

Non-Extensive Entropy for CAD Systems of Breast Cancer Images

Paulo S. Rodrigues and Gilson A. Giraldi
National Laboratory for Scientific Computing
Petrópolis, Rio de Janeiro, Brazil
{pssr,gilson@lncc.br}

Ruey-Feng Chang
Dept of C.S. and Information Engineering
National Chung Cheng University,
Chiayi, Taiwan, dlchen88@ms13.hinet.net

Jasjit S. Suri
Biomedical Technologies Inc., CO, USA
Biomedical Research Institute, ID, USA., suri0256@msn.com

Abstract

Recent statistics show that breast cancer is a major cause of death among women in all of the world. Hence, early diagnostic with Computer Aided Diagnosis (CAD) systems is a very important tool. This task is not easy due to poor ultrasound resolution and large amount of patient data size. Then, initial image segmentation is one of the most important and challenging task. Among several methods for medical image segmentation, the use of entropy for maximization the information between the foreground and background is a well known and applied technique. But, the traditional Shannon entropy fails to describe some physical systems with characteristics such as long-range and long-time interactions. Then, a new kind of entropy, called non-extensive entropy, has been proposed in the literature for generalizing the Shannon entropy. In this paper, we propose the use of non-extensive entropy, also called q -entropy, applied in a CAD system for breast cancer classification in ultrasound of mammographic exams. Our proposal combines the non-extensive entropy, a level set formulation and a Support Vector Machine framework to achieve better performance than the current literature offers. In order to validate our proposal, we have tested our automatic protocol in a data base of 250 breast ultrasound images (100 benign and 150 malignant). With a cross-validation protocol, we demonstrate system's accuracy, sensitivity, specificity, positive predictive value and negative predictive value as: 95%, 97%, 94%, 92% and 98%, respectively, in terms of ROC (Receiver Operating Characteristic) curves and A_z areas.

1 Introduction

As per American Cancer Society [1], breast cancer ranks second in the list of women's cancer. Even though the rate

of breast cancer has risen since 1980, the mortality rates have declined by 2.3% since 1990. The reduction in mortality rate is due to early detection and improvement in technology for treatment. Under this context and due to amount of patient data size, CAD systems has become fundamental for ultrasound exams of breast cancer. CAD systems are a set of tools with the goal of helping the diagnostic of hard lesions. These systems are built based on several steps, each one with different techniques, however, interdependent.

In a general manner, the images are acquired with a 2D or 3D US (Ultrasound) scanners and then the lesion is extracted from its background so that the probability of being malignant or benign will be evaluated. The lesion extraction is one of the most important task as it is an early process. Then CAD systems generally need the user intervention in order to correctly extract the lesion boundary.

However, a totally automatic process is still not a reality and so much work is necessary, mainly in the early steps which may involve segmentation, recognition and extraction of lesion from its background. Specifically speaking, in the case of ultrasound images, the initial segmentation is a complex process since the images have a low SNR (Sign to Noise Ratio), low resolution, low contrast and several small and spurious regions. There are several challenges in all steps yet. Particularly, the initial segmentation is a fundamental task since the success of the remainder ones depends on it.

Traditional algorithms for image segmentation such as fuzzy c-means (FCM), k-means (KM), Self-Organized Maps (SOM) and Bootstrap have been used in this task and generally give good results as they can achieve several clusters adequately. Nonetheless, they have high computational time as they generally use procedures with some kind of energy minimization, generally in an iterative loop. Other algorithms, such as those based on watershed, may also get several regions but their disadvantage is the oversegmenta-

tion. Since algorithms based on thresholding do not have such problems, they may be a good alternative. On the other hand, the choice of the threshold is not always an easy task as they are generally based on some kind of contrast enhancement between the background and the foreground. Since in the narrow boundary around the lesion the threshold is not simple to compute and these algorithms generally do not get a good boundary location.

Regarding that patterns of ultrasound images are probability distributions of gray scales, the entropy associated to these distributions may be measured in order to maximize the differences inter-clusters composed by the foreground and background's pixels. In this direction, algorithms of Pun [6], Kapur *et al.* [7], Pal [8] and Sahoo *et al.* [9] may be used to achieve the desired threshold. These traditional form of entropy has been applied based on the assumption that the probability distribution of the gray scale image represent extensive systems, which means that there is short-time and short-range interactions between the states. However, recent developments in statistical mechanics have been shown that, when these interactions are of long-time and long-range type, this traditional formalism fails to describe the physical system. Then, another formalism, called non-extensive entropy, has been proposed [10].

