. Either linearly or non-linearly, it maps the input vectors into a high dimensional feature space with the help of kernel functions. Some common kernels using SVM are Polynomial (homogeneous), Polynomial (inhomogeneous), Gaussian radial basis function and Sigmoid. The goal of SVM is finding the optimum separating hyper plane that separate the feature vectors into two classes.

Database indexing: The classified images and their corresponding features are stored into three different databases representing each of the three classes namely dense, glandular and fatty. In order to reduce the computation time for searching images relevant to the query image, the database of each class is first grouped into 2 clusters using K-means algorithm.

Kmeans clustering is an unsupervised classifier, which classifies the features into K number of clusters, where K is a positive integer. Clustering is done by minimizing Euclidean distance between data and corresponding cluster centroids.

Kmeans clustering algorithm:

1. Initialise K centroids
2. Compute the distance between each feature vector to the centroids
3. Assign the feature vector to the centroid whose distance is minimum
4. Re-estimate the centroids
5. Repeat the steps 2-4 until there is no change in centroids or for a fixed number of iterations

Retrieval: The query image is classified into any one of the three classes. The search is limited to the

corresponding query class database alone for retrieving similar images. The feature vector of the query image is compared with both the cluster centroid of the corresponding class. The minimum distance cluster is alone further searched to retrieve similar mammogram images. Top five relevant images are retrieved using k-NN algorithm. Euclidean distance is used as the distance metric in k-NN algorithm.

EXPERIMENTS AND RESULTS

Dataset: The Mammography Image Analysis Society (MIAS) has created a mammogram database with ground truth such as character of background tissue, class of abnormality and severity of abnormality and location of abnormality. This database is used in this study. It contains both the right and left breast images of the same patient and each image in the database is 1024×1024 pixels (Suckling *et al.*, 1994).

Classification and retrieval results: The MIAS database contains 322 mammogram images including normal and abnormal mammograms. The entire database is chosen for this study which contains 112 dense mammograms, 104 glandular mammograms and 106 fatty mammograms. Statistical features namely mean, standard deviation, skewness, kurtosis and entropy are obtained from the mammograms are fed to SVM for classification. Leave one out procedure has been adopted in testing the performance of the SVM classifier. The SVM with Gaussian kernel is trained to provide a value of 0 for dense tissue, 1 for glandular tissue and 2 for fatty tissue mammograms. The classification accuracy obtained is shown in Table 2.

From the table the overall classification rate of the classifier is calculated as 91.54 and 93.75 and 93.39% classification accuracy was obtained in the case of dense and fatty breast respectively because the skewness produces negative results for fibro glandular breast and positive for fatty breast. (87.5%) accuracy is obtained in the case of glandular. It is misclassified as dense because there is only a slight intensity variation between the two tissues. After classifying the mammograms based on the tissue type the mammogram image along with its feature vector is stored in the database corresponding to its class.

In the second step, the three databases thus obtained after classification of the mammograms is subjected to K-means clustering algorithm to form 2 clusters. This is done to speed up the retrieval process.

Table 2: Accuracy of SVM classifier

Type	Dense	Glandular	Fatty
Total images taken	112	104	106
Correct classification	105	91	99
Classification rate	93.75%	87.50%	93.39%

