



# Fitness Functions Evaluation for Segmentation of Lymphoma Histological Images Using Genetic Algorithm

Thaína A. A. Tosta<sup>1(✉)</sup>, Paulo Rogério de Faria<sup>2</sup>, Leandro Alves Neves<sup>3</sup>,  
and Marcelo Zanchetta do Nascimento<sup>1,4</sup>

<sup>1</sup> Center of Mathematics, Computing and Cognition, Federal University of ABC,  
Santo André, Brazil

[tosta.thaina@gmail.com](mailto:tosta.thaina@gmail.com)

<sup>2</sup> Department of Histology and Morphology, Institute of Biomedical Science,  
Federal University of Uberlândia, Uberlândia, Brazil

<sup>3</sup> Department of Computer Science and Statistics,

São Paulo State University, São José do Rio Preto, Brazil

<sup>4</sup> Faculty of Computer Science, Federal University of Uberlândia, Uberlândia, Brazil

**Abstract.** For disease monitoring, grade definition and treatments orientation, specialists analyze tissue samples to identify structures of different types of cancer. However, manual analysis is a complex task due to its subjectivity. To help specialists in the identification of regions of interest, segmentation methods are used on histological images obtained by the digitization of tissue samples. Besides, features extracted from these specific regions allow for more objective diagnoses by using classification techniques. In this paper, fitness functions are analyzed for unsupervised segmentation and classification of chronic lymphocytic leukemia and follicular lymphoma images by the identification of their neoplastic cellular nuclei through the genetic algorithm. Qualitative and quantitative analyses allowed the definition of the Rényi entropy as the most adequate for this application. Images classification has reached results of 98.14% through accuracy metric by using this fitness function.

**Keywords:** Nuclear segmentation · Lymphoma histological images  
Genetic algorithm · Fitness function evaluation

## 1 Introduction

Lymphomas are cancers that develop in the cellular components called lymphocytes [1]. Their subtypes are divided into Hodgkin's lymphoma (HL) and non-Hodgkin's lymphoma (NHL), which is responsible for 85% of lymphoma cases [2]. The wide variety of NHL subtypes makes their classification and segmentation complex tasks and a challenge for images analysis.

NHL diagnoses are performed by pathologists who analyze tissue samples stained with hematoxylin-eosin (H&E) [1]. The digitization of these samples allows the application of computational methods that aid specialists in their clinical decisions. Segmentation methods are applied to identify regions of interest (ROIs) that indicate the NHL incidence. The segmentation application allows the extraction of highly specific features from ROIs, which can improve the classification performance [1]. Using these features, classification techniques can aid in diagnoses definition [3].

There are two classes of the NHL that are investigated in literature and addressed by this work: chronic lymphocytic leukemia (CLL) and follicular lymphoma (FL). CLL is a different manifestation type of small lymphocytic lymphoma. It is the most frequent case of leukemia in western countries, with an incidence of 30%. FL is the second most common type of B-cells lymphoma in the classification defined by the world health organization [4].

Some studies propose the segmentation of histological images of these lesions. [5,6] presented segmentation methods of lymphocytes on blood images of CLL with 100× magnification. Both works were evaluated with ROIs localized in the center of the images. In [5], the authors also considered an uniform illumination condition and [6] presented limitations for overlapping cells segmentation.

For FL images, there are many works proposed due to its high incidence rate [4], such as [7–9]. Segmentation of these images is divided into the identification of centroblasts and follicular regions, with some limitations addressed by this work. [9] presented limitations due to the merging of different identified ROIs. This is a common condition, as indicated by [5,8–10]. Moreover, studies as [7,11–13] presented methods that employed images stained with IHC and H&E for FL segmentation. Thus, these methods require these two types of images for application.

Due to empirical definition of threshold values, [14–17] become not sufficiently robust for application on different images. In the studies of [3,9], the authors indicated limitations for processing low magnification images and low quality images resulting from the tissue sample preparation. Using private images datasets, the correlated works can present poor performance when applied on public images that have great variations of contrast and illumination. Thus, it is important to use public images to demonstrate the robustness of new proposals [18].

Thresholding is a major method for efficient segmentation of different histological structures [19]. The optimization method of genetic algorithm (GA) is a powerful technique to define the best threshold values due to its efficiency in complex combinatorial problems. Besides, GA explores its search space in a parallel way with no local convergence of its result [20]. Using this technique, it is possible to efficiently explore the parameters used for the definition of threshold values.

The information used by the GA method in its search for the best solution is the fitness function [21]. This function is responsible for the evaluation of the individuals, playing an essential role in the algorithm by defining the best solutions through its maximization. Since this function defines the algorithm

execution, explore different fitness functions is necessary so the best threshold values of segmentation can be obtained.

This paper presents an algorithm to aid specialists in diagnoses of CLL and FL using segmentation and classification steps. Neoplastic nuclei were segmented through the evaluation of the metrics of Fisher information and the entropies of Rényi, Shannon and Tsallis as GA fitness function for quantifying extracted information. Intensity and texture features were classified by the support vector machine (SVM) method to obtain objective diagnoses. The main contribution of this study is the evaluation of these different quantitative metrics to best associate the image intensity levels to the neoplastic nuclei of CLL and FL histological images. Besides, the evaluation of the proposed method was performed on images from a public domain that are characterized by color variations found in clinical practices.

