2010 23rd SIBGRAPI - Conference on Graphics, Patterns and Images

Segmentation and classification of histological images - application of graph analysis and machine learning methods

Francisco de Assis Zampirolli, Beatriz Stransky, Ana Carolina Lorena, Fábio Luis de Melo Paulon
Universidade Federal do ABC
Santo André–SP, Brazil
Email: {fzampirolli, beatriz.stransky, ana.lorena, fabio.paulon}@ufabc.edu.br

***Abstract*—The characterization and quantitative description of histological images is not a simple problem. To reach a final diagnosis, usually the specialist relies on the analysis of characteristics easily observed, such as cells size, shape, staining and texture, but also depends on the hidden information of tissue localization, physiological and pathological mechanisms, clinical aspects, or other etiological agents. In this paper, Mathematical Morphology (MM) and Machine Learning (ML) methods were applied to characterize and classify histological images. MM techniques were employed for image analysis. The measurements obtained from image and graph analysis were fed into Machine Learning algorithms, which were designed and developed to automatically learn to recognize complex patterns and make intelligent decisions based on data. Specifically, a linear Support Vector Machine (SVM) was used to evaluate the discriminatory power of the used measures. The results show that the methodology was successful in characterizing and classifying the differences between the architectural organization of epithelial and adipose tissues. We believe that this approach can be also applied to classify and help the diagnosis of many tissue abnormalities, such as cancers.**

***Keywords*-image analysis; mathematical morphology; graph analysis; machine learning; tissue.**

I. Introduction

In Biology, the association between tissue architecture and function was established a long time ago. The development, function and regeneration of tissues, both normal and abnormal, are determined by the spatial localization and distribution of various cell types and phenotypes within the tissue. The central role of the cellular arrangement in tissues integrity and homeostasis maintenance has been extensively studied by experimental and theoretical models [1], [2], [3], [4]. For example, the loss of tissue architecture is a prerequisite for, and one of the defining characteristics of most cancers. Conversely, normal organ architecture can act as a powerful tumour suppressor, preventing malignant phenotypes even in cells stricken with gross genomic abnormalities [5]. The analysis of histological images is an important factor in clinical procedures, where a pathologist tries to correlate the pattern of tissue disturbance with the pathology stage, like, for example, in neoplasm classification. To reach a definitive diagnosis, the pathologist analyzes slides of microscopic sections and, based on his/her own experience, considers different features observed in the image. Acquiring proper experience usually results from a long-term training process, which is not so easy to formalize [6]. To make things more complex, the subjective method takes into account not only characteristics directly derived from the image, such as cells size, shape and number, but also hidden features such as clinical information, pathological mechanisms and etiological agents [7].

The Mathematical Theory of Graphs has been widely employed to analyze or solve diverse biological questions [8], [9], [10], [11], [12], [13]. Image analysis, in particular, can profit from strategies where the neighborhood relation between objects in the image, such as cells, can be represented by neighbors graphs. Furthermore, Mathematical Morphology (MM) is an elegant form to solve image-processing problems using of a consistent theoretical base, that is the theory of sets [14], [15]. In MM the transformations between images, which are called morphological operators, can be defined by the structuring functions. In biology and medicine, the combination of advanced technologies of image processing with quantitative image analysis has been used to describe and characterize tissues in various conditions [16], [17], [18], [19], [20]. Specially, the application of this approach in cancer diagnosis and prognosis has been considerably improved, as indicated by many reports [7], [21], [22], [23], [24].

Thus, the identification of reliable markers that could be used in order to reduce the subjectivity of interpretation of microscopical specimens is one of the main interests of histophatologists. The application of Machine Learning methods, which are algorithms designed and developed to automatically learn to recognize complex patterns and make intelligent decisions based on data, can help to achieve this goal. Specifically, Support Vector Machines (SVMs) are a set of related supervised learning methods used for classification and regression, which are receiving increasing attention for their good predictive performance in several areas [25], [26]. In such classification tasks, in particular, the SVMs seek for an hyperplane in a high dimensional space which maximizes the margin of separation between data from different classes. According to the Statistical Learning Theory, the use of a large-margin hyperplane implies in a better generalization

1530-1834/10 $26.00 © 2010 IEEE
DOI 10.1109/SIBGRAPI.2010.51

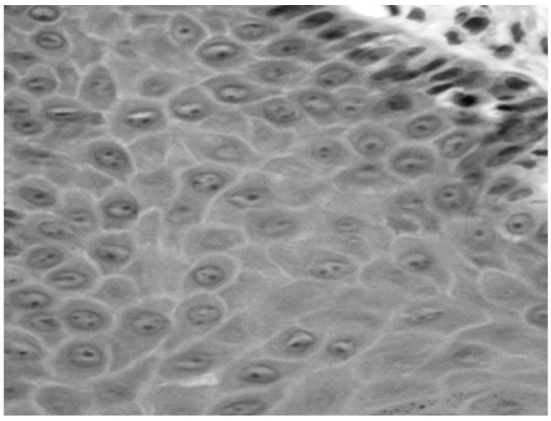


Figure 1. Microscopic image of epithelial tissue.

