Comparison of Methods for Classification of Breast Ductal Branching Patterns

Predrag R. Bakic[1], Despina Kontos[2], Vasileios Megalooikonomou[2], Mark A. Rosen[1], and Andrew D.A. Maidment[1]

[1] Department of Radiology, University of Pennsylvania,
3400 Spruce St., Philadelphia, PA USA 19104
{Predrag.Bakic, RosenMar, Andrew.Maidment}@uphs.upenn.edu
[2] Computer and Information Sciences Department, Temple University,
1805 N.Broad St., Philadelphia PA USA 19122
{DKontos, Vasilis}@temple.edu

Abstract. Topological properties of the breast ductal network have shown the potential for classifying clinical breast images with and without radiological findings. In this paper, we review three methods for the description and classification of breast ductal topology. The methods are based on ramification matrices and symbolic representation via string encoding signatures. The performance of these methods has been compared using clinical x-ray and MR images of breast ductal networks. We observed the accuracy of the classification between the ductal trees segmented from the x-ray galactograms with radiological findings and normal cases in the range of 0.86-0.91%. The accuracy of the classification of the ductal trees segmented from the MR autogalactograms was observed in the range of 0.5-0.89%.

1 Background

The vast majority of breast cancers originate from the epithelial tissue of breast ducts. Due to low radiographic contrast, ducts are barely visible in mammograms. However, the breast ducts contribute to the complexity of the parenchymal pattern, which has been used in computer algorithms for early cancer detection and cancer risk estimation [1].

Breast ductal branching patterns have been previously analyzed by manually tracing ductal trees from galactograms, 2D x-ray images of contrast-enhanced ducts. That preliminary analysis, performed using ramification matrices (R matrices), was applied to classify galactograms with radiological findings and normal cases (i.e., no radiographic findings) [2]. More recently, the analysis has been extended to include other descriptors of ductal topology [3,4]. This paper compares three methods for describing and classifying breast ductal topology. The performance of these methods is compared using breast ductal networks as visualized in clinical x-ray and MR images.

2 Methods

In this section, we describe methods to acquire clinical images of the ductal network and to extract ductal topology descriptors from clinical images.

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 634 – 641, 2006.
© Springer-Verlag Berlin Heidelberg 2006

2.1 Data Acquisition

We have traced ductal topology in clinical x-ray galactograms and magnetic resonance (MR) autogalactograms (see Fig. 1). Galactograms are x-ray images of the breast, in which a small amount of contrast material has been injected into a nipple opening leading to a ductal lobe (subtree). The ductal subtrees have been segmented manually.

Autogalactograms refer to breast MR images of women in which portions of their ductal network enhanced due to the presence of protein or blood in the ducts [5]. The enhanced portions of the ductal tree were segmented in MR slices acquired with a 3D GRASS pulse-sequence [5]. A semi-automated region growing algorithm was used for segmentation. The 3D ductal topology was manually reconstructed from the segmented portions in each slice.

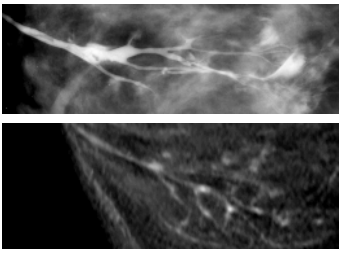


Fig. 1. Breast ductal network visualized in a galactogram (top) or an autogalactogram (bottom)

In this project we analyzed 22 clinical x-ray galactograms obtained retrospectively from 14 women (mean age 49.2 years, range 29–75 years), examined at the Thomas Jefferson University Breast Imaging Center, Philadelphia, PA, during the period of June 1994 through January 2001. Of these, seven women (13 images) had radiological findings corresponding to benign abnormalities, and eight women (12 images) had no findings; no malignant cases were available.

We also analyzed nine clinical autogalactograms obtained retrospectively from eight women (mean age 53.1 years; range 40-72 years), who had their breast MR studies at the Hospital of the University of Pennsylvania between June 2000 and April 2005. The five of eight women had radiological findings (four benign and one malignant; the latter with two identifiable ductal subtrees) and three cases were normal.