This paper proposed the use of non-extensive entropy in a CAD system for breast cancer mammography in ultrasound exams. Our paper has the following contributions. a) we proposed a natural extension to the algorithm proposed in [11] for segmentation of ultrasound images with a non-extensive approach. Our algorithm is a recursive version of [11] and achieves more than two regions in ultrasound images besides having a low computational time. b) The proposed algorithm, called Non-Extensive Segmentation Recursive Algorithm (NESRA) is a first step for a five-step methodology for a CAD system for breast lesion classification. c) This methodology uses a SVM (Support Vector Machine) framework in order to classify the breast regions between malignant or benign lesion.

The paper is organized as follows. In the next section we treat of related works. In Section 3 we present the theoretical background underlining the non-extensive entropy. The proposed method are outlined in Section 4. In Section 5 we show the practical results and its analysis and offer some discussion in Section 6.

2 Related Works

The process of breast lesion classification and efficient CAD-based methods from ultrasound is generally categorized into several techniques such as: Bootstrap [2], SVM [3] and Neural Networks [4, 12]. On these lines, Sawaki *et al.* [13] proposed a CAD system using fuzzy inference for breast sonography and adopted six different criteria to

classify lesions such as: lesion shape, border, edge shadows, internal echoes, posterior echoes, and halo. However, their system accuracy, sensitivity and specificity were only 60.3%, 82.1% and 42.9%, respectively. Garra *et al.* [14] analyzed the breast lesions in ultrasound images using a co-occurrence matrices of ultrasound images to represent their texture information. Chen *et al.* [4] proposed an autocorrelation coefficients to analyze the texture information. Since a CAD system trained by images from one ultrasound machine needs to be trained again for a different ultrasound machine due to different image resolution and image quality, Chen *et al.* [12] proposed nearly setting-independent features based on shape information. Their system was very robust and powerful because the statistical data using ROC curve were all greater than 0.95.

Recently, 3-D ultrasound (Chang *et al.* [15], Chen *et al.* [16] and Chen *et al.* [5]) has shown promising signs that overcome the limitations of traditional 2-D ultrasound, allowing physicians to view the anatomy in 3-D interactively, instead of assembling the sectional images in their minds. However, this texture-based CAD is highly machine dependent and utilizes these machine settings all the time.

The five step methodology proposed in this paper is the following: (a) application of non-extensive entropy for initial segmentation, (b) morphologic cleaning and (c) accurate region and boundary extraction in level set framework, (d) extraction of five lesion features such as area, circularity, protuberance, homogeneity, and acoustic shadow; (e) application of a SVM classifier with a B-Spline kernel for breast cancer classification between malignant and benign cases.

In our experiments, the accuracy of our proposed protocol for classify malignancies is 95%, specificity is 97%, positive predictive value is 94% and negative predictive value is 92%. Besides, we have an *Az* area of 92%.

3 Theoretical Background

3.1 Tsallis Entropy

The traditional equation for entropy is defined as:

$$S = - \sum_i p_i \ln(p_i) \quad (1)$$

Generically speaking, systems which can be described by Equation (1) are called extensive systems and have the following additive property: Let P and Q be two random variables, with probability densities functions $P = (p_1, \dots, p_n)$ and $Q = (q_1, \dots, q_m)$, respectively, and S be the entropy associated with P or Q . If P and Q are independent, under the context of the Probability Theory, the entropy of the composed distribution¹ verify the so called

¹we define the composed distribution, also called direct product of $P = (p_1, \dots, p_n)$ and $Q = (q_1, \dots, q_m)$, as $P * Q = \{p_i q_j\}_{i,j}$, with $1 \leq$

additivity rule:

$$S(P * Q) = S(P) + S(Q) \quad (2)$$

This traditional form of entropy is well known and for years has achieved relative success to explain several phenomenon *if* the effective microscopic interactions are *short*-ranged (i.e., close spatial connections) *and* the effective spatial microscopic memory is *short*-ranged (i.e., close time connections) *and* the boundary conditions are *non(multi)fractal*. Roughly speaking, the standard formalism are applicable whenever (and probably only whenever) the relevant space-time is non(multi)fractal. If this is not the case, some kind of extension appears to became necessary.

Then, recent developments, based on the concept of non-extensive entropy, also called Tsallis entropy, have generated a new interest in the study of Shannon entropy for Information Theory [10, 17]. Tsallis entropy (or q -entropy) is a new proposal for the generalization of Boltzmann/Gibbs traditional entropy applied to non-extensive physical systems.

The non-extensive characteristics of Tsallis entropy has been applied through the inclusion of a parameter q , which generates several mathematical properties and the general equation is the following:

$$S_q(p_1, \dots, p_k) = \frac{1 - \sum_{i=1}^k p_i^q}{q-1} \quad (3)$$

where k is the total number of possibilities of the whole system and the real number q is the entropic index that characterizes the degree of non-extensiveness. In the limit $q \rightarrow 1$, Equation (3) meets the traditional BGS entropy defined by Equation (1). These characteristics give to q -entropy flexibility in explanation of several physical systems. On the other hand, this new kind of entropy does not fail to explain the traditional physical systems since it is a generalization.

