Classifying Ductal Tree Structures Using Topological Descriptors of Branching

Angeliki Skoura[1], Vasileios Megalooikonomou[1], Predrag R. Bakic[2] and Andrew D.A. Maidment[2],

[1]Department of Computer Engineering and Informatics, University of Patras, 26504, Rio, Greece
[2]Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce St, Philadelphia, Pennsylvania 19104, USA

{skoura, vasilis}@ceid.upatras.gr, predrag.bakic@uphs.upenn.edu, andrew.maidment@uphs.upenn.edu

Abstract. We propose a methodological framework for the classification of the tree-like structures of the ductal network of human breast regarding radiological findings related to breast cancer. Initially we perform the necessary preprocessing steps such as image segmentation in order to isolate the ductal tree structure from the background of x-ray galactograms. Afterwards, we employ tree characterization approaches to obtain a symbolic representation of the distribution of trees' branching points. Our methodology is based on Sholl analysis, a technique which uses concentric circles that radiate from the center of the region of interest. Finally, we apply the k-nearest neighbor classification scheme to characterize the tree-like ductal structures in galactograms in order to distinguish among different radiological findings. The experimental results are quite promising as the classification accuracy reaches up to 82% indicating that our methods may assist radiologists to identify image biomarkers in galactograms.

Keywords: medical image classification, galactography, tree-like structures.

1 Introduction

Several structures in human body follow a tree-like topology. Characteristic examples include the bronchial tree, the blood vessel network, the nervous system and the breast ductal network (Fig. 1). Nowadays, medical imaging modalities, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and radiographs have made available large series of images that visualize the above mentioned structures. Properties of the tree topology including spatial distribution of branching, tortuosity and asymmetry have been analyzed in literature so far and have been associated with altered function and pathology [1]. This type of information can be used to facilitate more accurate medical diagnoses.

For example regional changes in vessel tortuosity were utilized to identify early tumor development in human brain [2]. Moreover, the three dimensional analysis of the airway tree inside the lungs was used to distinguish pathologic formations from

normal lung structures [3]. Similarly, careful examination of the morphology of the ductal network visualized in galactograms provided valuable information related to the risk of breast cancer [4].
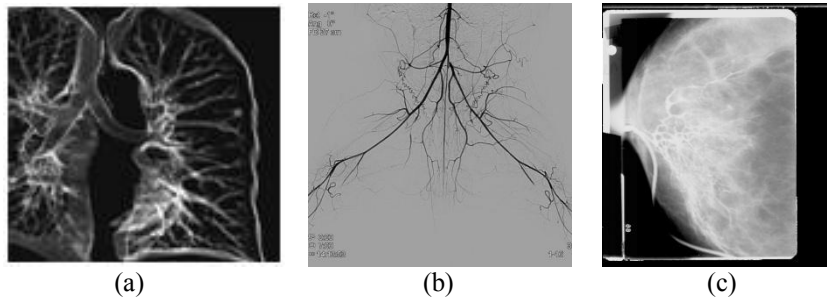

(a)   (b)   (c)

Fig. 1. Characteristic examples of tree shaped structures in medical images: (a) the human airway tree, (b) an angiography visualizing the hindlimb arterial network in a rabbit, (c) the network of human ducts depicted in a galactogram.

The motivation of this study is to provide insight into the discriminative characteristics of ductal networks in normal galactograms and in galactograms with reported radiological findings. Here, we propose descriptors for trees' topological analysis: the Sholl Analysis, the Sectoring Analysis and the Combined Method. These descriptors focus on the spatial distribution of tree branching points and we employ this information to characterize normal or disease states of the breast ductal tree. In order to evaluate the new descriptors, we use the classification scheme of k-nearest neighbors. Our approach has the potential benefit of advancing our understanding in breast anatomy and physiology and could assist early cancer detection and cancer risk estimation. Moreover, the proposed methodology could also be applied to other tree like structures of the human body such us the blood vessel network in order to detect ischemic or thrombotic patterns, or the bronchial tree to provide a primary detection of lung cancer or other malformations.

