Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

SVM Based System for classification of Microcalcifications in Digital Mammograms

Sukhwinder Singh, Vinod Kumar, *Senior Member, IEEE*, H.K. Verma, and Dilbag Singh

***Abstract*— This paper presents a SVM based computer-aided diagnosis (CAD) system for the characterization of clustered microcalcifications in digitized mammograms. First, the region of interest (ROI) in mammogram is enhanced using morphological enhancement (MORPHEN) method. Second, pixels in potential microcalcification regions are segmented out by using edge detection and morphological operations. Third, features based on shape, texture and statistical properties are extracted from each region. Finally, these features are fed to a SVM based classifier for identifying the clusters as either benign or malignant. The SVM with RBF kernel gave A_z = 0.9803 with 97% accuracy and the SVM with polynomial kernel gave A_z = 0.9541 with 95% accuracy.**

***Index Terms*—Computer aided diagnosis, contrast enhancement, microcalcification, support vector machine.**

I. Introduction

MAMMOGRAPHY is one of the most reliable methods for early detection of breast cancer. However, it is difficult for radiologists to make an accurate evaluation of the enormous number of mammograms generated in extensive screening. Some 10–30% of breast lesions are missed during routine screening [1]. With the advances of digital image processing, pattern recognition and artificial intelligence, radiologists have an opportunity to improve their diagnosis with the aid of computer systems [2-5].

The basic objective of computer-aided diagnosis (CAD) is to provide a "second opinion" on radiologists' image readings, with a view to improve the quality and productivity by improving the accuracy of radiological diagnosis and reducing the image reading time.

In a CAD scenario, computerized image analysis is used to suggest regions of suspicion (region of interest: ROI) in the mammogram so that the radiologist can examine these regions more critically. Additionally, there is a need to suppress the noise, to enhance the contrast between the ROI and background, and to extract the features from suspected microcalcifications (MCCs) to enable an accurate classification. The automated prompting and the additional information provided by computerized image analysis results in greater repeatability and uniformity in the standard of care.

On an average, the readers' sensitivity can be increased by 10% with the assistance of CAD systems [6]. Automated detection/classification of MCCs can be very helpful for breast cancer control because of high correlation between the form of the microcalcification clusters and the diseases. Although computer-aided mammography has been studied over two decades, yet automated analysis of microcalcifications remains a challenge. It is mainly due to their fuzzy nature, low contrast and low discrimination from their surroundings [1].

Dr. Sukhwinder Singh is with the Department of Computer Science and Engineering at Sant Longowal Institute of Engineering and Technology, Longowal, 148 106 (INDIA) (e-mail: sukhdalip@yahoo.com).

Dr. Vinod Kumar is with Department of Electrical Engineering at Indian Institute of Technology Roorkee, Roorkee 247667 (INDIA) (e-mail: vinodfee@iitr.ernet.in).

Dr. H.K. Verma is with Department of Electrical Engineering at Indian Institute of Technology Roorkee, Roorkee 247667 (INDIA) (e-mail: hkvfee@iitr.ernet.in)

Dilbag Singh is with the Department of Instrumentation & Control Engineering, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar – 144 011, Punjab INDIA (e-mail: singhd@nitj.ac.in)

II. Proposed CAD System

The problem of mammogram interpretation can be decomposed into two sub-problems. The first deals with the detection and localization of regions of interest (ROIs) containing microcalcifications. The second, and more difficult sub-problem, is the characterization of the suspected microcalcifications as malignant or benign.

In the proposed method, first step is selection or marking of region of interest (ROI) on mammogram. Then, the ROI selected from the digitized mammogram is de-noised and enhanced. The next stage is designed to find suspicious areas containing MCCs, and to separate the MCCs from the background that will be used for extracting features of MCCs. Next the features of MCCs are extracted and selected, and finally MCCs are classified into benign and malignant.