Furthermore, a generalization of some theory may suppose the violation of one of its postulates. In the case of the generalized entropy proposed by Tsallis, the additive property described by Equation (2) is violated in the form of Equation (4), which apply if the system has a non-extensive characteristic. In this case, the Tsallis statistics is useful and the q -additivity describes better the composed system. In our case, the experimental results (Section 5) show that it is better to consider our systems as having non-extensive behavior.

$$S_q(P * Q) = S_q(P) + S_q(Q) + (1 - q)S_q(P)S_q(Q) \quad (4)$$

In this equation, the term $(1 - q)$ stands for the degree of non-extensiveness. Note that, when $q \rightarrow 1$, this equation meets the traditional Equation 2.

$$i \leq n \text{ and } 1 \leq j \leq m$$

Albuquerque *et al.* [11] proposes an algorithm using the concept of q -entropy to segment US images. Since this concept may be naturally applied over any statistical distribution, in this paper we propose a natural extension of the algorithm proposed by Albuquerque *et al.* [11]. Our proposal is a recursive procedure of [11] as, for each distribution P and Q , we applied again the concept of q -entropy. We named our extended algorithm as NESRA (Non-Extensive Segmentation Recursive Algorithm) and also propose to apply it in an initial segmentation of our proposed five step methodology for CAD systems.

The motivations to use the q -entropy are: 1) managing only a simple parameter q yields to a more controllable system; 2) as suggested in [11], the mammographic images and possible several others natural images have a non-extensive profile 3) it is simple and makes the implementation easy, having a low computational overload.

In the following section, we fully describe the NESRA proposal.

3.2 The Non-Extensive Segmentation Recursive Algorithm (NESRA)

Suppose an image with k gray-levels. Let the probability distribution of these levels be $P = \{p_i = p_1; p_2; \dots; p_k\}$. Then, we consider two probability distribution from P , one for the foreground (P_A) and another for the background (P_B). We can make a partition at luminance level t between the pixels from P into A and B . In order to maintain the constraints $0 \leq P_A \leq 1$ and $0 \leq P_B \leq 1$ we must re-normalize both distribution as: $P_A : \frac{p_1}{p_A}, \frac{p_2}{p_A}, \dots, \frac{p_t}{p_A}$ and $P_B : \frac{p_{t+1}}{p_B}, \frac{p_{t+2}}{p_B}, \dots, \frac{p_k}{p_B}$, where $p_A = \sum_{i=1}^t p_i$ and $p_B = \sum_{i=t+1}^k p_i$.

Now, following the Equation (3), we calculate the a priori Tsallis entropy for each distribution as $S_A = \frac{1 - \sum_{i=1}^t (\frac{p_i}{p_A})^q}{q-1}$ and $S_B = \frac{1 - \sum_{i=t+1}^k (\frac{p_i}{p_B})^q}{q-1}$. Allowing the pseudo-additive property given by Equation (4), for two statistically independent systems, we can compute the pseudo-additive property of systems A and B as:

$$S_{A+B}(t) = \frac{1 - \sum_{i=1}^t (\frac{p_i}{p_A})^q}{q-1} + \frac{1 - \sum_{i=t+1}^k (\frac{p_i}{p_B})^q}{q-1} + (1 - q) \frac{1 - \sum_{i=1}^t (\frac{p_i}{p_A})^q}{q-1} \frac{1 - \sum_{i=t+1}^k (\frac{p_i}{p_B})^q}{q-1} \quad (5)$$

To accomplish the segmentation task, in [11] the information measure between the two classes (foreground and background) is maximized. In this case, the luminance level t is considered to be the optimum threshold value (t_{opt}), which can be achieved with a cheap computational effort of

$$t_{opt} = \operatorname{argmax}[S_A(t) + S_B(t) + (1 - q)S_A(t)S_B(t)] \quad (6)$$

Note that the value t which maximizes Equation (6) depends on mainly the parameter q . This is an advantage due to its simplicity. Furthermore, the value of q which generates t_{opt} is not explicitly calculated and, in this work, it was empirically defined as that which generates the best CAD system performance. Since $q = 1$ when the system behavior is extensive, in our experiments we have achieved a non-extensive system behavior for our data base since the better results, in terms of precision of classification, were obtained for $q < 1$. An example of application of $q \neq 1$ (non-extensive system) against $q = 1.0$ (traditional extensive system) is showed in the Figure 1.

