AN IMPROVED ASYMMETRY MEASURE TO DETECT BREAST CANCER

Dave Tahmoush and Hanan Samet
University of Maryland, College Park

Abstract

Radiologists can use the differences between the left and right breasts, or asymmetry, in mammograms to help detect certain malignant breast cancers. An image similarity method has been improved to make use of this knowledge base to recognize breast cancer. Image similarity is determined using computer-aided detection (CAD) prompts as the features, and then a cluster comparison is done to determine whether there is asymmetry. We develop the analysis through a combination of clustering and supervised learning of model parameters. This process correctly classifies cancerous mammograms 95% of the time, and all mammograms 84% of the time, and thus asymmetry is a measure that can play an important role in significantly improving computer-aided breast cancer detection systems. This technique represents an improvement in accuracy of 121% over commercial techniques on non-cancerous cases.

Most computer-aided detection (CAD) systems are tested on images which contain cancer on the assumption that images without cancer would produce the same number of false positives. However, a pre-screening system is designed to remove the normal cases from consideration, and so the inclusion of a pre-screening system into CAD dramatically reduces the number of false positives reported by the CAD system. We define three methods for the inclusion of pre-screening into CAD, and improve the performance of the CAD system by over 70% at low levels of false positives.

1. Introduction

Breast cancer remains a leading cause of cancer deaths among women in many parts of the world. In the United States alone, over forty thousand women die of the disease each year [1]. Mammography is currently the most effective method for early detection of breast cancer [2]. For two-thirds of the women whose initial diagnosis of their mammogram is negative but who actually have breast cancer, the cancer is evident upon a second diagnosis of their mammogram [2]. Computer-aided detection (CAD) of mammograms could be used to avoid these missed diagnoses, and has been shown to increase the number of cancers detected by more than nineteen percent [3]. Improving the effectiveness of CAD could improve the detection of breast cancer, and could improve

[λ] The support of the National Science Foundation under Grants EIA-00-91474 and CCF-0515241, Microsoft Research, and the University of Maryland Graduate Research Board is gratefully acknowledged.

the survival rate by detecting the cancer earlier.

An automated pre-screening system classifies a mammogram as either normal or suspicious, while CAD picks out specific points as cancerous [4]. Incidence rates of breast cancer are less than one percent [1], so the majority of screening mammograms are of women without cancer. Incorporating the findings of a pre-screening system could then reduce the percentage of false positives by a large percentage. This would drastically reduce the cost of both CAD analysis and screening of mammograms, but the danger is that some cancers would be missed. This paper defines and compares several methods for the inclusion of pre-screening results into CAD, and analyzes the effectiveness of these methods as well as their associated cost in missed cancers on cases from the Digital Database for Screening Mammography [5].

The majority of work on CAD analysis of mammograms has focused on determining the contextual similarity to cancer, finding abnormalities in a local area of a single image [6,7]. This paper focuses on combining this with a spatial comparison in order to complete an asymmetry measure. The previous work has used methods ranging from filters to wavelets to learning techniques, but a detailed discussion of various imaging techniques is beyond the scope of this paper. Problems arise in using filter methods [6] because of the range of sizes and morphologies for breast cancer, as well as the difficulty in differentiating cancerous from non-cancerous structures. The size range problem has been addressed by using multi-scale models [7]. Similar issues affect wavelet methods, although their use has led to reported good results [8] with the size range issue being improved through the use of a wavelet pyramid [9]. Learning techniques have included support vector machines [10] and neural networks [8].

Detecting breast cancer is challenging because cancerous structures have many features in common with normal breast tissue. This means that a high number of false positives or false negatives are possible. Asymmetry can be used to help reduce the number of false positives so that true positives are more obvious. Previous work utilizing asymmetry has used wavelets or structural clues to detect asymmetry with correct results as often as 77% of the time [11,12]. Additional work has focused on bilateral or temporal subtraction, which is the attempt to subtract one breast image from the other [13,14]. This approach works well because it utilizes multiple images taken with the same machine by the same technician and analyzed using the same process in an effort to reduce the systematic differences that can be introduced. However, bilateral subtraction is hampered by the necessity of exact registration and natural asymmetry of the breasts.

