Mammogram Analysis: Tumor Classification

Term Project Report

Geethapriya Raghavan

geeragh@mail.utexas.edu

EE 381K - Multidimensional Digital Signal Processing

Spring 2005

Abstract

Breast cancer is the main cause of cancer related mortality among American women. The use of screening mammography as the most reliable method for early diagnosis of breast cancer is widely recommended with the introduction of several Computer Aided diagnosis (CAD) techniques. I am presenting some of the pattern recognition techniques that have been most effective in classifying tumors as benign or malignant – support vector machine (SVM), kernel Fisher discriminant (KFD), relevance vector machine (RVM) and a multiresolution pattern recognition method using wavelet transforms and am implementing SVM and the method using wavelet transforms. These methods have been developed and implemented in statistical learning theory over the past decade and they give promising classification results for efficient tumor diagnosis.

1 Introduction

Breast cancer is among the most common and deadly of all cancers. Mammography is a uniquely important type of medical imaging used to screen for breast cancer. All women at risk go through mammography screening procedures for early detection and diagnosis of tumor. A typical mammogram is an intensity X-ray image with gray levels showing levels of contrast inside the breast that which characterize normal tissue and different calcifications and masses. The contrast level of a typical mammogram image is proportional to the difference in X-ray attenuation between different tissues. Important visual clues of breast cancer include preliminary signs of masses and calcification clusters (Fig.1a and 1b). A mass is a localized collection of tissue seen in two different projections, and calcifications are small calcium deposits. Unusually smaller and clustered calcifications are associated with malignancy while there are other calcifications (diffuse, regional, segmental and linear) that are typically benign. Such calcifications are termed as *microcalcifications (MC).* In the early stages of breast cancer, these signs are subtle and hence make diagnosis by visual inspection difficult. With millions undergoing mammography procedures, the need for quick and reliable computer based tools is strongly desired.

2 Background

The role computers play in mammogram analysis is threefold: detection, diagnosis and noise cancellation. Detection involves identifying cancerous tissues in a mammogram. It provides spatial information about the MCs. Some of the more important pitfalls in the diagnosis of breast cancer encountered two decades ago - low contrast and poor image quality in mammography are presented in [1]. Diagnosis is about classifying the detected
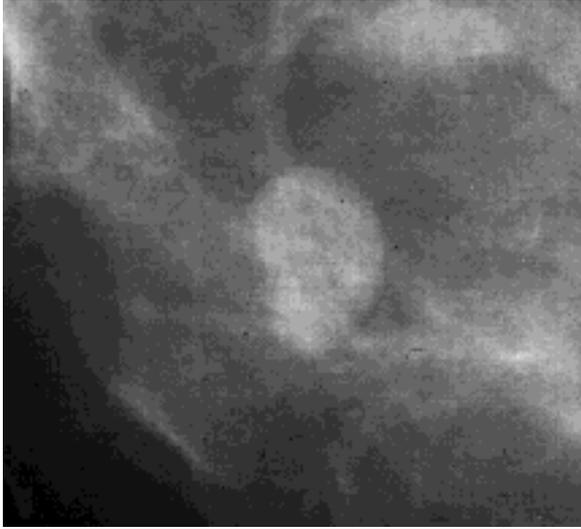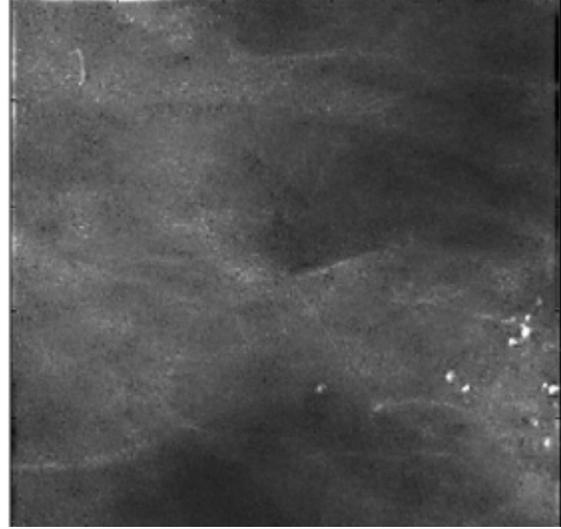
Fig 1(a)   **Fig 1(b)**

Fig1. (a) Mammogram showing a big mass and (b) a clustered microcalcification

cancerous regions as benign or malignant. A review of how CAD techniques help in diagnosis is presented in [2]. They have concluded that an effective diagnosis algorithm that runs at high speed and in less time is yet to be designed.