2.2 Description of Ductal Topology

R matrices. Elements of R matrices represent probabilities of branching at different levels of a ductal tree, computed following the Strahler labeling of individual ducts (see Fig. 2) [6]. Each R-matrix element $r_{k,j}$ can be expressed as [2]:

$$r_{k,j} = b_{k,j} / a_k , \tag{1}$$

where a_k is the total number of branches at the same level of the tree (those branches are identified by label k) and $b_{k,j}$ is the number of branches with label k, where the child branches are labeled k and j. The lateral branching is identified by labels $j \neq k$, while $j=k$ identifies bifurcation into child branches of the same order. The method for R matrix estimation from ductal trees has been described previously [7]. The R matrices estimated from clinical images have been used to realistically generate synthetic ductal network [8]. In addition, such estimated branching probabilities have been used for classification of galactograms with radiological findings and normal cases [2]. In this paper, we have extended that classification approach to include MR autogalactograms.

String encoding based descriptors. Another approach to represent the branching topology of the ductal network is to apply string encoding techniques. These techniques transform the initial ductal tree to a corresponding string signature. Further analysis is applied to these characterization signatures to investigate the properties of the branching topology. To avoid the problem of tree isomorphism, the ductal trees must be normalized to a canonical form [9]. The next step is to label the nodes (or branches) of the tree.

Prüfer encoding and tf-idf weighting. *Prüfer* encoding is a tree encoding scheme that reflects branching frequencies of the tree nodes [3]. This encoding constructs a unique string representation for each tree-like structure. The algorithm visits each node of the tree following an in-order traversal and depth-first search. During this process the encoding string is constructed; for each non-root node, the label of its parent is used to represent it. Fig. 2 shows the Prüfer encoding string for a sample labeled tree.

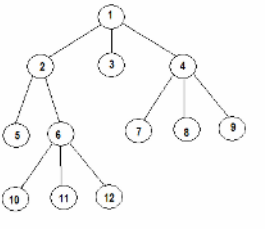


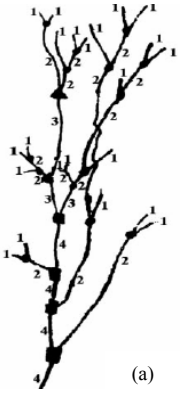
Fig. 2. An example of a labeled rooted tree. The corresponding: *Prüfer* encoding representation {*1 2 2 6 6 6 1 1 4 4 4*}, DFSE representation {*1 2 5 6 10 11 12 3 4 7 8 9*}.

The *tf-idf* weighting text mining technique can be further applied to the Prüfer encoding signatures to assign a significance weight to each string term (i.e., node label) and construct corresponding vectors of significance weights for each ductal tree. The cosine similarity metric can be applied to the tf-idf vectors in order to perform classification of the initial ductal trees [3].

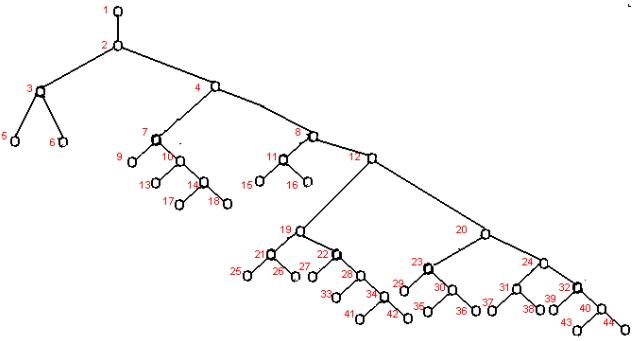
Depth-first encoding and fractal dimension. *Depth-first string encoding* (DFSE) is a straightforward encoding scheme that constructs a string representation for a tree by visiting each node following an in-order depth-first traversal. During this process each node is represented in the string by its label. Fig. 2 shows the DFSE for a sample labeled tree.

These DFSE signatures can be used for investigating the fractal properties of the ductal branching topology [4]. The *regularization dimension* [10] of the signatures is computed, which detects self-similar properties of the signature by looking into the scaling behavior of the lengths of less and less regularized versions of the string encoding representation. Classification is performed by thresholding the fractal dimension values. The performance can be assessed by ROC analysis [4].

Fig. 3 illustrates computation of the classification features for the three methods of ductal topology description, applied to the clinical galactogram from Fig. 1.



