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

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

[1]Data Engineering Laboratory (DEnLab), Computer and Information Sciences Department, Temple University, 1805 N.Broad St., Philadelphia PA 19122, USA
[2]Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, Pennsylvania 19104, USA

ABSTRACT

We propose a multi-step approach for representing and classifying tree-like structures in medical images. Examples of such tree-like structures are encountered in the bronchial system, the vessel topology and the breast ductal network. We assume that the tree-like structures are already segmented. To avoid the tree isomorphism problem we obtain the *breadth-first canonical form* of a tree. Our approach is based on employing tree encoding techniques, such as the *depth-first string encoding* and the *Prüfer encoding*, to obtain a symbolic representation. Thus, the problem of classifying trees is reduced to string classification where node labels are the string terms. We employ the *tf-idf* text mining technique to assign a weight of significance to each string term (i.e., tree node label). We perform similarity searches and *k*-nearest neighbor classification of the trees using the *tf-idf* weight vectors and the cosine similarity metric. We applied our approach to the breast ductal network manually extracted from clinical x-ray galactograms. The goal was to characterize the ductal tree-like parenchymal structures in order to distinguish among different groups of women. Our best classification accuracy reached up to 90% for certain experimental settings (k=4), outperforming on the average by 10% that of a previous state-of-the-art method based on ramification matrices. These results illustrate the effectiveness of the proposed approach in analyzing tree-like patterns in breast images. Developing such automated tools for the analysis of tree-like structures in medical images can potentially provide insight to the relationship between the topology of branching and function or pathology.

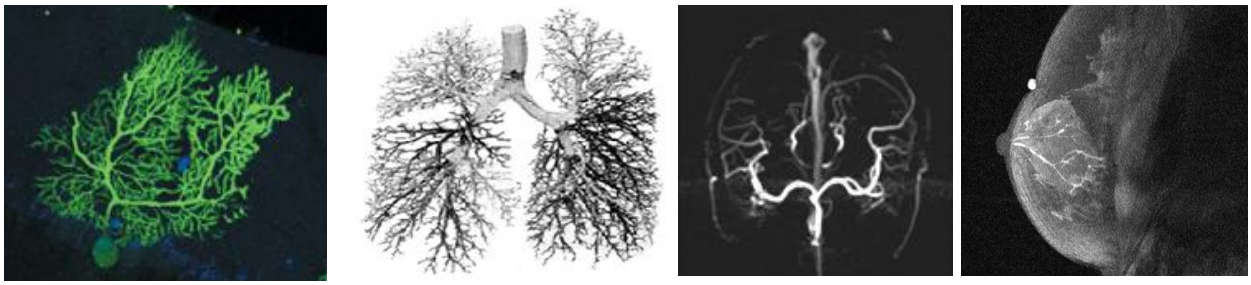
Keywords: Tree-like structures, Characterization, Branching pattern analysis, Classification, X-ray galactograms

1. INTRODUCTION

Several structures in the human physiology follow a tree shaped topological morphology. Examples of such tree-like anatomical constructs are dendritic extensions of neurons[1], intrathoracic airway trees[2], the blood vessel system[3] and the breast ductal network[4] (see Figure 1). Modern medical imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and x-ray mammography have made available large collections of two-dimensional (2D) and three-dimensional (3D) image datasets that visualize the structure and function of these tree-like anatomical constructs. A challenging issue when analyzing the morphological variability of these tree-like structures in 2D or 3D medical images is the automated extraction of descriptive features that correspond to topological patterns and discriminative characteristics among groups of images. These features characterize and represent the original trees in medical images, capturing properties such as the branching frequency, the tortuosity, and the spatial distribution of branching[3, 5]. During a pre-processing step, these tree-like structures are traced on medical images and extracted, preferably using automated procedures[6]. Computerized image processing techniques are then used to extract descriptors and analyze properties of the tree-like structures. In medical image analysis, these properties are usually associated with function, pathology, or the development stages of a disease and can be used to assist medical diagnosis. For example, regional changes in vessel tortuosity have been studied to identify early tumor development in the human brain[5].

