CLASSIFICATION OF GALACTOGRAMS USING FRACTAL PROPERTIES OF THE BREAST DUCTAL NETWORK

Despina Kontos[1], Vasileios Megalooikonomou[1], Ailar Javadi[1], Predrag R.Bakic[2], Andrew D.A.Maidment[2]

[1]Computer and Information Sciences Department, Temple University, Philadelphia PA 19122
[2]Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104

ABSTRACT

Several types of breast carcinomas tend to spread along the surface of the ductal lumen. Spontaneous nipple discharge can be an early symptom of such cancer development that does not otherwise result in visible mammographic changes. An imaging procedure that can visualize such symptoms is galactography. We focus on characterizing the topology of the ductal network in galactograms based on fractal properties. Statistically significant differences of fractal properties were detected among healthy subjects and patients with reported galactographic findings. We performed receiver operating characteristic (ROC) curve analysis in order to assess the accuracy of using the *regularization dimension* values for separating among ductal trees. The area under the ROC curve observed was 0.86.

1. INTRODUCTION

Nipple discharge, usually associated with benign changes, can indicate underlying malignant lesions in up to 15% of cases [1]. This type of spontaneous discharge is frequently caused by papilloma or ductal ectasia, symptoms that do not usually show recognizable changes in mammograms. An imaging procedure that can visualize such symptoms is galactography, during which x-ray mammography is performed after injecting a contrast agent into the lactiferous ducts [1, 2].

Several forms of cancerous lesions tend to exhibit a superficial spread along the surface of the ductal lumen or lobules. Studies have demonstrated that examining the morphology of the ductal network can provide valuable insight to the development of breast cancer and assist in diagnosing abnormal breast tissue [3]. Studies on mice have also supported the hypothesis that a particular relation exists between the branching of the ducts and cancerous pathology [4].

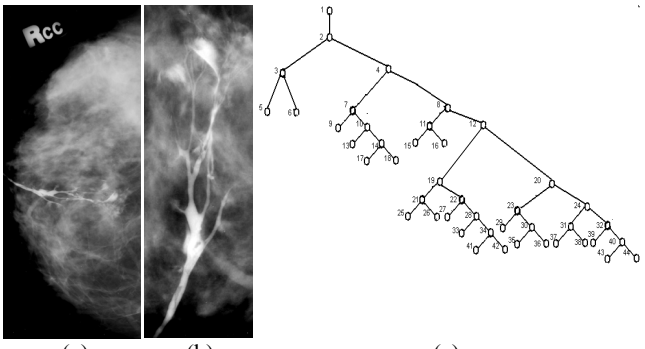
In order to evaluate the morphology of the ducts and its association to breast cancer symptoms, Bakic *et al.* [5] proposed a three-dimensional simulated model of the ductal network and a quantitative approach to classify galactograms based on ductal branching properties [6]. Focusing on characterizing tree-like structures in medical images, Megalooikonomou *et al.* [7] proposed a representation and classification scheme for tree-like structures such as the ductal network visualized by galactography. The latter approach employs the *Prüfer* encoding to obtain a symbolic representation of trees and a *tf-idf* mining technique for classification. In this paper we extend previous work [6, 7] by further examining fractal properties of the ductal topology in galactograms. In addition to proposing a quantitative classification scheme such as in previous reported approaches [6, 7], we interpret how fractal properties actually relate to the differentiation of ductal topology among healthy subjects and patients with reported galactographic findings of ductal ectasia, cysts or papilloma.

Fractal analysis can be used to describe properties that are not interpretable by the traditional Euclidean geometry [8]. In the field of mammography, fractal analysis has been employed for characterizing the parenchymal pattern [9], distinguish architectural distortion [10] and detect microcalcifications [11]. Most of these approaches utilize fractal analysis to characterize texture properties. We are interested in characterizing the actual topology of the ductal network. Some initial attempts to study self-similar properties of the breast parenchyma have been performed by fractal analysis of the lengths and cross-sectional areas of the ducts [12]. To the best of our knowledge, this is the first report on a fractal analysis of the breast ductal network topology, as visualized by galactography.