(a)



(c)

$$\mathbf{R}=\begin{bmatrix} 0.43 & 0.57 & \cdot & \cdot \\ 0 & 0.33 & 0.67 & \cdot \\ 0 & 0.75 & 0 & 0.25 \end{bmatrix}$$

(b)

Prüfer encoding = [1 2 3 3 2 4 7 7 10 10 14 14 4 8 11 11 8 12 19 21 21 19 22 22 28 28 34 34 12 20 23 23 30 30 20 24 31 31 24 32 32 40 40]

(d)

Depth-first string encoding= [1 2 3 5 6 4 7 9 10 13 14 17 18 8 11 15 16 12 19 21 25 26 22 27 28 33 34 41 42 20 23 29 30 35 36 24 31 37 38 32 39 40 43 44]

(e)

Fig. 3. An example of (a) manually traced ductal tree, (b) the corresponding R matrix, (c) the canonical form the tree labeled in a breadth-first manner, and the corresponding (d) Prüfer- and (e) Depth-first string encoded signatures. We computed tf-idf weighted vector and regularization dimension based on the signatures in (d) and (e), respectively.

3 Results

R matrices. Fig. 4 shows the range of values of the R matrix element used as the classification feature. The feature values were averaged separately over the auto-galactograms with (MR_F+) and without (MR_F-) findings, and over the

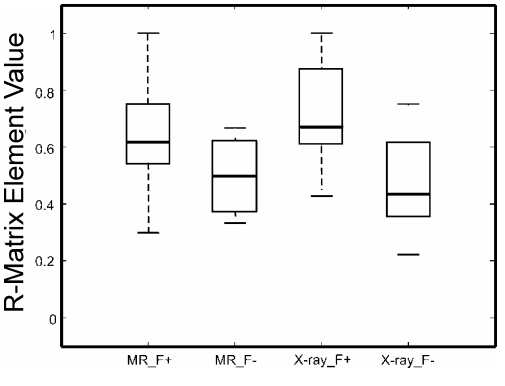


Fig. 4. Box-whisker plots of R-matrix based feature values used for classification of ductal trees. The whiskers indicate maximum and minimum feature values and the box indicates 25-, 50-, and 75-percentile values.

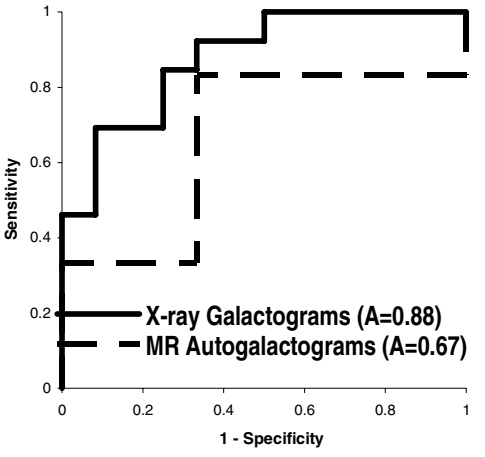


Fig. 5. The ROC curves corresponding to the class-ification of x-ray galacto-grams and MR auto galacto-grams, based on the values of R-matrix elements

galactograms with (X-ray_F+) and without (X-ray_F-) findings. The corresponding ROC curves are shown in Fig. 5.

Prüfer encoding and tf-idf weighting. Table 1 lists the accuracies of classifying x-ray galactograms and MR autogalactograms, based on the string representations computed using the Prüfer encoding and the tf-idf weighting. Leave-one-out k-nearest neighbor classification was performed based on the cosine similarity metric. The maximum accuracy was observed for x-ray galactograms at k=4. As there were only three MR autogalactograms with radiological findings, we were restricted to $k \leq 2$.