3 Diagnosis Tools

The diagnosis task is modeled as a two-class classification task. Features are extracted from Regions of Interest (ROIs - the region containing the masses or the microcalcifications) containing the abnormality whose spatial information is provided by detection algorithms and each ROI is classified using a classification algorithm which is a supervised method that is first trained on a set of sample images (whose classification is known) called the *training set*. The performance of the algorithm is then tested on a separate *testing set*. The metrics used to report the accuracy of these algorithms are sensitivity and specificity. *Sensitivity* is defined as a lesion for which the CAD predicts that it is cancerous and it is actually found to be malignant. *Specificity* is the fraction of

benign lesions that are correctly identified by the CAD as being benign. A plot of sensitivity versus specificity is called a *Receiver Operating Characteristic (ROC)* curve and this is used to report the performance of the CAD technique used [3]. The main parameter studied is the area under the ROC curve, A_z. The higher the value of A_z, the better is the performance of the CAD technique used. A good CAD technique has A_z values closer to one. Classifying a mammogram with microcalcifications is more challenging than with masses because of their erratic shapes, size, density and texture. Due to their high success rates [4], I investigate the following contemporary methods.

a. Support Vector Machine (SVM)

In recent years, SVM learning has found a wide range of real-world applications, including object recognition [5], speaker identification [6] and face detection in images [7]. The formulation of SVM learning is based on the principle of structural risk minimization. Instead of minimizing an objective function based on the training samples [such as mean square error], the SVM attempts to minimize a bound on the generalization error (the error made by the learning machine on test data not used during training). As a result, an SVM tends to perform well when applied to data outside the training set.

The nonlinear SVM classifier is defined by the function *y* given as

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b \tag{1}$$

such that such that for each training example, $\mathbf{x}_i$, the function yields

$$f(\mathbf{x}_i) \geq 0 \text{ for } y_i = +1, \text{ and } f(\mathbf{x}_i) < 0 \text{ for } y_i = -1 \tag{2}$$

This means that the two classes in the training set are separated by the hyperplane

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0. \tag{3}$$

Among the possible separating hyperplanes, the one from which the distant to the closest point is maximal is the optimal separating hyperplane (OSH). Following nonlinear transformation, the parameters w and b of the decision function **y** *(support vectors)* are determined by the minimization of $\|w\|^2$ which is proportional to the cost function to find the OSH [8]. Autocovariance textures based on the varying density of microcalcifications are used features for diagnosing tumors in mammograms [4] and in ultrasonic images [8].

b. Relevance vector machine (RVM)

This method, developed by Tipping [9], is based on Bayesian estimation for classification problems. It is proved to be faster than SVM since it yields an optimum solution with few training samples. These are called *relevance vectors*. The classification function is

$$f_{\text{RVM}}(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}, \mathbf{x}_i)$$

(4)

where **K** is a kernel function, and x_i, are the training samples. According to Tipping [20], the parameters, α_i, i = 1,2,3,…N, are determined using Bayesian estimation. The parameters in (4) are then obtained by maximizing the posterior distribution given the input vectors. This is equivalent to maximizing the following objective function:

$$J(\alpha) = \sum_{i=1}^{N} \log p(d_i|\mathbf{x}_i) + \sum_{i=1}^{N} \log p(\alpha_i|\lambda_i^*)$$

(5)

Only those samples associated with nonzero coefficients of α_i, which are the relevance vectors, will contribute to the decision function (4).

c. Kernel Fisher Discriminant (KFD)

KFD is an extension to Fisher's linear discriminant method [10]. This method is based on the principle of projecting data onto a one-dimensional space after initially mapping the

input vectors into higher dimensional space. This effectively yields a nonlinear

discriminant with respect to the original vector x. It has a classification function similar

to SVM. The coefficients, α_i, i = 1, 2, 3,…N, are determined by maximizing the Rayleigh

coefficient corresponding to the decision function $f_{KFD}(\mathbf{x})$ [11]. The main difference

between KFD, SVM and RVM is that KFD uses all the training samples. The

performance summary of the methods analyzed is presented in the following Table.

