The Detection of Micro-Calcifications in Mammographic Images Using High Dimensional Features

Jeffrey L. Solka,[1] Wendy L. Poston,[2] Carey E. Priebe,[1] George W. Rogers,[1] Richard A. Lorey,[1] David J. Marchette,[1] Kevin Woods,[3] Kevin Bowyer[3]
1 - The Naval Surface Warfare Center
Systems Research and Technology Department
Computational Statistics Group - Code B10
Dahlgren VA 22448-5000

2- The Naval Surface Warfare Center
Weapons Systems Department
Smart Munitions Development Branch -Code G33
Dahlgren, VA 22448-5000

3 - Department of Computer Science and Engineering
The University of South Florida
Tampa Florida 33612

Abstract

This paper examines techniques for the efficient use of high dimensional feature sets in the detection of micro-calcifications in mammograms. The paper focuses on techniques for dimensionality reduction and discriminant analysis. The paper examines the use of principal components and Fisher's linear discriminant for dimensionality reduction along with parametric and nonparametric statistical techniques for discriminant analysis.

1: Introduction

It is estimated that there are approximately 150,000 new cases of breast cancer diagnosed annually[1]. Besides the large toll in human suffering that this represents, the diagnosis and treatment of these cases is a burden on the health care system. As with many medical conditions a favorable prognosis for the patient is based on the earliest possible detection of the abnormality.

There has been a great deal of recent interest in using computer techniques in image processing and pattern recognition to help the radiologist's diagnostic efforts[2]. This work, along with work in medical expert systems, has come under the general heading of computer aided diagnosis (CAD). Much of the CAD work seeks to help the radiologist by drawing attention to regions of

the image that the computer has identified as suspicious. This would be an extremely valuable tool to aid in the timely and accurate detection of anomolous tissue. CAD techniques have been applied to such diverse areas as liver ultrasound[3], chest radiography[4], and x-ray mammography[5].

One of the early indicators of breast cancer is the radiographic presence of clusters of malignant microcalcifications. In many cases these clusters are located in the midst of dense parenchymal tissue and it takes very careful inspection by the radiologist to locate them. Their prominent role in diagnosis and difficult detectability make their CAD an important goal.

The next section focuses on our techniques used to extract features from the microcalcifications. This is followed by a brief discussion of the Fisher linear discriminant and principal component analysis approaches to dimensionality reduction. Next we turn our attention to a few words on parametric and nonparametric Bayesian based classification techniques. The fruits of the application of these techniques to the detection problem at hand is presented in the results section and we conclude the paper with a few words on future directions.

2: Feature Extraction

The first step in the feature extraction process is the identification of regions of interest from the image. It is expected that these extracted regions contain both true groups of malignant microcalcifications and groups of benign microcalcifications or other anatomical features. It is the function of the classification portion of the system to sort out the true malignancies from the false positives, which will be discussed below. The segmentation is done by thresholding a difference image which is created by subtracting from every pixel the average of a 15x15 pixel (1.13mm) window surrounding it. Region growing is performed on all the pixels left in order to group them into candidate objects for input into the feature extraction system. This segmentation routine can produce hundreds of candidate objects due to its sensitivity. To reduce the number of candidate objects only 3% of the total number of segmented objects with the highest contrast (with a minimum of 100) are retained. The twenty-one extracted features along with references to where they first appeared in the literature are listed in Table 1. A designator of "Standard" indicates a standard image processing technique.

Table 1: Microcalcification Features.

Ref. Number	Name	Description
[6]	Contrast 1	Avg. Obj. - Avg. Bck.
[6]	Contrast 2	Avg. Obj. / Avg. Bck.
[7]	Contrast 3	Max. in Obj. - Avg. Bck.
Standard	Smoothness B	Computed on Background
[8]	Area	Number of Pixels
[8]	Smoothness 1	RMS Noise Fluctuations in Object
[7]	Smoothness 2	RMS Noise Fluctuations in 51x51 windowed BCK.
[9]	Edge Strength	Avg. of Robert's Gradient for Perimeter Pixels

Table 1: Microcalcification Features.