We introduce an improved measure of asymmetry that is more approximate in nature and seems more robust to the large amount of noise in the data, using learning to determine a highly constrained number of model parameters. Minimizing the number of

parameters that are learned makes the model less subject to overfitting the noise in the data at the possible expense of accuracy. We then incorporate the asymmetry measurement into the CAD system using several different approaches.

The rest of this paper is organized as follows. Section 2 describes the new asymmetry measurement, while Section 3 discusses the evaluation of the measure and the results. Section 4 discusses techniques for the incorporation of the asymmetry measure into CAD. Section 5 describes the results of incorporating the measure into CAD, while Section 6 draws conclusions and discusses future work.

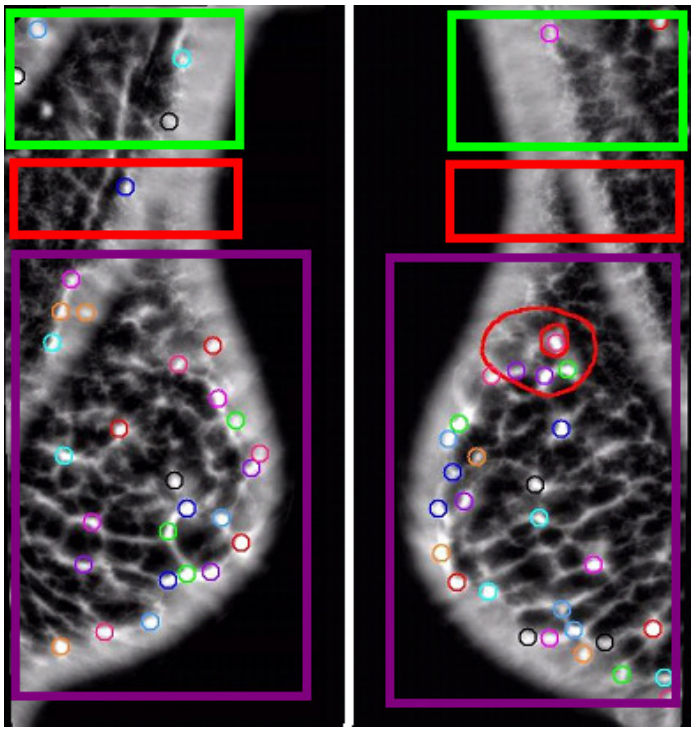

Figure 1. The cluster comparison method for measuring asymmetry. The tiny colored circles are the CAD prompts that are used to build the measure. The larger red circles are the radiologist's biopsy-confirmed diagnosis. The large colored boxes are example clusters that are learned in the determination of asymmetry. An occupancy threshold is learned for each cluster as well, and that is used for the determination of asymmetry.

2. Asymmetry Measurement

Our analysis starts with CAD prompts to find the contextually similar suspicious points that could be cancers in the mammograms. Points with a high CAD prompt have a higher chance of corresponding to an occurrence of cancer. This collection of CAD prompts is then sorted in decreasing order of suspicion. All suspicious sites that are closer than 5mm from a more suspicious site are removed to prevent multiple reporting of

the same site. This yields a set of potential detection sites that can be analyzed for asymmetry. Although this may not be the optimal choice of either CAD prompt or ranking, the spatial analysis that we used can be applied to any technique that can rank the suspiciousness of areas. We limit the analysis to the top thirty-two suspicious points. The number of points used is one of the variables that can be adjusted to optimize the analysis. Alternatively, we can also make use of a threshold on the CAD prompt instead of taking the top few. However, we chose to take the top few in order to be insensitive to image processing choices that might bias the analysis.

The analysis for asymmetry that we used does a comparison of the locations and values of the sets of suspicious points. Clusters in the 3D space are learned with a training set of images. Identical clusters are formed in images of the left and right breasts. In our earlier work [15] we had compared the populations in the clusters using a distance function, but found that using a threshold on the cluster populations themselves improved the performance significantly on the cancerous cases. The clustering is shown on two images in Figure 1. If the populations in the clusters exceed a learned limit, then the case is recognized to be cancerous. We used parametric learning techniques to determine the optimal structure and parameters for the clusters from the data.