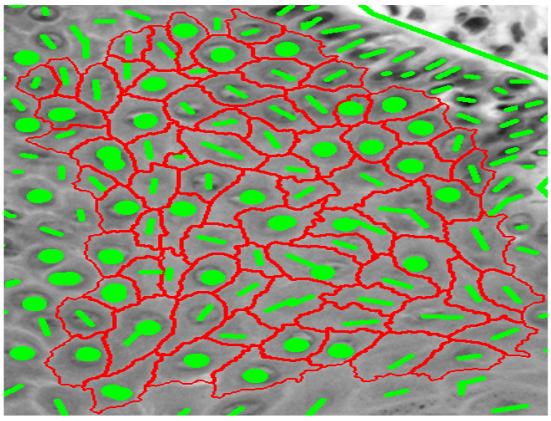


Figure 2. Epithelial image with markers and cell boundaries.

ability. In the present study, we developed a methodology to analyze topological arrangements on tissue images, based on image processing and machine learning techniques. As a benchmark, we successfully used this methodology to characterize and to classify the differences between the architectural organization of the distinct tissues - the epithelial and adipose, as observed in 2-D histological images. We believe that this approach can be also applied to classify and help the diagnosis of many tissue abnormalities, such as cancers.

II. Methods

A. Data Set and Image Acquirement

The image data set was obtained from the archives of Biology Laboratory of Digestive Epithelium (University of São Paulo). Our data set contains 5 images of tick epithelial tissue and 5 images of mamary gland adipose tissue from mice. The material consisted of 5-m-tick histological sections stained with hematoxylin and eosin. The digitalized images genetared a 300×300 dpi resolution files, as shown in Fig. 1.

B. Segmentation

The segmentation process using Mathematical Morphology [27], is a multi-step process, where the first one is to make an automatic segmentation to find markers in each cell of the image. After that, a specialist uses a MATLAB interface to create new markers and remove others incorrectly marked. Ideally this method should be an automatic segmentation for multiple images application, minimizing the specialist intervention. These experiments were performed in the SDS Morphology Tooblox for MATLAB [28].

The segmentation of the epithelial tissue begins with the reading of the image, converting to gray scale, see Fig. 1, and with a threshad of the gradient, followed by closing the holes and an operator opening with a structuring function [29] defined by a Euclidean circle with a radius of 5 pixels, as shown in the Matlab code below[1]:

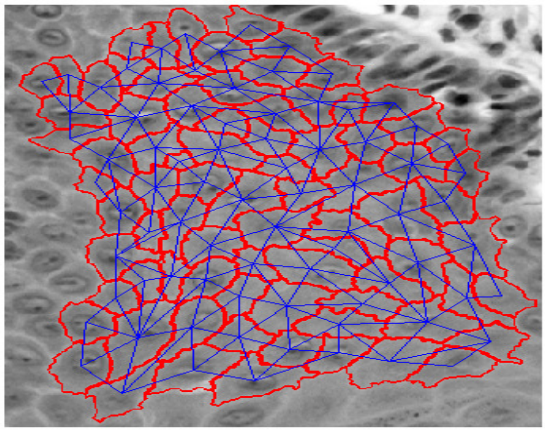
```
% reads the RGB image
img3=imread('epithelial01.tif');
img=rgb2gray(img3);
f1 = mmgradm(img);
f2 = mmthreshad(f1,5);
f3 = mmclohole(f2);
b1 = mmsedisk(5,'2D','Euclidean');
markers = mmopen(f3,b1);
```

The segmentation for adipose tissue begins with the reading of the image and converting to gray scale. The image is inverted (mmneg) for better visualization of this paper. Then followed by a dilate with a structuring function defined by a Euclidean circle with a radius of 9 pixels, in order to highlight the lighter regions. The next filter is the Area Opening, aiming to eliminate the regions (flats zone) with areas less than 370 pixels. Then was performed the threshad and regional minimum that correspond to the markers for each cell, as shown in the Matlab code below:

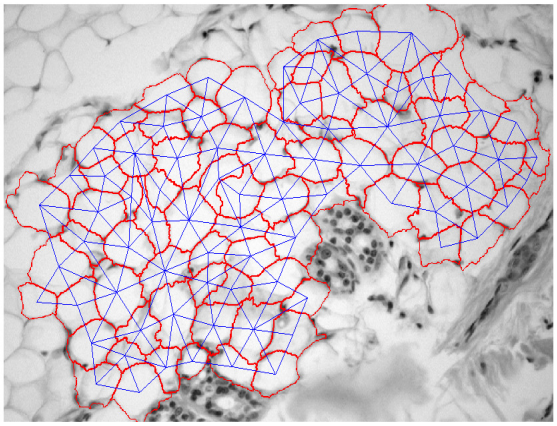
```
img3 = imread('adipose01.tif');
img = rgb2gray(img3); % gray scale
img = mmneg(img);
b1 = mmsedisk(9,'2D','Euclidean');
f1 = mmdil(img,b1);
f2 = mmareaopen(f1,370);
f3 = mmthreshad(f2,70);
f4 = mmregmax(f3);
markers = mmintersec(f3,f4);
```

The correct definition of markers is essential for the segmentation of the cells of epithelial and adipose tissue. Thus, after the automatic segmentation, a specialist uses an interface to create new markers and remove others. After correction of the markers, the watershed is calculated [30], creating the outline of each cell, see an example in Fig. 2.

[1]The community can validate these results and applied in other problems in image processing (where the neighborly relationship between objects is significant) using functions in Matlab, available at: http://professor.ufabc.edu.br/~fzampirolli/cells.



(a)



(b)

Figure 3. Epithelial (a) and adipose (b) tissue with the graph and cell boundaries.

C. Construction of Neighborhood Graphs

Using a strategy developed by Zampirolli [31] a graph using morphological operations in images was created, as illustrated in Fig. 3. The vertices of the graph are defined by the centroids of each cell, and the neighborhoods of each vertice (edges) are defined by the neighborhoods between the cells.

D. Calculation of Measures

Measures calculated using histological images from epithelial and adipose tissues (beyond centroid and the neighbor of each cell) are presented briefly in this section. These measures can be classified as global, individual and also measures in graphs [32], [33]. Examples of the global measures are: sum of pixels in the entire image, number of cells (vertices or objects) and perimeter of cells in the entire image. Examples of measure in graphs are: neighborhoods number and mean distance between cells. Using the *regionprops* command of the MATLAB, you can also calculate the following topological measures (individual measures) for each cell (details can be obtained in the MATLAB documentation):

- $Perimeter$ = distance around the boundary of the cells;
- $Area$ = number of pixels in the cells;
- $MajorAxisLength$ = major axis of the ellipse containing the cell;
- $MinorAxisLength$ = minor axis of the ellipse containing the cell;
- $Orientation$ = angle (-90 to 90 degrees) between the x-axis and the major axis of the ellipse containing the cell;
- $ConvexArea$ = area of the convex hull of the cells;
- $Eccentricity$ = ratio of the distance between the foci of the ellipse and its major axis length;
- $EquivDiameter$ = diameter of a circle with the same area as the cell;
- $Extent$ = ratio of pixels in the cell to pixels in the total bounding box;
- $Solidity = Area/ConvexArea$.

Other measures derived from [33] for diagnosis of solid breast tumours, can also be calculated for cells:

- $FormFactor = \frac{4*\pi*Area}{Perimeter^2}$;
- $Roundness = \frac{4*Area}{\pi*MajorAxisLength^2}$;
- $AspectRatio = \frac{MajorAxisLength}{MinorAxisLength}$;
- $Convexity = \frac{ConvexPerimeter}{Perimeter}$;
- $Solidity2 = \frac{ConvexArea-Area}{\sum_{i=1}^{N}(ConvexArea_i-Area_i)/N}$,

where $ConvexPerimeter$ is the perimeter of the convex hull of a cell, and N is the number of cells in a tissue.

E. Classification using Support Vector Machines

The SVM is a Machine Learning (ML) technique based on concepts of the Statistical Learning theory [34]. In classification problems, as the one considered in this paper, usually a supervised training data set must be provided to the technique. This data set must be composed of pairs of examples in the form $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i$ is a data item and y_i corresponds to its known label (classification). For instance, in the distinction of epithelial and adipose histological images, each $\mathbf{x}_i$ corresponds to one cell - in fact, the vector $\mathbf{x}_i$ will have as coordinates the features extracted from the cell image - and y_i represents whether $\mathbf{x}_i$ is an epithelial or an adipose cell.

