Journal of Physics: Conference Series **574** (2015) 012135 doi:10.1088/1742-6596/574/1/012135

# Automatic Method to Classify Images Based on Multiscale Fractal Descriptors and Paraconsistent Logic

**E Pavarino[1], L A Neves[1], M Z Nascimento[1], M F Godoy[3,4], P F Arruda[4], D S Neto[5]**

[1] São Paulo State University (UNESP), Department of Computer Science and Statistics (DCCE), Brazil
[2] Federal University of Uberlândia, FACOM, Brazil
[3] FAMERP - São José do Rio Preto, Brazil
[4] Transdisciplinary Center for Study of Chaos and Complexity (NUTECC), Brazil
[5] Hospital de Base de São José do Rio Preto, Department of Pathology, Brazil

E-mail: pavarinodu@gmail.com

**Abstract.** In this study is presented an automatic method to classify images from fractal descriptors as decision rules, such as multiscale fractal dimension and lacunarity. The proposed methodology was divided in three steps: quantification of the regions of interest with fractal dimension and lacunarity, techniques under a multiscale approach; definition of reference patterns, which are the limits of each studied group; and, classification of each group, considering the combination of the reference patterns with signals maximization (an approach commonly considered in paraconsistent logic). The proposed method was used to classify histological prostatic images, aiming the diagnostic of prostate cancer. The accuracy levels were important, overcoming those obtained with Support Vector Machine (SVM) and Best-first Decicion Tree (BFTree) classifiers. The proposed approach allows recognize and classify patterns, offering the advantage of giving comprehensive results to the specialists.

## 1. Introduction

The biological vision system is one of the most important alternative for human beings to explore the world, smoothly performing complex functions, such as analysis, recognition and classification of patterns. Many researches try to produce computational systems with the same efficiency of the biological system [1]. This task is complex, especially to transpose one of the most evident problem, the quantification and classification of information represented in different domains, such as intensity of gray levels, borders, contours and texture. A computer-aided diagnosis (CAD) can be an alternative. The desired specificity and sensibility for a CAD can be obtained if produced with techniques inspired in natural phenomena, capable to treat spatial and temporal data efficiently [2]. Thus, it is possible to highlight fractal approaches to quantify attributes [3,4] and paraconsistente logic to obtain classifications compatible with human reasoning [5,6].

In this context, an automatic method is presented to classify images from fractal descriptors, such as multiscale fractal dimension and lacunarity, and paraconsistent logic. As a first application, the proposed method was used to classify histological prostatic images, aiming the diagnostic of prostate cancer. The proposed approach allows recognize and classify patterns in the studied context, offering the advantage of giving comprehensive models to the specialists.

## 2. Methodology

The proposed method was organized as follows: formalization of multi-scale fractal dimension and lacunarity, which were used to quantify images given as input; definition of the classifier core, using the values of the descriptors as references; definition of the quantitative evaluation applied to validate the proposed method; and, description of the features from prostatic images representing normal, hyperplastic and tumor/cancer groups (the first application).

From each image given as input, the first step consisted in segmentation and quantification of the regions of interest. The method chosen to segment images was presented by [4]: the algorithm is an automatic technique and does not require specific knowledge about the explored context. Then, each segmented region was quantified from multiscale fractal dimension and lacunarity techniques. The fractal dimension was applied as a measure of the complexity of the organization of pixels in the region of interest or image [2-4]. The multiscale approach was obtained by overlapping a mesh of squares in the analyzed image. The objective was to obtain the number of frames needed to cover it. Therewith, considering an image $I$ given as input, the number of squares that contains part of the form $I$, $N_l(I)$, depended of the size of the $l$ box. This relation allowed estimating the fractal dimension $D$, equation (1).

$$D = -\lim_{l \to 0} \frac{\ln(N_l(I))}{\ln(l)}. \tag{1}$$

The multiscale approach requires a limit, which is the count of frames, $N_l(I)$, performed for different values of $l$ ($l_0$=max(height, width) and $l_i$+1=$l_i$ / 2). The approximation of a straight line was obtained by the regression (log-log graphic) of $N_l(I)$ (number of occupied boxes) by $l$ (size of the side of this box). This allowed to define $D$=–θ as the fractal dimension of $I$ in the multiscale approach.

The lacunarity was applied to quantify the pixel's organization in a determined region of the image (quantifying how the space is filled). The method used was the *Gliding-box* [2-4]. The lacunarity complemented the evaluations made by fractal dimension, because shapes with the same value of fractal dimension can present different values of lacunarity. The process was started with a box of size $l$ positioned in the upper left corner of the image and counting the number of image points. This process was repeated to all rows and columns of the image, producing a distribution of frequency of the image mass. The number of boxes with size $l$ containing a value of mass $M$ of the interest region was designated by $n(M,l)$. The total counted boxes in the image was designated by the probability $P(M,l)$, equation (2). The first ($A^1$) e second ($A^2$) moments of this distribution were determined in the equations (3) e (4). The lacunarity $L$ for a box of size $l$ is defined in the equation (5).