A. Image Enhancement

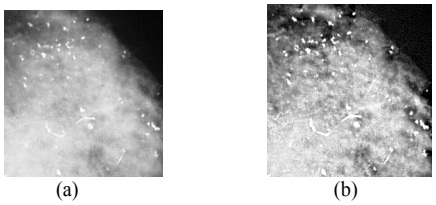
Enhancement of mammograms is a delicate process, which involves improving contrast while preserving visual acuity. Too little enhancement may rule out the detection of minor microcalcification peaks, while too much enhancement may significantly increase the amplitude of background noise leading to a large number of false detections.Morphological enhancement (MORPHEN) is derived by calculating the dual area top-hats in parallel [8]. The effect of enhancement can be visualized from the images shown in Fig 1.

B. Detection of Microcalcifications

Fig. 2 shows the process of detection of suspected microcalcifications. After enhancement of ROI with MORPHEN method, the enhanced image undergoes Sobel

and Canny edge detection [9] to segment the enhanced borders from the background image.

The Sobel and Canny binary output images are further subjected to the flood-filling operation to fill all objects with closed borders. Both flood-filled images are then logically ORed together to produce a new image that possibly includes false positives but discourages false negatives. Finally, morphological closing is applied on this new image to eliminate noise.



(a) (b)

Fig. 1. ROIs of digital mammograms containing microcalcifications (c45bmlo) (a) Original image, (b) image enhanced with MORPHEN.

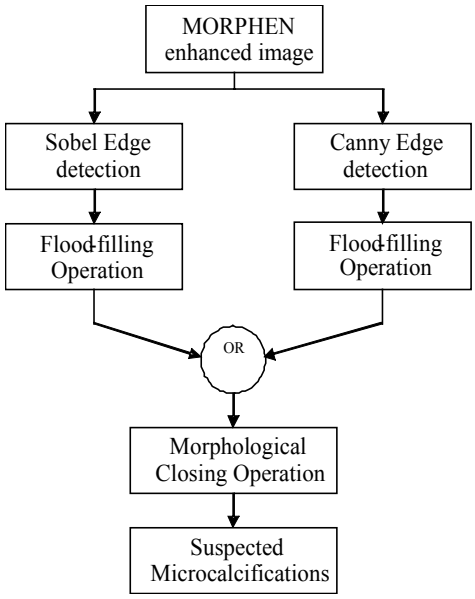


Fig. 2 Process of detection of suspected microcalcifications from enhanced mammogram ROI.

C. Feature Extraction

A typical mammogram contains a vast amount of heterogeneous information that depicts different tissues, vessels, ducts, chest skin, breast edge, the film, and the X-ray macine characteristics. In order to build a robust diagnostic system towards correctly classifying abnormal and normal regions of mammograms, there is a need to present all the available information that exists in mammograms to the diagnostic system so that it can easily discriminate between the abnormal and the normal tissue. In this work, features based on shape, texture, and spectral domain are extracted.

1) Shape Based Features

Shape features, which have proved effective in classifying microcalcifications, are moments [10]. The moments used here are moments of region boundaries. In this definition, boundaries are characterized by an ordered sequence, which represents the Euclidean distances between the centroid of the region and all contour pixels. If z(i), i = 1, 2,…, N are the Euclidean distances of the contour pixels to the centroid (where N is he number of these pixels) the pth moment is defined as

$$m_p = \frac{1}{N}\sum_{i=1}^{N}\left[z(i)\right]^p \qquad (1)$$

and the pth central moment is defined as

$$M_p = \frac{1}{N}\sum_{i=1}^{N}\left[z(i) - m_1\right]^p \qquad (2)$$

In addition, low order moments are the ones less sensitive to noise. These are named as F1, F2, F3. It has been shown [10] that F3–F1 is the best moment to represent roughness of microcalcifications. Maximum value of F3–F1 and mean of F3–F1 in each cluster have been used as features for representing the shape of microcalcifications for classification.

The various shape features extracted from each region are listed below.