Table 1

Authors	Image type and number	Method	A_z
Wei et al., 2005 [4]	Mammogram, 697	SVM	0.8545
		KFD	0.8303
		RVM	0.8421
Chang et al., 2003 [8]	Ultrasonic image, 250	SVM	0.9396

d. Wavelet Analysis

This approach, termed as a multiresolution pattern recognition approach [12], aims at

extracting localized features from the Regions of Interest (ROIs). Wavelet transforms can

decorrelate the image, are invertible and have flexibility to discretize scale and

orientation. The wavelets are functions used as basis for representing other functions, and

once a so called mother wavelet is fixed, a family can be generated by translations and

dilations of it. If we denote a mother wavelet as $(\mathbf{H_x})$, its dilations and translations are

$$F(\mathbf{H_x}) \quad = \quad \mathbf{H_x} \,((x-a)/b) \tag{6}$$

where $a = 2^{-j}$ and $b = k \times 2^{-j}$ and k, j are integers. Applying convolution of low and high

pass filters, the signal can be decomposed in specific sets of coefficients, at each level of

decomposition (Fig 2) as : low frequency coefficients $(\mathbf{A^d_{2j}\,f})$, vertical high frequency
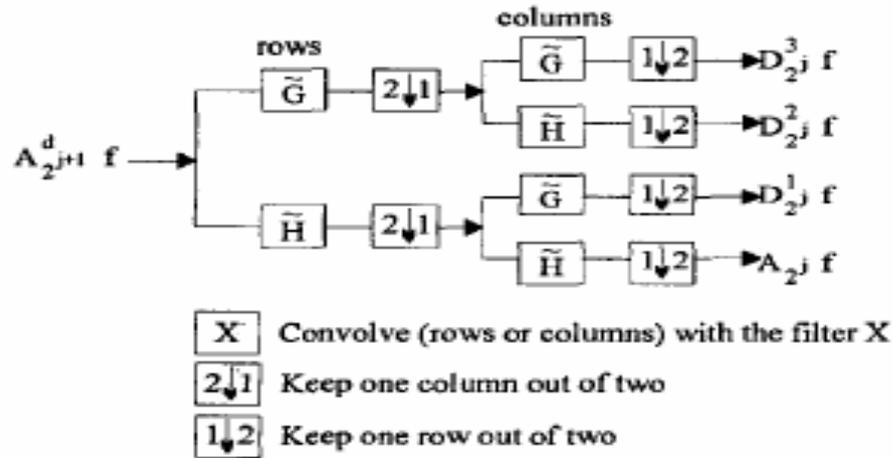
Fig (2)

coefficients ($D^1_{2^j}f$), horizontal high frequency coefficients ($D^2_{2^j}f$), and high frequency coefficients in both directions ($D^3_{2^j}f$).

4 Implementation and Results

Mammograms for simulation were obtained from the MIAS (Mammographic Image Analysis Society) database. MIAS is an organization from UK interested in the understanding of mammograms. The number of benign and malignant images used for implementation was 40 and 31 respectively. Most diagnosis algorithms begin with a region of interest (ROI) containing a cluster of calcifications. Precisely, the mammogram images picked from the MIAS database were the ROIs of mammograms [13]. The steps involved in the diagnosis of calcifications are 1) segmentation of individual calcifications, 2) feature extraction, and 3) classification.

Method I

Chang et al. performed SVM on Ultrasonic images with selective calcification properties as features. The features were microcalcification area and contrast. Method I that I

followed implements the same algorithm on mammograms. Both linear (3) and non linear SVM (1) were performed on the mammograms with the two features considered individually. The nonlinear function used to run the nonlinear SVM algorithm was a parabolic function. The performances of the techniques were measured in terms of the success percentage (Table 2) and an ROC plot (Fig 3). The software used to perform SVM was *SVM Light*- an open source software available on the internet [14].

