Improved Breast Cancer Diagnosis through Decision Fusion

Anna Leigh Rack-Gomer

Department of Electrical Engineering
Duke University

Advisor: Dr. Loren Nolte, Professor of Electrical and Biomedical Engineering

Abstract

With the introduction of several new modalities for the detection of breast cancer, it has become even more important to implement computer-aided diagnostic models to help generate the best decisions from multiple tests, particularly when including multimodal data sets such as a mammogram and gene expression profile. Decision fusion provides a statistical model that can best combine information from multiple tests, by taking into account the performance of each diagnostic as a detector of breast cancer. To show this, the promoter methylation levels of five specific genes (proven biomarkers of breast cancer) were analyzed in 19 known tumor tissue samples and 22 known normal tissue samples. The individual performance of each gene as a detector of cancer was measured using Receiver Operating Characteristic (ROC) analysis. The performance was also evaluated by calculating the area under the ROC curve (AUC). The information from each gene was then combined using three fusion algorithms: decision fusion, a summation approach, and the linear discriminate analysis (LDA), which is a commonly used computer-aided diagnostic model for breast cancer. All fusion methods resulted in improved diagnostic performance over that associated with any individual gene, emphasizing the need for computer-aided models that are able to fuse multiple diagnostic tests for breast cancer in the clinic. Furthermore, decision fusion had the largest area under the curve (AUC) of all three models, proving it to be the best model in this case.

Introduction

The incidence of breast cancer in women has steadily increased over the last 40 years, but with improved screening methods and treatment procedures the overall survival rate is now 88% [1]. The regular use of mammography in the clinic has enabled breast cancer to be detected much earlier, and women diagnosed with tumors still in their earliest stages (localized or smaller than 2cm) have a five year relative survival rate of 98% [2]. With these encouraging statistics, much research has been done in the development of innovative screening methods with higher sensitivity. These improved screening methods are gradually being introduced into the clinic. The overall accuracy of breast cancer diagnosis is extremely important, both in sensitivity and specificity. While the need for detection, especially at early stages, has already been emphasized, it is also important to prevent false alarms that will put patients through unnecessary, costly, and painful treatments.

Computer-aided diagnostic (CAD) models serve as valuable assistants to clinicians in making decisions based on screening data; and they can significantly improve both sensitivity and specificity. Linear discriminate analysis (LDA) is a well established statistical technique, which has been successfully used in detecting masses in mammography [3]. However, LDA and other CAD models are generally used with only one set of data, such as a single mammogram image. There has been very little emphasis on CAD models that will separate cancerous cases from normal case using multiple sets of data, potentially taken from different testing modalities. This is unfortunate as it has been shown by Sahiner, et al. [4] that the fusion of extracted features from multiple modalities, in this case a mammogram and a 3D ultrasound, will increase the overall classification accuracy compared to single-modality classifiers.

Monitoring epigenetic alterations is a novel technique for the early detection of breast cancer, and it requires a multiple detector solution. Modification in the expression of key regulatory genes is an early and frequent event in the development of breast cancer [5] and can therefore serve as a biomarker for the disease. Promoter hypermethylation is a significant contributor to carcinogenesis through the resulting inactivation of tumour-supressor genes [6]. A DNA-based approach for early detection of breast cancer has the potential to be very effective, since DNA extracted from patient's plasma, serum or other body fluids can be amplified by PCR technology and achieve very high sensitivity [7]. Since only small amounts of fluid are required, it would be a convenient clinical diagnostic test.

There has been a large number of published studies on specific genes with altered expression in the presence of cancer. Widshcwendter and Jones [7] reviewed more than 40 genes whose expression is lost in breast cancer because of promoter hypermethylation. In one study, Fackler, et al. [8] measured quantitatively the promoter methylation levels of RASSF1A, TWIST, Cyclin D2, HIN1 and RARβ [4] in the presence and absence of breast cancer. They were able to show that the average level of methylation among tumor samples was significantly higher, although with varying degrees, for each of these five genes. It was then necessary to combine information from all five genes to make the best decision on the presence of cancer. The group chose a cumulative approach and summed the methylation levels across all five genes. However, a quantitative assessment of promoter methylation conducted by Lehmann et al. [5] has clearly shown huge gene-specific differences in the extent of methylation in tumor tissue. By summing methylation levels, these differences are ignored and important diagnostic information is lost.