## 2 Materials and Methods

This section describes the used image dataset, the proposed algorithm, the analyzed fitness functions and the quantitative evaluation metrics.

### 2.1 Images Dataset

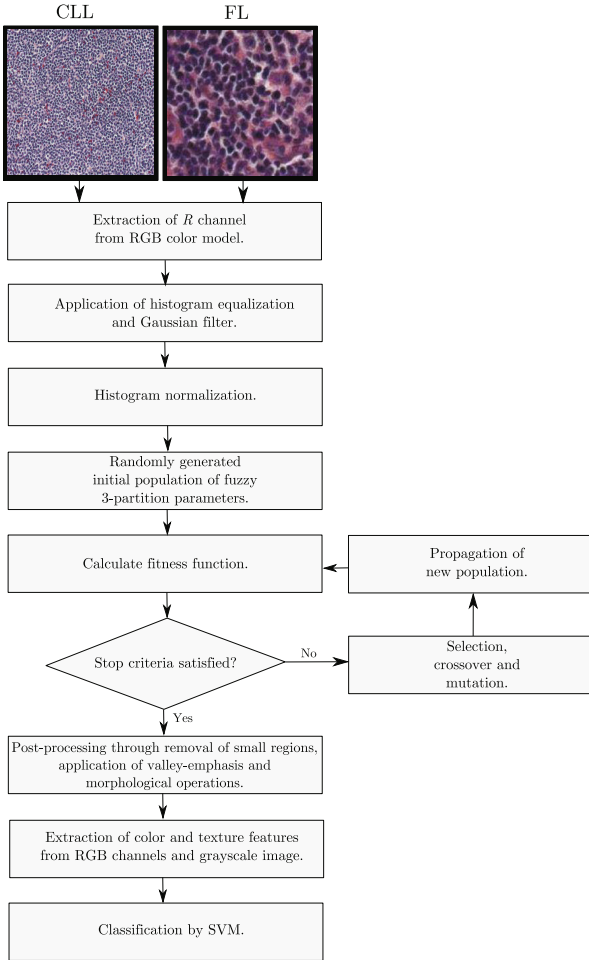
The used lymphoma cases were digitized with a Zeiss AxioScope microscope with white light, objective lens of 20 $\times$  and a CCD AxioCam MR5 color camera. All images were obtained under the same configuration of the used equipment for digitization of histological samples stained with H&E. The resulting images are represented by the RGB color model with 24 bits of quantization, available for download at [22].

The public images used for validation of the proposed method compose a set of 12 and 62 images of CLL and FL, respectively. In both classes, each case has almost 2,000 cells, a quantity close to other studies dedicated to segmentation of histological and cytological images [12,23].

### 2.2 Proposed Algorithm

Figure 1 illustrates the methodological sequence used for development of the proposed method. This work used the MATLAB<sup>®</sup> language for implementation of the proposed method.

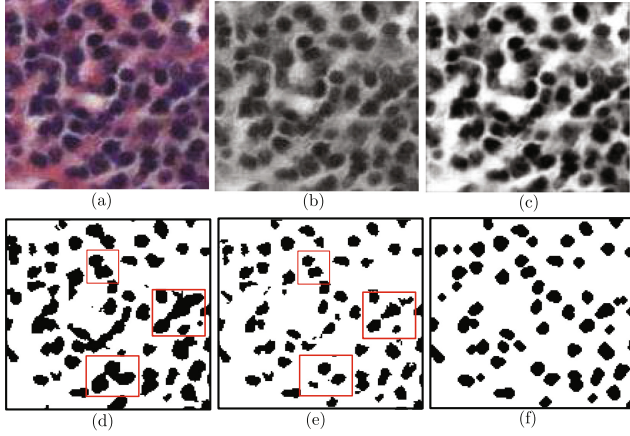
In the preprocessing step, the  $R$  channel from RGB color model was extracted for having the greatest contrast difference in relation to the image background [24]. The histogram equalization technique was applied to deal with variations of contrast and illumination. This method allows the redistribution of image intensity levels, leading to a histogram with an uniform distribution. Thus, the images are now characterized by a greater contrast [25]. Subsequently, the Gaussian filter was used for small noise removal and image smoothing. This method consists of a convolution process with a mask characterized by



**Fig. 1.** The methodological sequence used by the proposed algorithm for segmentation and classification of CLL and FL images.

its elements distribution defined by a 2-dimensional Gaussian function [25]. For its application, the mask size was assigned to  $3 \times 3$  pixels and the  $\sigma$  variable was given by the value 2, which were empirically defined. The results of the channel extraction and application of histogram equalization and Gaussian filter are illustrated by Fig. 2(b) and (c), respectively.

Following, a thresholding method based on a fuzzy 3-partition technique [26] segmented the neoplastic nuclear regions. For this purpose, the  $S$  and  $Z$  functions, represented by Eqs. 1 and 2, define a membership degree of each intensity level ( $k$ ) to the investigated structures: neoplastic nuclei, cytoplasm and background.



**Fig. 2.** Example of application of preprocessing, segmentation and post-processing steps on a subimage of *sj-05-5389-R1\_012* case from FL class: (a) original image, (b) *R* channel from the previous image, (c) resulting image of preprocessing, (d) resulting image of segmentation using GA and Shannon entropy, with more than one nucleus identified as one object indicated by red rectangles, (e) result of valley-emphasis application with indication of corrected regions, and (f) result of morphological operations of dilation and opening.