Ref. Number	Name	Description
Standard	Gradient 1	Avg. of 3x3 Sobel Gradient for Perimeter Pixels
[10]	Moment F1	About Centroid
[10]	Moment F2	About Centroid
[10]	Moment F3	About Centroid
[10]	Moment (F3-F1)	About Centroid
[10]	Compactness	P^2/A, P=Perimeter, A=Area
Standard	Shape 1	Rmax/Rmin
[8]	Shape 2	Std. Dev. from Perimeter Pixels to centroid
Standard	Shape 3	Area/Rmax
[8]	Average Gray	Computed on Object
[10]	Gray Level Moment 1	Computed on Perimeter Pixels
[10]	Gray Level Moment 2	Computed on Object Pixels
Standard	Average Edge Step	Computed on Perimeter Pixels

The reader is referred to Woods et al[11], [12] for a more complete description of the feature extraction procedure.

3: Dimensionality Reduction

Given our set of extracted features in R^{21}, Bellman's curse of dimensionality [13] dictates that we project our features to as low a dimensional space as possible before we parametrically or nonparametrically model their underlying density. There are many techniques for dimensionality reduction. Since our ultimate goal is classification, it is beneficial to consider those transformations that enhance the separation between the probability density function of the features computed on the malignant microcalcifications and the one computed on the other nonmalignant tissue.

The two techniques for dimensionality reduction that were used here are the Fishers linear discriminant (FLD) [14] and principal component analysis (PCA) [13]. The FLD procedure seeks the best projection from the feature space to R^1, i. e. it seeks the best line to project upon. In this case optimality is measured by J(w), the ratio of between-class scatter, S_B, to within-class scatter, S_W, given by

$$J(\vec{w}) = \frac{\vec{w}^t S_B \vec{w}}{\vec{w}^t S_W \vec{w}}. \quad (1)$$

In eq. 1 $S_B = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^t$ where $\vec{m}_1$ is the mean vector for class 1, and $\vec{m}_2$ is the mean for class 2. The within class scatter matrix $S_W=S_1+S_2$, where $S_i = \sum_{\vec{x} \in X_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^t$.

In PCA one seeks to rotate the original data axes so that they align with the eigenvectors of the covariance matrix Σ. In this manner we seek a lower dimensional space that accounts for the variance of the features. In our case, in order to put the PCA approach on equal footing with the FLD we seek the best 1-d subspace to project upon. In order to account for the largest variance in the features we project upon the eigenvector corresponding to the largest eigenvalue.

4: Discriminant Analysis

Given the projected features in R^1 we proceed with a Bayesian approach to the classification problem. We seek parametric or nonparametric models of the class-conditional probability densities $p(x|c_i)$. Once we have modeled these class-conditional or state-conditional probability densities we employ a standard likelihood ratio hypothesis testing procedure[15] to determine the class of an unknown observation.

In the simplest approach each of the two class-conditional probability density functions is modeled as $N(\mu_i,\Sigma_p)$, where μ_i is the mean for the i-th class and Σ_p is the pooled covariance matrix. This procedure is known in the literature as a linear classifier. If we allow each class-conditional density to have its own covariance matrix then the resulting classifier is known as a quadratic classifier. The reader is referred to Duda and Hart [14] for a good treatment of these topics.

If on the other hand there is reason to believe that the underlying class-conditional densities are inherently nonnormal, then it is beneficial to employ a nonparametric estimator. One of the simplest such procedures is the kernel estimator[16]. In this case given $X=\{x_i\}^n_{i=1}$ where x_i~i.i.d $f(x)$, we model the unknown distribution as

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} g\left(\frac{x-x_i}{h}\right), \tag{2}$$

where the smoothing function $g(z)$ is chosen to be $N(0,1)$, and h is referred to as the kernel bandwidth. This procedure has not only the benefits of computational simplicity, but it possesses powerful asymptotic properties that provide it robustness to the nonnormality of the class-conditional probability density functions. In our analysis h was chosen optimal based on the normal assumption[16].

