ELSEVIER SCIENTIFIC PUBLISHERS IRELAND

Cancer Letters 77 (1994) 183–189

CANCER LETTERS

The application of fractal analysis to mammographic tissue classification

Carey E. Priebe*[a], Jeffrey L. Solka[a], Richard A. Lorey[a], George W. Rogers[a], Wendy L. Poston[a], Maria Kallergi[b], Wei Qian[b], Laurence P. Clarke[b], Robert A. Clark[b]

[a]*Naval Surface Warfare Center, Dahlgren Division, Advanced Computation Technology Group, Code B10, Dahlgren, VA 22448, USA*
[b]*H. Lee Moffitt Cancer and Research Center, University of South Florida, Tampa, FL 33612, USA*

(Received 19 September 1993; revision received 25 October 1993; accepted 29 October 1993)

Abstract

As a first step in determining the efficacy of using computers to assist in diagnosis of medical images, an investigation has been conducted which utilizes the patterns, or textures, in the images. To be of value, any computer scheme must be able to recognize and differentiate the various patterns. An obvious example of this in mammography is the recognition of tumorous tissue and non-malignant abnormal tissue from normal parenchymal tissue. We have developed a pattern recognition technique which uses features derived from the fractal nature of the image. Further, we are able to develop mathematical models which can be used to differentiate and classify the many tissue types. Based on a limited number of cases of digitized mammograms, our computer algorithms have been able to distinguish tumorous from healthy tissue and to distinguish among various parenchymal tissue patterns. These preliminary results indicate that discrimination based on the fractal nature of images may well represent a viable approach to utilizing computers to assist in diagnosis.

Key words: Computational statistics; Computer assisted diagnosis; Wolfe patterns; Feature extraction; Pattern recognition; Fractals; Probability density function

1. Introduction

The development of computer assisted diagnosis (CAD) as a physician's tool is widely recognized as a desirable goal [4]. The application of CAD to mammography could well lead to broader based screening without increased medical costs. Additionally, a successful CAD system could aid radiologists in evaluating the myriad normal mammograms and segregate the questionable ones for further diagnosis. Any successful CAD system applied to medical imagery must necessarily be concerned with such issues as image segmentation, feature extraction, and pattern recognition. This

* Corresponding author.

SSDI 0304-3835(94)03275-N

paper addresses one method of feature extraction and pattern recognition based on image texture. The feature extraction technique is derived from the theory of fractals [2]. Related work in fractal analysis of mammographic images can be found in references [3,7,9]. The pattern recognition system we have developed is based on a branch of statistics known as Computational Statistics [19].

Computational Statistics is most useful in dealing with extremely large data sets and data sets that cannot be represented by usual statistical models such as normal (or other closed form) distributions. The texture information in a digitized medical image such as a mammogram easily represents a data set of 100 000 local observations. Additionally, our observations indicate this data is not well represented by a normal (or gaussian) distribution. With the advent of current computer technology, manipulation of data sets of this size or larger is quite feasible. Hence, the application of Computational Statistics to medical imagery and CAD represents a quite natural blend of technologies.

In dealing with the issues raised above we need to address and define several mathematical terms. The first is a probability density function or pdf. The pdf is analogous to a mass density which is described with respect to some variable (e.g., the mass density of the earth as a function of distance from the center). Here, the probability density will be a function of the image data. To estimate the pdf, we use a technique such as adaptive mixtures [12,13]. Adaptive mixtures is a means of calculating the pdf without making strict assumptions about the actual statistical distribution of the data and is a hybrid approach designed to maintain the best features of both kernel estimation [16] and finite mixture models [18]. That is, we do not assume that the data is normally distributed.

In order to compute the pdf estimate, we need to extract image features. For this work we have used features based on the concept of fractals which can be used to represent various textures [11]. Fractals have been used to describe many common images such as coastlines, clouds, and fern leaves [8]. In other words, fractals are geometric objects which have a non-Euclidean (non-man-made) character which can be described in part by a property called the fractal dimension which is different than the normal Euclidean dimension.