In this case, it is necessary to compute two threshold values to separate these three regions of the images. So, two pairs of these functions were calculated.

$$S(k, u, v, w) = \begin{cases} 1, & k \leq u \\ 1 - \frac{(k - u)^2}{(w - u) \cdot (v - u)}, & u < k \leq v \\ \frac{(k - w)^2}{(w - u) \cdot (w - v)}, & v < k \leq w \\ 0, & k > w \end{cases} \quad (1)$$

$$Z(k, u, v, w) = 1 - S(k, u, v, w). \quad (2)$$

These functions allow the assignment of pixels to three fuzzy sets, defined by Eqs. 3, 4 and 5, that correspond to membership degrees to neoplastic nuclei, cytoplasm and image background, respectively.

$$M_n(k) = S(k, u_1, v_1, w_1), \quad (3)$$

$$M_c(k) = \begin{cases} Z(k, u_1, v_1, w_1), & k \leq w_1, \\ S(k, u_2, v_2, w_2), & k > w_1, \end{cases} \quad (4)$$

$$M_b(k) = Z(k, u_2, v_2, w_2), \quad (5)$$

where,  $u_1, v_1, w_1, u_2, v_2$  and  $w_2$  ( $0 \leq u_1 < v_1 < w_1 < u_2 < v_2 < w_2 \leq 255$ ) are parameters that determine the membership degrees distribution. The threshold

values ( $t_1$  and  $t_2$ ) are defined by the intersection points of  $M_n$  and  $M_c$ , and  $M_c$  and  $M_b$ .

The GA algorithm was used to assign intensity values to  $u_1, v_1, w_1, u_2, v_2$  and  $w_2$  with an initial random generation of 60 possible solutions coded by six values using the preprocessed image normalized histogram. Each individual was evaluated by a fitness function. The Fisher Information metric (FIM), defined by [27], was one of the evaluated GA fitness functions ( $H$ ):

$$H(u_1, v_1, w_1, u_2, v_2, w_2) = w_n \cdot I_n + w_c \cdot I_c + w_b \cdot I_b, \quad (6)$$

$$I_n = \frac{1}{w_n} \sum_{i=1}^{t_1} \frac{(P_{i+1} - P_i)^2}{P_i}, \quad (7)$$

$$I_c = \frac{1}{w_c} \sum_{i=t_1+1}^{t_2} \frac{(P_{i+1} - P_i)^2}{P_i},$$

$$I_b = \frac{1}{w_b} \sum_{i=t_2+1}^{255} \frac{(P_{i+1} - P_i)^2}{P_i},$$

where,  $w_i$  represents the probabilities of intensity levels of the investigated structures (neoplastic nuclei, cytoplasm and background):

$$w_n = \sum_{i=0}^{t_1} P_i, \quad w_c = \sum_{i=t_1+1}^{t_2} P_i, \quad w_b = \sum_{i=t_2+1}^{255} P_i, \quad (8)$$

where,  $P_i$  is the probability of the intensity level  $i$ , i.e. the number of pixels with intensity  $i$  divided by the total number of pixels in the image.

The Rényi entropy metric (REM) was also used, as expressed by [28]:

$$\begin{aligned} H(u_1, v_1, w_1, u_2, v_2, w_2) = & \frac{1}{1-\alpha} \left[ \ln \sum_{i=0}^{t_1} \left( \frac{P_i}{P_n} \right)^\alpha \right] \\ & + \frac{1}{1-\alpha} \left[ \ln \sum_{i=t_1+1}^{t_2} \left( \frac{P_i}{P_c} \right)^\alpha \right] \\ & + \frac{1}{1-\alpha} \left[ \ln \sum_{i=t_2+1}^{255} \left( \frac{P_i}{P_b} \right)^\alpha \right], \end{aligned} \quad (9)$$

with the empirical assignment of 2 to  $\alpha$  and the probabilities  $P_n, P_c$  and  $P_b$  representing the probabilities of each investigated structure:

$$P_j = \sum_{k=0}^{255} h(k) \cdot M_j(k), \quad (10)$$

where,  $j = \{n, c, b\}$ ,  $M_j(k)$  corresponds to the fuzzy sets defined by Eqs. 3, 4 and 5 and  $h(\cdot)$  represents the image normalized histogram.

The Shannon entropy metric (SEM) [29] also evaluated the values of  $u_1$ ,  $v_1$ ,  $w_1$ ,  $u_2$ ,  $v_2$  and  $w_2$ , defined by:

$$H(u_1, v_1, w_1, u_2, v_2, w_2) = -P_n \cdot \log(P_n) - P_c \cdot \log(P_c) - P_b \cdot \log(P_b). \quad (11)$$