Method II

The second algorithm extends from the diagnosis algorithm provided by *Borges* et al. in [12]. This method involves the following steps:

1. Apply Wavelet Transform on the ROIs and get the image frequencies (6)

2. Repeat the decomposition on only the low frequency coefficients, *A(i, j)* (Fig. 2)

3. Design a Nearest Neighborhood Classifier with Euclidean distance as a metric between correspondent normalized wavelet coefficients. When M is the prototype of a class (represent the ground truth about the class being malignant or benign), Euclidean distance is computed as

$$D_{Euclidiana} = \sqrt{\sum_{i,j}(A(i, j) - M(i, j))^2}$$

(7)

The success percentages for the two methods are presented in Table 2

Table 2

Classifier	Microcalcification Contrast	Microcalcification Area
Non-linear SVM	67.7 %	78 %
Linear SVM	42.8 %	70.4 %
NNR	72 %	76.2 %

An ROC plot for the methods implemented is given in Figure 3.
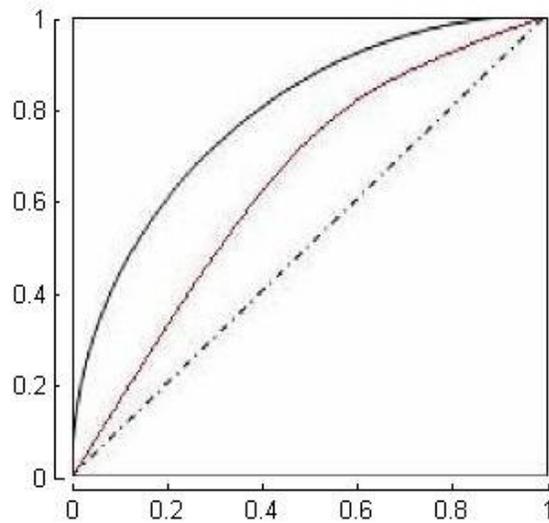
Fig 3 – ROC plot with sensitivity on the y axis and specificity on the x axis. *Dotted line –* performance threshold for any diagnosis algorithm. ROC curve below this line indicates a bad classifier design; *Red Curve –* ROC for Method II; *Black curve –* a smooth parabolic curve for SVM (linear)

5 Conclusion

The performance of the different classifier models analyzed in this report show that they are very promising to achieve higher efficiency and speed in lesser time in diagnosing tumor as benign or malignant in mammograms so that they can provide reliable assistance to radiologists. Comparatively, with the features selected as MC area and contrast (density), SVM is more suited to computer aided diagnosis since it requires lesser training samples to train the learning machine and get the support vectors and perform as good as other methods discussed. Diagnosing tumor using microcalcifications has always been a challenging task.

REFERENCES

[1] M. J. Homer**, "**Breast imaging: pitfalls, controversies, and some practical thoughts", *Radiologic Clinics of North America,* vol. 3, pp. 459-472. Sep 1985

[2] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," *Academic Radiology*, vol. 6, pp. 22–33, Jan 1999

[3] C. E. Metz, "ROC methodology in radiologic imaging", *Investigative Radiology*, vol. 21, pp. 720-33, Sep 1986

[4] L. Wei, Y. Yang, R. M. Nishikawa, Y. Jiang, "A Study on Several Machine-Learning Methods for Classification of Malignant and Benign Clustered Microcalcifications", *IEEE Transactions on Medical Imaging*, pp. 1-10, Jan 2005

[5] M. Pontil and A. Verri, "Support vector machines for 3-D object recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 637–646, June 1998

[6] V. Wan and W. M. Campbell, "Support vector machines for speaker verification and identification", *Proc. IEEE Workshop on Neural Networks for Signal Processing*, Sydney, Australia, pp. 775–784, Dec. 2000

[7] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: Application to face detection", *Proc. Computer Vision and Pattern Recognition*, Puerto Rico, pp. 130–136, June 1997

[8] R. Chang, W. Wu, W. K. Moon, Y. Chou and D. Chen, "Support vector machines for Diagnosis on breast tumors on US images", *Academic Radiology*, vol. 10(2), pp. 189-197, Feb 2003

[9] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, Aug 2001

[10] R. Duda, P. Hart, D. Stork, "*Pattern Recognition*", 2nd ed., Wiley, 2002

[11] B. Scholkopf and A. Smola, "*Learning With Kernels—Support Vector Machines, Regularization, Optimization and Beyond*", Cambridge, MIT Press, 2002.

[12] C. B. Ferreira and D. L. Borges, "Automated mammogram classification using a multiresolution pattern recognition approach", *Proc. XIV Brazilian Symposium on Computer Graphics and Image Processing*, pp. 76-83, Oct. 2001

[13] Mammographic Image Analysis Society (MIAS) – *http://wiau.man.ac.uk/services/MIAS*

[14] SVM Light Software - *http://svmlight.joachims.org/*