2   Background

Several studies in the literature have demonstrated that examining the morphology of the ductal network can provide valuable insight into the development of breast cancer and assist in diagnosing pathological breast tissue. For example, Bakic et al. [5] proposed a quantitative method based on Ramification matrices (R-matrices) to classify galactograms regarding radiological findings. An R-matrix represents a description of branching structures at the topological level and the elements of such a matrix represent the probabilities of branching at various levels.

More recently, Megalooikonomou et al. [6] proposed a multi-step approach for representing and classifying trees in medical images including galactograms. The authors employed tree encoding techniques such as the depth-first string encoding and the Prüfer encoding to obtain a symbolic string representation of the tree's branching topology. Based on these encodings, the problem of comparing tree structures was

reduced to the problem of string comparison. Moreover, they used the tf-idf text mining technique to assign a weight of significance to each string term. The classification was performed using the k-nearest neighbor scheme and the cosine similarity metric.

In the field of mammography, fractal analysis was also employed for characterizing the parenchymal pattern, distinguishing architectural distortion and detecting microcalcifications [7]. Fractal analysis can be used to describe properties that are not interpretable by the traditional Euclidean geometry and most of the above approaches utilize fractal analysis to characterize texture properties. In another approach, the tree asymmetry index was employed to characterize the topology of the ductal network in order to classify galactograms [8]. Finally, a combination of texture and branching descriptors was investigated to deepen the understanding of the relationships among the morphology, function and pathology of the ductal tree [9].

In this paper, we compare the proposed descriptors to the state-of-the-art Prüfer encoding approach and show that our methods compare favorably to it. A detailed description of the Prüfer encoding technique applied in galactograms can be found in [6].

3 Methodology

The goal of this paper is to investigate whether the distribution of branching points of ductal trees consists a discriminative factor concerning radiological findings in galactograms. Towards this direction, we present three methods that capture topological information regarding the tree branching points. The proposed methodology for analysis begins with the preprocessing of the images to segment the tree-like structures from the rest of the tissue depicted in the medical images. Then, we perform characterization of the extracted tree topologies applying the proposed techniques. Given a collection of such structures and the query tree, we perform similarity searches in order to find the ductal tree structures that are most similar to a query. Finally, we use the k-nearest neighbor classification scheme to evaluate the accuracy of the methods.

3.1 Image Preprocessing

Image preprocessing includes several steps needed to extract the tree structures from the background of the original medical images (Fig. 2(a)) and transform them into a suitable form for analysis. Initially, the Region Of Interest (ROI) of the galactograms, which is the part of the medical image where the ductal tree appears, is recognized (Fig. 2(b)). After focusing on the ROI, the segmentation of tree-like structures follows. Image segmentation is the boundary identification of objects or regions of special interest from the rest of the displayed objects or the background. In the application presented here, the task of segmentation has been done manually by medical experts (Fig. 2(c)). Then, the tree-like structures are reconstructed by identifying true positive branching points often resolving potential ambiguities such as anastomoses occurring mostly as a result of two-dimensional acquisition artifacts

[10]. Although more sophisticated and fully-automated methods of reconstruction could potentially be applied, such an approach is beyond the scope of this work as our main objective is the feature analysis and the classification of the tree-like structures. The final preprocessing step is the skeletonization of the ductal structures. This thinning process that reduces most of the original foreground pixels in a binary image to obtain a skeletal remnant that preserves the extent and the connectivity of the original structure is necessary in order to detect the exact positions of the branching points.
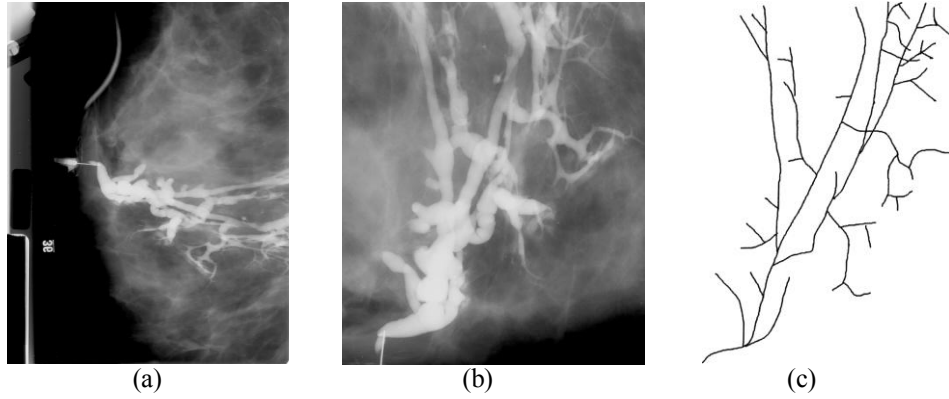