The Tsallis entropy metric (TEM) was also used, defined by Eqs. 12 and 13 [30]:

$$\begin{aligned} H(u_1, v_1, w_1, u_2, v_2, w_2) &= S_q^n + S_q^c + S_q^b \\ &+ (1 - q) \cdot (S_q^n \cdot S_q^c + S_q^n \cdot S_q^b + S_q^c \cdot S_q^b) \\ &+ (1 - q)^2 \cdot S_q^n \cdot S_q^c \cdot S_q^b, \end{aligned} \quad (12)$$

$$S_q^j = \frac{1 - \sum_{i=lb}^{ub} (P_i^j)^q}{q - 1}, \quad (13)$$

where, the constant  $q$  was empirically assigned to 10 and it represents an index that denotes the degree of nonextensivity.  $P_i^j$  represents the normalization of the probabilities of each investigated structure ( $j = \{n, c, b\}$ ) defined by the intensity values between its lower bound ( $lb$ ) and upper bound ( $ub$ ):

$$P^n = \frac{P_1, \dots, P_{t_1}}{w_n}, \quad P^c = \frac{P_{t_1+1}, \dots, P_{t_2}}{w_c}, \quad P^b = \frac{P_{t_2+1}, \dots, P_{255}}{w_b}. \quad (14)$$

Then, the GA used an elitism process in which only 30% of the individuals that have reached the best results of the fitness function were preserved in subsequent iterations. For the population to have 60 individuals again, the crossover step was applied selecting individuals for combination through a crossover probability of 0.65. The crossover point, that defines the combination point between two individuals, was randomly obtained in the interval [1, 6], corresponding to the parameters of the fuzzy 3-partition technique. The populational diversity was then guaranteed by the mutation step. This process randomly changes some parameters of the individuals through the definition of a mutation probability equal to 0.01.

The termination condition of the GA method was based on the approach of [31]. In this method, if the average of intensity levels of each identified structure was the same over two consecutive iterations, the algorithm execution is interrupted. The segmentation result is exemplified by Fig. 2(d), where it is noticeable some limitations.

To refine the segmentation results, a post-processing step was necessary. Initially, segmented regions with areas smaller than 10 pixels were removed due to their correspondence to false positive regions. The segmentation was not able to individually identify some neoplastic nuclei. These regions, highlighted by red rectangles in Fig. 2(d), were characterized by inter-nuclear regions with intensity levels brighter than the ROIs. Thus, the valley-emphasis method [32] was locally

applied on segmented regions with areas bigger than 80 pixels, an indicative condition of representation of more than one segmented nucleus as one object. This method defined as its threshold value ( $T$ ) the valley regions of the preprocessed image histogram to identify neoplastic nuclei and inter-nuclear regions. Its application is based on Eq. 15:

$$T = \max\{(1 - p_t) \cdot (\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t))\}, \quad (15)$$

where,  $\omega$  corresponds to the probability of neoplastic nuclei and inter-nuclear regions and  $\mu$  represents the average of the intensity levels of these regions. The result of this application is presented in Fig. 2(e), where regions in red rectangles can be compared with their correspondences in the segmentation result.

Finally, the morphological operations of opening and dilation were applied. The opening allows to increase space between objects, smooth contours and remove small noises, meanwhile, the dilation increases the objects areas and fills small holes [25]. The structuring elements of opening and dilation operations, in this study, had disk and square distributions, respectively, with their parameters assigned to the value 2. The result of these operations is illustrated in Fig. 2(f), where it is possible to note a better representation of nuclear contours and their internal regions.

Using the segmented regions, intensity features (mean, median, standard deviation, kurtosis, skewness, variance, 1-norm, 2-norm and entropy) were extracted from the  $R$ ,  $G$  and  $B$  channels and grayscale images, as explored in [33]. Measures of median, standard deviation, entropy and energy were also extracted from the diagonal, horizontal and vertical sub-bands of Daubechies4 wavelet, which has a perfect reconstruction compared to other wavelet types [34], composing a texture descriptor of the image [35]. The classification step was performed through the SVM method with the radial basis function and the cross-validation method with 10 folds.

### 2.3 Evaluation Metrics

The images manually segmented by a pathologist were used to evaluate the results. It was possible to obtain the measures of accuracy ( $Ac$ ), sensitivity ( $Se$ ) and specificity ( $Sp$ ) by analyzing the pixels identified by both segmentations. These pixels were divided into four possible classes: true positive ( $T_P$ ), corresponding to correct identifications, true negative ( $T_N$ ) that represents the pixels correctly undetected, false positive ( $F_P$ ) that presents the amount of incorrectly detected pixels, and false negative ( $F_N$ ) that denotes incorrectly undetected pixels [36]. These metrics are expressed by:

$$Ac = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, \quad (16)$$

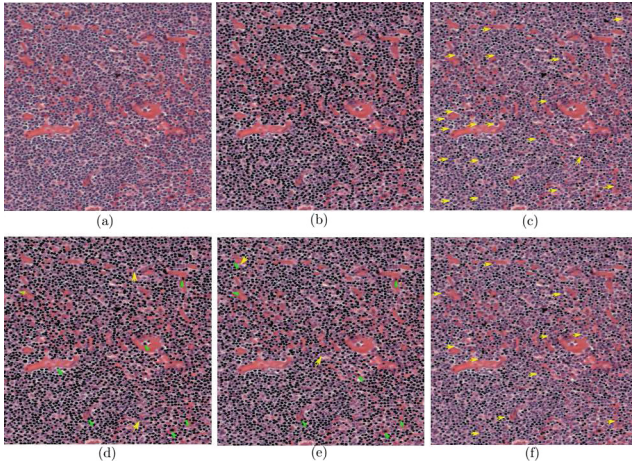
$$Se = \frac{T_P}{T_P + F_N}, \quad Sp = \frac{T_N}{T_N + F_P}. \quad (17)$$



These metrics were also used for classification evaluation. High values of Eqs. 16 and 17 represent better results of segmentation and classification. For classification evaluation, the metric of concordance probability ( $C_p$ ), defined by the product between the sensitivity and specificity, was also used for expressing the area of a rectangle associated with the receiver operating characteristic curve [37,38]. In this way, the higher the rectangle area, the better the classification.