3. Evaluation of the Asymmetry Measure

The analysis was done with cases that were normal mammograms and mammograms with malignant spiculated lesions from the Digital Database for Screening Mammography. Spiculated lesions are a type of breast cancer with central areas that are usually irregular, ill-defined borders, and lines radiating out from their central core. The training set had 39 non-cancerous cases and 37 cancerous cases, while the test set had 38 non-cancerous cases and 40 cancerous cases. The data is roughly spread across the density of the breasts and the subtlety of the cancer. The breast density and subtlety were specified by an expert radiologist. The subtlety of the cancer shows how difficult it is to determine that there is cancer. The training data set was used to determine the parameters for the clusters as well as the number of CAD prompts to use and the threshold for the comparison of the clusters. The other approaches are tested against the same test set in order to be unbiased.

Our results are summarized in Table 1. The results are good on all cases of the test set, correctly classifying 84% of the mammograms in the test set. However, it is much more important to correctly classify the cancerous cases, and we correctly classify 95% of the cancerous cases. Neither the subtlety nor the density of the cancer had an effect on the results. The comparison with a commercial system shows that the results are surprisingly good. Correct classification results of 96% of the cancerous cases and 33% of non-cancerous cases are

possible using the R2 ImageChecker system [16]. Our method showed an improvement on the non-cancerous cases of over 121%. This demonstrates the importance of asymmetry in pre-screening, since using only asymmetry achieves a better performance than a complete commercial system. The inclusion of additional factors other than asymmetry in the method should improve the results. However, the data sets used are different, as the R2 ImageChecker data contains all cancer types and our method has only the difficult to detect spiculated lesions. The R2 ImageChecker data set also had a much higher proportion of non-cancerous mammograms to cancerous cases.

Our method makes use of a spatial clustered analysis of the suspicious points, counting the number of suspicious points in the groups. Its success is an encouraging sign for the investigation and utilization of more complicated analysis techniques in medical imaging and analysis. It is also an improvement over our previous work [15].

Method	Performance on Cancerous Cases	Performance on Non-Cancerous Cases	Total Performance
Our New Cluster Technique	95%	73%	84%
Our Previous Asymmetry Technique	87%	71%	80%
R2 Image Checker	96%	33%	65%*
Wavelet	77%	77%	77%

Table 1: Results Table. The results compare favorably against the R2 ImageChecker system and other techniques. The total performance for the R2 ImageChecker system is not reported. Note that the R2 performance is extrapolated onto the data set used in this paper while the actual performance is significantly lower due to a higher proportion of non-cancerous cases.

4. Methods for the Incorporation of Pre-screening into CAD

There are three basic methods for including pre-screening into CAD analysis. The first is the strict method, where the pre-screening removes the non-cancerous cases entirely from the consideration of the CAD software. The second is probabilistic, where the probability of the case being cancerous or non-cancerous is determined by the pre-screening system and then incorporated into the CAD analysis. We also describe an improvement on our technique that we call an optimal approach, where a learning approach is used to try to determine the optimal factors for the inclusion of the pre-screening results into the CAD analysis. These methods will be defined and compared below.

The strict method is the simplest to define. Images that are screened as normal are removed from consideration by the CAD analysis. Since there are no false positives drawn from these cases, the number of false positives per image decreases. This is the most effective technique at reducing the number of false

positives, but it is also the most dangerous as mistakes by the pre-screening system cannot be rectified by the CAD system.

The probabilistic method relies on the statistics of the pre-screening method to adjust the output of the CAD system. To incorporate prescreeing into a CAD system, we made use of Bayes Theorem, P(CancerSite | Pre-screen) = {P(Pre-screen | CancerSite) P(CancerSite) / P(Pre-screen)}. The sites where pre-screening indicates cancer are thus given an increased probability of being cancerous, while sites where pre-screening does not indicate cancer are given a reduced probability of being cancerous. Since the pre-screening measurement is done on an entire case, all of the sites in those cases are affected similarly. This was the method that was employed in [17].

The optimal approach is a variant of the probabilistic approach, but instead of deriving the change from the underlying probabilities, the change is learned on a training set of cases. In theory, this approach can optimize the incorporation of pre-screening into CAD, but can be difficult in practice. In this case, P(CancerSite | Pre-screen) = A(Pre-screen) P(CancerSite), where A(Pre-screen) is the learned adjustment factor. This approach has more flexibility than the probabilistic approach, but is mush harder to implement. The choice of what to optimize is also a concern. There are two main options, optimizing the area under the ROC curve or optimizing the accuracy of the CAD results in a certain range of specificity. Both approaches were attempted and will be discussed.