$$P(M,l) = \frac{n(M,l)}{N(l)}. \tag{2}$$

$$A^{(1)} = \sum MP(M,l). \tag{3}$$

$$A^{(2)} = \sum M^2 P(M,l). \tag{4}$$

$$L(l) = \frac{A^{(2)}}{\left(A^{(1)}\right)^2}. \tag{5}$$

The classifier allows to treat the descriptors as feature vector ($L$-dimensional vector) of the type: $A$ of order [$n$, $d$], where $n$ is the number of analyzed images and $d$ is the quantity of considered descriptors. In this work two descriptors (fractal dimension and lacunarity) were used: $d$=2. So, each pair of columns from the vector/matrix $A$ contains the limiting values of the intervals of each descriptor considered. These values were verified to determine the interval that best represents each descriptor, inclusive treating possible overlaps. Therefore, the core classifier was composed by treating two types of overlaps between the intervals A e B: (1) A e B overlapped, but A $\not\subset$ B and B $\not\subset$ A; and (2) A e B overlapped, but A $\subset$ B and/or B $\subset$ A. For both types of overlapping, the classifier uses a numeric value α called "cut-off point (threshold value)", which is responsible for indicating the most appropriate value to classify the classes. The difference between the cases (1) and (2) it is the manner that the value of α is defined. Being $S$ the overlapping interval between A and B, in overlaps of type (1), α was obtained as follows: (a) minA < minB and minB < maxA < maxB → S = [minB, maxA] → α = minB. Thus, A e B will be redefined as A = [minA, α[ and B = [α, maxB]; and, (b) minB < maxA e min A < maxB < maxA → S = [minA, maxB] → α = maxB. Thus, B and A will be redefined as B = [minB, α[ e A = [α, maxA]. For overlaps of type (2), in which the approach (1) does not apply, the value of α was defined as a cut-off point calculated from the Receiver Operating Characteristics (ROC) [7].

From the intervals calculated previously, the classifier core compares each row of the matrix, that represents a case to be studied, and assigns binary values called degree of certainty (*Gc*) and degree of uncertainty (*Gi*): this degrees define if a given row *i* (case) belongs or not to the interval of certain descriptor *d*. If it belongs, $Gc(i, d) = 1$ and $Gi(i, d) = 0$; otherwise, $Gc(i, d) = 0$ e $Gi(i, d) = 1$. Then, for each case *i*, the values of *Gc* where maximized (applying logic gate OR) and the values of *Gi* where minimized (applying logic gate AND): this procedure is commonly explored in paraconsistent logic [6,7]. The classification of each row *i* was performed comparing the degrees of certainty and uncertainty, allowing to identify the class of a specific case, made as follows: being $\mathrm{Gc}^{X}(i,d)$, $\mathrm{Gi}^{X}(i,d)$, $\mathrm{Gc}^{Y}(i,d)$ e $\mathrm{Gi}^{Y}(i,d)$ the degrees of certainty and uncertainty of an image *i* belong to the groups X e Y, according a descriptor $d$=2, then *i* will be classified as belonging to a class X if, and only if, $\mathrm{Gc}^{X}(i,d) > \mathrm{Gc}^{Y}(i,d)$ e $\mathrm{Gi}^{X}(i,d) < \mathrm{Gi}^{Y}(i,d)$.

As a first application context, the proposed method was tested using images with recognized scientific importance: detection of prostate cancer [4]. The bank of prostatic images used in this work was composed of 99 histological slides, obtained from 35 patients' prostates, carriers of prostate cancer, aged between 50 and 75 years (64.4 ± 5.9 years). The extraction of each prostate allowed producing histological slides from fragments removed from normal, hyperplastic and tumor areas of the same prostate. The microscopic images were acquired digitally, using a trinocular microscopic Olympus BX41, with planacromatic objective of 10x coupled to a digital camera Color Samsung SCC-31, with an adapter Olympus U-TV1X-2. Each slide was photographed digitally with 40x of magnification and saved in the jpeg file format. The images constituted the groups: normal, hyperplastic and tumor/cancer, with 33 cases each.

The discriminative power from the proposed method was evaluated applying known artificial intelligence algorithms, such as support vector machine (SVM) and Best-First Decision Tree (BFTree) [8]. The features matrixes, with values of fractal dimension and lacunarity, were compared two by two (normal *versus* hyperplastic, normal *versus* cancer and hyperplastic *versus* cancer), considering all the classifiers and the segmented regions stroma and lumen. This procedure resulted in 396 comparisons. The performance between groups and techniques were measured from accuracy, which is defined as the proportion of true results (both true positives (TP) and true negatives (TN)) in relation to all positive (P) and all negatives (N), equation (6) [7,8].

$$Accuracy = \frac{TP+TN}{P+N} . \qquad (6)$$

## 3. Results

The performance of the method was evaluated from the previously described strategies and using prostatic images segmented in regions of stroma and lumen (application context). The values of accuracy are presented in figure 1, involving the regions of stroma and lumen of the groups normal (N), hyperplastic (H) and tumor (T). The values were obtained from the features matrixes composed by values of multiscale fractal dimension and lacunarity, which were used as input to the classifiers SVM, BFTree and that proposed in this work.