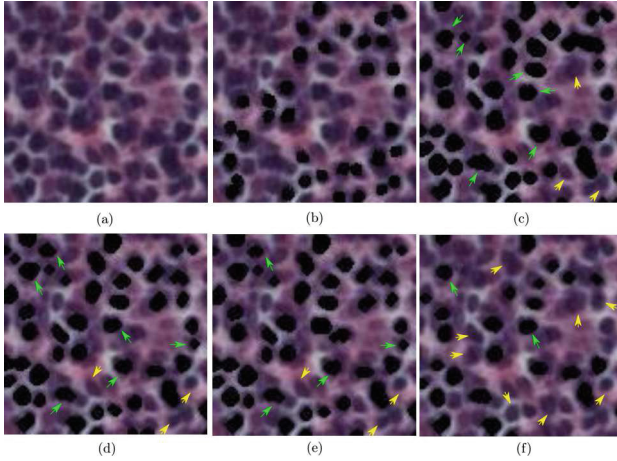
### 3 Results and Discussion

Figures 3 and 4 exemplify the results of CLL and FL images segmentation, respectively, where yellow arrows were used to indicate some false negative regions and green arrows for some false positive identifications. Application of the Fisher information (Figs. 3(c) and 4(c)) and the entropy of Tsallis (Figs. 3(f) and 4(f)) presented an expressive amount of false negative regions, mainly in the CLL lesion. Due to their similar amounts of false positive and false negative regions identified and the fewer number of over or under segmentation regions, the results achieved with the Rényi and Shannon entropies are the most adequate.



**Fig. 3.** Application of the proposed method on a subimage of *sj-03-852-R2.012* case from CLL: original image (a), manually segmented image (b), segmentation result using the Fisher information (c) and the entropies of Rényi (d), Shannon (e) and Tsallis (f).

The accuracy, sensitivity and specificity results of segmentation are presented by Table 1 by the metrics of mean and standard deviation of five executions of the proposed segmentation algorithm. For both lesions, the best results of accuracy and specificity were obtained by the Tsallis entropy. However, this function has reached the worst results of sensitivity. In comparison to the best



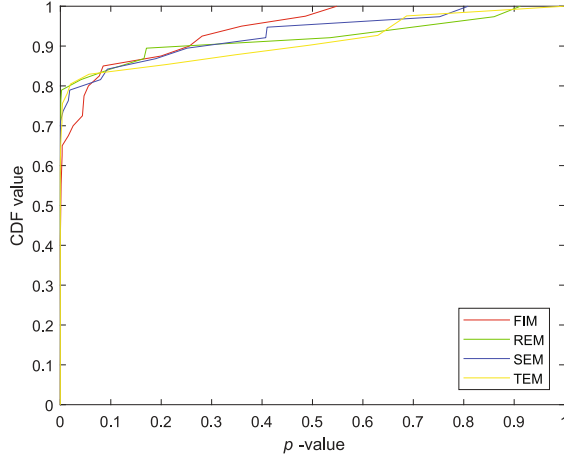
**Fig. 4.** Application of the proposed method on a subimage of *sj-05-5389-R1-012* case from FL class: original image (a), manually segmented image (b), segmentation result using the Fisher information (c) and the entropies of Rényi (d), Shannon (e) and Tsallis (f).

**Table 1.** Mean and standard deviation results of five executions of CLL and FL segmentations using the fitness functions of FIM, REM, SEM and TEM through accuracy (%), sensitivity (%) and specificity (%) metrics.

Fitness functions	CLL segmentation			FL segmentation		
	Ac	Se	Sp	Ac	Se	Sp
FIM	79.42 (0.26)	<b>45.22</b> <b>(1.08)</b>	87.97 (0.58)	81.82 (0.30)	<b>55.55</b> <b>(0.96)</b>	85.64 (0.46)
REM	80.42 (0.14)	42.48 (1.04)	88.15 (0.60)	82.39 (0.25)	54.35 (1.93)	85.31 (2.38)
SEM	80.55 (0.22)	43.00 (0.79)	88.10 (0.19)	82.78 (0.26)	53.24 (0.78)	86.97 (0.32)
TEM	<b>82.82</b> <b>(0.33)</b>	17.34 (1.27)	<b>95.79</b> <b>(0.43)</b>	<b>86.72</b> <b>(0.09)</b>	22.68 (0.92)	<b>95.57</b> <b>(0.17)</b>

results of this metric, differences of 27.88% and 32.87% were obtained for the CLL and FL lesions, respectively. Despite its good sensitivity results, the Fisher information had a poor performance in the proposed segmentation, as illustrated by the Figs. 3 and 4. The analysis of the obtained results points out the good performance of Rényi entropy. In comparison with the best results of accuracy, sensitivity and specificity, this function has obtained differences of 2.40%, 2.74% and 7.64% for CLL segmentation and 4.33%, 1.20% and 10.26% for the FL lesion. Through this analysis, it is possible to define the Rényi entropy as the most adequate fitness function for the proposed algorithm.