(a) (b) (c)

Fig. 2. Segmentation of a ductal tree: (a) an original galactogram, (b) the Region Of Interest of the medical image showing enlarged the ductal network, (c) the segmented tree structure manually traced by medical experts.

3.2 Spatial Distribution of Tree Branching Points

3.2.1 Sholl Analysis. The first method that we propose for capturing information about the spatial distribution of branching points is based on Sholl analysis [11]. According to Sholl's methodology, a number n of concentric circles having as center the mass center of the ROI is applied. As the common center of all circles, we consider the root node of the tree structure. More specifically if the first circle has a radius b, the radii of the concentric $2^{nd}, 3^{rd}, 4^{th}, ..., n^{th}$ circle is $2 \cdot b, 3 \cdot b, 4 \cdot b, ..., n \cdot b$ correspondingly. Based on these circles, a vector that counts the number of tree branching points between any two successive concentric circles is computed. The vector's length is equal to the number n of circles of Sholl's analysis. The element i of the Sholl's vector is computed as follows:

Sholl (i) = the number of branching points existing between the ith circle and $(i+1)^{th}$ circle. (1)

For example, the application of Sholl Analysis Fig. 3(b) with radius $b = 50$ pixels to the image Fig. 3(a) with dimensions 1150x790 pixels, produces the vector $Sholl = [0, 0, 1, 1, 0, 1, 0, 1, 2, 1, 2, 0, 2, 3, 1, 1, 1, 6, 1, 3, 0, 2, 4, 3, 2, 3, 0, 0]$.
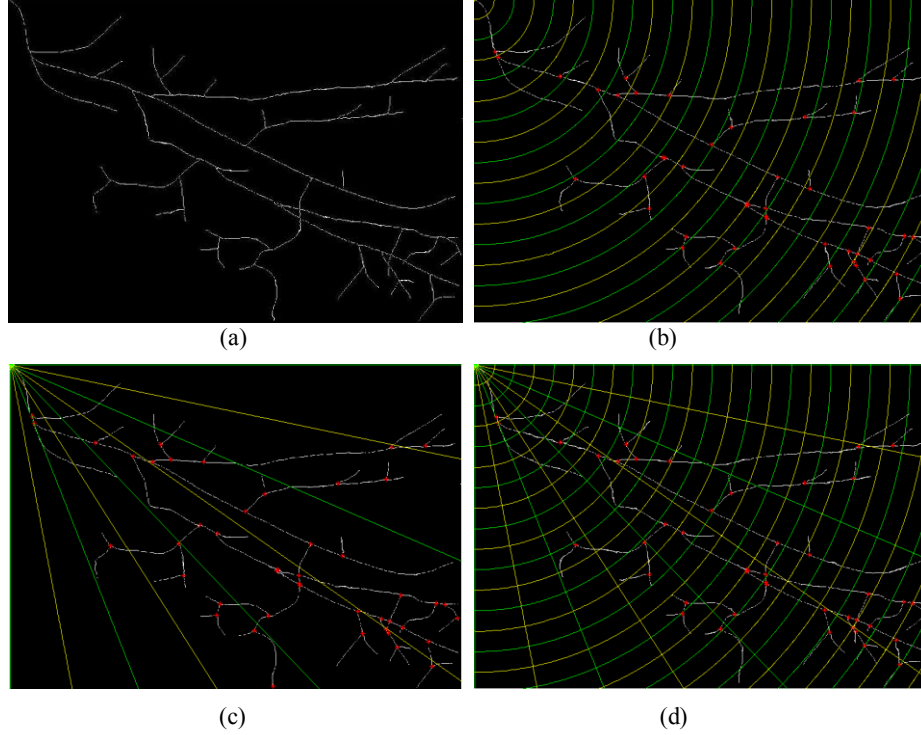
Fig. 3. Applying three methods to characterize the spatial distribution of tree's branching points: (a) the initial delineated ductal structure, (b) application of Sholl Analysis, (c) employment of Sectoring Analysis, (d) application of Combined Method.