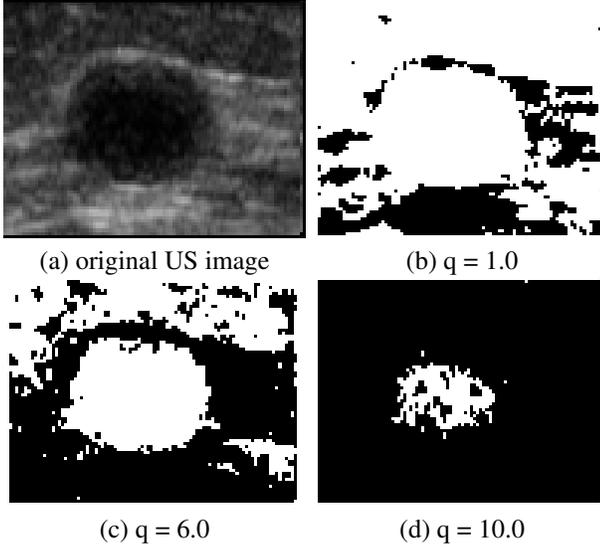


Figure 1. Original ultrasound benign image (a) and the non-extensive entropy segmentation results for different q values: (b) 1.0, (c) 6.0 and (d) 10, respectively.

Among the results presented in Figure 1, only the (c) and (d) seems to be useful in a post-processing module. With (c) we can extract the central region and used its boundary's curve to initialize some deformable model framework (level set, snake, dual snake, etc.) in order to get a better tumor's boundary location. Since in the ultrasound images we have several important regions, an image binarization does not guarantee an accurate analysis of these regions. For this reason we create an extension of the algorithm proposed by Albuquerque *et al.* [11] which is able to obtain the same result as in Figure 1 but with more levels of segmentation. This is a multi-segmentation algorithm.

Following the definitions above, we can take each distribution P_A and P_B and subdivide them into two news distribution, P_{A1}, P_{A2}, P_{B1} and P_{B2} , as following. $P_{A1} : \frac{p_1}{p_{A1}}, \frac{p_2}{p_{A1}}, \dots, \frac{p_t}{p_{A1}}$, $P_{A2} : \frac{p_{t+1}}{p_{A2}}, \frac{p_{t+2}}{p_{A2}}, \dots, \frac{p_\theta}{p_{A2}}$, $P_{B1} : \frac{p_{\theta+1}}{p_{B1}}, \frac{p_{\theta+2}}{p_{B1}}, \dots, \frac{p_\nu}{p_{B1}}$ and $P_{B2} : \frac{p_{\nu+1}}{p_{B2}}, \frac{p_{\nu+2}}{p_{B2}}, \dots, \frac{p_k}{p_{B2}}$, hav-

ing the constraints $p_{A1} = \sum_{i=1}^t p_i$, $p_{A2} = \sum_{i=t+1}^\theta p_i$, $p_{B1} = \sum_{i=\theta+1}^\nu p_i$, $p_{B2} = \sum_{i=\nu+1}^k p_i$. For each one of these four distributions we can compute its respective non-extensive entropy as follows: $S_{A1} = \frac{1 - \sum_{i=1}^t (\frac{p_i}{p_{A1}})^q}{q-1}$, $S_{A2} = \frac{1 - \sum_{i=t+1}^\theta (\frac{p_i}{p_{A2}})^q}{q-1}$, $S_{B1} = \frac{1 - \sum_{i=\theta+1}^\nu (\frac{p_i}{p_{B1}})^q}{q-1}$ and $S_{B2} = \frac{1 - \sum_{i=\nu+1}^k (\frac{p_i}{p_{B2}})^q}{q-1}$.

Then, computing $S(A) = S(A1 + A2)$ and $S(B) = S(B1 + B2)$, as the same way as Equation (5), we have a new additive property, which, to find the optimal luminance level, such as the equivalent in the Equation (6), we take the argument that maximizes the following expression:

$$t_{opt} = \operatorname{argmax}[(S_{A1} + S_{A2} + (1 - q) \cdot S_{A1} \cdot S_{A2}) \cdot (S_{B1} + S_{B2} + (1 - q) \cdot S_{B1} \cdot S_{B2}) \cdot (1 - q) \cdot (S_{A1} + S_{A2} + (1 - q) \cdot S_{A1} \cdot S_{A2}) \cdot (S_{B1} + S_{B2} + (1 - q) \cdot S_{B1} \cdot S_{B2})] \quad (7)$$

In this case, instead of Equation (6), $t_{opt} = \{t_{opt1}, t_{opt2}, t_{opt3}\}$ is a set of threshold values for two recursions, where t_{opt1} is the first threshold and t_{opt2} and t_{opt3} are new thresholds for P_A and P_B distributions, respectively.

At each iteration the NESRA generates 2^{r+1} regions, where r is the number of recursions. Note that for $r = 0$, the NESRA is the simple image binarization algorithm.

After this segmentation, it is necessary to extract from the background the tumor's boundary. To accomplish this task, we apply a mathematical morphology approach. Since this approach may generate a coarse contour of the tumor, we apply a level set framework to smooth the contour. In the next section, we describe our proposed protocol for breast cancer classification.