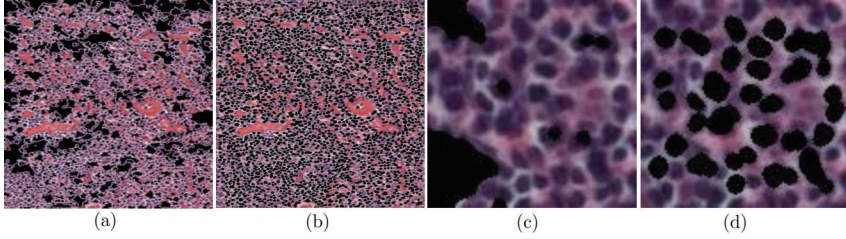
The Wilcoxon rank sum test was performed with a significance level of 5% for evaluation of the color and texture features obtained by the different fitness functions used in the segmentation step. The null hypothesis is that the features from CLL and FL segmented nuclei compose samples from continuous distributions with equal medians. As it can be seen on Fig. 5, all investigated functions have presented a number of relevant features higher than 50%.



**Fig. 5.** Wilcoxon rank sum test for CLL×FL classification using GA segmentation with Fisher information and entropies of Rényi, Shannon and Tsallis.

The nuclear segmentation methods for H&E images of [39,40] were used for comparison with the proposed segmentation technique. The segmentations proposed by these studies do not use any search method, like the GA. The study of [39] presented a segmentation method of cellular nuclei with the extraction of the  $R$  channel from RGB color model and the application of the Gaussian filter and the morphological operations of erosion and dilation in the preprocessing step. In the segmentation, this proposal used the radial symmetry transform and the Otsu thresholding to reduce the over-segmentation of the watershed algorithm applied. The post-processing was composed of the removal of false positive regions by morphological operations. [40] initially used a deconvolution method for application of the opening reconstruction on the hematoxylin channel obtained. The segmentation of cellular nuclei was performed by a multilevel thresholding based on the Otsu method. Finally, the morphological operations of opening and hole filling were applied in addition to the removal of false positive regions by its area.

As illustrated by Fig. 6, the results of [39] are not satisfactory for this application. For the CLL segmentation, this technique has reached 77.07%, 21.01% and 88.10% of accuracy, sensitivity and specificity, respectively. In the FL segmentation, the results of accuracy, sensitivity and specificity of [39] were



**Fig. 6.** Segmentation results of [39](a)(c) and [40](b)(d) applied to CLL and FL images.

79.38%, 25.66% and 86.78%, respectively. [40] obtained better results but with more false positive regions, as indicated by its specificity, with values of 72.57% and 75.30% for CLL and FL images, respectively. In combination with sensitivity of 70.39% for CLL and 69.95% for FL, lower accuracy values were obtained: 71.97% and 74.56% for CLL and FL, respectively.

Table 2 presents the quantitative results of CLL $\times$ FL classification. Among the evaluated fitness functions, the Rényi entropy has reached the best results in the metrics of accuracy and specificity. The worst results were reached by the Fisher information and intermediate results were obtained by the entropies of Shannon and Tsallis. Through the concordance probability metric, it is noticeable the relevant performance of the Rényi entropy applied in the segmentation, which allowed to obtain the most discriminant features between CLL and FL neoplastic nuclei.

**Table 2.** Results of CLL $\times$ FL classification using the fitness functions of FIM, REM, SEM and TEM through accuracy (%), sensitivity (%) and specificity (%) metrics computed by the 10-fold cross-validation, and the concordance probability.

Fitness functions	Classification			
	Ac	Se	Sp	Cp
FIM	90.54	<b>100.00</b>	41.67	0.4167
REM	<b>98.14</b>	99.33	<b>83.33</b>	<b>0.8277</b>
SEM	91.89	<b>100.00</b>	50.00	0.5000
TEM	93.24	<b>100.00</b>	58.33	0.5833

Table 3 presents the classification results of  $T_P$ ,  $F_P$ ,  $T_N$  and  $F_N$ , with a high rate of true positive samples identified by the analyzed fitness functions. Considering false positive and false negative rates, the Rényi and Tsallis entropies have reached the highest values of these metrics, respectively. This indicates that the regions identified by these entropies contribute to misclassification, demanding a detection step to remove false positive and false negative regions segmented.

**Table 3.** Confusion matrix of the proposed classification computed by the 10-fold cross-validation using different GA fitness functions for neoplastic nuclei segmentation.