3.2.2 Sectoring Analysis. While Sholl's method captures the spatial distribution of branching points radiating out of the tree root, the second method we propose, the Sectoring Analysis, reflects the distribution of branching points based on consecutive clockwise sectors. As a typical ductal tree covers at most one out of the four quadrants of the plane, we focus only on the quadrant that contains all tree nodes. We omit the remaining three quadrant of the plane to avoid zero elements in the final vector. The quadrant, where the tree structure is depicted, is divided into a number m of consecutive sectors. The m sectors have as common starting point the root of the ductal tree and each triangular sector covers a part of the plane of a central angle equal to $90°/m$. According to the above partition of the plane, the Sectoring vector counts the number of tree branching points that belong to each sector. The vector's length is equal to the number m of Sectoring analysis. The ith element of the Sectoring vector is computed according to following formula:

Sectoring (i) = the number of branching points existing inside the ith sector. (2)

An example of the application of Sectoring method using $m = 8$ sectors is presented in Fig. 3(c). The resulted vector is $Sectoring = [0, 0, 3, 5, 13, 17, 3, 1]$.

3.2.3 Combined Method. The idea of combining the Sholl and Sectoring methods resulted in the third method we propose. According to this approach, we apply both Sholl and Sectoring Analysis to a tree structure. The center of Sholl as well the center of Sectoring Analysis is the root of the ductal tree. The application of the Combined method to an image divides one plane quadrant to a number of $n \cdot m$ cells, where n is the total number of circles of Sholl Analysis and m is the total number of sectors produced by Sectoring Analysis. The vector that reflects the spatial distribution of branching points using the Combined Method is constructed as Fig. 3(d) shows. The ith element of the Combined vector counts the branching points inside the ith cell where the ith cell is the cell produced by the $[(i\ div\ n)]$ circle of Sholl Analysis and the $(i\ mod\ m)$ sector of Sectoring Analysis. More formally, the ith element of the Combined vector is:

$$Combined\ (i) = the\ number\ of\ branching\ points\ existing\ inside\ the\ cell\ generated\ by\ the\ [i\ div\ n]^{th}\ circle\ and\ the\ (i\ mod\ m)^{th}\ sector. \qquad (3)$$

Appling the Combined Method to the initial image Fig. 3(a) is presented in Fig. 3(d) and the Combined vector becomes
$Combined\ =\ [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,....]$
containing a total of $m \cdot n =\ 8 \cdot 28 =\ 224$ elements.

3.3 Galactogram Characterization

In order to characterize a clinical x-ray galactogram, we perform classification experiments considering two classes: galactograms with no radiological finding and galactograms with radiological findings. Using the above mentioned methods, the problem of galactogram classification is reduced to the problem of classification of the visualized tree structure. We employ the k-nearest neighbor classification scheme, which assigns the test tree to the class that appears most frequently among its neighbors. As the three methods presented above represent the spatial distribution of branching points with a numeric vector, we utilize the cosine similarity metric to retrieve the k closest neighboring tree structures. Given the vectors a and b of length t that represent the branching point's distribution of two trees correspondingly, the cosine similarity between them is computed by the following formula:

$$CosSim(a,b) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} = \frac{\sum_{i=1}^{t} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{t} a_i^2 \cdot \sum_{i=1}^{t} b_i^2}}. \qquad (4)$$

4 Results

We applied the proposed methodology to a clinical galactogram dataset and we performed similarity search experiments and classification experiments to evaluate the accuracy of our methods. A detailed description of the dataset and the experimental results are presented below.