SVMs are receiving an increasing attention in the last years, mainly for their appealing theory, which controls the capacity of the classification model for a better generalization ability. Given a training data set where $\mathbf{x}_i \in \Re^m$ and $y_i \in \{-1, +1\}$ - with m real-valued characteristics

describing data and two classes, named -1 and $+1$ for mathematical purposes (for instance, -1 can be class epithelial and $+1$ class adipose), SVMs seek for an hyperplane $(\mathbf{w}\cdot\mathbf{x}+b=0)$ able to separate data with a maximal margin. In order to perform this task, they solve the following optimization problem [25]:

$$\textbf{Minimize: } \|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$
$$\textbf{Restricted to: } \begin{cases} \xi_i \geq 0 \\ y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) \geq 1 - \xi_i \end{cases}$$

where C is a constant that imposes a tradeoff between training error and generalization and the ξ_i are slack variables. The restrictions are imposed in order to avoid training examples between the margins. The slack variables relax these restrictions by allowing some data items to lie within the margins and also the presence of some training errors. Their sum is minimized for controlling the training error.

The decision border obtained for the classification of new data is given by

$$f\left(\mathbf{x}\right) = \sum_{\mathbf{x_i}\in \mathrm{SV}} y_i\alpha_i\mathbf{x_i}\cdot\mathbf{x} + b, \qquad (1)$$

where the constants α_i are named Lagrange multipliers and are determined in the optimization process. SV corresponds to the set of support vectors (SVs), data items for which the associated Lagrange multipliers are larger than zero. These data are those closest to the optimal hyperplane. For all other data, the associated Lagrange multiplier is null, so they do not contribute to the determination of the final hypothesis. In order to obtain the final classification using Equation 1, a signal function is applied to $f(\mathbf{x})$, such that, if $f(\mathbf{x}) > 0$, the predicted class is $+1$ and if $f(\mathbf{x}) < 0$, the class is -1.

It is also possible to build non-linear decision functions with SVMs, by the use of Kernel functions. Nevertheless, based on the characteristics of the data set employed in the experiments, linear SVMs were used in this paper.

III. Results and Discussion

The aim of this study was to investigate the potential of characteristic measures in characterizing different tissues. In our first experimental approach, we used these measures to characterize two types of tissues: epithelial and adipose. It must be noticed that the distinction of such images can be considered quite easy for a histologist, since epithelial cells are more elongated and adipose cells tend to be more round. Since the distinction of epithelial and adipose cells can be considered a simple problem, we expect the extracted measures will be able to fully characterize the given cells, validating their use for this type of problem. To evaluate the discriminatory power of the measures considered in this paper, a data set of characteristic measures, also referred as features, were fed into a linear Support Vector Machine (SVM). The objective was to evaluate whether these characteristics would provide a fair description of these two types of tissues, by observing the predictive performance of a linear SVM in their recognition when using the given features as input.

For such, first a data set composed of epithelial and adipose cell images was built. It contains a total of 810 cells, equally divided in 405 epithelial cells and 405 adipose cells. The balancing was ensured so that an unbalancing would not harm the classifier induction with SVMs. In unbalanced problems, the classification technique tends to favor the majority class, in detriment of the minority class. For each cell, 16 real-valued attributes or features were extracted: solidity2, perimeter, area, mean distance to neighbors, major axis length, minor axis length, orientation, convex area, eccentricity, diameter, extent, solidity, form factor, roundness, aspect ratio and convexity. Tables I and II presents the minimum, maximum, average and standard deviation of these attribute values considering the whole data set. The percentage shown represents the standard deviation from the average. We can see from this table that there are differences in the numerical ranges of the attributes. Therefore, we had to normalize the data in order to feed them to SVMs. The use of attributes with such range discrepancies could lead the classifier to misleading conclusions regarding the relative importance of the attributes (it would tend to favor attributes in higher numerical ranges). It is also possible to note the presence of attributes with low variations in their overall values, as extent, solidity and convexity, which all showed quite low standard deviation values.

For a better visualization on how the feature values vary for the different cell types, we built the histograms presented in Fig. 4, with the aid of the Weka Data Mining and Machine Learning tool [35]. In this figure, values obtained for epithelial cells are plotted in red, while those for adipose cells are plotted in blue. For all features we can see an overlap in the values. Nevertheless, this overlap is less pronounced for some features, as perimeter, area, eccentricity, diameter, solidity, roundness and mainly for mean distance to neighbors. For this last feature in particular, the only obtained using graph measures, there is already a good separation in the feature values for the two considered cell types.

Although there are already some indicatives of which should be the best features describing the two classes of cells, we decided to use a linear SVM for weighting the importance of the given set of attributes in recognizing the epithelial and adipose cells. By observing: (1) the predictive performance of the SVMs in the recognition of the cells; (2) how SVMs weight the attribute values in the hyperplane equation (for instance, the vector $\mathbf{w}$ in $\mathbf{w}\cdot\mathbf{x} + b$); we can provide further conclusions regarding the ability and importance, respectively, of the morphological measures in characterizing these data. It should be noticed that estimating

Table I
MEASURES CONSIDERING THE DATA OF THE EPITHELIAL TISSUE.