- Number of microcalcifications in the cluster
- Maximum size of calcifications in cluster
- Area
- Eccentricity
- Major Axis length
- Minor Axis length
- Maximum Value of F3-F1

2) Textural Analysis Based Features

Texture information is used in a wide range of applications including natural scene, remotely sensed data, and biomedical modalities. The perception of texture is believed to play an important role in human visual system for recognition and interpretation. Despite the lack of a complete and formal definition of texture, numerous methods for texture classification have been suggested [11,12]. Specifically, texture features have been utilized for many medical image applications, including mammography [13], and have proven to be useful in discriminating texture classes in X-ray mammography, as well as in other modalities for breast cancer detection.

The spatial gray level dependence (SGLD) matrix is constructed by counting the number of occurrences of pixel pairs at a given displacement. In particular, to compute the SGLD matrix for an image $H(i,j)$, a displacement vector $d = (x,y)$ is defined. The (i,j)th element of the SGLD matrix P is defined as the joint probability that gray levels i and j occur at a distance d and angle θ relative to H [14]. Here matrix H is considered to be of 32x32 pixels, and the centre of matrix coincides with the centre of mass of the suspected microcalcification.

$$P(i,j,d,\theta^o) = Count\{((u,v),(s,t) \in W \times W \mid (((f(u,v)=i \wedge (f(s,t)=j)) \vee ((f(u,v)=j) \wedge (f(s,t)=i))) \wedge (s-u=x) \wedge (t-v=y)\} \quad (3)$$

where Count{x}is the number of elements of set x, $W = \{0,1,...,n_{gl}-1\} \times \{0,1,...,n_{gl}-1\}$, and n_{gl} is the number of gray levels in the image. The matrix is then normalized so that it can be treated as a probability density function (pdf).

$$p(i,j,d,\theta^o) = \frac{P(i,j,d,\theta^o)}{\sum_{i,j} P(i,j,d,\theta^o)} \quad (4)$$

The parameter values are d =1, and θ = 0, 45, 90 and 135^o.

Various texture features [14] extracted for each (d,θ) pair are difference entropy, difference variance, entropy, sum entropy, sum variance, sum average, inverse difference moment, variance, and correlation

For each suspected microcalcification, four (d,θ) pairs yields 44 texture features. Then the range values are taken as features to be used for classification of microcalcifications. This reduces the number of texture feature to half.

3) Spectral domain features

Spectral domain features describe the frequency characteristics of the input image. In this paper, two spectral domain features are extracted from each 32x32 suspected microcalcification. They are the energy or block activity (En) and the spectral entropy (Ent), given by [15]:

$$En = \sum_{u=1}^{32}\sum_{v=1}^{32} |X(u,v)| \quad (5)$$

$$Ent = -\sum_{u=1}^{32}\sum_{v=1}^{32} \bar{X}(u,v)\ln(\bar{X}(u,v)) \quad (6)$$

where X(u,v) is the DCT coefficient in row u and column v of the DCT matrix and $\bar{X}(u,v) = \frac{|X(u,v)|}{En}$.

D. Classification

Several methodologies have been proposed for the microcalcification characterization problem, such as decision trees, linear discriminant analysis and artificial neural networks (ANNs) [16-18]. In general it is very difficult to compare the efficiency of the above methods since they have been tested in different mammographic datasets using different performance measures. A review of the existing detection and classification methodologies can be found in [1].

Recently support vector machines (SVMs) [19] have been used for classification for mammographic microcalcifications in [20-24]. In the present work, SVM based classifier has been implemented that uses features which result from the shape descriptors extracted from ROI, statistical texture analysis, and parameters in frequency domain capable of distinguishing between benign and malignant microcalcifications.

The performance of the classifiers has been evaluated using the receiver operating characteristic (ROC) methodology [25-28] and the classification rate. The obtained results provide high classification performance and thus our method can be considered quite promising.