4 The Proposed Five Steps Methodology

As outlined, we put all the above theoretical background together to generate a five step protocol in order to classify 3D breast images. After application of NESRA algorithm, we extract the lesion area from the background. The Figure 2 shows an original benign image example (left) and the NESRA result (right). This result was obtained with one NESRA recursion only, which means that the gray scale distribution was partitioned into two new distributions and each one was partitioned into two others, generating four distinct regions. Although we wish to split the image into two regions only – and in this case no recursion will be needed – the use of 1 (one) recursion in this case helps to delimit a tumor's nucleus and also a boundary as a narrow region around the nucleus. Note in the Figure 2 the white region around the image center (tumor) and the intermediate gray-level region around the tumor's region (the transition

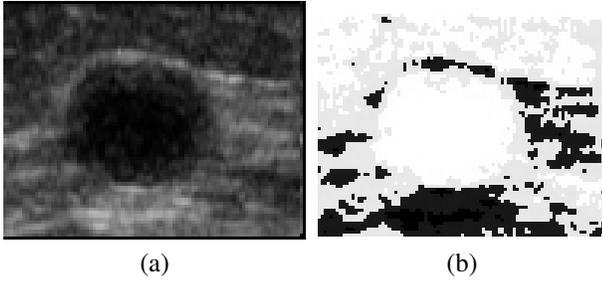


Figure 2. Original ultrasound benign image (left) and the NESRA results (right) with two recursions and $q = 0.5$.

region between the tumor's region and background). This transition region can be isolated and helps to delimit accurately the lesion boundary. Generally, this transition region is made of a mixing of dead and alive cells. These dead cells are generally swallowed to inside the lesion region as tumor growing. For the result of Figure 2-b we used $q = 0.5$. In this case, the ROI (Region Of Interest) is both, the tumor's nucleus and its transition region.

In the second step we use a morphological chain approach to extract the ROI from the background. This is accomplished through the following rule. Considering the binary image (tumor's nucleus) generated by NESRA (e.g Figure 2-b), let α and β be the total ROI's area and the total image area, respectively. If $\alpha \geq \xi\beta$ an erosion is carried out; and if $\alpha \leq \delta\beta$ a dilation is carried out. After, assuming that the ROI has a geometric center near to the image center, we apply a region growing algorithm which defines the final ROI's boundary. In our experiments, we fixed $\xi = 0.75$ and $\delta = 0.25$ to correctly extract most of all ROIs. The result of this morphological rule applied in the image of Figure 2-b is shown in Figure 3-a.

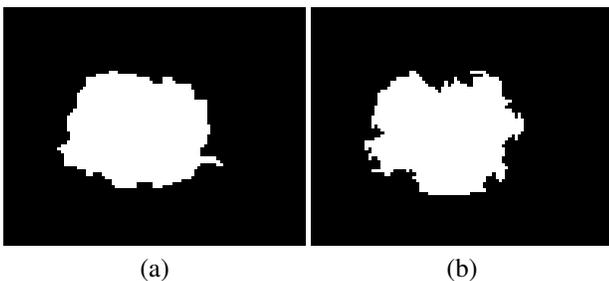


Figure 3. (a) ROI after morphological chain application. (b) ROI after level set application.

The region generated by the morphological chain rule is a coarse representation of the lesion region. Then, as a third

step of our method, we apply a level set framework using as initialization this region's boundary [18]. However, since this initial ROI is near to the real boundary, there is no need for several iterations, as normally occurs in several level set applications. Then, in our experiments, we have applied ten iterations only. It was sufficient to get good results, which generates low computational overhead. The Figure 3-b is the final lesion boundary after the level set computation, which is a more smoothing boundary for the lesion than that in Figure 3-a.

After the above boundary tracing, the next step is the feature extraction of the ROI. Three radiologists have defined five features which have high probability to work well as a discriminator between malignant and benign lesions. In our work, we have used these five features and tested them in order to achieve some combination which yields to a good empirical result.

- **AR:** The first feature is the lesion area. Since malignant lesions generally have large areas in relation to benign ones, this characteristic is a power discriminant. We have normalized it by the total image area.
- **CT:** The second characteristic is related to the region circularity. Since benign lesions generally have more circular areas compared with the malignant ones, also this can be a good discriminant. Then, we take the ROI's geometric center (GC) point and compute the distance from each boundary point (x_i, y_i) to the GC. Malignant lesions tend to have high standard deviations of the average distances in relation to the benign ones. Also, this feature is normalized by total image area.
- **PT:** The third feature is the size distribution of the lobes in a lesion. A boundary's lobe is a protuberant region on the boundary. We compute the convex hull of the ROI and the lobe as a protuberance between two valleys. The lobe areas are computed and only those greater than 10% of the lesion area are considered. This feature is taken as the average area of the lobes. Malignant lesions have high average area in relation to benign ones.
- **HT:** The next feature is related to the homogeneity of the lesion. Malignant lesions tend to be less homogeneous than benign ones. Then, we take the BGS entropy – taken over the gray scale histogram – relative to the maximum entropy as the fourth discriminant feature. As higher the relative entropy more homogeneous is the lesion region and, consequently, higher is the chance to be a benign lesion.
- **AS:** The last feature is related with an interesting characteristic of the lesions: the acoustic shadow. When