Attribute	Minimum	Maximum	Mean	Standard Deviation	
Perimeter	73,00	317,00	168,61	40,92	(24%)
Area	283,00	3560,00	1334,98	502,39	(37%)
MeanDistNeighbors	25	65	40,914	6,364	(15%)
MajorAxisLenght	22,00	112,00	53,94	13,06	(24%)
MinorAxisLenght	13,00	56,00	33,43	7,41	(22%)
Orientation	-90,00	89,00	-6,29	57,41	(-912%)
ConvexArea	316,00	4187,00	1557,48	618,17	(39%)
Eccentricity	0,22	0,96	0,74	0,15	(19%)
EquivDiameter	19,00	67,00	40,49	7,80	(19%)
Extent	0,36	0,80	0,61	0,08	(12%)
Solidity	0,62	0,96	0,87	0,06	(6%)
FormFactor	0,32	0,86	0,59	0,10	(17%)
Roundness	0,23	0,93	0,59	0,15	(24%)
AspectRation	1,03	3,53	1,66	0,42	(25%)
Convexity	0,66	0,89	0,79	0,04	(5%)
Solidity2	0,12	5,18	1,00	0,69	(69%)

Table II
MEASURES CONSIDERING THE DATA OF THE ADIPOSE TISSUE.

Attribute	Minimum	Maximum	Mean	Standard Deviation	
Perimeter	154,00	643,00	354,97	75,52	(21%)
Area	1018,00	17102,00	5932,77	2333,47	(39%)
MeanDistNeighbors	56	110	83,88	10,09	(12%)
MajorAxisLenght	53,00	185,00	106,14	20,99	(19%)
MinorAxisLenght	28,00	133,00	73,86	17,55	(23%)
Orientation	-89,00	89,00	4,71	43,80	(929%)
ConvexArea	1217,00	18852,00	6780,09	2654,86	(39%)
Eccentricity	0,20	0,94	0,68	0,15	(21%)
EquivDiameter	36,00	148,00	85,24	16,99	(19%)
Extent	0,33	0,77	0,61	0,80	(131%)
Solidity	0,66	0,97	0,88	0,06	(6%)
FormFactor	0,32	0,81	0,58	0,10	(17%)
Roundness	0,27	0,97	0,66	0,14	(21%)
AspectRation	1,02	2,97	1,49	0,34	(23%)
Convexity	0,64	0,88	0,78	0,04	(5%)
Solidity2	0,15	4,19	1,00	0,68	(68%)

a function weighting the attribute values as performed can be considered a suitable procedure for understanding how they jointly contribute to the recognition problem.

We used the Weka Machine Learning tool for the induction of the SVM classifiers [35]. Default parameter values were maintained in the experiments ($C = 1$ and linear Kernel) and a normalization of the features was performed with the aid of the tool. The stratified 10-fold cross-validation procedure was employed for estimating the performance of the classifier. Accordingly, the data set was divided into 10 folds, maintaining the proportion of examples per class (that is, 50% of epithelial cells and 50% of adipose cells). Ten training and test rounds are then performed. In each one of them, nine folds are used for training (inducing) the classifier and the remaining fold is used for testing its performance for new data. The performance of the classifier is then given by the mean and standard deviation of the performances registered in the ten testing rounds. Three performance measures were calculated: (1) accuracy rate, which gives the total percentage of correct predictions for test data; (2) specificity, which measures the proportion of correct predictions for epithelial cells, that is, the proportion epithelial cells correctly classified; (3) sensitivity, which corresponds to the proportion of correct predictions for adipose cells. The results obtained are shown in Table III. The values shown correspond to the averages and standard deviations calculated for the cross-validation test partitions.

All performance measures calculated showed high values, specially for epithelial class, for which SVM had 100% of correct predictions. By observing the weight vector of the SVM hyperplane, we could identify the prominence of five features in determining the decision frontier: mean distance to neighbors, solidity2, solidity, minor axis length and major axis length. On the other hand, the features convexity, roundness, aspect ration, form factor and specially eccentricity, had little influence on the final decision. In fact,