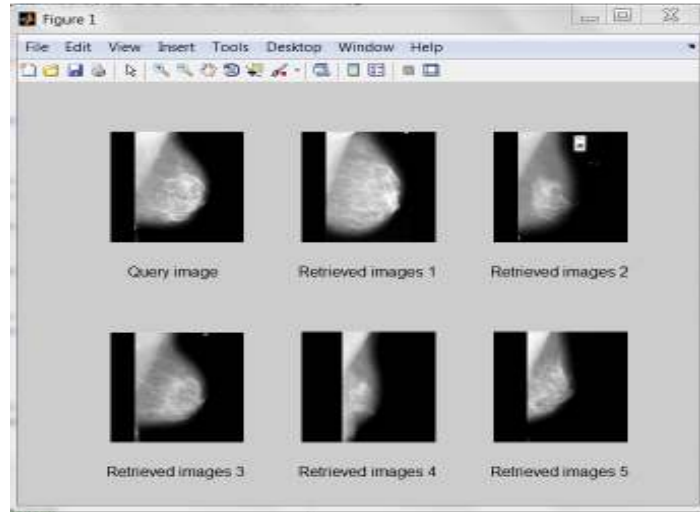


Fig. 4: Glandular images

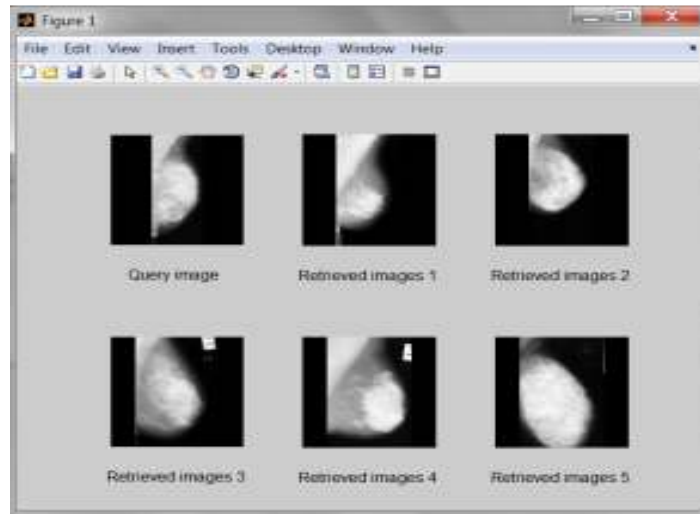


Fig. 5: Dense images

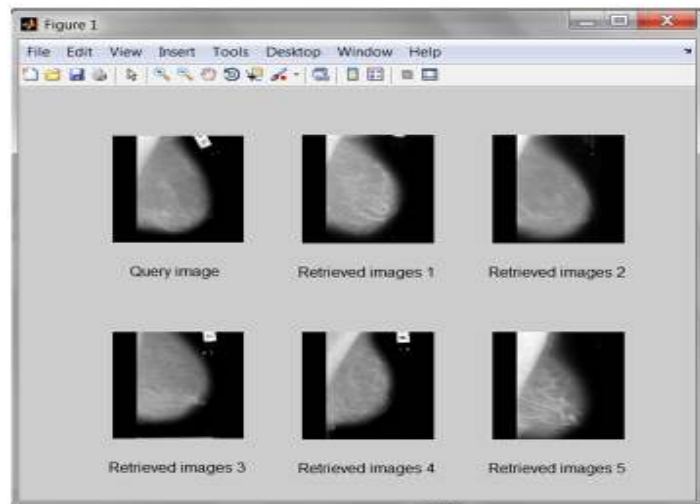


Fig. 6: Fatty images

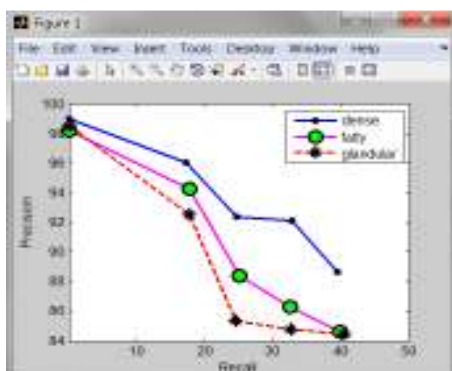


Fig. 7: Precision-recall curve of three different tissues class

The query image is first classified into any of the three tissue class and the similar image is searched only in the corresponding database. The database of the query feature vector is compared with the 2 cluster centroids to find the nearest cluster to the query image. Then k-NN using Euclidean distance as the distance metric is used to retrieve the top 5 similar images. The 5 retrieved results are found which are relevant to the query image of each class is shown in the Fig. 4 to 6.