benign lesions have many water particles, the formation of an acoustic reinforcement below it is more probable. On the other hand, when the lesion is more solid (a malignant characteristic), there is a tendency in forming an acoustic shadow. Then, by comparing the lesion region and the region below the lesion may give an idea if the lesion is a benign or malignant one. If the region has acoustic shadow it tends to be more white than if it has an acoustic reinforcement. Then, we compute the gray scale histograms of both regions and compare them. When more darkness is the region below the lesion more is the acoustic reinforcement and, consequently, higher is the probability to have a benign lesion. We have computed the relative darkness between both areas (tumor's area and area below the tumor) and take it as the fifth lesion feature.

These five features are combined into a 5-dimensional feature vector. The space of all feature vectors are, finally, input for a Nonlinear Support Vector Machine (SVM) [19] in order to classify the lesion area as a being or a malignant region. SVM is very good discriminant tool, mainly when we do not have a linear separation between the data. Since this feature space does not have a linear separation, we have used a B-Spline curve as a kernel. Since some characteristics are better discriminant than others, we also have combined them into the SVM framework. To justify that using a B-Spline as a kernel for the SVM improves the classification under this non-linear space, we compare in the Figure 4 the performance of our proposal under three different kernels: B-Spline, polynomial and exponential. In this

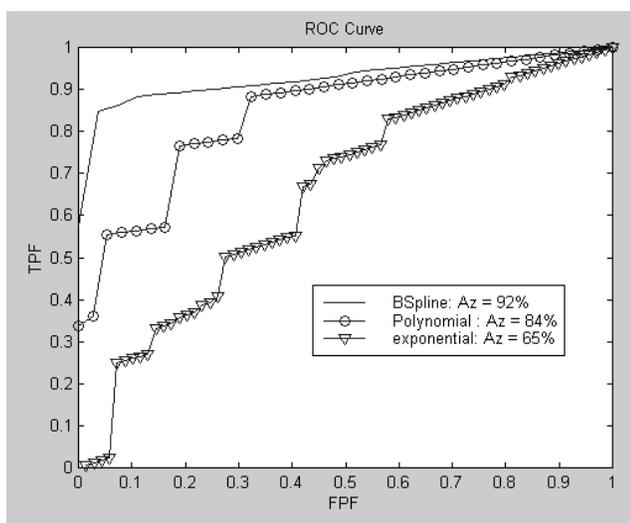


Figure 4. ROC curves for the proposed CAD System for three different kernel functions used in the SVM classifier.

figure, the area under the ROC curve (92% of the total) is clearly superior when the kernel is polynomial (area 84%) and exponential (area 65%). Also, we show in Figure 5 how a polynomial kernel separates our feature space for two features only: AR and HT. On the other hand, when we use

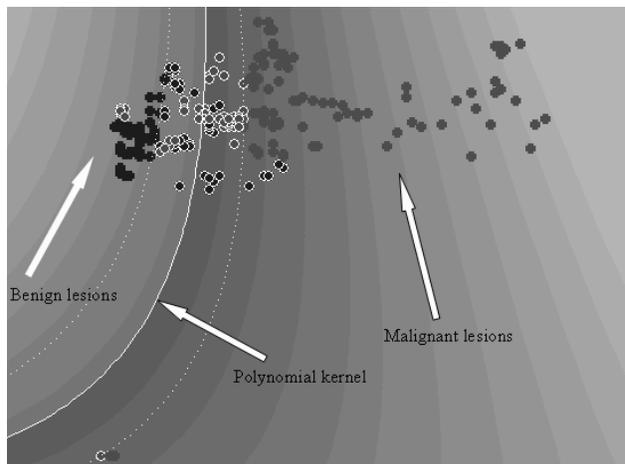


Figure 5. SVM output with a polynomial kernel. The white line is the polynome separating malignant tumors (gray circles) from benign ones (black circles). The circles with white boundaries are those inside confident region.

a B-Spline kernel instead of polynomial, we have the aspect shown in Figure 6. In the case of B-Spline kernel, we can note in Figure 6, a better separation of the feature space compared with the polynomial-kernel shown in Figure 5.

In the Figure 7 we show five ROC curves for five different combinations of the defined tumor's features. According to these curves, the best performance is reached when it is combined area (AR) + heterogeneity (HT) + acoustic shadow (AS). This combination generated an $A_z = 92\%$. All other combinations have resulted in similar performance behavior and an A_z inferior to AR + HT + AS combination. In the next section, we offer a better discussion of these results.