4.1 Dataset

Our dataset consisted of 50 x-ray galactograms performed at Thomas Jefferson University Hospital and the Hospital of the University of Pennsylvania. From these images, 28 corresponded to women with no reported galactographic findings (class NF) and 22 to women with reported galactographic findings (class RF). The ductal trees visualized in the original galactograms were manually extracted and delineated by medical experts.

4.2 Similarity Experiments

We applied the proposed methods to the dataset and performed similarity search experiments. In similarity experiments, we considered each tree and its corresponding vector of spatial distribution of branching points as a query and retrieved the k-most similar trees based on the cosine similarity metric. The parameter k ranged from 1 to 5. To evaluate the methodology we reported the precision i.e., the percentage of neighboring images belonging to the same class as the query averaged over the entire dataset. As the two classes in our dataset are unbalanced, we randomly under-sampled the NF class to the size of the RF class and averaged the results over 100 sampling iterations. The following tables illustrate the precision obtained using the Sholl Analysis (Table1), the Sectoring Analysis (Table1), the Combined Method (Table2) and the tree characterization technique presented in [6] (Table2). We mention that in the experiments the parameter b of Sholl Analysis was equal to 10 pixels, the parameter m of Sectoring Analysis was equal to 16 sectors and the parameters b and m of Combined Method were 10 pixels and 16 sectors correspondingly.

Table 1. The obtained precision of similarity searches for Sholl and Sectoring Analysis.

	Precision of Similarity Searches					
k	**Sholl Analysis**			**Sectoring Analysis**		
	NF	RF	Total	NF	RF	Total
1	100%	100%	100%	100%	100%	100%
2	98.44%	74.06%	86.25%	99.06%	78.44%	88.75%
3	86.04%	65.62%	75.83%	87.29%	57.50%	72.40%
4	82.34%	65.31%	73.83%	83.13%	55.31%	69.22%
5	74.00%	60.75%	67.38%	74.00%	55.63%	64.81%

Table 2. The obtained precision of similarity searches for Combined and Prüfer/tf-idf methods.

	Precision of Similarity Searches					
k	**Combined Method**			**Prüfer/tf-idf method**		
	NF	RF	Total	NF	RF	Total
1	100%	100%	100%	100%	100%	100%
2	96.25%	86.56%	91.41%	92.94%	74.53%	83.73%
3	85.21%	68.13%	76.67%	77.67%	66.92%	72.29%
4	78.91%	65.00%	71.95%	67.72%	66.70%	67.21%
5	70.13%	61.88%	66.00%	71.16%	60.16%	65.66%

4.3 Classification Experiments

Here, we performed leave-one-out k-nearest neighbor classification experiments using the cosine similarity metric. Also in this case, the values of k ranged from 1 to 5 and the NF class was under-sampled to the size of the smaller RF class. To evaluate the classification, we compared the accuracy obtained using the Sholl Analysis (Table3), the Sectoring Analysis (Table3), the Combined Method (Table4) and the state-of-the-art Prüfer/tf-idf method presented in [6] (Table4). The values of parameter b and m were equal to the values in similarity search experiments.

Table 3. The obtained classification accuracy using Sholl and Sectoring technique.

	Classification Accuracy					
k	**Sholl Analysis**			**Sectoring Analysis**		
	NF	RF	Total	NF	RF	Total
1	95.00%	61.25%	78.13%	93.75%	46.88%	70.31%
2	96.88%	48.13%	72.50%	98.12%	56.87%	77.50%
3	69.58%	41.04%	55.31%	73.54%	39.37%	56.46%
4	73.39%	53.70%	63.54%	74.27%	40.36%	57.32%
5	66.63%	52.25%	59.44%	65.50%	42.13%	53.81%

Table 4. The obtained classification accuracy using Combined and Prüfer/tf-idf methods.