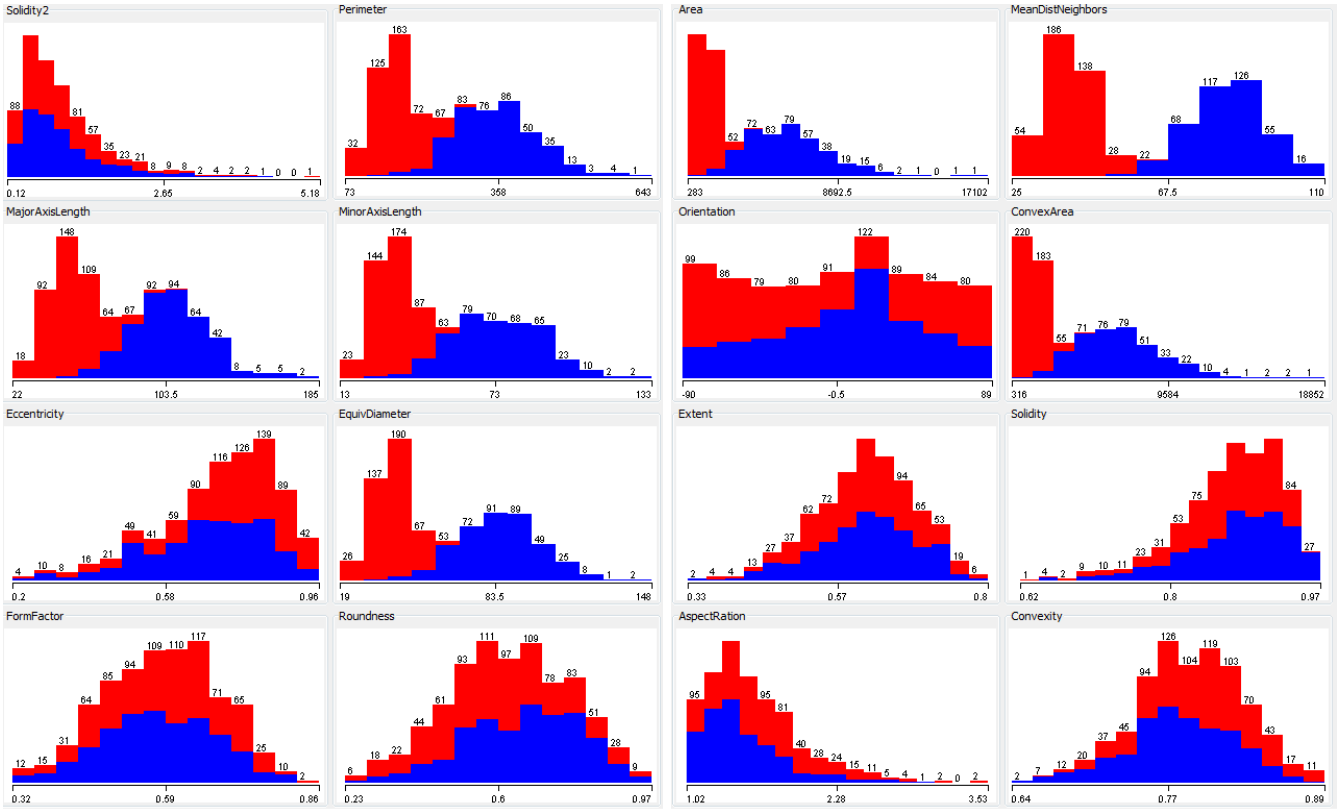


Figure 4. Histogram of the 16 measures (adipose tissue in blue and epithelial tissue in red).

Table III
PERFORMANCE OF SVM CLASSIFIER

Performance measure	Mean	Standard deviation
Accuracy rate	99.88	0.39
Specificity	100.00	0.00
Sensitivity	99.76	0.77

using only the five most prominent features, the SVM is already able to achieve the performance shown in Table III.

Below we show the weight vector of the SVM hyperplane, reported by Weka, using only the five most prominent features as input, where we can see a higher absolute weigth value attributed to the meanDistNeighbors feature:

```
-4.6801 * (normalized) meanDistNeighbors
-1.6845 * (normalized) majorAxisLength
-1.9868 * (normalized) minorAxisLength
 1.3429 * (normalized) solidity
 3.8776 * (normalized) solidity2
```

We can also see in Fig. 5 that there is a good separation between epithelial and adipose data when we consider the features meanDistNeighbors and solidity2.

As the mean distance to average feature was already discriminative for this data set, as shown in the histogram from Fig. 4, we also induced a SVM classifier using this feature as input alone. The accuracy rate obtained was of 99.38%, with an standard deviation of 0.83%. This high accuracy was expected, since this feature was already discriminative for the majority of the data items. However, it was not possible to classify the data using only solidity2 as input to the SVM method.

However, although the areas of adipose cells are generally larger than the areas of epithelial cells, this measure does not allow a direct differentiation among them. On the other hand, we believe that, by using morphological filters on graphs (such as opening or closing) employing as vertices values the areas of each cell and with edges connecting neighbor cells, we can get good classification results, since there is a relationship between cell areas and the average distances between neighbor cells.

Moreover, the best result occurred when a neighbohood relationship between vertices within graphs was considered.