5 Performance Evaluation

In order to test our proposed method we used a 50 pathology-proven cases data base – 20 benign and 30 malignant –. Each case is a sequence of 5 images of the same lesion. Then, we tested 100 images of benign lesion and 150 of malignant ones. Since the detection of a malignant lesion between five images of the same case indicates a malignant case, it is reasonable to consider 250 different cases.

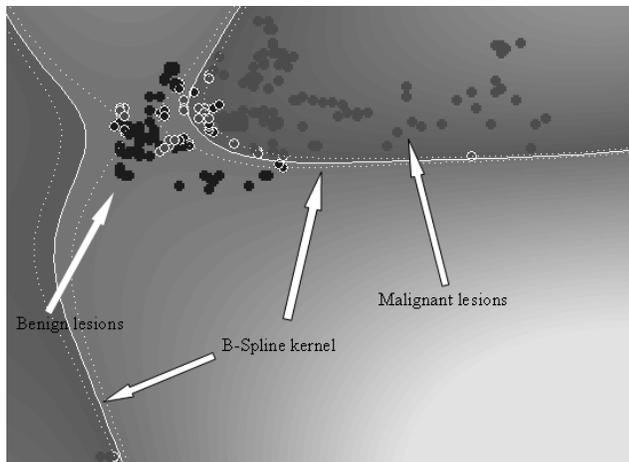


Figure 6. SVM output with a bspline kernel. The white line is the polynome separating malignant tumors (gray circles) from benign ones (black circles). The circles with white boundaries are those inside confident region.

Since our data base is small we have improved the results through a cross-validation method. Then, these ultrasonic images are randomly divided into five groups. We first set the first group as a testing group and use the remaining four groups to train the SVM. After training, the SVM is then tested on the first group. Then, we set the second group as a testing group and the remaining four groups are trained and then the SVM is tested on the second. This process is repeated until all five groups have been set in turn as testing group.

To estimate the performance of the experimental result, five objective indices are used. These indices are accuracy, sensitivity, specificity, positive predictive value and negative predictive value. In our experiment, accuracy of SVM with B-Spline kernel for classification malignancies is 95.2% (238/250), the sensitivity is 97% (97/100), the specificity is 94% (141/150), the positive predictive value is 91.51% (97/106) and the negative predictive value is 97.92% (141/144)².

In Figure 7 we show the complete SVM classification output for our methodology in terms of ROC curves.

²TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative; Accuracy = (TP+TN)/(TP+TN+FP+FN); Sensitivity = TP/(TP+FN); Specificity = TN/(TN+FP); Positive Predictive Value = TP/(TP+FP); and Negative Predictive Value = TN/(TN+FN).

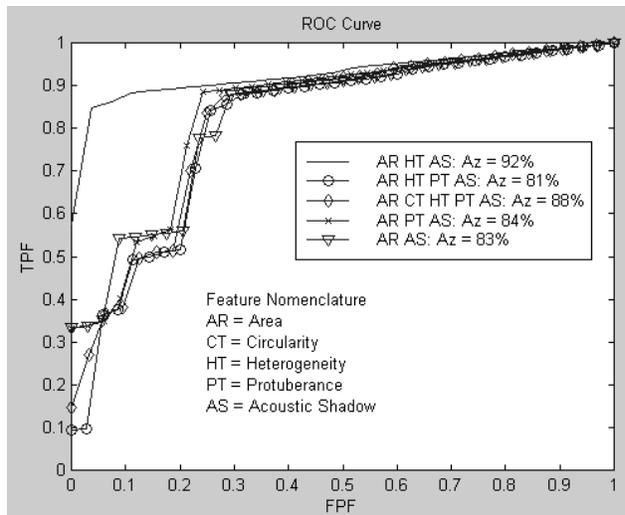


Figure 7. ROC curves for the proposed CAD System for several combinations of tumor features.

6 Conclusions

Geometrical and Textured information from lesion area in ultrasound images provide important discriminant for computer-aided diagnosis systems. Since ultrasound images generally have complex characteristics between pixels it is interesting to study them from the point of view of non-extensive entropy. Since that the traditional Shannon entropy has provided the image segmentation between foreground and background only, it does not guarantee that all regions will be adequately analyzed. Then, our proposed algorithm is a good option to initial segmentation. This NESRA algorithm includes the same result as the traditional entropy, but with the improvement that it is able to isolate intermediate regions, such as the lesion's boundary transitions. This is one of the advantages of the NESRA when applied on this kind of image. A CAD system can decide if the transition area will be included or not in the ROI based on a careful analysis.

On the other hand, the use a level set framework helps to find the correct boundary position without, however, increasing the computational time, but improving the system precision.