* vasilis@temple.edu; phone 1 215 204 5774; fax 1 215 204 5082

Medical Imaging 2006: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6144, 61441H, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.651426



(a) (b) (c) (d)

Figure 1: Examples of tree-like structures in medical images: (a) a dendritic brain neuron, (b) airway tree of the lungs, (c) vessel system and (d) the breast ductal network.

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 pathological breast tissue [7, 8]. Lung structure and function can be investigated based on the 3D analysis of pulmonary airway trees that are automatically extracted from medical images with intelligent computerized techniques[2, 6].

In this paper, we propose a multi-step approach for characterizing and classifying tree-like structures in medical images. Our approach employs tree encoding schemes to obtain a symbolic representation of the tree-like structures. Thus, the problem of classifying trees is reduced to string classification where node labels comprise the string terms. We utilize text mining techniques to assign a significance weight to each string term (i.e., node label), identifying corresponding discriminative tree branching patterns among groups of images. As a case study we consider the breast ductal network that has been manually extracted from clinical x-ray galactograms. The goal is to develop effective descriptors of ductal tree-like parenchymal structures for facilitating similarity searches and classification, distinguishing among women with reported galactographic findings and normal cases. We compare different tree encoding schemes that produce different descriptors. Our results demonstrate that developing such automated procedures for characterizing and classifying tree-like structures in medical images can potentially provide insight to the relationship between the topology of branching and function or pathology.

2. BACKGROUND

In graph theory, a tree is defined as a *directed acyclic graph* (DAG) in which there is only one path between any two nodes[9]. Each node has one or no parents, while the node at the top of the tree that has no parents is identified as the root of the tree. The path between two nodes has a defined one-way direction from the parent to the child. A binary tree is defined as a DAG in which each node has at most two successors or child nodes. When the nodes of the tree are assigned labels, then the tree is denoted as *labeled rooted tree.* Methods proposed in the literature for representing trees follow two approaches for characterization:

1. Topological modeling, where the tree is described in terms of connections between nodes (branching points).
2. Metrical level modeling, where the lengths and spatial directions of the branches are specified.

By focusing on features of the breast ductal network in mammographic images, the breast duct anatomy has been analyzed to understand normal breast development[10] and distinguish between groups of women with present and absent radiological findings [11, 12]. Taking into consideration that breast cancer is one of the leading causes of cancer-related mortality world wide, and that it originates in ductal and lobular epithelium, analysis of the breast ductal anatomy is essential for understanding cancer development and spread, and for estimating breast cancer risk. In order to evaluate ductal morphology with respect to breast cancer symptoms, Bakic *et al.* [4] proposed a three-dimensional simulated model of the ductal network based on *Ramification Matrices* and a quantitative approach to classify galactograms based on ductal branching properties [12].

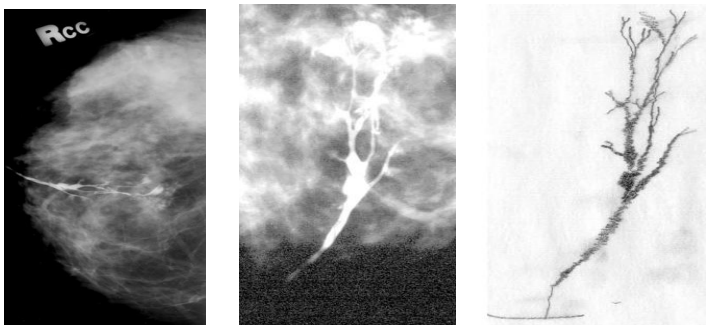
Ramification Matrices have been used in general for describing tree branching[12, 13]. The elements of a ramification matrix represent the probabilities of branching at various levels of a tree. More specifically, a *Ramification (R) matrix* represents a descriptor of branching structures at the topological level. The R-matrix elements are equal to the probabilities of different patterns of branching at various tree levels. The root, internal and terminal nodes, and branches are identified in a tree and the R-matrix elements are computed as follows: (a) all terminal branches have label 1, (b) a