2. METHODS

2.1 Data and Preprocessing

In our study the ductal trees were manually segmented from the x-ray galactograms [6] (Fig. 1.a-b). We considered 22 galactograms from a total of 14 patients. From these images 10 corresponded to subjects with no reported galactographic findings (NF) and 12 to subjects with reported findings (RF). Follow up data was available for 8 out of the 12 patients and confirmed absence of malignancy over a period of 5 years (on average) [6].



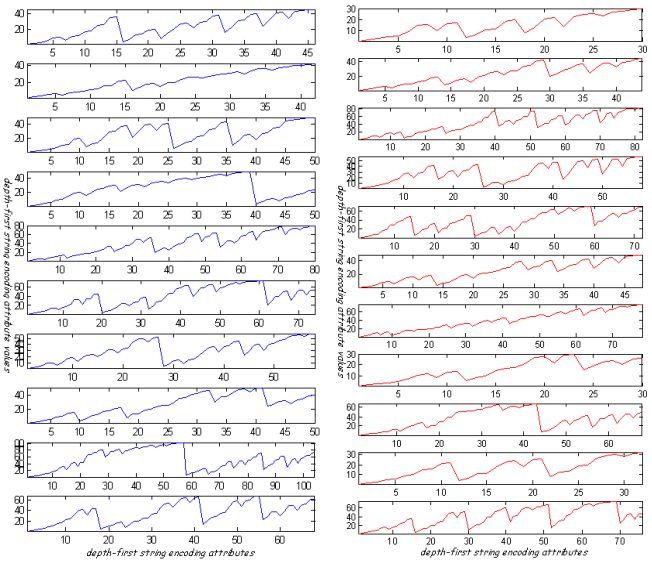
(a) (b) (c)

[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]

(d)

Fig. 1. (a) A real x-ray galactogram, (b) the ductal tree magnified, (c) the manually extracted ductal tree in a canonical form, and (d) the corresponding *depth-first string encoding*.

To avoid the problem of tree isomorphism we normalize the trees by applying a procedure for obtaining a canonical form of the tree [13]. The labeling of the trees was based on consecutive increasing integers assigned on a breadth-first search manner (Fig. 1.c). In order to study the fractal properties of the actual tree topology we encode the extracted ductal tree-like structures into a representation that reflects topological properties. We select the *depth-first string encoding*, which provides a one-to-one correspondence between the tree and the obtained string representation [13]. The *depth-first string encoding* constructs this unique string representation for each 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.1.d). These strings can be treated as *signatures* representing the original trees (Fig. 2).



(a) (b)

Fig. 2. The *depth-first string encoding* signatures of the ductal trees for (a) the NF class and (b) the RF class.

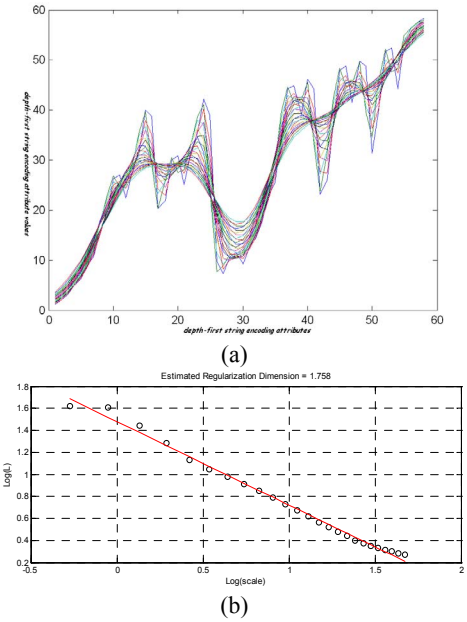
2.2 Fractal Analysis

In order to examine the fractal properties of the ductal branching we calculate the *regularization dimension* [14] of the 1D *signatures* corresponding to the initial ductal trees. The *regularization dimension* detects self-similar properties of the signature by looking into the scaling behavior of the lengths of less and less regularized versions of the 1D graph.