III. Results and Discussion

A. Database

Total numbers of 435 samples are obtained from mammograms containing microcalcifications from McGill University database [7]. These are divided into two sets: *training* and *testing*. The *training-se*t contained 300 samples of which 165 are benign and 135 are of malignant nature respectively. Remaining 135 samples are used for testing of SVM classifier. *Testing-set* contained 67 cases of benign and 68 cases of malignant calcifications respectively.

B. Parameter Selection and Training

Once the training samples are obtained, the next step is to determine the optimal parametric settings of SVM. In this process, the following variables: the type of kernel function, its associated parameter, and the regularization parameter C must be decided. To optimize these parameters, *six fold cross validation* [32] has been applied to the *training-set*.

Various parameters for the SVM like regularization parameter C, degree of polynomial, sigma of RBF etc. are varied as: C from 1 to 10^5, degree of polynomial (*p*) from 2 to 9, and Gamma (γ) from 0.13 to 2.5 to choose the best parameters for SVM. It is found that polynomial kernel with degree (*p*) to be 3 and regularization parameter (C) having value 100 gives highest area of 0.958 under ROC curve (A_z). For Gaussian RBF kernel it has been observed that gamma = 1.50 and C=10 gave highest area of 0.962 under ROC curve (A_z).

TABLE I
A_z VALUES OF ROC CURVES WITH DIFFERENT KERNEL OF SVM

Kernel	TN	FN	FP	TP	A_z	
					Training	Test
Polynomial	64	3	2	66	1.0000	0.9703
RBF	66	1	2	66	1.0000	0.9887

TABLE II
A_z VALUES OF ROC CURVES WITH DIFFERENT METHODS

Kernal	Ref. [20]	Ref. [22]	Proposed
Polynomial	0.9682	0.9716	0.9703
RBF	0.9834	0.9903	0.9887

The SVM classifier is trained with the *training-set* using the optimal parameters of polynomial and gaussian RBF kernel respectively.

C. Performance Evaluation

As ROC analysis is commonly used approach for

classification performance evaluation. Table I shows area under ROC curves (Az) of the proposed SVM approach with training and test data sets. In this study, cases actually benign and malignant are considered as true negative (TN) and true positive (TP) respectively. Table I shows that Az for Gaussian RBF kernel is more than that of polynomial kernel. The methods proposed in [20] and [22] are also applied to the same data set. The results obtained are given in Table II. It is observed that performance of the proposed method is better than Ref. [20] and almost similar to that of Ref. [22]. It has been found that sensitivity of 98.51%, and Specificity 97.06% has been obtained with RBF kernel as compared to sensitivity of 95.65% and specificity of 96.96% for polynomial kernel.

IV. Conclusion

Computer aided analysis of digitized mammographic images for the purpose of classifying microcalcifications has been presented in this paper. Shape, textural, and spectral domain statistical features have been extracted from each region containing suspected microcalcification. In all 435 samples have been found from MCCs containing mammograms in McGill University dataset to be used for support vector machine (SVM) based classifier. Two kernel functions: polynomial and Gaussian radial basis function (RBF) have been used here. Performance of classifier has been evaluated using area under receiver operating characteristic (ROC) curves (A_z) along with other indices. The SVM with RBF kernel gave A_z = 0.9887 and SVM with polynomial kernel A_z=0.9703. The perfromnace of the proposed classifier has been found to be better/comparable with other existing methods. The high accuracy of classification of MCCs obtained with the proposed SVM classifier can help radiologists in making an accurate diagnostic decision, which can reduce unnecessary biopsies.

References

[1] H.D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recognition*, vol. 36, pp. 2967–2991, 2003.

[2] S. Yu, L. Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 115-126, 2000.

[3] B. Verma and J. Zakos, "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 1, pp. 46-54, 2001.

[4] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Medical Imaging and Graphics,* vol. 26, pp. 309-319, 2002.

[5] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *The British Journal of Radiology,* vol. 78, pp. S3-S19, 2005.