The science of fractals was first pioneered by Mandelbrot [8]. Barnsley, et al. [1] showed that many biological systems could be described using fractal geometry. This quantity (fractal dimension f_d) has been shown to correlate with subjective human texture classification and is defined by Pentland [10] as

$$E\{\,|\Delta I(\Delta x)|\,\} = E\{\,|\Delta I(1)|\,\}\,\|\Delta x\|^{H}$$

Here H is the difference between the normal Euclidean dimension (e_d) and the fractal dimension ($H = e_d - f_d$). The Euclidean dimension for an image is, of course, two. $E\{\ldots\}$ is the expected value of the quantity in brackets. $E\{\,|\Delta I(1)|\,\}$ is the expected change in the intensity one unit away from a given point. $E\{\Delta I(\Delta x)\}$ is the expected change in the intensity Δx units away. The symbol $\|\ldots\|$ is the standard Euclidean norm or distance. For the case of a Euclidean object where the fractal and Euclidean dimensions are the same (i.e., a non-fractal object) the expected change in the image intensity at a distance $\times$ units from some point is the same as the expected change one unit away. This result is in keeping with our intuition since the grayscale value is constant for such an object. Previous work of Caldwell et al. [3] has shown a correlation between fractal dimension and the subjective human classification of parenchymal patterns.

Many of the methods for estimating f_d are based on Richardson's Law [8]

$$M(\epsilon) = K\epsilon^{(e_d - f_d)},$$

where $M(\epsilon)$ is the measured property of a fractal at a scale of ϵ and K is a constant of proportionality. The measured quantity at a given point is the change in image intensity as a function of scale. Using this equation, and the technique described in Solka et al. [17], we extract three features that describe the texture. The first is directly related to the fractal dimension, the second is a measure of how well the fractal model fits the data, and the third is related to the local degree of contrast in the image. As will be seen below, the use of all three features gives improved results over the use of any single feature.

The process of extracting fractal features and computing a pdf must be done for each class of data represented in the image. Once we have the estimates of the class pdfs our job is to discriminate each class from the other classes. If we plot the various pdfs and show that they are distinct from one another, we can calculate discrimination boundaries and the task of discriminating the various classes is a relatively straightforward application of Bayes' rule [5]. Once we have good estimates of the pdfs based on data from a number of images (i.e., from training data) and have determined appropriate discrimination boundaries, we can, in effect, turn the process around and use these boundaries to classify regions in new images.

As stated above, when we extract fractal features for a given class we determine three features for each observation. We can think of these three features as being the coordinates of a vector in three dimensional space. While more information is often contained in higher dimensional feature space, the difficulty associated with pdf estimation increases dramatically with any increase in the dimensionality of the observations [15]. To simplify our computations we can use the Fisher linear discriminant (FLD) [5] which allows us to project these three dimensions to the one dimension that is in some sense best for discrimination and thus decreases the computational complexity. Additionally, this projection eases the problem of illustration since we are able to plot the results.

For each class of training data we have some number of observations that we label as X_i which represents the observations from class i (where $i = 1,2$ for two classes). Using the fractal features as components we can represent each observation vector as $\underline{x} = [x_1, x_2, x_3]^T$, where each observation vector $\underline{x}$ is contained in the set X_i. The FLD yields a projection vector whose use results in a projected one dimensional observation y for each vector $\underline{x}$. Thus we obtain a set of observations Y_i for each class i. Then we employ a normalization transformation which yields Y'_1 with a mean of 0 and variance of 1. This same transformation is applied to class 2 of the projected observations yielding Y'_2 from Y_2. Next we estimate the pdfs for Y'_1 and Y'_2 using the adaptive mixtures technique and determine the discriminant boundary.

For a different image, which we label as testing data, we apply the FLD projection vector and the normalization transformation which were determined from the training data. We then estimate the pdfs for the two classes of data from the testing image using the projected and transformed data. This allows us to obtain performance estimates of this overall procedure and to judge its utility to CAD.