More specifically, consider the signal $S = s(t)$ as a function of time; in our case the *time* dimension is equivalent with the successive node labeling during the *depth-first string encoding* procedure. Also, consider a *kernel* function $X = x(t)$, such that $\int X = 1$. We define $X_a(t) = \frac{1}{a} x\left(\frac{t}{a}\right)$ as the dilated version of X at scale a; then $S_a = S \bullet X_a$ can be defined as the convolution of S with X_a. The length of the signal S on a finite interval T is given by $L_a = \int_T \sqrt{1 + S_a'(t)^2}\, dt$. According to the above assumptions and definitions, the *regularization dimension* of a signal S is defined as [14]:

$$\dim_R(S) = 1 - \overline{\lim_{a \to 0}} \frac{\log(L_a)}{\log(a)} \quad (1)$$

To calculate the actual *regularization dimension* we used the *FracLab* toolbox for Matlab developed at INRIA [15] (also available on-line at *http://www.irccyn.ec-nantes.fr/hebergement/FracLab/*). Figure 3.a illustrates an example of these successive dilations of the kernel and their convolution with the original 1D tree encoding signature.



(a) (b)

Fig. 3. An example of calculating the *regularization dimension* for the ductal tree 1D representations; (a) successive convolutions of the dilated Gaussian kernel with the original *depth-first string encoding* signature and (b) the corresponding log-log plot from which the *regularization dimension* is estimated as the slope.

We selected a Gaussian kernel and 64 progressive dilations of the kernel for the regularization process. To fit the log-log plot in order to calculate the actual *regularization dimension* (Eq.1) we selected a least square regression approach. Figure 3.b illustrates an example of a fitted log-log graph from which the *regularization dimension* is estimated. The selection of the kernel function should not affect the range of the *regularization dimension* difference between the classes.

Receiver operating characteristic (ROC) curve analysis [16] is further employed, in order to evaluate the accuracy of using the *regularization dimension* values of the two classes for separating NF from RF ductal trees. In order to interpret how the estimated fractal properties are related to the topology of the ductal network we take into consideration the nature of the *depth-first string encoding*.

3. RESULTS

We calculated the *regularization dimension* for all the ductal trees extracted from the available x-ray galactograms. Our initial focus was to investigate whether a statistically significant difference of this fractal dimension exists among the two classes. Table 1 illustrates the actual calculated *regularization dimension* for each subject and both classes.

We investigate the possibility of overlap between the two classes due to the fact that no malignancy existed for any of the subjects, regardless of the presence of galactographic findings. By performing a Lilliefors test for goodness of fit to a normal distribution we can confirm that the NF class follows a normal distribution. Nevertheless, this does not hold for the RF class when applying the same test. As the box-plot of illustrates in Figure 4, the distribution of the RF class is skewed, indicating the presence of outliers in the lower values of the distribution. Looking into the actual *regularization dimension* values in Table 1 we identified and removed View 17 and View 19 as outliers for the RF class. After removing these outliers the distribution of the RF class conforms to a normal distribution as illustrated in Figure 4. This is also confirmed by applying a Lilliefors test.

Regularization Dimension			
NF		**RF**	
View 1	1.79	**View 11**	1.76
View 2	1.33	**View 12**	1.67
View 3	1.66	**View 13**	1.81
View 4	1.28	**View 14**	1.76
View 5	1.62	**View 15**	1.69
View 6	1.67	**View 16**	1.74
View 7	1.46	**View 17**	1.39
View 8	1.65	**View 18**	1.66
View 9	1.44	**View 19**	1.49
View 10	1.63	**View 20**	1.58
		View 21	1.84
		View 22	1.92

Table 1. The fractional *regularization dimension* for every subject in each of the two classes.

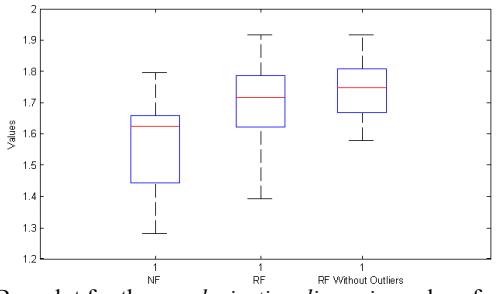


Fig. 4. Box-plot for the *regularization dimension* values for the NF class and the RF class before and after the removal of the outliers.