A more robust statistical approach is needed to incorporate each gene's performance as a detector of cancer in the final decision. This can be accomplished through a statistical algorithm called decision fusion. Decision fusion requires that binary decisions are made independently at each detector. These decisions are fused in the form of likelihood values that depend only on the known performance (sensitivity and specificity) of each detector. The result is a fused set of likelihood values that can be compared to a threshold to make a final decision. Decision fusion is advantageous because it handles heterogeneous data sources (multimodal data) well [9]. This is very

important in a clinical setting, where multimodal screening tests are quickly becoming a valuable option. It also reduces the problem of dimensionality, and it is easy to use and interpret in a clinical setting [9].

Methods

Data:

The methylation data used in this experiment were provided by Saraswati Sukumar and Mary Jo Fackler of the Breast Cancer Program at Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins. Sets of DNA from 9 normal mammoplasty, 13 benign, and 19 tumor specimens were analyzed by quantitative multiplex methylation-specific PCR (QM-MSP) for gene promoter hypermethylation of RASSF1A, TWIST, Cyclin D2, HIN1 and RARβ [8]. The relative amount of methylation in each unknown sample was calculated as % M = 100 x [no. of copies of methylated DNA / (no. of copies of methylated + unmethylated DNA)] [8]. The raw methylation data can be seen in Figure 1 below. Binary hypothesis testing was done, such that each sample fell in either H_1 (tumor tissue) or H_0 (normal/benign tissue). The true cases were known for each sample from biopsy. From this figure it is clear that higher levels of methylation generally occur in H_1 than in H_0.

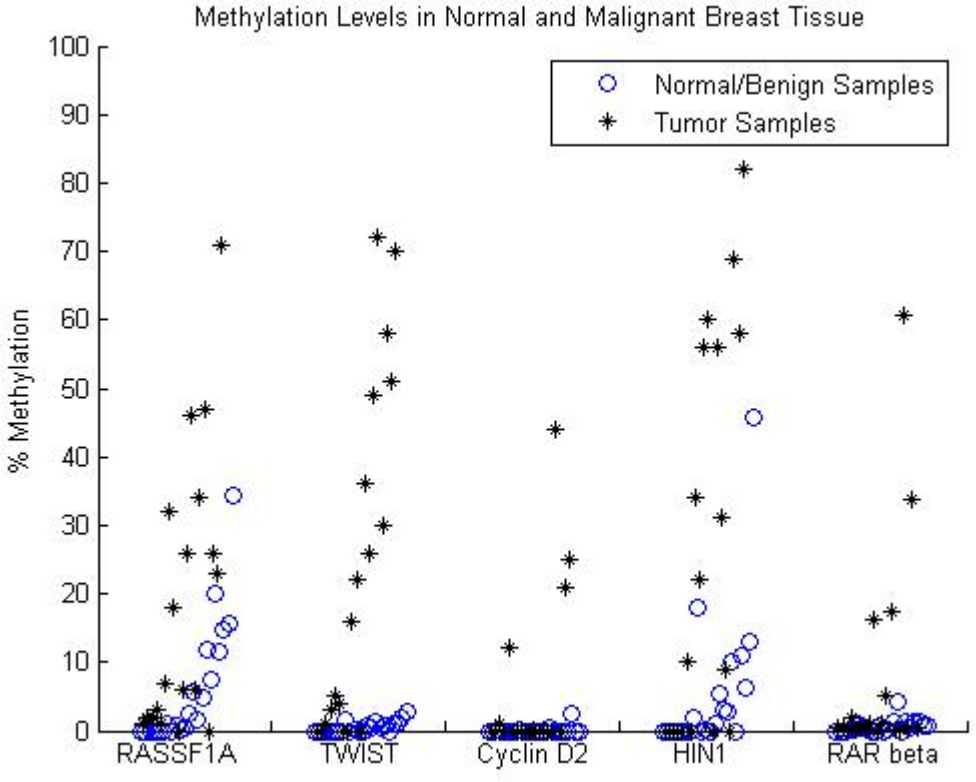

Figure 1. Comparison of gene promoter hypermethylation in each gene under the two cases H_1 and H_0.

ROC Analysis:

The performance of each gene as a detector of cancer was determined independently. Binary decisions were made on samples (blind as to the true case) for each gene such that if % M was greater than some threshold it was considered cancer (H_1), and if it was less than that threshold it was considered normal/benign (H_0). Receiver Operating Characteristic (ROC) analysis was used to measure the performance of each gene. The decisions made by thresholding the methylation levels in each gene were compared to the true cases of each sample to produce the curve in Figure 2.