2. Methods

Images used in this study were provided courtesy of the H. Lee Moffitt Cancer Center and Research Institute and the Department of Radiology at the University of South Florida. The acquisition of the mammograms was accomplished under the breast screening program of the Center. The selected cases represent a variety of cancers of different subtlety. All tumorous regions were biopsy proven. The mammograms were digitized at ~220 mm/pixel and 8 bit/pixel by the Mindax Corporation of Minneapolis, MN using a DuPont FD-2000 scanner.

Using the techniques described by Solka et al. [17], features for tumorous and healthy tissues are extracted (10 000 healthy tissue observations and 500 tumorous tissue observations) from mammogram A shown in Fig. 1. This represents the training data. We develop the FLD projection vector and the normalization transformation from this image (which contains a malignant stellate mass ~10 mm).

A second mammogram not pictured here (mammogram B) was used for testing (10 000 healthy tissue observations and 300 tumorous tissue observations). The mammogram contains a malignant stellate mass (~6 mm). We apply the FLD projection and transformation developed from mammogram A to the data from mammogram B. We next reverse the order of training and testing data. That is, we use mammogram B as training data and mammogram A as testing data.

At this point, one might well ask if the process of projection from three dimensions to one dimen-

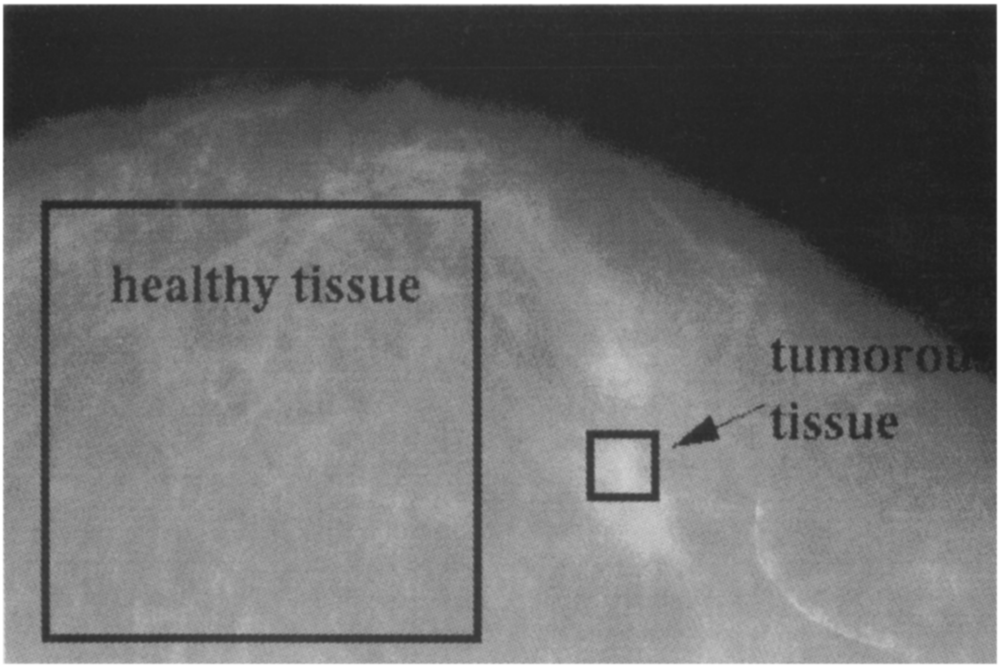

Fig. 1. Regions of interest in mammogram A. This image has been enhanced for presentation.

sion results in any degradation in discriminant capabilities. For the small number of images we are using in our work we can of course calculate the estimated pdfs using the full three dimensional vectors. We can also calculate the correlation matrices which show numbers which are a measure of how well any given feature correlates to the class variable for the two mammograms in this study.