Regarding the chosen features, this work does not issue why some feature combination, such as AR + HT + AS, generates better CAD System performance than others. Then, we choose to test empirically which combination is better, which can be seen in the graphic of Figure 7.

Besides the NESRA algorithm, our work proposes a SVM diagnostic system which uses as input a five-dimensional feature space characteristics. These character-

istics are based on geometrical and textured information and should be combined in order to tuning the system. In our experiments, the best combination was achieved when area, heterogeneity and shadow were used only. All these features are affected by the segmentation process. The use of a B-Spline as SVM kernel is one of the main contributions of our paper and one of the main responsible by the system performance (up to 92%), since that this kind of curve allows more flexibility over our feature space, as can be seen in Figure 6. The use of a B-Spline clearly improve the classification as can be seen in Figure 4.

Now, as a future work, we are investigating our proposed approach under ultrasound images of several lesion regions.

Acknowledgment

The authors are grateful to FAPERJ (Fundação de Amparo à Pesquisa no Estado do Rio de Janeiro) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the support of this work.

References

- [1] American Cancer Society, "Cancer Facts and Figures", 2004.
- [2] D. R. Chen and W. J. Kuo and R. F. Chang and W. K. Moon and C. C. Lee, "Use of the bootstrap technique with small training sets for computer-aided diagnosis in breast ultrasound", *Ultrasound Med.*, vol.28, no. 7, pp. 897-902, July 2002.
- [3] R. F. Chang and W. J. Wu and W. K. Moon and D. R. Chen, "Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis", *Ultrasound Med. Biol.*, vol. 29, no. 5, pp. 679-686, 2003.
- [4] D. R. Chen and R. F. Chang and Y. L. Huang, "Computer-aided diagnosis applied to US of solid breast nodules by using neural networks", *Radiology*, vol. 213, no. 2, pp. 407-412, November 1999.
- [5] W. M. Chen *et al.*, "3-D ultrasound texture classification using run difference matrix", *Ultrasound Med. Biol.*, vol. 31, no. 6, pp. 763-770, June 2005.
- [6] T. Pun, "Entropic thresholding: A new approach", *Comput. Graphics Image Process*, vol. 16, pp. 210-239, 1981.
- [7] J. N. Kapur and P. K. Sahoo and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram, *Comput. Graphics Image Process*, vol. 29, pp. 273-285, 1985.
- [8] N. R. Pal, "On minimum cross entropy thresholding", *Pattern Recognition*, vol. 26, pp. 575-580, 1996.
- [9] P. K. Sahoo and S. Soltani and A. K. C. Wong, "A survey of thresholding techniques", *Comput. Vis. Graphics Image Process*, vol. 41, pp. 233-260, 1988.
- [10] C. Tsallis, "Nonextensive Statistical Mechanics and its Applications", *Series Lecture Notes in Physics*, Springer, Berlin, 2001.
- [11] M. P. Albuquerque, M. P. Albuquerque, I. A. Esquef and A. R. G. Mello, "Image thresholding using Tsallis entropy, *Pattern Recognition Letters*, vol. 25, pp. 1059-1065, 2004.
- [12] C. M. Chen and Y. H. Chou and K. C. Han and G. S. Hung and C. M. Tiu and H. J. Chiou and S. Y. Chiou, "Breast Lesion on Sonograms: Computer-Aided Diagnosis with Neural Setting-Independent Features and Artificial Neural networks", *Journal of Radiology*, vol. 226, pp. 504-514, 2003.
- [13] A. Sawaki and K. Shikamoto and H. Satake and T. Ishigaki and S. Koyama and Y. Obata and M. Ikeda, "Breast Ultrasonography: diagnostic efficacy of computer-aided diagnostic system using fuzzy inference, *Radiat. Med.*, vol. 17, no. 1, pp. 41-45, January 1999.
- [14] B. S. Garra and B. H. Krasner and S. C. Horii and *et al*, "Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis", *Ultrasound Imaging*, vol. 15, no. 4, pp. 267-285, 1993.
- [15] R. F. Chang *et al.*, *Computer-aided Diagnosis for 2-D/3-D Breast Ultrasound*, in *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*, Suri, J. S. and Rangayyan, R. M. (eds.) Bellingham, WA: SPIE, 2006.
- [16] D. R. Chen, R. F. Chang, W. M. Chen, and W. K. Moon, "Computer-aided diagnosis for 3-dimensional breast ultrasonography", *Arch. Surg.*, vol. 138, no. 3, pp. 296-302, Mar. 2003.
- [17] C. Shannon and W. Weaver, "The Mathematical Theory of Communication", University of Illinois Press, Urbana, 1948.
- [18] J.S. Suri *et al.* "Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review," *IEEE Trans. Inf. Technol. Biomed.*, vol. 6, no. 1, pp. 8-28, Mar. 2002.
- [19] V. N. Vapnik, "Statistical Learning Theory, Wiley Inter-Science, New York, 1998.