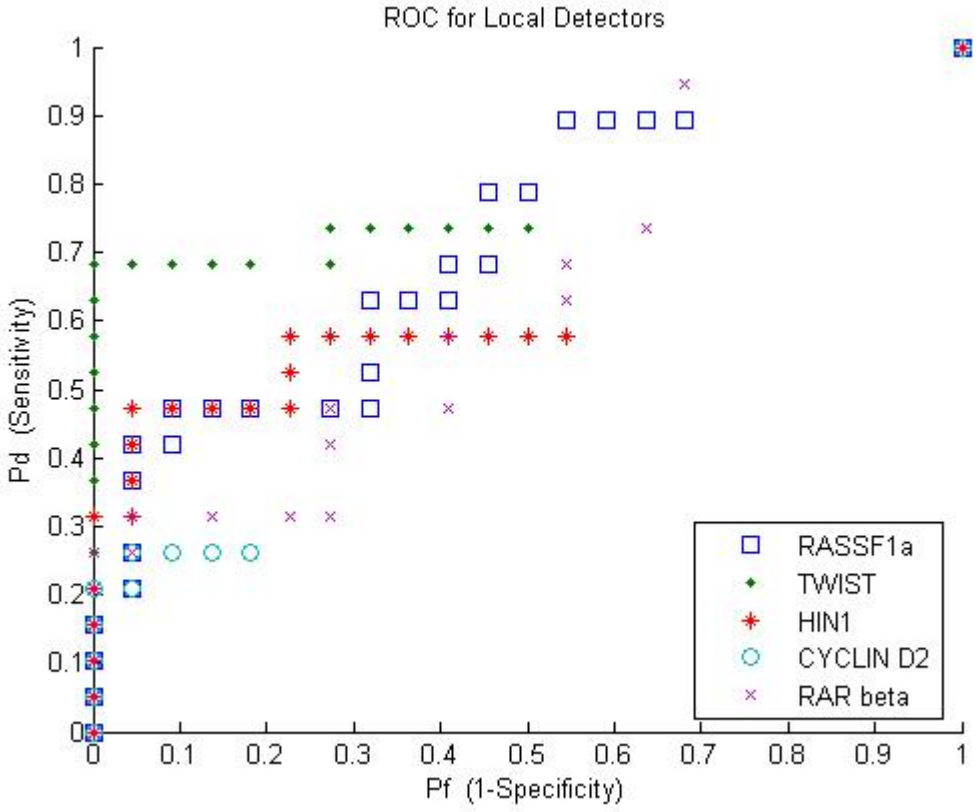

Figure 2. ROC analysis of each gene as a detector of breast cancer.

ROC analysis is shown graphically as a plot of the probability of a false alarm (1-specificity) verse the probability of detection (sensitivity). Measuring both sensitivity and specificity takes into account the limitations of diagnostic "accuracy" as a measure of decision performance [10]. However, the sensitivity and specificity alone do not provide a unique measurement of diagnostic performance because they depend upon the arbitrary selection of a decision threshold [10]. The receiver operating characteristic (ROC) curve has been shown to be a simple yet complete empirical description of this decision threshold effect, indicating all possible combinations of the relative frequencies of the various kinds of correct and incorrect decisions [10]. The closer the curve falls to the upper left-hand corner the better the diagnostic performance, whereas decisions being made entirely at random will produce the straight line: probability of detection (Pd) = probability of a false alarm (Pf).

Another measurement of performance is the area under the ROC curve (AUC), which approximates the probability that a randomly chosen sample will be correctly identified as either cancerous or benign [11]. The AUC can be seen for each gene in the table below.

Table 1. Area under the ROC curve for individual genes

Gene	RASSF1A	TWIST	Cyclin D2	HIN1	RARβ
AUC	0.72	0.79	0.56	0.64	0.64

Cumulative Approach:

Three decision algorithms were implemented to combine the methylation levels from all five genes and allow decisions to be made based on this multiple detector data set. The first algorithm was a cumulative approach in which the methylation levels (%M) for all five genes in a sample were summed. The final sums, one from each sample, were hypothesis tested such that any sample with a cumulative methylation value of greater than some threshold was considered to be in H_1, and any sample below that threshold was considered to be in H_0.