5. Evaluation of Incorporating Asymmetry into CAD

The analysis was done with cases that were normal mammograms and mammograms with malignant spiculated lesions from the Digital Database for Screening Mammography [5], the same data that had been used for the development of the asymmetry measure. The training data set was used to determine the parameter A(Pre-screen) for the optimal approach. The other approaches are tested against the same test set in order to be unbiased.

The results were good at low levels of false positives in all three techniques as is shown in Figure 2, and it is at high levels and medium levels of false positives where techniques distinguish themselves. Using the probabilistic approach to incorporate pre-screening into CAD is shown to work well at low numbers of false positives per image and can improve the performance by over 70%, but at high levels of false positives per image, this technique has minimal effect. This is expected since using Bayes Theorem merely reduces the probability of the false positives and does not eliminate them.

The results of the strict approach are identical to the results of the probabilistic approach at low levels of false positives, but diverge at higher levels of false positives. Since this approach eliminates the false positives instead of just diminishing them, the results at high levels of false positives per image are worse than the probabilistic approach

because true positives are eliminated. However, in medium levels of false positives, the performance is significantly better than the probabilistic approach.

The optimal approach was tuned to determine the best performance at both low levels of false positives and the overall area under the ROC curve. The performance under both converged to the strict approach; however, this may be due to the pre-screening technique that was chosen.

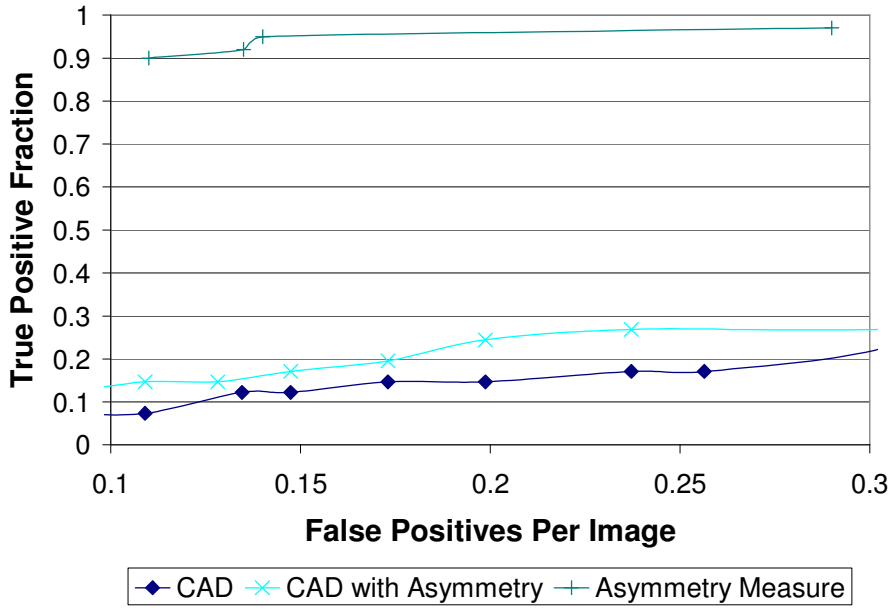

Figure 2. ROC curve comparison of the original CAD with the improved CAD with Asymmetry and the Asymmetry Measure. Incorporating asymmetry into CAD does improve the effectiveness. The asymmetry measure looks good because it is limited to producing only one false positive per image pair, while the CAD system can create many.

6. Conclusion

The overall results of using our asymmetry measurement techniques are good, our experiments on malignant cases yielded 95% accuracy suggesting that asymmetry is an important measure to incorporate into prescreening or CAD software. The technique can be tuned to be more effective at diagnosing cancerous cases, reaching 97% accuracy but at a significant loss of accuracy on non-cancerous cases. We suggest several ways to improve on the methods that we used to measure asymmetry. One method is to convert a mammogram into a connected graph structure of suspicious points and utilize known graph comparison methods for the measure. Alternatively, using non-space-filling or non-disjoint clusters could improve the method.

Our work has demonstrated the potential of utilizing techniques like image clustering and other methods with medical imaging. We have shown that we can effectively measure doctor-

defined quantities like asymmetry. We believe that in the future, the combination of capturing doctor-defined quantities like asymmetry and machine learning of parameters could be a powerful method for improving the quality of research in medical imaging, and this is one of the avenues of research that we intend to pursue.