"parent" branch whose "children" have labels i and j are labeled by $\max(i, j)$ if $i \neq j$ or by $(i+1)$ if $i=j$, and (c) the labeling procedure continues until the root branch is reached whose label s is called the *Strahler number* of the tree structure. The R matrix of a tree with Strahler number s is a lower triangular matrix, defined as:

$$R_{s-1,s} = \left[r_{k,j} = b_{k,j} / a_k \, , k \in (2, s), j \in (1, k) \right],$$

where a_k is equal to the number of branches with label k. For $j<k$, $b_{k,j}$ is the number of pairs of branches with labels k and j, while for $j=k$, $b_{k,j}$ is the number of pairs of branches both labeled k-1, descending from a node. Therefore, $r_{k,j}=b_{k,j}/a_k=p(b_{k,j}|a_k)$ is the probability that a branch with label k will bifurcate into branches with the appropriate labels.

An imaging procedure that can visualize the breast ductal network in mammographic images is galactography, during which x-ray mammography is performed after injecting a contrast agent into the lactiferous ducts [14-16]. Galactography can be useful for visualizing early symptoms of papilloma or ductal ectasia, which cause spontaneous nipple discharge, without usually showing recognizable change in screen-film x-ray mammograms. In the studies performed by *Bakic et al*[12] galactographic images were used to manually extract the breast ductal tree-like structures in order to perform branching pattern analysis.



(a) (b) (c)

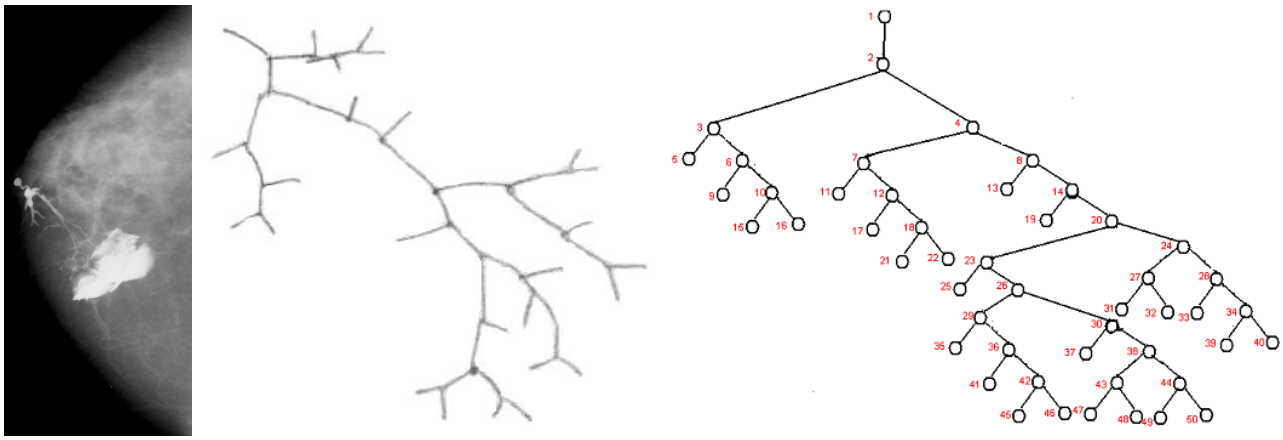
Figure 2: Segmentation of a ductal tree, showing (a) a galactogram with a contrast–enhanced ductal network, (b) part of the galactogram showing enlarged the ductal network, (c) the manually traced network of larger ducts from the contrast–enhanced portion of the galactogram.

3. METHODS

Our methodology is based on combining tree encoding schemes with text mining techniques in order to analyze tree-like structures visualized in medical images. The problems we consider in this type of analysis are:

1. *Characterization*: Encoding and representing trees in an appropriate form so that storage, indexing and retrieval are facilitated.
2. *Similarity searches*: Given a collection of tree structures and a query tree find the trees that are most similar to the query one.
3. *Classification*: Given classes of labeled tree structures build a model that correctly identifies the class of a new previously unseen tree.