Linear Discriminate Analysis:

The second approach used was Linear Discriminate Analysis, a commonly used computer-aided diagnostic (CAD) model. This algorithm forms an optimal linear combination of the multi-dimensional data that maximizes the group mean separation of the tumor and normal objects [12]. The LDA algorithm was executed in Matlab to find the linear function that best separated H_1 and H_0 and to provide the probability that each sample fell within H_1, given its position relative to the separating function. These probabilities were compared to a threshold and a decision was made of H_1 or H_0.

Decision Fusion:

Decision fusion was also performed on the data. In this procedure, separate decisions were made using the data from each gene independently. Five individual thresholds were selected such that each gene/detector had approximately a 10% false alarm rate, or 90% specificity. The 90^{th} quantile of the raw data known to be in H_0 was used as the threshold. Each sample was compared to the threshold and individual decisions were made for each gene: d = 0 for H_0 and d = 1 for H_1. By comparing the decision made for each sample to its true state, the probability of detection and false alarm for each gene were calculated. The five decisions (k = 5) made for each sample were fused in the form of a likelihood ratio.

Local detectors: decision $d = 0$ or $d = 1$ (P_d and P_f known)

$$\lambda(d_1 \ldots d_k) = \frac{P(d_1 \ldots d_k | H_1)}{P(d_1 \ldots d_k | H_0)}$$

Each decision was assumed to be statistically independent. This was shown to be a reasonable assumption by Yuwei Liao, even in the case of dependence [13]. The decision fusion algorithm thus became the product of the individual likelihood ratios from each detector, which depended only on the performance (Pd and Pf) of each detector.

$$\lambda(d_1 \ldots d_k) = \frac{P(d_1|H_1)P(d_2|H_1)\ldots P(d_k|H_1)}{P(d_1|H_0)P(d_2|H_0)\ldots P(d_k|H_0)}$$

$$= \prod_{\substack{i \\ d=1}}^{k} \frac{P_{d_i}}{P_{f_i}} \prod_{\substack{i \\ d=0}}^{k} \frac{1 - P_{d_i}}{1 - Pf_i}$$

The likelihood values were compared to a threshold, and a final decision for each sample was made. A schematic of the decision fusion method can be seen in Figure 3.