Given models of the class conditional probability densities, one is interested in assessing the performance of the classifier. One approach to this task is through the production of receiver operator characteristic (ROC) curves. These curves typically display some function of the probability of misclassifying a target (microcalcification in our case) as a function of false alarm rate. The probability of the first type of error is designated as α and $(1-\alpha)$ is often referred to in the medical community as the sensitivity. A false alarm of course is when the classifier incorrectly identifies normal tissue as tumorous. In order to produce the most accurate estimation of the ROC curve of our classifiers we employ a variant of the "leave one out" procedure[17]. In this procedure the features from a given image are left out and the FLD and PCA projections along with the class-conditional probability densities are built on the features from the remaining images. The projections are applied to the features in the image that was left out and the constructed class-conditional probability densities are used to build the likelihood ratios that are eventually used to determine class membership at each point. This procedure is repeated using the other images. At the end of the procedure the full set of likelihood ratios is used to perform the ROC analysis.

5: Results

The 24 images in the study were digitized at 70 micron resolution with 8 bit gray levels. The location of the malignant microcalcifications in the image was determined by having the groups of microcalcifications first identified by a radiologist and then the subsequent exact location of each microcalcification labeled by a trained technician. After passing the images through the segmentation portion of the algorithm there were 5730 nonmalignant regions and 542 malignant regions produced. The twenty-one features were computed for each of these regions.

Let's first examine the structure of the data for the two classes. Figure 1 represents kernel estimator approximations of the probability densities for each of the two classes' feature sets in FLD space. The normal appearance of the densities in Figure 1 is to be contrasted with the kernel estimated densities in PCA space shown in Figure 2.

Figures 3, 4, and 5 show the ROC curves for the linear, quadratic, and kernel based classifiers for the FLD and PCA features. The curves plot α as a function of false alarm rate. In each case the classifiers perform better using the FLD feature then the PCA one. It is also interesting to note that the performance of the three classifiers is virtually identical. To provide a figure for comparison we note that at a false alarm rate of 10% the classifiers possesses an α level of 20% accross the set of images. This value is in the same ballpark as some of our previous work utilizing seven dimensional features[11].

6: Conclusions

The preliminary results put forth in this paper indicate the superiority of the Fisher linear discriminate transformation over projection on the strongest principal component as a means for dimensionality reduction prior to Bayesian based discriminant analysis. These results are based on one area of application and future work is needed to determine if this trend will hold in general. With the curse of dimensionality in mind it is also important to reemphasize the comparable performance of the classifiers using the one dimensional features and our previous higher dimensional work. Future work will focus on "optimal" projections to R^2 and R^3 along with the application of our semi-parametric adaptive mixtures density estimator[18].

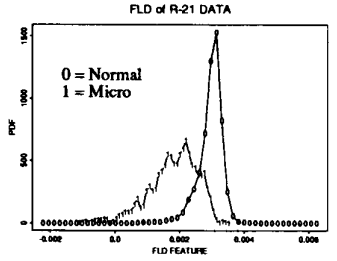

Figure 1. Kernel estimates for the microcalcification and non-microcalcification probability densities in Fisher linear discriminant space.

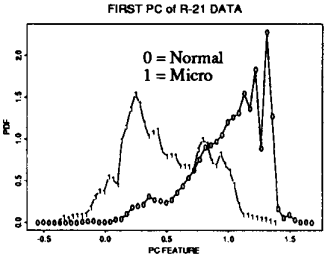

Figure 2. Kernel estimates for the microcalcification and non-microcalcification probability densities in the first principal component space.

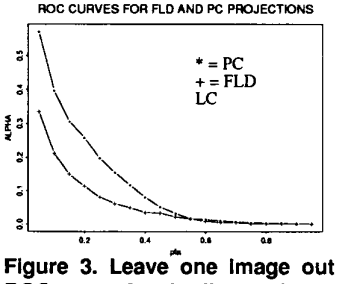

Figure 3. Leave one image out ROC curve for the linear classifier on the FLD and PC features.

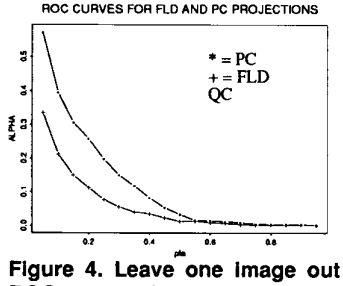

Figure 4. Leave one image out ROC curve for the quadratic classifier on the FLD and PC features.