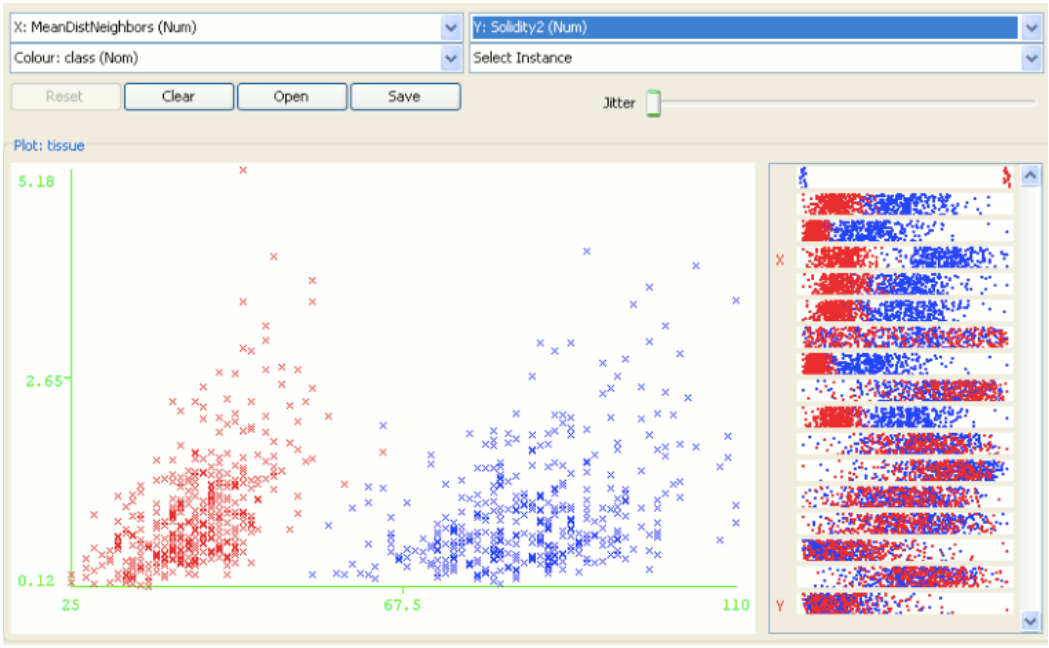


Figure 5. MeanDistNeighbors and solidity2 features (adipose tissue in blue and epithelial tissue in red).

IV. Final Remarks

In this paper, we investigate the potential use of some measures, associated to morphological and topological features, to characterize different tissues. It should be noticed that the aim of this study was to make an initial evaluation of the features described in Section II-D on classifying different cell types. We may conclude from the results shown that these features are indeed good descriptors of the images. In the next approach, we intend to use these measures on a more challenging classification problem, related to the recognition of different types of histological images, as in the distinction of normal from tumour images from the same tissue. It would be interesting to apply other techniques of Mathematical Morphology using neighbourhood graphs, that can generate other positive results in the classification of images, as evidenced in this paper. Furthermore, we intend to explore more carefully which subsets of features should be considered for each particular classification problem - the topological, both individual and global ones versus the graphical measures. For such, we intend to employ feature selection techniques from the Data Mining and Machine Learning literature [36] to better evaluate the importance of subsets of attributes and to develop a robust tissue characterization and classification method.

Acknowledgment

The first author is supported by FAPESP, process: $2009/14430-1$. The third author is supported by CNPq.

References

[1] N. Zahir and V. M. Weaver, "Death in the third dimension: apoptosis regulation and tissue architecture," *Curr Opin Genet Dev*, vol. 14, no. 1, pp. 71–80, 2004.

[2] N. L. Komarova and P. Cheng, "Epithelial tissue architecture protects against cancer," *Math Biosci*, vol. 200, no. 1, pp. 90–117, 2006.

[3] J. Galle, M. Hoffmann, and G. Aust, "From single cells to tissue architecture-a bottom-up approach to modelling the spatio-temporal organisation of complex multi-cellular systems," *J Math Biol*, 2008.

[4] D. E. Ingber, "Can cancer be reversed by engineering the tumor microenvironment?" *Semin Cancer Biol*, vol. 18, no. 5, pp. 356–64, 2008.

[5] C. M. Nelson and M. J. Bissell, "Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer," *Annu Rev Cell Dev Biol*, vol. 22, pp. 287–309, 2006.

[6] J. Jelonek and J. Stefanowski, "Feature subset selection for classification of histological images," *Artif Intell Med*, vol. 9, no. 3, pp. 227–39, 1997.

[7] G. Landini and I. E. Othman, "Architectural analysis of oral cancer, dysplastic, and normal epithelia," *Cytometry A*, vol. 61, no. 1, pp. 45–55, 2004.