The last part of our analysis utilized the four tissue patterns (labeled as N1, P1, P2, DY) which have been distinguished by Wolfe [20] and correspond to increasing breast tissue density and different morphology. The relationship of these patterns with breast cancer risk is the subject of much discussion. See, for instance, Saftlas and Szklo [14]. To determine the applicability of this technique to the discrimination of Wolfe patterns, we analyzed an additional eight mammograms. Our method consisted of determining pdfs for two Wolfe patterns, using this as training data, and using two other Wolfe patterns as testing data.

3. Results

3.1. Two class data discrimination

Fig. 2 is a plot of the pdfs of the projected data showing the separation of the healthy and tumorous classes. This figure indicates that for this

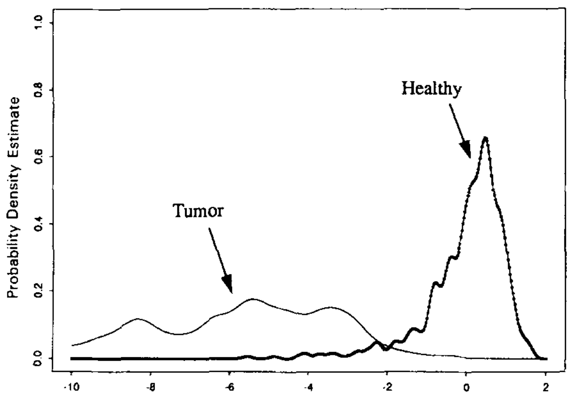

Fig. 2. Fisher linear discriminant pdfs for mammogram A.

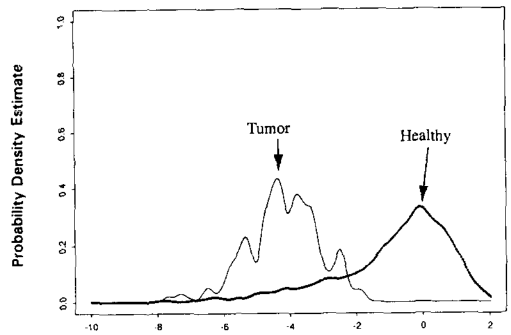

Fig. 3. Fisher linear discriminant pdfs for mammogram B using the independent projection.

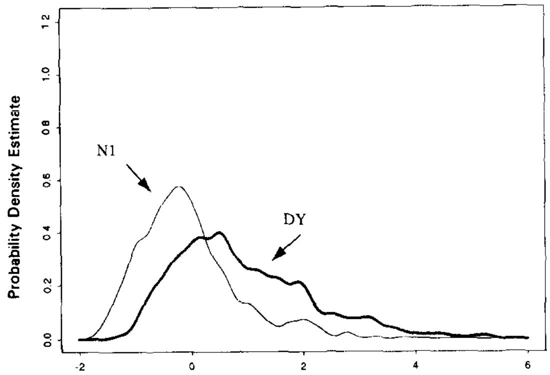

Fig. 4. FLD pdfs for mammogram N1 vs. mammogram DY.

image, the fractal dimension features can be used to distinguish healthy tissue from tumorous tissue.

We apply the FLD projection and transformation developed from mammogram A to the data from mammogram B. The estimated pdfs from this testing data are shown in Fig. 3. The discrimination boundary is clearly evident. Furthermore, this boundary appears to be invariant. That is, the boundaries in Figs. 2 and 3 are in the same place.

When the roles of mammograms A and B are reversed, the plots obtained exhibit the same behavior but with a different discriminant boundary. However, the results indicate that once a projection is chosen the discriminant boundary is invariant from training to testing data. This indicates that the discriminant boundary obtained from training images can be successfully applied to new test images. We hasten to add that since this is based on just two mammograms this represents an indication of what might be possible.