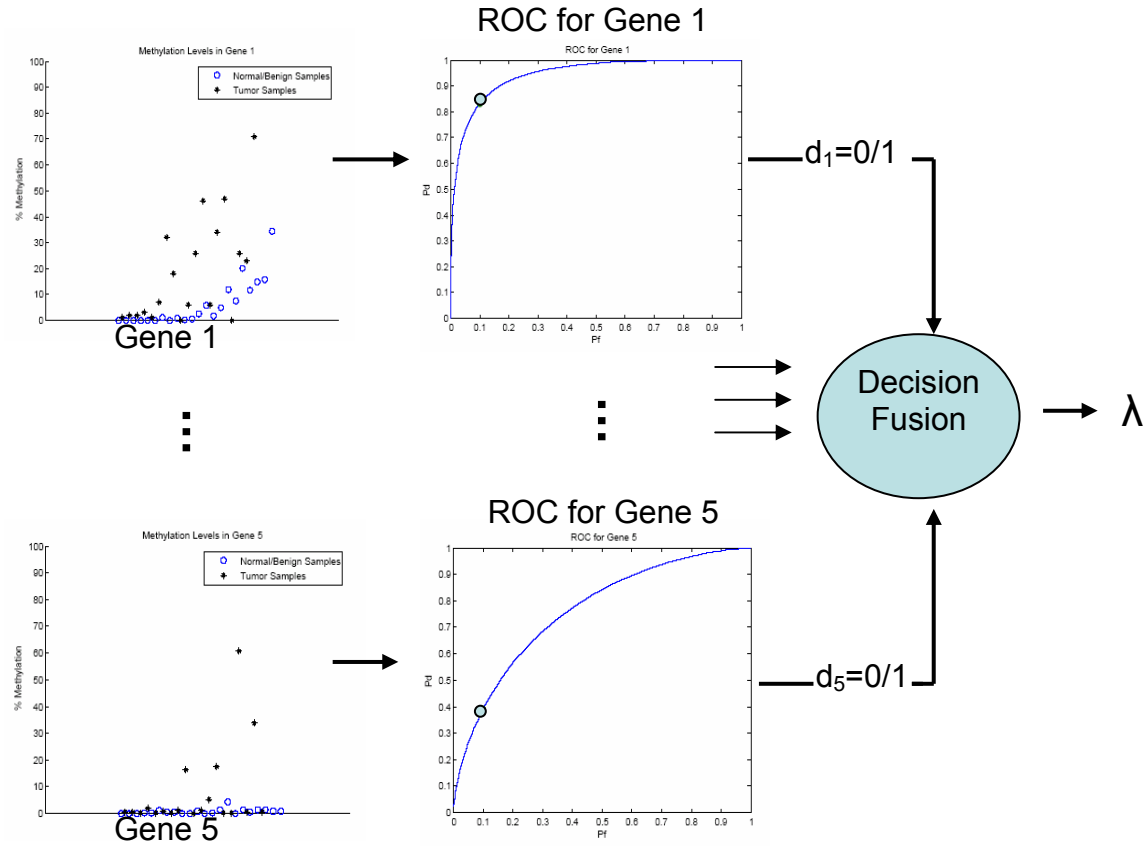

Figure 3. Schematic of the Decision Fusion Process

Results

The ROC analysis of the cumulative method, LDA, and decision fusion can be seen in Figure 4. The decision fusion method has two parts shown, an experimental result and an upper bound. For the experimental decision fusion curve, the final decisions made using decision fusion were compared to their true cases, and the algorithm's performance was evaluated with ROC analysis. The upper bound is the ideal curve that would have resulted using decision fusion, given the same detector performance if the local decisions were truly statistically independent. The two decision fusion curves are similar, but the ideal curve has a slightly better performance. The discrepancy between the two may be a result of some degree of dependence between genes, but it is not clear since such limited data were used. However, it is clear from the ROC curve that decision fusion was able to achieve a higher sensitivity at a higher specificity than either LDA or the cumulative method.

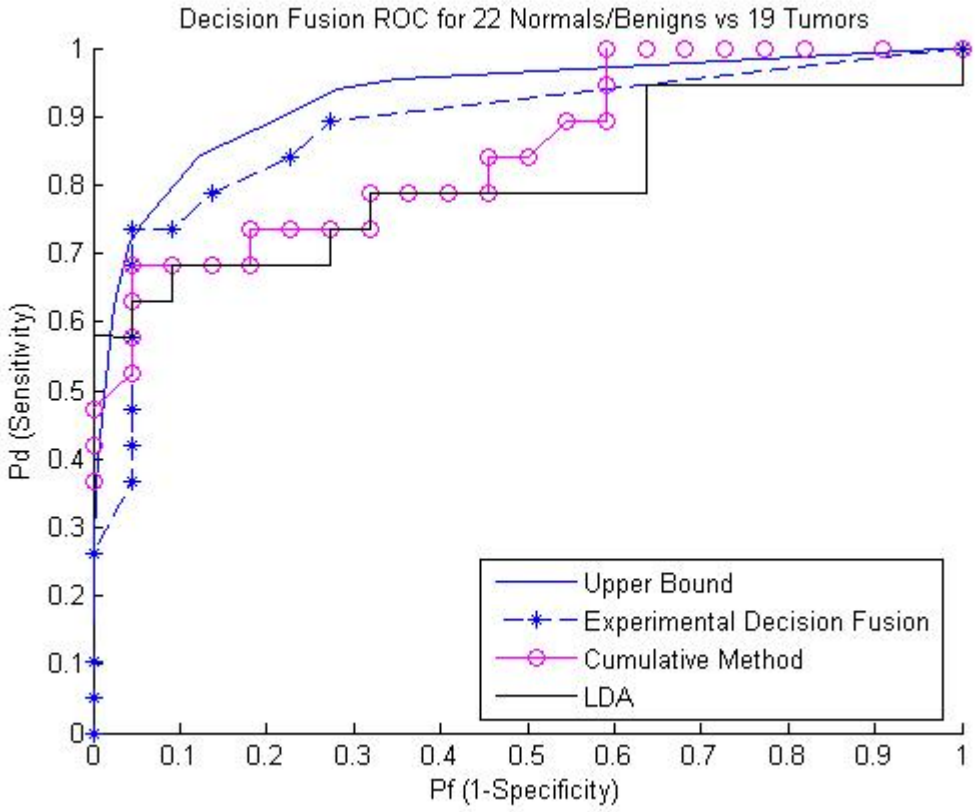

Figure 4. ROC analysis of the performance of decision fusion, LDA, and the cumulative method.