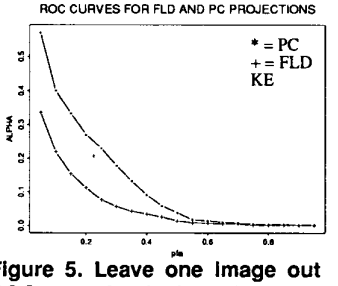

Figure 5. Leave one image out ROC curve for the kernel estimator on the FLD and PC features.

7: References

1. The Technology Applications Team of the Research Triangle Institute (1992) "Digital mammography system," *Problem Statement Produced by the Research Triangle Institute for the National Cancer Institute.*

2. Shtern, F. (1992). "Digital mammography and related technologies: a perspective from the National Cancer Institute." *Radiology*, 183(3), pp 829-830.

3. C. Wu, Y. Chen, and K. Hsieh (1992)"Texture features for classification of ultrasonic liver images," *IEEE Transactions on Medical Imaging*, Volume 11,No. 2, pp 141-152.

4. Sanada, S., Doi, K., and MacMahon, H. (1992) "Image feature analysis and computer-aided diagnosis in digital radiography: Automated Detection of pneumothorax in chest images," *Medical Physics*, Vol. 19, No. 5, pp. 1153-1160.

5. Priebe, C. E., Solka, J. L., Lorey, R. A., Rogers, G. W., Poston, W. L., Kallergi, M., Qian, W., Clarke, L. P., Clark, R. A. (1994) "The application of fractal analysis to mammographic tissue classification", *to appear in Cancer Letters.*

6. Fam, B. W., Olson, S. L., Winter, P. F., and Scholz, F. J. (1988) "Algorithm for the Detection of Fine Clustered Calcifications on Film Mammograms", *Radiology*, 169: pp. 333-337.

7. Chan, H. P., Doi, K., Vyborny, C. J., Lam, K. L., and Schmidt, R. A. (1988) "Computer-aided detection of microcalcifications in mammograms, methodology and preliminary clinical study", *Investigative Radiology*, 23(9): pp. 664-671.

8. Wee, W. G., Moskowitz, M., Chang, N. C., Ting, Y. C., and Pemmeraju, S. (1975) "Evaluation of Mammographic Calcifications Using a Computer Program", *Radiology*, 116: pp. 717-720.

9. Davies, D. H. and Dance, D. R. (1990) "Automatic Computer Detection of Clustered Calcifications in Digital Mammograms". *Physics in Medicine and Biology*, 35(8): pp. 1111-1118.

10. Shen, I., Rangayan, R. M., and Desautels, J. E., "Shape Analysis of Mammographic Calcifications". In Proceedings of the Fifth IEEE Symposium on Computer-Based Medical Systems, June 1992.

11. Woods, K. S., Solka, J. L., Priebe, C. E., Doss, C. C., Bowyer, K. W., Clarke, L. P. (1993) "Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalcifications," *Biomedical Image Processing and Biomedical Visualization, SPIE Electronic Imaging Science and Technology, 1-4 Feb.*

12. Woods, K. S., Solka, J. L., Priebe, C. E., Kegelmeyer, P. K., Jr., Doss, C. C., Bowyer, K. W. (1994) "Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalcifications in Mammography", scheduled to appear in the *International Journal of Pattern Recognition and Artificial Intelligence.*

13. Scott, D. W. (1992) *Multivariate Density Estimation.* John Wiley & Sons, New York, NY.

14. Duda, R.O., and Hart, P.E. (1973) *Pattern Classification and Scene Analysis.* John Wiley & Sons, New York, NY.

15. Lehmann, E. L. (1991) *Testing Statistical Hypothesis.* Wadsworth and Brooks/Cole, Pacific Grove, CA.
Priebe, C.E. and Marchette, D.J. (1991)

16. Silverman, B.W. (1986) *Density Estimation.* Chapman and Hall. New York, NY.

17. Lachenbruch, P. A. and Mickey, M. R. (1968) "Estimation of error rates in discriminant analysis," *Technometrics, 10,* pp. 1-11.

18. Priebe, C. E., and Marchette, D. J. (1991) "Adaptive Mixtures: Recursive Nonparametric Pattern Recognition." *Pattern Recognition,* Vol. 24, No. 12, 1197-1209.