Several preprocessing steps are necessary before the tree-like structures are available for analysis. First the boundary of these structures needs to be traced to distinguish these structures from the rest of the tissue. This process of segmentation can be performed manually, automatically or semi-automatically. Since here we concentrate on the analysis of the branching pattern we examine the medial axis of branching structures which can be derived using a thinning process. Because the main focus of the work presented in this paper is the representation and classification of tree-like structures we performed both of these preprocessing steps manually (Figure 2 illustrates an example of a hand-traced tree that has been manually extracted from a clinical x-ray galactogram). After the tree structure has been extracted the next step is to label the nodes (or branches) of the tree. This is done using consecutive increasing integers assigned in a breadth-first search manner. The reason for employing this labeling approach is that it creates more robust representation schemes by dealing better with cases where branches at the lower levels may not be visible due to image acquisition problems or the use or not of contrast agents. To avoid the tree isomorphism problem as a last step in this preprocessing phase we normalize the trees by applying a procedure for obtaining the *breadth-first canonical form* (BFCF) of a tree [17]. Figures 3a-c show the procedure of applying the BFCF and labeling a hand traced ductal tree.



(a) (b) (c)

[1 2 3 5 6 9 10 15 16 4 7 11 12 17 18 21 22 8 13 14 19 20 23 25 26 29 35 36 41 42 45 46 30 37 38 43 47 48 44 49 50 24 27 31 32 28 33 34 39 40]

(d)

[1 2 3 3 6 6 10 10 2 4 7 7 12 12 18 18 4 8 8 14 14 20 23 23 26 29 29 36 36 42 42 26 30 30 38 43 43 38 44 44 20 24 27 27 24 28 28 34 34]

(e)

Figure 3: (a) A real x-ray galactogram visualizing the breast ductal network for a case with no reported radiological findings, (b) the corresponding manually traced ductal tree, and (c) the labeled tree normalized to a canonical form. The obtained (d) *depth-first string* encoding and (e) the *Prüfer encoding* is shown.

Starting with a labeled tree we propose to use two different encodings: the *depth-first string encoding* and the *Prüfer* encoding. By using either one of the proposed encoding schemes the problem of classifying the trees is reduced to string classification where node labels comprise the string terms. These characterization strings capture properties of the branching patterns and the topological structure of the corresponding tree.

The *depth-first string encoding* is a rather simple and straightforward encoding scheme, which constructs a unique string representation for each tree by visiting each node following a pre-order depth-first traversal. During this process each node is represented in the string by its label. These encoding strings can be treated as *signatures* representing the original trees. Figure 3.d shows the *depth-first string encoding* obtained for the hand traced labeled tree in Figure 3.c. It has been proven in the literature that the *depth-first string encoding* provides a one-to-one correspondence between a rooted labeled tree and the obtained string representation [17].

A more sophisticated tree encoding scheme that reflects branching frequencies of the tree nodes is the *Prüfer* encoding. The *Prüfer* encoding scheme constructs a unique string representation for each tree-like structure. The algorithm visits each node of the tree following a pre-order traversal and depth-first search. During this process the encoding (characterization) string is constructed, using for each non-root node the label of its parent to represent it (see Figure 4 for a labeled tree and its *Prüfer* encoding).