Table 1. Comparative x-ray galactogram and MR autogalactogram classification accuracies for Prüfer string encoding, assuming leave-one out k-nearest neighbor classifier based on cosine similarity

	Galactogram Classification Accuracy		
k	NF	RF	Total
1	80 %	41.67 %	59.09 %
2	80 %	66.67 %	72.73 %
3	80 %	50 %	63.64 %
4	100 %	83.3 %	**90.91 %**

	Autogalactogram Classification Accuracy		
k	NF	RF	Total
1	66.67 %	66.67 %	66.67 %
2	66.67 %	100 %	**88.89 %**
	--	--	--

Depth-first encoding and fractal dimension. Fig. 6 shows the range of the regularization dimension values computed for the DFSE signatures, used as the classification feature. The feature values were averaged separately over the auto-galactograms with (MR_F+) and without (MR_F-) findings, and over the galactograms with (X-ray_F+) and without (X-ray_F-) findings. The corresponding ROC curves are shown in Fig. 7. Fig. 7 shows also the ROC curve obtained after

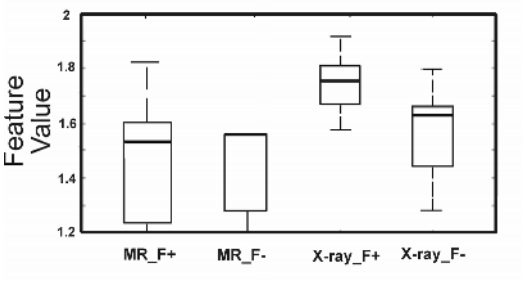


Fig. 6. Box-whisker plots of fractal based regularization dimension values used for classification of ductal trees. The whiskers indicate maximum and minimum feature values and the box indicates 25-, 50-, and 75-percentile values.

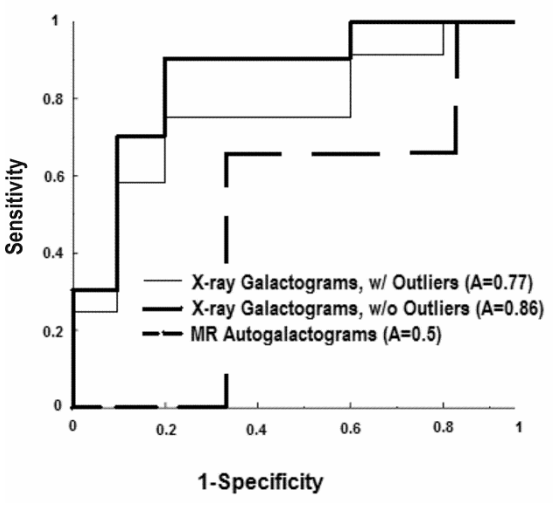


Fig. 7. The ROC curves corresponding to the class-ification of x-ray galacto-grams and MR auto galacto-grams, based on the regular-ization dimension values

removing two x-ray galactograms, whose regularization dimension values were identified as statistical outliers [4]. The two images were removed from the set of galactograms with radiological findings.

4 Discussion

The three methods for describing breast ductal topology compared are inherently different. The R-matrix method is based on the probabilistic nature of R matrices. Elements of an R matrix represent probabilities of branching at various levels of the ductal tree. Thus, a single matrix could be used to describe a family of the trees, with characteristic topological properties. We used this feature to generate synthetic ductal networks with realistic topological properties [8].

The string encoding methods analyzed in this paper have the ability to generate unique signature for each ductal tree, which is not possible using R matrices. Such a unique representation is useful for indexing, similarity retrieval, and similarity searches in large databases of tree structures. The string encoding signatures are usually vectors of variable length, where the length depends on the number of nodes in the analyzed tree. Direct use of such signatures in classification is a problem due to the high-dimensionality (i.e., number of features). In this work we considered two solutions: (1) the k-nearest neighbor classification of high-dimensional tf-idf weighted Prüfer strings and (2) DFSE signature vector dimension reduction by fractal analysis. Table 1 shows that relatively high classification accuracy (90.91% for x-ray galactograms and 88.89% for MR autogalactograms) was achieved using tf-idf weighted Prüfer strings signatures and k-nearest neighborhood classifier. This classification performance however, depends on the available sample size.

We evaluated the classification performance of the method based on R matrices and the fractal analysis of DFSE signatures using an ROC approach. Figs. 5 and 7 show that it can be seen that the two methods perform similarly for the x-ray galactograms; A=0.88 for the R-matrix based method, and A=0.77 and A=0.86 for the fractal based method, with and without the outliers, respectively. For the MR autogalactograms, the R-matrix method performed better than the fractal method (A=0.67 vs A=0.5, respectively). One reason for this difference in performance may be attributable to the small sample size (three autogalactograms with radiological findings and five normal cases).