The performance of the three decision algorithms also was evaluated using the area under the curve metric, which can be seen in Table 2. The results for all three fusion methods were better than the performance of any individual gene. TWIST, which was

the best local detector, had an AUC that was ten percent lower than experimental decision fusion. The cumulative method out-performed TWIST by a very small margin and had the lowest performance of the fusion algorithms, of which decision fusion again performed the best. Although a cross-validation of these fusion methods would be desirable, the limited amount of data (31 samples) in this preliminary study did not lend itself to separate training and testing sets of reasonable size.

Table 2. AUC for each Decision Algorithm

Decision Algorithm	Ideal Decision Fusion	Experimental Decision Fusion	LDA	Cumulative Method
AUC	0.93	0.89	0.85	0.81

Conclusion

Breast cancer is still a serious health issue and its diagnosis is extremely important. The many screening options for breast cancer, whether currently in the clinic or still being researched, should be taken advantage of to give patients the most accurate diagnosis possible. Our results have clearly shown that fusing multiple detector data can significantly increase diagnostic sensitivity and specificity. Decision fusion provides a statistical approach that can optimally combine multiple detector data by including the performance of each local detector. A summation approach, which weights the performance of each detector equally, loses important diagnostic information, evidenced by the lower performance of this method, which achieved only a slight improvement over that of a single detector. Diagnostic information is not lost using linear discriminate analysis, but it is difficult to implement on multimodal data without alterations to the raw data. Finally, decision fusion proved to be the best fusion algorithm in this case, when its performance was measured via ROC analysis.

Acknowledgements

I thank Mary Jo Fackler and Saraswati Sukumar for generously sharing their methylation data, and Dr. Loren Nolte and Jonathan Jesneck for their invaluable advice and time.

References

[1] American Cancer Society, Inc. "Cancer Statistics 2005 Presentation." [Online Document]. 2005, http://www.cancer.org/downloads/STT/Cancer_Statistics_2005_Presentation.ppt

[2] American Cancer Society, Inc. "Breast Cancer Facts and Figures." [Online Document]. 2006, http://www.cancer.org/downloads/STT/CAFF2005BrF.pdf

[3] Chan H, Wei D, Helvie M, Sahiner B, Adler D, Goodsitt M, Petrick N. "Computer-Aided Classification of Mammographic Masses and Normal Tissue: Linear Discriminate Analysis in Texture Feature Space." Phys. Med. Biol. 40: 857-876, 1995.

[4] Sahiner B, Chan H-P, Hadjiiski LM, et al. "Multimodality CAD: combination of computerized classification techniques based on mammograms and 3D ultrasound volumes for improved accuracy in breast mass characterization." Medical Imaging 2004: Image Processing. SPIE: 5370;67-74, 2004.

[5] Lehmann U, Länger F, Feist H, Glöckner S, Hasemeier B, Kreipe H. "Quantitative Assessment of Promoter Hypermethylation during Breast Cancer Development." American Journal of Pathology. 160:2, February 2002.

[6] Jones PA, Laird PW. "Cancer epigenetics comes of age." Nat Genet. 21:163-167, 1999.

[7] Widschwendter M, Jones PA. "DNA methylation and breast carcinogensis." Oncogene. 21:5462-82, 2002.

[8] Fackler M, McVeigh M, Megrotra J, Blum M, Lange J, Lapides A, Garrett E, Argani P, Sukumar S. "Quantitative Multiplex Methylation-Specific PCR Assay for the Detection of Promoter Hypermethylation in Multiple Genes in Breast Cancer." Cancer Research. 64(13):4442-52, 2004.

[9] Jesneck J, Nolte L, Baker J, Floyd CE, Lo J. "Optimized Decision Fusion of Heterogenous Data for Breast Cancer Diagnosis" Med Phys. (submitted November 2005).

[10] Metz, CE. "Basic principles of ROC analysis." Semin Nuc Med. 8(4):283-98, October 1978.

[11] Hanley J, Mcneil B. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." Radiology. 143:23-36, 1982.

[12] Petrick N, Chan H, Sahiner B, Wei D. "An Adaptive Density-Weighted Contrast Enhancement Filter for Mammographic Breast Mass Detection." IEEE Transactions on Medical Imaging. 15:1, 1996.

[13] Liao, Yuwei, "Distributed Decision Fusion in Signal Detection -- A Robust Approach", PhD Thesis, Duke University, 2005.