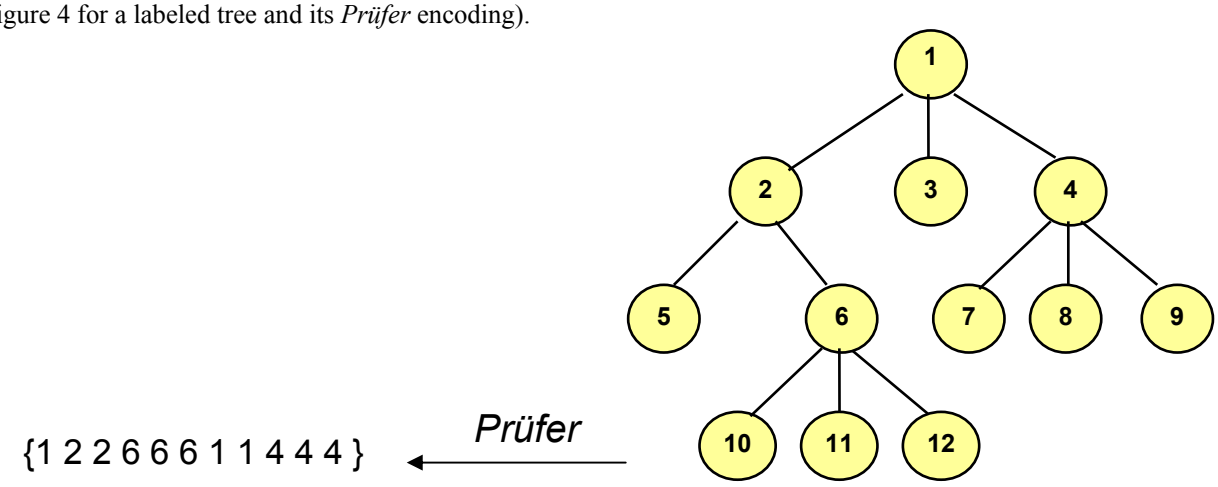


Figure 4: A simple tree represented with a string based on the *Prüfer* encoding scheme.

These characterization strings capture properties of the branching patterns and the topological structure of the corresponding tree. *Prüfer*, in his proof of Cayley's theorem regarding the number of labeled trees on *n* vertices, showed that there exists a 1-1 correspondence between *(n-2)*-length sequences of integers from the set $\{1,2,...,n\}$ and labeled trees on *n* vertices. Further, if an integer *k* occurs exactly *m* times in a sequence corresponding to a tree *T*, then the vertex in *T* with label *k* has degree *m+1*. The greater the maximum degree of a tree the more occurrences of a node's label in the code. Figure 3.e shows the *Prüfer* encoding of a ductal tree illustrated in Figures 3.a-c. Employing *Prüfer* encoding results in obtaining unique characterization strings for each tree. These strings capture important information with respect to the spatial arrangement of the structure as well as the branching patterns of the nodes and are used to represent the initial trees in further analysis.

In order to analyze patterns in the tree-like structures after representing them using strings we utilize text mining techniques, namely the *tf-idf* weighting, to assign a significance weight to each string term (i.e., node label). We perform similarity searches and classification by employing the cosine similarity distance metric on the string representations. The following paragraphs describe in detail these steps of our proposed approach.

Vector normalization and weighting: We employ the *tf-idf* text mining technique to assign a weight of significance to each string term (i.e., tree node label), indicating terms that form discriminative branching patterns. The string representations constructed by applying the *depth-first string encoding* or the *Prüfer* encoding can be viewed as document vectors. In this case, the features of the vector (string elements) are considered to be a collection of terms (such as the labels shown in Figures 3d-e and Figure 4). A technique that is applicable in this case to further normalize these vectors is the *tf-idf* weighting. According to this, each term in the document vector can be assigned a weight. The new representation becomes a vector with the corresponding weight at each terms feature position $d_j = (w_{1j}, w_{2j}, ..., w_{tj})$, where t = |vocabulary|=dimension. More specifically, the main idea of *tf-idf* weighting is that:

(i) more frequent terms in a document are more important, i.e. more indicative of the topic,

(ii) we may want to normalize term frequency (*tf*) across the entire corpus and

(iii) terms that appear in many different documents are less indicative of overall topic.

The weights derived by this approach are given by the following formula:

$$w_{ij} = tf_{ij}\, idf_i = tf_{ij} \log_2 (N/ df_i), \qquad \textbf{(3.1)}$$

where

f_{ij} is the frequency of term i in document j,
$tf_{ij} = f_{ij} / max\{f_{ij}\}$,
df_i = document frequency of term i = number of documents containing term i,
idf_i = inverse document frequency of term $i = \log_2 (N/ df_i)$ and
N is the total number of documents.