Folds	FIM				REM				SEM				TEM			
	$T_P$	$F_P$	$T_N$	$F_N$	$T_P$	$F_P$	$T_N$	$F_N$	$T_P$	$F_P$	$T_N$	$F_N$	$T_P$	$F_P$	$T_N$	$F_N$
Fold 1	6	1	0	0	6	1	0	0	7	1	0	0	7	0	0	1
Fold 2	12	1	0	1	12	1	0	1	13	1	0	1	13	1	0	1
Fold 3	18	1	0	2	18	2	0	1	19	1	0	2	19	1	0	2
Fold 4	24	1	0	3	25	2	0	2	25	2	0	2	25	2	0	2
Fold 5	30	1	0	4	31	4	0	2	32	2	0	3	31	4	0	2
Fold 6	37	2	0	4	37	6	0	2	38	3	0	3	37	4	0	3
Fold 7	43	3	0	4	44	6	0	3	44	4	0	4	43	4	0	4
Fold 8	49	4	0	4	50	7	0	3	50	5	0	5	49	4	0	5
Fold 9	56	5	0	5	56	8	0	3	56	5	0	6	56	5	0	6
Fold 10	62	7	0	5	62	8	0	4	62	6	0	6	62	5	0	7

## 4 Conclusion

To define diagnoses of different types of cancer, the identification of indicative regions of these diseases in histological images is an essential task. When manually performed by pathologists, this task becomes time-consuming and subjective. In order to obtain objective diagnoses, segmentation and classification techniques are able to aid specialists.

In this study, an unsupervised segmentation algorithm of neoplastic nuclei of CLL and FL was presented with an evaluation of different GA fitness functions. Experimental results defined the Rényi entropy as the most adequate for this application. In comparison to Fisher information and the entropies of Shannon and Tsallis, the Rényi entropy was the one that best quantitatively evaluated the membership degrees of the intensity levels, which allowed to obtain the most adequate threshold values for the segmentation of CLL and FL neoplastic nuclei. Through intensity and texture features extracted from the segmented regions, it was possible to classify the images using the SVM method. With 98.14% of accuracy in the images classification, the proposed algorithm indicates the importance of the segmentation step for definition of diagnoses of CLL and FL.

A limitation of the proposed method was the false positive and false negative regions identified in the segmentation. Thus, in future works, additional features to the intensity information used in the segmentation will be explored to correctly identify the ROIs of CLL and FL images.

**Acknowledgments.** T.A.A.T. and M.Z.N. thank to CAPES (1575210) and FAPEMIG (TEC - APQ-02885-15 project) for financial support.

## References

1. Orlov, N.V., Chen, W.W., Eckley, D.M., Macura, T.J., Shamir, L., Jaffe, E.S., Goldberg, I.G.: Automatic classification of lymphoma images with transform-based global features. *IEEE Trans. Inf. Technol. Biomed.* **14**(4), 1003–1013 (2010)
2. Lowry, L., Linch, D.: *Non-Hodgkin's lymphoma* (2013)
3. Belkacem-Boussaid, K., Samsi, S., Lozanski, G., Gurcan, M.N.: Automatic detection of follicular regions in H&E images using iterative shape index. *Comput. Med. Imaging Graph.* **35**(7), 592–602 (2011)
4. Canellos, G.P., Lister, T.A., Young, B.: *The Lymphomas*, 2nd edn. Saunders Elsevier, Philadelphia (2006)
5. Mohammed, E.A., Far, B.H., Naugler, C., Mohamed, M.M.A.: Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding. In: 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1–5. IEEE (2013)
6. Mohammed, E.A., Far, B.H., Naugler, C., Mohamed, M.M.A.: Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation. In: 15th International Conference on e-Health Networking, Applications and Services. IEEE (2013)
7. Sertel, O., Kong, J., Lozanski, G., Catalyurek, U., Saltz, J.H., Gurcan, M.N.: Computerized microscopic image analysis of follicular lymphoma. In: *Medical Imaging*, vol. 6915. International Society for Optics and Photonics (2008)
8. Kong, H., Belkacem-Boussaid, K., Gurcan, M.: Cell nuclei segmentation for histopathological image analysis. In: *SPIE Medical Imaging*, p. 79622R. International Society for Optics and Photonics (2011)
9. Belkacem-Boussaid, K., Prescott, J., Lozanski, G., Gurcan, M.N.: Segmentation of follicular regions on H&E slides using a matching filter and active contour model. In: *SPIE Medical Imaging. International Society for Optics and Photonics* (2010)
10. Oztan, B., Kong, H., Gurcan, M.N., Yener, B.: Follicular lymphoma grading using cell-graphs and multi-scale feature analysis. In: *SPIE Medical Imaging*, vol. 8315 (2012)
11. Oger, M., Belhomme, P., Gurcan, M.N.: A general framework for the segmentation of follicular lymphoma virtual slides. *Comput. Med. Imaging Graph.* **36**(6), 442–451 (2012)
12. Dimitropoulos, K., Barmpoutis, P., Koletsa, T., Kostopoulos, I., Grammalidis, N.: Automated detection and classification of nuclei in pax5 and h&e-stained tissue sections of follicular lymphoma. *Signal Image Video Process.* 1–9 (2016)
13. Dimitropoulos, K., Barmpoutis, P., Koletsa, T., Kostopoulos, I., Grammalidis, N.: Classification of nuclei in follicular lymphoma tissue sections using different stains and bayesian networks. In: Kyriacou, E., Christofides, S., Pattichis, C.S. (eds.) *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*. IP, vol. 57, pp. 234–238. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-32703-7\\_47](https://doi.org/10.1007/978-3-319-32703-7_47)
14. Dimitropoulos, K., Michail, E., Koletsa, T., Kostopoulos, I., Grammalidis, N.: Using adaptive neuro-fuzzy inference systems for the detection of centroblasts in microscopic images of follicular lymphoma. *Signal Image Video Process.* **8**(1), 33–40 (2014)
15. Luo, Y., Celenk, M., Bejai, P.: Discrimination of malignant lymphomas and leukemia using radon transform based-higher order spectra. In: *Medical Imaging*, pp. 61445K-1–61445K-10. International Society for Optics and Photonics (2006)