In order to assess the statistical significance of the fractal dimension divergence among the two classes, we applied the unpaired two-sampled t-test. The mean *regularization dimension* values were 1.55 ± 0.17 for the NF class and 1.74 ± 0.01 for the RF class, with p-value=0.01 and confidence intervals CI= [-0.32 -0.06]. This confirms the initial indication from Figure 2 that both classes seem to express self-similarity to some degree, with the RF class having a slightly more complex and noisy 1D representation.

We performed a receiver operating characteristic (ROC) curve analysis [16] in order to assess the accuracy of using the *regularization dimension* values for separating NF from RF ductal trees. For this analysis we used the ROC toolbox for MATLAB developed by G.C. Cawley (publicly available at http://theoval.cmp.uea.ac.uk/~gcc /matlab/default.html). The software also computes the *convex hull* of the ROC curve. The ROC curves obtained before and after removing outliers from the RF class are illustrated in Figure 5. The corresponding area under the ROC curve was equal to A_1=0.77 before removing the outliers. The performance improved up to A_2=0.86 when the outliers were removed from the RF class.

In order to provide an interpretation of the statistically significant difference of the *regularization dimension* between the two classes, we need to take into consideration the nature of the *depth-first string encoding* representation. As the encoding algorithm traverses the ductal tree, each node is represented in the string by its label. The labeling of the trees is based on consecutive increasing integers assigned on a breadth-first search manner. For this reason it is expected that the 1D *depth-first string encoding* signatures will resemble a positive signal with a trend for increasing values (Fig. 2).

In general, a fractal dimension indicates the degree of self-similarity of a structure. In that sense, both classes exhibit self-similarity since their *regularization dimension* is between the traditional Euclidean geometry dimension values of 1 and 2. As the scale of the convolution with the regularization Gaussian kernel changes, the 1D ductal tree representations exhibit self-similar variations. Also the *signatures* of the RF class seem to have a higher degree of self-similarity than the NF trees (i.e. higher *regularization dimension*).

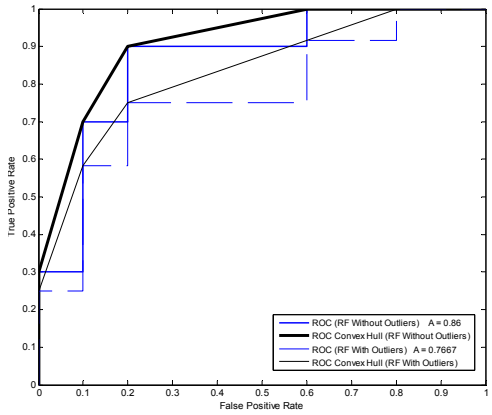


Fig. 5. The ROC curves for the *regularization dimension*.

These findings can be interpreted in the two following ways:

1. The local branching of the smaller ducts in the ductal network reflects in both structure and topology the branching of the core ducts.
2. The topology of the ducts in subjects with reported galactographic findings tends to be more complex indicating possible distortion of the normal parenchyma.

These are preliminary results. Larger populations of women need to be studied, including verified malignant cases, in order to derive more specific conclusions on the association between the ductal branching topology and the pathology of breast cancer. To perform this type of large-scale analysis, we intend to investigate methods for automating the process of tracing and extracting the ductal trees from contrast enhanced mammograms. This would be a critical step for introducing this type of analysis into clinical practice.

4. CONCLUSION

In this paper we presented a study of the fractal properties of the breast ductal network based on real ductal trees extracted from x-ray galactograms. We focused on characterizing the actual topology of the ductal network using the *regularization dimension*. We examined possible associations of the detected self-similar properties and symptoms of breast pathology. We also performed a receiver operating characteristic (ROC) curve analysis in order to assess the accuracy of fractal properties for discriminating among healthy subjects and patients with reported galactographic findings. The best ROC performance was A_2=0.86 after removing two outliers from our dataset.

5. ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation under grant IIS-0237921, by the Toshiba America Medical Systems Inc./Radiological Society of North America Research Seed Grant SD0329, and by the National Cancer Institute Program Project Grant PO1 CA85484. The funding agencies specifically disclaim responsibility for any analyses, interpretations and conclusions.

6. REFERENCES

[1] M. F. Hou, T. J. Huang, and G. C. Liu, "The diagnostic value of galactography in patients with nipple discharge," *Clinical Imaging*, vol. 25, pp. 75-81, 2001.

[2] H.P. Dinkel, A. Trusen, A. M. Gassel, M. Rominger, S. Lourens, T. Muller, and A. Tschammler, "Predictive value of galactographic patterns for benign and malignant neoplasms of the breast in patients with nipple discharge," *The British Journal of Radiology*, vol. 73, pp. 706-714, 2000.

[3] D. Moffat and J. Going, "Three dimensional anatomy of complete duct systems in human breast: pathological and developmental implications," *Journal of Clinical Pathology*, vol. 49, pp. 48-52, 1996.

[4] C. Atwood, R. Hovey, J. Glover, G. Chepko, E. Ginsburg, W. Robison, and B. Vonderhaar, "Progesterone induces side-branching of the ductal epithelium in the mammary glands of peripubertal mice," *Journal of Endocrinology*, vol. 167, pp. 39-52, 2000.

[5] P. R. Bakic, M. Albert, D. Brzakovic, and A. D. Maidment, "Mammogram synthesis using a three-dimensional simulation. III. Modeling and evaluation of the breast ductal network," *Medical Physics*, vol. 30, pp. 1914-1925, 2003.

[6] P. R. Bakic, M. Albert, and A. D. Maidment, "Classification of galactograms with ramification matrices: preliminary results," *Academic Radiology*, vol. 10, pp. 198-204, 2003.

[7] V. Megalooikonomou, D. Kontos, J. Danglemaier, A. Javadi, P. R. Bakic, and A. Maidment, "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. of SPIE Medical Imaging, San Diego, CA, 2006.

[8] B. Mandelbrot, *Fractal Geometry of Nature*. New York: W. H. Freeman, 1977.

[9] C. B. Caldwell, S. J. Stapleton, D. W. Holdsworth, R. A. Jong, W. J. Weiser, G. Cooke, and M. J. Yaffe, "Characterization of Mammographic Parenchymal Pattern by Fractal Dimension," *Physics in Medicine and Biology*, vol. 35, pp. 235-247, 1990.

[10] G. Tourassi, N. Eltonsy, A. Elmaghraby, and C. Floyd, "Detection of architectural distortion in mammograms using fractal analysis," in Proc. of SPIE Medical Imaging: Image Processing, San Diego, CA, vol. 5747, pp.930-938, 2005.

[11] L. Bocchi, G. Coppini, J. Nori, and G. Valli, "Detection of Single and Clustered Microcalcifications in Mammograms using Fractal Models and Neural Networks," *Medical Engineering and Physics*, vol. 26, pp. 303-312, 2004.

[12] P. Taylor, R. Owens, and D. Ingram, "3D fractal modelling of breast growth," in Proc. of 5th International Conference on Digital Mammography, Toronto, Canada, 2000.

[13] Y. Chi, Y. Yang, and R. Muntz, "Canonical forms for labeled trees and their applications in frequent subtree mining," *Knowledge and Information Systems*, vol. 8, pp. 203-234, 2005.

[14] F. Roueff and J. L. Vehel, "A Regularization Approach to Fractional Dimension Estimation," in Proc. of Fractals 98, Malta, 1998.

[15] J. L. Vehel and P. Legrand, "Signal and Image processing with FracLab," in Proc. of FRACTAL04, Complexity and Fractals in Nature, 8th International Multidisciplinary Conference, Vancouver, Canada, 2004.

[16] N.A. Obuchowski, "Receiver Operating Characteristic Curves and Their Use in Radiology," *Radiology*, vol. 229, pp. 3-8, 2003.