Similarity searches and Classification: As shown in the previous step of our proposed approach, each term i, in a document, j, is given a real-valued weight, w_{ij}, thus each document can be expressed as a t-dimensional vectors: $d_j = (w_{1j}, w_{2j}, ..., w_{tj})$, where t = |vocabulary|=Dimension. For example, let us consider three documents, $D_1 = 2T_1 + 3T_2 + 5T_3$, $D_2 = 3T_1 + 7T_2 + 1T_3$, $Q = 0T_1 + 0T_2 + 2T_3$. There are many ways to tell whether two of these documents are similar; here we use the cosine value of the two vectors, which is the called cosine similarity measure and is computed as follows:

$$\text{CosSim}(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{\left|\vec{d}_j\right| \cdot \left|\vec{q}\right|} = \frac{\sum_{i=1}^{t} (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^{\,2} \cdot \sum_{i=1}^{t} w_{iq}^{\,2}}} \qquad \textbf{(3.2)}$$

Figure 5 illustrates the main notion of the cosine similarity metric. Note that when encoding the objects to strings the encoding should contain, in some sense, information about the properties we want to study.

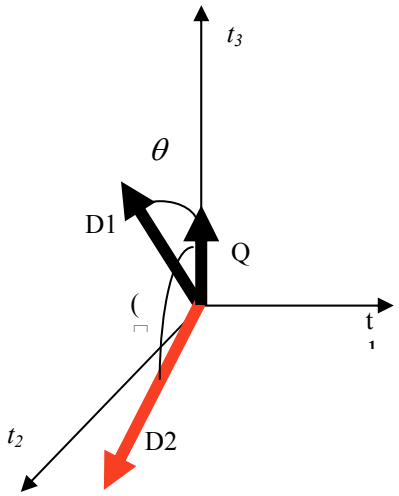


Figure 5: Cosine similarity.

4. RESULTS

We considered 22 x-ray galactograms from which the ductal trees were manually delineated and extracted. An example of a galactogram along with the corresponding hand-traced tree is shown in Figures 3a-3c. These views were acquired from a total of 14 women. From these images, 10 corresponded to women with no reported galactographic findings (NF) and 12 to women 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)[12]. The nodes of each tree were labeled with a unique positive integer number that was assigned following a *breadth-first search* traversal. The labeling begins from the root of each tree assigning the integer '1' and continues in an increasing manner until all nodes are labeled. To avoid the problem of tree isomorphism, we normalized these hand-traced trees by applying a procedure for obtaining a canonical form[17]. Figure 3.c illustrates an example of such a resulting labeled tree normalized to a canonical form.

We applied the *depth first string encoding* and the *Prüfer encoding* to obtain the string representations corresponding to the original ductal trees. These encoding strings uniquely characterize each initial tree. Figures 3.d and 3.e show examples of such characterization strings. We further employed *tf-idf* weighting to assign a weight of significance to each string term (node label) of the characterization strings. We performed *tf-idf* weighting separately for the *depth first string encoding* and the *Prüfer* encoding datasets. In each of these two cases, we considered both classes of ductal trees (NF and RF) as one group (i.e., forest) of trees and applied the *tf-idf* weighting to this combined dataset of encoding strings. When applying the *tf-idf* weighting the unequal lengths of the encoding strings were handled by padding the end of the characterization strings with a very small value of $1.00e^{-013}$ to avoid numerical errors when calculating the cosine similarity distance. By performing the *tf-idf* weighting we obtained two datasets (one from the *depth first string encoding* and one from the *Prüfer encoding*) of *tf-idf* weights indicating the significance of each encoding string term (i.e., node label) in each characterization vector. Using the obtained *tf-idf* weight vectors we performed similarity searches and classification experiments based on the cosine similarity distance.