To examine the feature correlation we can consider the correlation against class obtained from the mammogram A data. The correlation number for the single projected dimension is −0.749. The correlation numbers for the three original features are 0.128, 0.367, and 0.094. Comparing absolute values, we find in each case that the number associated with the projected single dimension correlates better with class than any of the three original features. For the data from mammogram A, the FLD projection correlation of −0.749 is significantly higher than any of the class correlations for the original features. Although the magnitude of the correlations are smaller for the data from mammogram B, this same pattern still holds. This result is the motivation for using all three features with the FLD rather than any single feature such as the fractal dimension by itself. We can also compute correlations based on one image used as a training set and the other image used as a testing set. Comparing the correlations of the projected data, there is no serious degradation in the correlation when the projection is obtained independently. For example, we obtain a correlation of −0.749 for the projected mammogram A data when the projection is obtained using this same data. When the projection used is obtained from the mammogram B data this correlation is nearly unchanged (−0.724). The numbers obtained for the mammogram B data follow this same pattern for the correlation. This gives further credence to the generalization property, i.e., the projection obtained from one image can be used on data extracted from a separate image.

3.2. Wolfe's pattern analysis

The pdfs for the Wolfe patterns N1 and DY are shown in Fig. 4. This figure, along with Fig. 5 and Fig. 6, indicates the ability to discriminate among the various Wolfe patterns. These particular combinations of patterns were chosen simply for illustration and are not meant to be exhaustive. To demonstrate the same kind of generalization shown above, FLD projection vectors were determined using the data drawn from these new mammograms and were used in producing the pdfs for

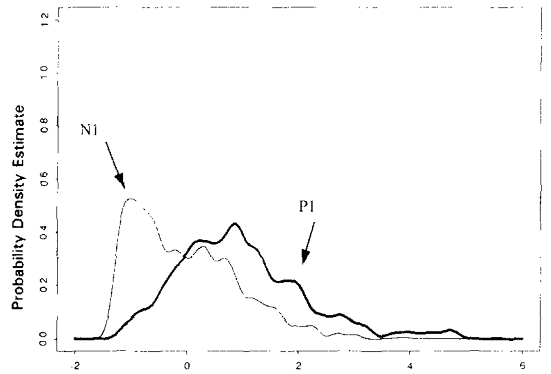

Fig. 5. FLD pdfs for mammogram N1 vs. mammogram P1.

the combinations N1–DY, N1–P1, and N1–P2 represented by Figs. 4, 5, and 6. In all cases, the patterns could be discriminated from one another and the discriminant boundaries generalize from training to testing image.

The pdfs here have a greater overlap than those discussed previously and certainly this overlap represents a measure of the amount of error, or probability of error, of incorrect classification of a single observation associated with a testing set. However, in the case of spatial data for which multiple observations can be assumed to be from a single class, this probability of error is drastically reduced by considering the joint pdfs as a product of the (independent) individual pdfs. In fact, as few as 10 observations, which is certainly a reasonable number for this application, could reduce the probability of error significantly.

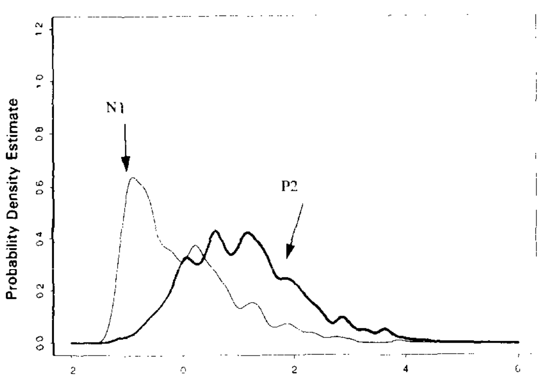

Fig. 6. FLD pdfs for mammogram N1 vs. mammogram P2.

4. Conclusions

Based on the limited number of cases studied, the preliminary indications are that the extraction of features based on the fractal nature of images, the reduction of dimensionality employing the Fisher Linear Discriminant projection vector, and the estimation of probability density functions using adaptive mixtures, represents a viable approach to pattern recognition that may be useful in the application of Computer Assisted Diagnosis to mammography. If the results of the efforts involving the Wolfe patterns can be extended to include non-malignant abnormal tissue this technique may aid in distinguishing these tissue types.