16. Sertel, O., Kong, J., Lozanski, G., Catalyurek, U., Saltz, J.H., Gurcan, M.N.: Histopathological image analysis using model-based intermediate representations and color texture: follicular lymphoma grading. *J. Signal Process. Syst.* **55**(1–3), 169–183 (2009)
17. Sertel, O., Kong, J., Lozanski, G., Shana'ah, A., Catalyurek, U., Saltz, J., Gurcan, M.: Texture classification using nonlinear color quantization: application to histopathological image analysis. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 597–600. IEEE (2008)
18. McCann, M.T., Ozolek, J.A., Castro, C.A., Parvin, B., Kovacevic, J.: Automated histology analysis: opportunities for signal processing. *Signal Process. Mag.* **32**(1), 78–87 (2015)
19. Oswal, V., Belle, A., Diegelmann, R., Najarian, K.: An entropy-based automated cell nuclei segmentation and quantification: application in analysis of wound healing process. *Comput. Math. Meth. Med.* **2013**, 1–10 (2013)
20. Tang, L., Tian, L., Steward, B.L.: Color image segmentation with genetic algorithm for in-field weed sensing. *Trans. ASAE* **43**(4), 1019 (2000)
21. Maulik, U.: Medical image segmentation using genetic algorithms. *IEEE Trans. Inf. Technol. Biomed.* **13**(2), 166–173 (2009)
22. Shamir, L., Orlov, N., Eckley, D.M., Macura, T.J., Goldberg, I.G.: IICBU 2008: a proposed benchmark suite for biological image analysis. *Med. Biol. Eng. Comput.* **46**(9), 943–947 (2008)
23. Wang, P., Hu, X., Li, Y., Liu, Q., Zhu, X.: Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Sig. Process.* **122**, 1–13 (2016)
24. Zorman, M., Kokol, P., Lenic, M., Rosa, J.L.S., Sigut, J.F., Alayon, S.: Symbol-based machine learning approach for supervised segmentation of follicular lymphoma images. In: *20th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 115–120. IEEE (2007)
25. Gonzalez, R.C., Woods, R.E.: *Processamento de Imagens Digitais*. Edgard Blucher, São Paulo (2000)
26. Yin, S., Zhao, X., Wang, W., Gong, M.: Efficient multilevel image segmentation through fuzzy entropy maximization and graph cut optimization. *Pattern Recogn.* **47**(9), 2894–2907 (2014)
27. Abo-Eleneen, Z., Abdel-Azim, G.: A novel algorithm for image thresholding using non-parametric fisher information. In: *International Electronic Conference on Entropy and Its Applications*, vol. 1, p. 15. Multidisciplinary Digital Publishing Institute (2014)
28. Sarkar, S., Das, S., Chaudhuri, S.S.: Hyper-spectral image segmentation using rényi entropy based multi-level thresholding aided with differential evolution. *Expert Syst. Appl.* **50**, 120–129 (2016)
29. Shannon, C.E.: A mathematical theory of communication. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **5**(1), 3–55 (2001)
30. Zhang, Y., Wu, L.: Optimal multi-level thresholding based on maximum tsallis entropy via an artificial bee colony approach. *Entropy* **13**(4), 841–859 (2011)
31. Hammouche, K., Diaf, M., Siarry, P.: A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation. *Comput. Vis. Image Underst.* **109**(2), 163–175 (2008)
32. Ng, H.: Automatic thresholding for defect detection. *Pattern Recogn. Lett.* **27**(14), 1644–1649 (2006)

33. Dong, F., Irshad, H., Oh, E., Lerwill, M.F., Brachtel, E.F., Jones, N.C., Knoblauch, N.W., Montaser-Kouhsari, L., Johnson, N.B., Rao, L.K., et al.: Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. *PLoS ONE* **9**(12), e114885 (2014)
34. Abbas, Q.: Segmentation of differential structures on computed tomography images for diagnosis lung-related diseases. *Biomed. Signal Process. Control* **33**, 325–334 (2017)
35. Semler, L., Dettori, L., Furst, J.: Wavelet-based texture classification of tissues in computed tomography. In: *IEEE Symposium on Computer-Based Medical Systems*, pp. 265–270. IEEE (2005)
36. Chang, V., Saavedra, J.M., Castañeda, V., Sarabia, L., Hitschfeld, N., Härtel, S.: Gold-standard and improved framework for sperm head segmentation. *Comput. Methods Programs Biomed.* **117**(2), 225–237 (2014)
37. Unal, I.: Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput. Math. Meth. Med.* **2017**, 1–14 (2017)
38. Liu, X.: Classification accuracy and cut point selection. *Stat. Med.* **31**(23), 2676–2686 (2012)
39. Vahadane, A., Sethi, A.: Towards generalized nuclear segmentation in histological images. In: *International Conference Bioinformatics and Bioengineering*, pp. 1–4. IEEE (2013)
40. Phoulady, H.A., Goldgof, D.B., Hall, L.O., Mouton, P.R.: Nucleus segmentation in histology images with hierarchical multilevel thresholding. In: *SPIE Medical Imaging*, p. 979111. International Society for Optics and Photonics (2016)