We believe that R-matrix based classification may have a basis in the ductal biology. The ductal branching morphology is known to be influenced by variations in hormonal stimuli and interactions with the extracellular matrix. These factors alter the probability of lateral branching [11-13]. The elements of a R-matrix are able to quantify and distinguish the probability of lateral branching. We observed a significant difference between the matrix elements, consistent with our hypothesis about their biological correlation. Similar hypotheses about the string encoding based methods will be tested in our future experiments.

5 Conclusions

Classification results have been obtained using the three methods of description of the breast ductal branching topology. The methods were applied on two sets of ductal trees, extracted from clinical x-ray galactograms and MR autogalactograms. The R-matrices offer a higher-level representation of the tree branching topology. We hypothesize that such a representation may be related to the biological nature of breast pathology. On the other hand, string encoding based descriptors introduce a transformation of the tree topology from the 2D or 3D image space to the 1D signal (signature) space. A number of 1D signal processing methods could be then applied. Such methods could be advantageous for indexing and similarity retrieval in large databases of tree-like structures.

Acknowledgement

The work was funded by the Toshiba America Medical Systems Inc./Radiological Society of North America Research Seed Grant SD0329, by the National Cancer Institute Program Project Grant PO1 CA85484, and by the National Science Foundation under grant IIS-0237921. The authors are grateful to Catherine Piccoli, M.D. and Andrea Frangos, M.S. from Thomas Jefferson University, Philadelphia, PA, for providing anonymized x-ray galactographic images.

References

1. Li, H., Giger, M.L., Huo, Z., Olopade, O.I., Lan, L., Weber, B., Bonta, I.: Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location. Med. Phys. 31 (2004) 549-555
2. Bakic, P.R., Albert, M., Maidment, A.D.A.: Classification of Galactograms with Ramification Matrices: Preliminary Results. Acad Radiol. 10 (2003) 198-204
3. Megalooikonomou, V., Kontos, D., Danglemaier, J., Javadi, A., Bakic, P.R., Maidment, A.D.A.: A Representation and Classification Scheme for Tree-like Structures in Medical Images: An Application on Branching Pattern Analysis of Ductal Trees in X-ray Galactograms. In Proc. SPIE. 6144 (2006) 488-496
4. Kontos. D, Megalooikonomou, V., Javadi, A., Bakic, P.R., Maidment, A.D.A.: Classification of Galactograms using Fractal Properties of the Breast Ductal Network. In Proc. 3rd IEEE International Symposium on Biomedical Imaging. 2006 ISBI (2006)
5. Bakic, P.R., Rosen M.A., Maidment A.D.A.: Comparison of breast ductal branching pattern classification using x-ray galactograms and MR autogalactograms. In Proc. SPIE 6144 (2006) 707-716
6. Strahler, A.N.: Hypsometric area-altitude analysis of erosional topography. Bull. Geol. Soc. Am. 63 (1952) 1117–1142
7. Viennot, X.G., Eyrolles, G., Janey, N., Arques, D.: Combinatorial analysis of ramified patterns and computer imagery of trees. Comput. Graph. 23 (1989) 31–40
8. Bakic, P.R., Albert, M., Brzakovic, D., Maidment, A.D.A.: Mammogram synthesis using a 3-D simulation. III. Modeling and evaluation of the breast ductal network. Med. Phys. 30 (2003) 1914-1925
9. Chi, Y, Yang, Y, Muntz, R.: Canonical forms for labeled trees and their applications in frequent subtree mining. Knowledge and Info Systems 8 (2005) 203-234
10. Roueff, F., Vehel, J.L.: A regularization approach to fractional dimension estimation," in Proc. of Fractals 98, Malta, 1998.
11. Fata, J.E., Werb, Z., Bissell, M.J.: Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. Breast Cancer Research 6 (2004) 1-11
12. Robinson, G.W., Hennighausen, L., Johnson, P.F.: Side-branching in the mammary gland: the progesterone-Wnt connection. Genes & Devel 14 (2000) 889-894
13. Sternlicht, M.D., Key stages in mammary gland development: The cues that regulate ductal branching morphogenesis. Breast Cancer Research 8 (2006) 201-211