	Classification Accuracy					
k	**Combined Method**			**Prüfer/tf-idf method**		
	NF	RF	Total	NF	RF	Total
1	91.87%	70.63%	81.25%	73.61%	53.61%	63.61%
2	92.50%	73.12%	82.81%	85.88%	49.06%	67.47%
3	68.75%	49.38%	59.06%	58.73%	50.92%	54.82%
4	73.59%	52.86%	63.23%	55.56%	56.60%	56.08%
5	62.87%	51.00%	56.94%	63.56%	50.20%	56.88%

5 Conclusion

We presented a new methodology for capturing branching properties of tree-like structures in medical images. Based on Sholl Analysis, we proposed Sectoring Analysis and we combined the two methods to obtain a more detailed analysis of the spatial distribution of branching points. We applied our approach to a dataset of clinical galactograms attempting to correlate the topological features of branching points with the underlying pathology of breast ductal trees. Our experimental results suggest that it is possible to deduce valuable information about the radiological findings from the topology of the ductal trees. Within our future research plans is the fine-tuning of the values of parameters b and m, as well as the application of statistical tests in order to check the statistical significance of the results. In addition, the study of other topological descriptors and the combination of all of them is a promising perspective to achieve even higher classification accuracy.

References

1. Bullit, E., Muller, M.E., Jung, I., Lin, W., Aylward, S.: Analyzing attributes of vessel populations, Medical Image Analysis, 24, 39--49 (2005).
2. Bullit, E., Zeng, D., Gerig, G., Aylward, S., Joshi, S., Smith, J.K., Lin, W., Ewend, M.G.: Vessel Tortuosity and Brain Malignancy: A Blinded Study. Academic Radiology 12, 1232--1240 (2005)
3. Tschirren, J., McLennan, G., Palagyi, K., Hoffman, E.A., Sonka, M.: Matching and anatomical labeling of human airway tree. IEEE Transactions in Medical Imaging, 24, 701--716 (1996)
4. Dinkel, H.P., Trusen, A., Gassel, A.M., Rominger, M., Lourens, S., Muller, T., Tschammler, A.: Predictive value of galactographic patterns for benign and malignant neoplasms of the breast in patients with nipple discharge, The British Journal of Radiology, 73, 706—714 (2000)
5. Bakic, P.R., Albert, M., Maidment, A.D.: Classification of galactograms with ramification matrices: preliminary results, Academic Radiology, 10, 198--204 (2003)
6. Megalooikonomou, V., Barnathan, M., Kontos, D., Bakic, P.R., Maidment, A.D.: A Representation and Classification Scheme for Tree-like Structures in Medical Images: Analyzing the Branching Pattern of Ductal Trees in X-ray Galactograms, IEEE Transactions in Medical Imaging, 28 (2009)
7. Kontos, D., Megalooikonomou, V., Javadi, A., Bakic, P.R., Maidment, A.D.: Classification of galactograms using fractal properties of the breast ductal network. Proc. IEEE International Symposium on Biomedical Imaging, 1324--1327 (2006)
8. Skoura, A., Barnathan, M., Megalooikonomou, V.: Classification of ductal tree structures in galactograms. Proc. 6th IEEE International Symposium on Biomedical Imaging (2009)

9. Barnathan M., Zhang, J., Kontos, D., Bakic, P., Maidment, A.D., Megalooikonomou, V.: Analyzing Tree-Like Structures in Biomedical Images Based on Texture and Branching: An application to Breast Imaging, LNCS (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, vol. 5116, pp. 25-32. Springer, Heidelberg (2008)

10. Moffat, D., Going, J.: Three dimensional anatomy of complete duct systems in human breast: pathological and developmental implications. Journal of Clinical Pathology, 49, 48--52 (1996)

11. Sholl, D.: Dendritic organization in the neurons of the visual and motor cortices of the cat. J. Anat., 87, 387--406, (1953)