Research is continuing in this area and we have begun the digitization of a larger number of mammograms. Our plans are to develop an FLD projection vector based on training data from healthy tissue and tumorous tissue extracted from a substantial quantity of mammograms. The normalization transformation will be calculated and the pdfs of these two classes will be estimated. This model will then be tested on a sizable number of distinct mammograms to determine its utility. We will continue our research and testing until there exists sufficient evidence to support the viability of the approach. Additionally, we will extend our efforts to include non-malignant abnormal tissue.

For any system to be of practical value, the model must be formed from a sufficiently large pool of mammograms to instil confidence in the method. The pdfs formed from this large pool will form the baseline system and will be stored for archiving and reference purposes. The adaptive mixtures approach to pdf estimation will give the system the capability to update the pdfs with each new mammogram that is tested. Finally, our previous work indicates that the performance of our system can be improved through the use of wavelet transformations to enhance the areas of interest in the images [6].

5. Acknowledgments

The authors would like to thank an anonymous reviewer for helpful comments.

6. References

1 Barnsley, M.F., Massopust, P., Strickland, H. and Sloan, A.D. (1987) Fractal modeling of biological structures. Ann. NY Acad. Sci., 504, 179–194.

2 Barnsley, M.F. (1988) Fractals Everywhere. Academic Press, San Diego, CA.

3 Caldwell, C.B., Stapleton, S.J., Holdsworth, D.W., Jong, R.A., Weiser, W.J., Cooke, G. and Yaffe, M.J. (1990) Characteristics of mammographic parenchymal pattern by fractal dimension. Phys. Med. Biol., 35, 235–247.

4 Chan, H.P., Doi, K, Vyborny, C.J., Schmidt, R.A., Metz, C.E., Lam, K.L., Ogura, T. Wu, Y. and MacMahon, H. (1990) Improvements in radiologists' detection of clustered microcalcifications on mammograms. Invest. Radiol., 25, 1102–1110.

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

6 Li, H.D., Qian W., Clarke, L.P., Kallergi, M. and Clark, R.A. (1993) Mammographic image feature extraction and enhancement by wavelet transform. Proceedings of the 35th Annual Meeting of the AAPM (Abs), Washington DC, August 8–12.

7 Magnin, I.E., Cluzeau, F., Odet, C.L. and Bremond, A. (1986) Mammographic texture analysis: an evaluation of risk for developing breast cancer. Opt. Eng., 25, 780–784.

8 Mandelbrot, B. (1977) The Fractal Geometry of Nature. W.H. Freeman and Company, New York, NY.

9 Miller, P. and Astley, S. (1992) Classification of breast tissue by texture analysis. Image Vision Comput., 10, 277–282.

10 Pentland, A.P. (1986) From Pixels to Predicates. Ablex Publishing Corporation, Norwood, NJ.

11 Pietgen, H. and Saupe, D. (Eds.) (1988) The Science of Fractal Images. Springer-Verlag, New York, NY.

12 Priebe, C.E. and Marchette, D.J. (1991) Adaptive mixtures: recursive nonparametric pattern recognition. Pattern Recognition, 24, 1197–1209.

13 Priebe, C.E. and Marchette, D.J. (1993) Adaptive mixture density estimation. Pattern Recognition, 26, 771–785.

14 Saftlas, A.F. and Szklo, M. (1987) Mammographic parenchymal patterns and breast cancer risk. Epidemiol. Rev., 9, 146–174.

15 Scott, D.W. (1992) Multvariate Density Estimation. John Wiley and Sons, New York, NY.

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

17 Solka, J.L., Priebe, C.E. and Rogers, G.W. (1992) An initial assessment of discriminant surface complexity for power law feature. Simulation, 58, 311–318.

18 Titterington, D.M., Smith A.F.M. and Makov, U.E. (1985) Statistical Analysis of Finite Mixture Distributions. John Wiley and Sons, New York, NY.

19 Wegman, E.J. (1988) Computational Statistics: a new agenda for statistical theory and practice. J. Wash. Acad. Sci., 78, 310–322.

20 Wolfe, J.N. (1976) Breast patterns as an index of risk for developing breast cancer. Am. J. Radiol., 126, 1130–1139.