Similarity Searches: For similarity searches, we calculated the pairwise cosine distance matrix for all the *tf-idf* vectors. We considered each tree (i.e., *tf-idf* vector) as the query subject and retrieved the k most similar trees based on the cosine distance matrix. Considering the small size of our datasets, the k parameter ranged from 1 to 5. We report the percentage of relevant trees among the retrieved trees (i.e., precision) averaged over all the similarity queries performed. As relevant trees we consider the trees belonging to the same class as the query tree (NF vs. RF). Table 1 illustrates the similarity searches results obtained when using the *depth first string encoding* and the *Prüfer encoding*. As shown from these results, the *Prüfer encoding* performs better than the *depth first string encoding* by an average of approximately 11% over all the different values of k. These results illustrate the potential of our proposed framework to be employed for efficient indexing in medical databases where images visualizing tree-like structures need to be stored and managed.

Depth-First String Encoding			
k	Precision		
	NF	RF	Total
1	90.00 %	100.00 %	**95.45 %**
2	65.00 %	62.50 %	63.64 %
3	60.00 %	61.11 %	60.61 %
4	50.00 %	58.33 %	54.55 %
5	54.00 %	60.00 %	57.27 %

(a)

Prüfer Encoding			
k	Precision		
	NF	RF	Total
1	100%	100 %	**100%**
2	90.00 %	70.83 %	**79.55 %**
3	80.00 %	66.67 %	**72.73 %**
4	72.50 %	64.58 %	68.18 %
5	68.00 %	65.00 %	66.36 %

(b)

Table 1: The obtained precision for similarity searches experiments based on the cosine similarity distance metric when using (a) the *depth-first string encoding* and (b) the *Prüfer encoding.*

Depth-First String Encoding			
k	Classification Accuracy		
	NF	RF	Total
1	40.00 %	25.00 %	31.82 %
2	50.00 %	66.67 %	59.09 %
3	40.00 %	41.67 %	40.91 %
4	60.00 %	83.33 %	**72.73 %**
5	60.00 %	75.00 %	68.18 %

(a)

Prüfer Encoding			
k	Classification Accuracy		
	NF	RF	Total
1	80.00 %	41.67 %	59.09 %
2	80.00 %	66.67 %	**72.73 %**
3	80.00 %	50.00 %	63.64 %
4	100.00 %	83.33 %	**90.91 %**
5	70.00 %	58.33 %	63.64 %

(b)

Table 2: The obtained accuracy for the classification experiments based on the cosine similarity distance metric when using (a) the *depth-first string encoding* and (b) the *Prüfer encoding.*

Classification: For classification, we performed leave-one-out *k*-nearest neighbor experiments. For each test tree we retrieved the *k* closest neighbor trees (i.e., *tf-idf* vectors), based on the cosine similarity distance. Considering the small size of our dataset, the *k* parameter ranged from 1 to 5. The label of the majority of the *k* neighbors was assigned as the label of the test tree. Table 2 illustrates the classification accuracy obtained when using the *depth first string encoding* and the *Prüfer encoding*. As, shown in Table 2, *Prüfer encoding* outperformed on the average *depth first string encoding* by approximately 15 %.

These experimental results illustrate the potential of the proposed tree characterization and classification framework to be employed for the analysis of topological patterns in medical images. Our best results reached up to 91% classification accuracy, on the average over the two classes, when using the *Prüfer encoding* to represent the original trees. These results outperform by approximately 10% previous experimental results reported in the literature, in which R-matrix elements computed from the breast ductal trees were used to distinguish among the two classes (NF vs. RF) in the same dataset of galactographic images[12]. Moreover, our approach has the advantage of constructing characterization strings that uniquely represent the tree structures and can be utilized as signatures for the corresponding original trees.