Performance analysis: Precision and recall are the standard performance metrics used to measure the effectiveness of the CBIR system in retrieving most similar images. In the P-R graphs, the x-axis represents the recall and y-axis represents the precision. A precision-recall graph with higher initial value that tails off more quickly indicates that the corresponding algorithm performs relatively better (El-Naqa *et al.*, 2004). Precision and recall are computed as shown in Eq. (1) and (2):

$$\text{Precision} = \frac{\text{No. of relevant images retrieved}}{\text{No. of images retrieved}} \times 100 \quad (1)$$

$$\text{Recall} = \frac{\text{No. of relevant images retrieved}}{\text{No. of relevant images in the database}} \times 100 \quad (2)$$

The class wise precision and recall rate of 5 different k values are shown in the Fig. 7. Thirty query images were randomly selected from each of the three class. The system returns (retrieves) five similar images (descending rank). In a precision-recall graph, the first point, which is the left most point in a curve, represents the highest precision rate. The three tissue classes shows the highest precision ranging between 98 to 99% with k = 1. Dense tissue produces slightly better retrieval performance than the other two tissue classes.

CONCLUSION

In summary, we investigated and assessed the relationship between the performance of the CBIR based CAD scheme in this study. In our previous study we used 43 normal images of MIAS database for

density classification, which gave an accuracy of 95.44%. In this study all the 322 images of the Mini MIAS was tested for evaluating the efficiency of the proposed methodology for CBIR. The classification accuracy obtained was 91.54% and the precision ranges between 98 to 99% depending upon the tissue type.

The proposed study was carried out using MATLAB. Unlike our previous study which was based on the features extracted from the entire breast parenchyma for tissue classification. In this study, the features obtained from the interleaved sub blocks of the breast region, were able to capture the local features and were effective in classification of the tissue type. After classifying the breast tissue into any of the three types namely dense, glandular and fatty and the content based mammogram retrieval is achieved using k-NN algorithm with Euclidean distance as the distance measure. The query image is first characterized into any of the three class before searching and retrieving images similar to the query from the corresponding class database. To speed up retrieval, the three class database is first clustered into 2 groups using K-means algorithm. The precision obtained is 98 to 99% depending upon the type of the breast tissue. This indicates that the proposed system based on breast tissue density can be effectively used in content based mammogram image retrieval.

ACKNOWLEDGMENT

The authors would like to thank Dr. M.K. Sivakkolunthu, Professor of Radiology, Raja Muthiah Medical College Hospital, Annamalai Nagar for his valuable help and comments in carrying out this study. The study has been done under University Grants Commission (UGC) Major Research Project. The financial support of UGC is greatly acknowledged with appreciation.

REFERENCES

- Bozek, J., M. Mario, D. Kresimir and G. Mislav, 2009. A survey of image processing algorithms in digital mammography. In: Grgic, M. *et al.* (Eds.), *Rec. Advan. in Mult. Sig. Process. and Commun.*, SCI 231, Springer-Verlag, Berlin, Heidelberg, pp: 631-657.
- Chandy, D.A., J.S. Johnson and S.E. Selvan, 2014. Texture feature extraction using gray level statistical matrix for content-based mammogram retrieval. *Multimed. Tools Appl.*, 72(2): 2011-2024.
- Da Silva, S.F., M.X. Ribeiro, J.D.E.S. Batista Neto, C. Traina-Jr. and A.J.M. Traina, 2011. Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decis. Support Syst.*, 51(4): 810-820.