The results indicate that the proposal is relevant, providing accuracy values higher than those obtained from other recognized classifiers. The developed method overcame the SVM classifier in all the realized testes, considering the regions of interest stroma and lumen. This has not occurred regarding the classifier BFTree, which overcame the described approach in the comparisons between the classes: normal and hyperplastic. Nevertheless, the developed method showed better performance for the other testes (normal *versus* tumor and hyperplastic *versus* tumor), considering the regions of stroma and lumen, too. This is an important advantage, demonstrating that the presented technique allowed better distinction between the classes that requires therapeutic procedures totally different in the clinical practice. Another important observation for the clinical practice is that, in a 40x magnification, the developed method indicated lumen as the most relevant region to distinguish the classes normal, hyperplastic and tumor. Regarding the descriptors used, the fractal dimension was decisive in 66.67% of the tests to classify the groups from lumen, mainly between normal *versus* hyperplastic and hyperplastic *versus* tumor. This percentage was obtained with lacunarity too, but for comparisons with the region of stroma and highlighting to classify the groups normal *versus* tumor and hyperplastic *versus* tumor.

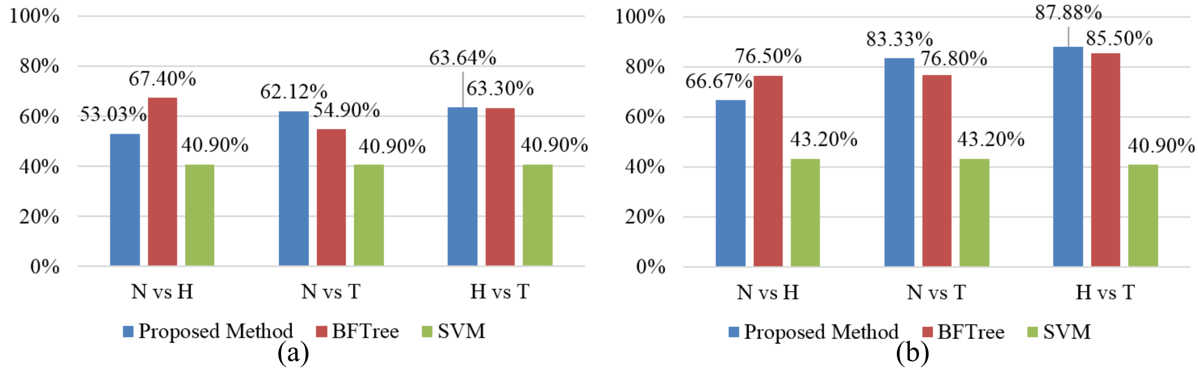


**Figure 1.** Accuracy levels obtained with the proposed method and BFTree and SVM classifiers, considering lacunarity and fractal dimension as descriptors. The results obtained with stroma are represented in (a) and with lumen in (b).

## 4. Conclusion

This work presented a new technique to classify images using fractal descriptors (fractal dimension and lacunarity) and paraconsistent logic. In a first context, the proposed model was subjected to tests with prostatic images with 40x of magnification to prove your efficiency. Through the values adopted as reference for the studied classes (normal, hyperplastic and tumor/cancer), the classifications reached higher levels of accuracy than other recognized classifiers in the majority of tests performed. The highest accuracy was approximately 88% and the method indicated lumen as the most relevant region to distinguish groups. By the results achieved, the proposed technique can be considered as a viable alternative to classify signals, with the advantage of making the results understandable to specialists.

## Acknowledgements

This work was financially supported by PROPe/UNESP (Pró-Reitoria de Pesquisa/UNESP).

## References

[1] Jain A K and Duin R 2000 Statistical pattern recognition: A review *IEEE Trans. on Patt. An. and Mac. Intel.* **22** 1 4–37
[2] Neves L A et al 2011 Maximum entropy, fractal dimension and lacunarity in quantification of cellular rejection in myocardial biopsy of patients submitted to heart transplantation *J. of Phys.: Conf. Series* **285** 1
[3] Wan-rong S, Bian-zhang Y, Xiao-jing Z and Zheng-hui, Z 2005 Fractal feature extraction of marrow cell images *Proceeedings of the 2005 IEEE*, *Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China*
[4] Neves L A et al 2014 Multi-scale lacunarity as an alternative to quantify and diagnose the behavior of prostate cancer *Expert Systems with Applications* **41.11** 5017-29
[5] Da Costa N C, Krause D and Bueno O 2007 Paraconsistent Logics and Paraconsistency *Handbook of the Philosophy of Science, Philosophy of Logic, Elsevier*
[6] Lopes H F S, Abe J M and Anghinali R 2009 Application of Paraconsistent Artificial Neural Networks as a Method of Aid in the Diagnosis of Alzheimer Disease *J. Med. Syst.*
[7] Fawcett T 2006 An introduction to ROC analysis *Pattern recognition letters* **27.8** 861-74
[8] Duda R O, Hart P E and Stork D G 2012 Pattern classification *John Wiley & Sons*