5. CONCLUSION

We present a novel methodology for characterizing and classifying tree-like structures in medical images. Our approach combines symbolic graph representation with text mining techniques. We employ the idea of applying a string encoding algorithm, such as the *depth-first* string encoding and the *Prüfer* encoding to construct a unique characterization string for each tree-like structure. We further perform *tf-idf* weighting, to assign a significance weight to each string term (i.e., node label). Our methodology was applied to breast ductal trees manually extracted from real x-ray galactograms. The images were divided into two groups; those with no reported galactographic findings (NF) and those with reported findings (RF). We performed similarity searches and leave-one-out *k*-nearest neighbor classification based on the cosine similarity distance metric. The obtained results demonstrated the effectiveness of the proposed methodology. Our best results outperformed by almost 10% on average previous results obtained by a state-of-the-art method applied to the same dataset. Considering the small size of our dataset, experiments on larger collections of clinical data need to be performed in order to further evaluate the applicability of the proposed framework for the analysis of topological patterns of tree-like structures in medical images. Our approach can potentially assist in investigating associations between the topological structure and branching of the tree and their function or pathology.

ACKNOWLEDGEMENTS

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

REFERENCES

1. R. Yuste and D. W. Tank, *Dendritic integration in mammalian neurons, a century after Cajal*, Neuron, 16 (1996), pp. 701-716.
2. J. Tschirren, G. McLennan, K. Palagyi, E. A. Hoffman and M. Sonka, *Matching and anatomical labeling of human airway tree*, IEEE Transactions in Medical Imaging, 24 (2005), pp. 1540-7.
3. E. Bullitt, K. E. Muller, I. Jung, W. Lin and S. Aylward, *Analyzing attributes of vessel populations*, Medical Image Analysis, 9 (2005), pp. 39-49.
4. 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, 30 (2003), pp. 1914-1925.
5. E. Bullitt, D. Zeng, G. Gerig, S. Aylward, S. Joshi, J. K. Smith, W. Lin and M. G. Ewend, *Vessel Tortuosity and Brain Tumor Malignancy: A Blinded Study*, Academic Radiology, 12 (2005), pp. 1232-1240.
6. W. Park, E. A. Hoffman and M. Sonka, *Segmentation of Intrathoracic Airway Trees:A Fuzzy Logic Approach*, IEEE Transactions in Medical Imaging, 17 (1998), pp. 489-497.
7. B. Pereira and K. Mokbel, *Mammary ductoscopy: past, present, and future*, International Journal of Clinical Oncology, 10 (2005), pp. 112-116.
8. D. Moffat and J. Going, *Three dimensional anatomy of complete duct systems in human breast: pathological and developmental implications*, Journal of Clinical Pathology, 49 (1996), pp. 48-52.
9. T. H. Cormen, C. E. Leiserson and R.L.Rivest, *Introduction to Algorithms*, MIT Press, Cambridge, 2001.
10. J. J. Going and D. F. Moffat, *Escaping from Flatland: clinical and biological aspects of human mammary duct anatomy in three dimensions*, Journal of Pathology, 203 (2004), pp. 538-544.
11. D. Kontos, V. Megalooikonomou, A. Javadi, P. R. Bakic and A. D. Maidment, *Classification of Galactograms Using Fractal Properties of the Breast Ductal Network*, *3rd IEEE International Symposium on Biomedical Imaging (ISBI 2006)*, Arlington, VA, 2006 (to appear).
12. P. R. Bakic, M. Albert and A. D. Maidment, *Classification of galactograms with ramification matrices: preliminary results*, Academic Radiology, 10 (2003), pp. 198-204.
13. X. G. Viennot, G. Eyrolles, N. Janey and D. Arques, *Combinatorial analysis of ramified patterns and computer imagery of trees*, Computer Graphics, 23 (1989), pp. 31-40.
14. J. Peters, A. Thalhammer, V. Jacobi and T. J. Vogl, *Galactography: an important and highly effective procedure*, European Radiology, 13 (2003), pp. 1744-1747.

15. 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, 73 (2000), pp. 706-714.
16. M. F. Hou, T. J. Huang and G. C. Liu, *The diagnostic value of galactography in patients with nipple discharge*, Clinical Imaging, 25 (2001), pp. 75-81.
17. Y. Chi, Y. Yang and R. Muntz, *Canonical forms for labeled trees and their applications in frequent subtree mining*, Knowledge and Information Systems, 8 (2005), pp. 203-234.