- Deserno, T.M., M. Soiron, J.E.E. de Oliveira and A. de A. Araújo, 2012. Towards computer-aided diagnostics of screening mammography using content-based image retrieval. Proceeding of 24th SIBGRAPI Conferences on Graphics, Patterns and Images (Sibgrapi), pp: 211-219.
- El-Naqa, I., Y. Yang, N.P. Galatsanos, R.M. Nishikawa and M.N. Wernick, 2004. A similarity learning approach to content-based image retrieval: Application to digital mammography. *IEEE T. Med. Imaging*, 23(10): 1233-1244
- Jose, T.J. and P. Mythili, 2009. Neural network and genetic algorithm based hybrid model for content based mammogram image retrieval. *J. Appl. Sci.*, 9(19): 3531-3538.
- Kinoshita, S.K., P.M. de Azevedo-Marques, R.R. Pereira Jr, J.A.H. Rodrigues and R.M. Rangayyan, 2007. Content-based retrieval of mammograms using visual features related to breast density patterns. *J. Digit. Imaging*, 20(2): 172-190.
- Kwok, S.M., R. Chandrasekhar and Y. Attikiouzel, 2000. Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection. Proceeding of Information Systems Conference, Perth, Australia, pp: 67-82.
- Liasis, G., C. Pattichis and S. Petroudi, 2012. Combination of different texture features for mammographic breast density classification. Proceeding of IEEE 12th International Conference on Bioinformatics and Bioengineering (BIBE), pp: 732-737.
- Muštra, M., M. Grgić and K. Delač, 2012. Breast density classification using multiple feature selection. *Automatika*, 53(4): 362-372.
- Oliver, A., J. Freixenet, R. Marti, J. Pont, E. Perez, E.R.E. Denton and R. Zwiggelaar, 2008. A novel breast tissue density classification methodology. *IEEE T. Inf. Technol. B.*, 12(1): 55-65.
- Oliver, A., X. Lladó, E. Pérez, J. Pont, E. Denton, J. Freixenet and J. Martí, 2010. A statistical approach for breast density segmentation. *J. Digit. Imaging*, 23(5): 527-537.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE T. Sys. Man Cyb.*, 9(1): 62-66.
- Rangayyan, R.M., F.J. Ayres and J.E. Leo Desautels, 2007. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *J. Frankl. Inst.*, 344(3): 312-348.
- Sheshadri, H.S. and A. Kandaswamy, 2007. Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms. *Comput. Med. Imag. Grap.*, 31: 46-48.
- Srinivasan, G.N. and G. Shobha, 2008. Statistical texture analysis. Proceeding of World Academy of Science, Engineering and Technology, 36: 1264-1269, ISSN: 2070-3740.
- Subashini, T.S., V. Ramaingam and S. Palanivel, 2010. Automated assessment of breast tissue density in digital mammograms. *Comput. Vis. Image Und.*, 114: 33-43.
- Suckling, J., J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal and J. Savage, 1994. The mammographic images analysis society digital mammogram database. *Exerp. Med. Int. Congress Series*, 1069: 375-378.
- Tagliafico, A., G. Tagliafico, S. Tosto, F. Chiesa, C. Marinoli, L.E. Derchi and M. Calabrese, 2009. Mammographic density estimation: Comparison among BI-RADS categories, a semi-automated software and a fully automated one. *Breast*, 18(1): 35-40.
- Tzikopoulos, S.D., M.E. Mavroforakis, H.V. Georgiou, N. Dimitropoulos and S. Theodorides, 2011. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *Comput. Meth. Prog. Bio.*, 102: 47-63.
- Vaidehi, K. and T.S. Subashini, 2013. Automatic Identification and elimination of pectoral muscle in digital mammograms. *Int. J. Comput. Appl.*, 75(14): 15-18.
- Vállez, N., G. Bueno, O. Déniz-Suárez, J.A. Seone, J. Dorado and A. Pazos, 2011. A tree classifier for automatic breast tissue classification based on BIRADS categories. In: Vitria, J., J.M. Sanches and M. Hernandez (Eds.), *IbPRIA 2011*. LNCS 6669, Springer-Verlag, Berlin, Heidelberg, pp: 580-587.
- Vapnik, V., 1998. *Statistical Learning Theory*. Wiley, New York.
- Wang, X.H., W.F. Good, B.E. Chapman, Y.H. Chang, W.R. Poller, T.S. Chang and L.A. Hardesty, 2003. Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. *Am. J. Roentgenol.*, 180: 227-262.
- Wei, C.H., S.Y. Chen and X. Liu, 2012. Mammogram retrieval on similar mass lesions. *Comput. Meth. Prog. Bio.*, 106(3): 234-248.
- Wei, L., Y. Yang and R.M. Nishikawa, 2009a. Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recogn.*, 42(6): 1126-1132.
- Wei, L., Y. Yang, M.N. Wernick and R.M. Nishikawa, 2009b. Learning of perceptual similarity from expert readers for mammogram retrieval. *IEEE J. Sel. Top. Signal.*, 3(1): 53-61.
- Winsberg, F., M. Elkin and J. Macy, 1967. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*, 89: 211-215.
- Zheng, B., 2009. Computer-aided diagnosis in mammographic using content based image retrieval approaches: Current status and future perspectives. *Algorithms*, 2(2): 828-849.