We have described and evaluated several methods for the incorporation of pre-screening results into CAD systems, improving the accuracy of CAD by as much as 70% at low levels of false positives. Surprisingly, the comparison of the techniques demonstrates that the strict approach was nearly optimal. However, this may be due to the pre-screening technique. The overall performance is still strongly dependent on the effectiveness of the CAD system. The accuracy of the pre-screening is essential in order to prevent true positives from having their probabilities diminished, and the specificity is important for improving the effectiveness of the CAD system.

Bibliography

1. American Cancer Society. Breast Cancer Facts and Figures 2005-2006. American Cancer Society, Inc., Atlanta, GA, 2006.

2. J. Linda, W. Burhenne, S.A. Wood, C.J. D'Orsi, S.A. Feig, D.B. Kopans, K.F. O'Shaughnessy, E.A. Sickles, L. Tabar, C.J. Vyborny, and R.A. Castellino. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology, 215(554-562), 2000.

3. T.W. Freer and M.J. Ulissey. Screening mammography with computer-aided detection. Radiology, 220:781-786, 2001.

4. S. Astley, T. Mistry, C.R.M. Boggis, V. F. Hillier. Should we use humans or a machine to pre-screen mammograms? In Proc. of the 6th Int. Wks. Digital Mammography, pp 476-480, 2002.

5. M.D. Heath, K.W. Bowyer, D. Kopans et al. Current status of the digital database for screening mammography. Digital Mammography, Kluwer Academic Publishers, pp 457-60, 1998.

6. M.D. Heath and K.W. Bowyer, Mass detection by relative image intensity, In The Proc. of the 5th Int. Conf. on Digital Mammography, Medical Physics Publishing, Madison, WI, 2000.

7. P. Sajda, C. Spense, and L. Parra, Capturing contextual dependencies in medical imagery using hierarchical multi-scale models, In Proc. of the IEEE Int. Symp. on Biomedical Imaging, pp 165-168, 2002.

8. B.L. Kalman, S.C. Kwasny, and W.R. Reinus, Diagnostic screening of digital mammograms using wavelets and neural networks to extract structure, Technical Report 98-20, Washington University, 1998.

9. S. Lui, C.F. Babbs, and E.J. Delp, Multiresolution detection of spiculated lesions in digital mammograms, IEEE Trans. on Image Processing, vol. 6, pp 874–884, 2001.

10. R. Campanini, A. Bazzani, A. Bevilacqua, D. Bollini, D. Dongiovanni, E. Iampieri, N. Lanconelli, A. Riccardi, M. Roffilli, and R. Tazzoli, A novel approach to mass detection in digital mammography based on support vector machines, In Proc. of the 6th Int. Workshop on Digital Mammography, 2002.

11. R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, A.F. Frere, Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets, IEEE Trans. on Medical Imaging, 20(9), 2001.

12. P. Miller and S. Astley, Detection of breast asymmetry using anatomical features, In Proc. of the Int. Society for Optical Engineering Conf. on Biomedical Image Processing and Biomedical Visualization, vol. 1905, pp 433–442, 1993.

13. M.A. Wirth and A. Jennings, A nonrigid-body approach to matching mammograms, In Proc. of the IEEE Image Processing and its Applications, pp 484–7, 1999.

14. F.F. Yin, M.L. Giger, K. Doi, C.E. Metz, C.J. Vyborny, and R.A. Schmidt, Computerized detection of masses in digital mammograms: analysis of bilateral subtraction images, Medical Physics, vol. 18, pp 955–63.

15. D. Tahmoush and H. Samet, Using image similarity and asymmetry to detect breast cancer, Proceedings of SPIE – Volume 6144 Medical Imaging 2006: Image Processing, Joseph M. Reinhardt, Josien P. W. Pluim, Editors, pages 61441S-1 to 61441S-7, San Diego, CA, February 2006.

16. S. Astley and F.J. Gilbert, Computer-aided detection in mammography, Clinical Radiology, vol. 59, pp 390–9, 2004.

17. D. Tahmoush and H. Samet, Image Similarity and Asymmetry to Improve Computer-Aided Detection of Breast Cancer, Proceedings of the International Workshop on Digital Mammography (IWDM) 2006, pages 221–228, Manchester, UK, June 2006.