[6] R.M. Nishikawa, Detection of microcalcification, *in: N. Karssemeijer, M. Thijssen, J. Hendriks, L. van Erning (Eds.), Digital Mammography Nijmegen 98, Kluwer Academic Publisher, Amsterdam*, pp. 131–153, 1998.

[7] http://sprojects.mmi.mcgill.ca/mammography/cases.htm

[8] M. Wirth, M. Fraschini, and J. Lyon, "Contrast enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements," *Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems, 2004(CBMS 2004),* pp.134 – 139, 2004.

[9] R.C. Gonzalez and R.E. Woods, "*Digital image processing,*" Addison-Wesley, an imprint of Pearson Education, Fifth Indian reprint, 2000.

[10] L. Shen, R.M. Rangayyan, and J.E. Leo Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Transactions on Medical Imaging,* vol. 13, no. 2, pp. 263-274, 1994.

[11] M. Unser, "Sum and difference histograms for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 118-125, 1986.

[12] V. Vickers and J. Modestino, "A maximum-likelihood approach to texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, pp. 61-68, 1982.

[13] P. Miller and S. Astley, "Classification of breast tissue by texture analysis," *Image Vision Comput*, vol. 10, no. 5, pp. 277-282, 1992.

[14] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. SMC-3, no. 6, pp. 610-621, 1973.

[15] B. Zheng, W. Qian, and L.P. Clarke, "Digital mammography: mixed feature neural network with spectral entropy decision for detection of microcalcification," *IEEE Transactions on Medical Imaging,* vol. 15, no. 5, pp. 589–597, 1996.

[16] A.P. Dhawan, Y. Chitre, M. Moskowitz, and E. Gruenstein, "Classification of mammographic microcalcification and structural features using an artificial neural network," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* vol. 13, no. 3, pp. 1105–1106, 1991.

[17] Y. Chitre, A.P. Dhawan, and M. Moskowitz, "Artificial neural network based classification of mammographic microcalcifications using image structure and cluster features," *Proceedings of the Annual Conference on Engineering in Medicine and Biology,* vol. 15, pp. 298–309, 1993.

[18] A.P. Dhawan, Y. Chitre, C. Bonasso, and K. Wheeler, "Radial-basis-function based classification of mammographic microcalcifications using texture features," *IEEE Proceedings of the 1995 IEEE Engineering in Medicine and Biology 17th Annual Conference*, pp. 535–536, 1995.

[19] V. Vapnik, *Statistical Learning Theory*, New York: Wiley; 1998.

[20] A. Bazzani, A.Bevilacqua, D. Bolline, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani, "An SVM classifier to separate false signals from microcalcifications in digital mammograms," *Phys. Med. Bio.,* vol. 46, pp. 1651-1663, 2001.

[21] I. El-Naqa, Y. Yang, M.N. Wernick, N.P. Galatsanos, R.M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging,* vol. 21, no. 12, pp. 1552–1563, 2002.

[22] J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, and H.K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Computerized Medical Imaging and Graphics,* vol. 29, no. 6, pp. 419-429, 2005.

[23] A. Papadopoulos, D.I. Fotiadis, and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp.141-150, 2005.

[24] L. Wei, Y. Yang, R.M. Nishikawa, Y. Yang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Transactions on Medical Imaging,* vol. 24, no. 3, pp. 371-380, 2005.

[25] C.E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine,* vol. VIII, no. 4, pp. 283-298, 1978.

[26] C.E. Metz, "Evaluation of digital mammography by ROC analysis," *Proceedings of the 3rd International Workshop on Digital Mammography 1996 (Digital Mammography'96),* pp. 61-68, 1996.

[27] C.E. Metz, "ROC methodology in radiologic imaging," *Investigative* Radiology, vol. 21, no. 9, pp. 720-733, 1986.

[28] K.R. Muller, S. Mika, G. Ratch, K. Tsuda, and B. Scholkopf, " An introduction to kernel-based learning algorithms," IEEE Transactions on Neural Networks, vol. 12, pp. 188-201, 2001.