[8] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–5, 2002.

[9] C. Gunduz, B. Yener, and S. H. Gultekin, "The cell graphs of cancer," *Bioinformatics*, vol. 20, no. 1, pp. 145–I151, 2004.

[10] A. L. Barabasi and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nat Rev Genet*, vol. 5, no. 2, pp. 101–13, 2004.

[11] E. Lieberman, C. Hauert, and M. A. Nowak, "Evolutionary dynamics on graphs," *Nature*, vol. 433, no. 7023, pp. 312–6, 2005.

[12] T. Aittokallio and B. Schwikowski, "Graph-based methods for analysing networks in cell biology," *Brief Bioinform*, vol. 7, no. 3, pp. 243–55, 2006.

[13] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi, "The human disease network," vol. 104, no. 21, 2007, pp. 8685–90.

[14] J. Serra, *Image analysis and mathematical morphology*. London: Academic Press, 1982.

[15] ——, *Image analysis and mathematical morphology - Volume II: theoretical advances*. London: Academic Press, 1988.

[16] J. M. Geusebroek, A. W. Smeulders, F. Cornelissen, and H. Geerts, "Segmentation of tissue architecture by distance graph matching," *Cytometry*, vol. 35, no. 1, pp. 11–22, 1999.

[17] S. J. Keenan, J. Diamond, W. G. McCluggage, H. Bharucha, D. Thompson, P. H. Bartels, and P. W. Hamilton, "An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (cin)," *J Pathol*, vol. 192, no. 3, pp. 351–62, 2000.

[18] R. D'Antoni and A. D. Giusti, "Model based retinal analysis for retinopathy detection," 2007, pp. 6732–5.

[19] B. D. Khoodoruth, H. C. Rughooputh, and W. Lefer, "Semi-automatic integrated segmentation approaches and contour extraction applied to computed tomography scan images," *Int J Biomed Imaging*, 2008.

[20] V. Shrimali, R. S. Anand, V. Kumar, and R. K. Srivastav, "Medical feature based evaluation of structuring elements for morphological enhancement of ultrasonic images," *J Med Eng Technol*, vol. 33, no. 2, pp. 158–69, 2009.

[21] C. Demir and S. H. Gultekin, "Augmented cell-graphs for automated cancer diagnosis," *Bioinformatics*, vol. 21, no. 2, pp. ii7–12, 2005.

[22] R. Abu-Eid and G. Landini, "Morphometrical differences between pseudo-epitheliomatous hyperplasia in granular cell tumours and squamous cell carcinomas," *Histopathology*, vol. 48, no. 4, pp. 407–16, 2006.

[23] C. Bilgin, C. Demir, C. Nagi, and B. Yener, "Cell-graph mining for breast tissue modeling and classification," 2007, pp. 5311–4.

[24] C. Gunduz-Demir, "Mathematical modeling of the malignancy of cancer using graph evolution," *Math Biosci*, 2007.

[25] N. Cristianini and J. Shawe-Taylor, Eds., *An introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[26] E. R. Dougherty and J. Barrera, "Pattern recognition theory in nonlinear signal processing," *J. Math. Imaging Vis.*, vol. 16, no. 3, pp. 181–197, 2002.

[27] F. Meyer, "An overview of morphological segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 7, pp. 1089–1118, 1996.

[28] E. R. Dougherty and R. A. Lotufo, Eds., *Hands-on morphological image processing*. SPIE, 2003, vol. TT59.

[29] H. Heijmans, *Morphological image operators*. Boston: Academic Press, 1994.

[30] L. Vincent and P. Soille, "Watersheds in digital spaces : an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[31] F. A. Zampirolli, "Neighborhood graphs built with morphological operators," *Journal of Applied Computing*, vol. 4, pp. 5–12, 2008.

[32] J. Barrera, F. A. Zampirolli, and R. A. Lotufo, "Morphological operators characterized by neighborhood graphs," in *SIBGRAPI'97 - X Brazilian Symposium of Computer Graphic and Image Processing*. IEEE Computer Society, October 1997, pp. 179–186.

[33] R.-F. Chang, W.-J. Wu, W. K. Moon, and D.-R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," *Breast Cancer Research and Treatment*, vol. 89, pp. 179–18, 2005.

[34] V. N. Vapnik, Ed., *The nature of Statistical learning theory*. New York: Springer-Verlag, 1995.

[35] I. A. Witten and E. Frank, *Data mining: practical machine learning tools and techniques*. Elsevier, 2005.

[36] H. Liu and H. Motoda, *Feature selection for knowledge discovery and data mining*. Kluwer Academic Publishers, 1998.