

Using wavelet sub-band and fuzzy 2-partition entropy to segment chronic lymphocytic leukemia images

Thaína A. Azevedo Tosta^{a,*}, Paulo Rogério Faria^b, Valério Ramos Batista^a,
Leandro Alves Neves^c, Marcelo Zanchetta do Nascimento^{a,d}

^a Federal University of ABC, Center of Mathematics, Computer Science and Cognition, Av. dos Estados, 5001, 09210-580 Santo André, São Paulo, Brazil

^b Federal University of Uberlândia, Department of Histology and Morphology, Institute of Biomedical Science, Av. Amazonas, S/N, 38405-320 Uberlândia, Minas Gerais, Brazil

^c São Paulo State University (UNESP), Department of Computer Science and Statistics, R. Cristóvão Colombo, 2265, 15054-000 São José do Rio Preto, São Paulo, Brazil

^d Federal University of Uberlândia, Faculty of Computer Science, Av. João Naves de Ávila, 2121, 38400-902 Uberlândia, Minas Gerais, Brazil

ARTICLE INFO

Article history:

Received 22 October 2016

Received in revised form

20 September 2017

Accepted 22 November 2017

Available online 5 December 2017

Keywords:

Nuclei segmentation

Genetic algorithm

Wavelet transform

H&E-stained histological images

Chronic lymphocytic leukemia

ABSTRACT

Histological images analysis is an important procedure to diagnose different types of cancer. One of them is the chronic lymphocytic leukemia (CLL), which can be identified by applying image segmentation techniques. This study presents an unsupervised method to segment neoplastic nuclei in CLL images. Firstly, deconvolution, histogram equalization and mean filter were applied to enhance nuclear regions. Then, a segmentation technique based on a combination of wavelet transform, fuzzy 2-partition entropy and genetic algorithm was used, followed by removal of false positive regions, and application of valley-emphasis and morphological operations. In order to evaluate the proposed algorithm H&E-stained histological images were used. In the accuracy metric, the proposed method attained more than 80%, which can surpass similar methods. This proposal presents spatial distribution that has a good consistency with a manual segmentation and lower overlapping rate than other techniques in the literature.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Chronic lymphocytic leukemia (CLL) is a kind of blood cancer among the most frequent ones in western countries, where its incidence is about 25% of the total cases of leukemia in adults [1]. Estimates show that this neoplasia had over 14 thousand new cases and caused over 4 thousand deaths in 2015 [2]. Patients diagnosed with CLL are, on average, 70 years old and rarely under 40. This leukemia affects two men for each woman, and one of its most relevant risk factors is the family history. Approximately 20% of the patients with this illness have first-degree relatives also diagnosed with CLL, which has increased in three times the risk of developing it [3].

The diagnosis of this neoplasia requires histological sections analysis of either lymph nodes or blood. Such analysis is performed by a specialist in identifying cancer cells. This is crucial for the patient, since only a correct diagnosis can make it possible an adequate medical follow-up, a right identification of the illness degree,

and also the best guidance to efficient treatments [4]. However, analyses of specialists are partly subjective, highly time-consuming and also influenced by the fact that leukemia has many variations. This all makes diagnosing a complex task [5].

When digitized, histological images enable researches on computational techniques applied to improve diagnoses and prognoses. One of the essential tasks in histopathology is structures identification, which in computer science corresponds to image segmentation. Since CLL is a lymphoproliferative illness, methods that identify its neoplastic cells try to segment specific components called lymphocytes [4]. These are the second most populous type of white blood cells that take part in the immunological defence of the organism [6]. Visualizing these structures by the naked eye is a complex task [7]. Therefore, computational image analysis is able to help in the identification of these regions of interest (ROI) with precision, which is essential to correlate them with this pathology [8]. In order to overcome this general problem, several studies have been developed to integrate digital processing techniques with histological images.

Mohammed et al. [9] proposed a segmentation method of lymphocytes to identify their nucleus, cytoplasm and cell regions. Initially, the images were converted to grey levels and Otsu's

* Corresponding author.

E-mail address: tosta.thaina@gmail.com (T.A. Azevedo Tosta).

method was applied to segment nuclear regions. Still in this step, the canny edge detector technique and also the morphological operations of dilation, hole-filling and erosion were applied. For cellular segmentation, the same previous methods were used but then added the watershed algorithm, with a removal of 1% of its local minima to reduce over-segmentation effects, as proposed in [10]. Finally, cytoplasm segmentation was done by subtracting the results obtained in the previous steps. Applied to 132 images, this method achieved results of accuracy close to 100% in all the proposed segmentations.

In a second study of Mohammed et al. [11], a method based on support vector machine (SVM) algorithm was proposed in order to segment lymphocytes. Initially, Otsu's method was used to segment nuclear regions from the image background. This was done to a training set of images. Afterwards, from that previous segmentation, colour features from RGB colour channels were extracted. Then, the SVM used these features to classify pixels as nuclear or non-nuclear. In order to reduce the training set, the k-means algorithm was applied, as indicated by [12,13]. For cellular segmentation, the same methods were applied. Cellular regions were manually identified for the classifier training. Finally, cytoplasm segmentation was carried out as described in [9]. The accuracy of this system attained the average of 95.56% among the proposed segmentations, considering a specialist's segmentation of 440 cell images.

In the specific case of segmenting CLL images, there are still few articles in the literature. Nowadays, the only available studies are [9,11], considering that they treat limitations observed in previous works. Both [9,11] are applied to Giemsa-stained blood images with 100× magnification, which gives more information for effective analysis of complex features [14]. However, this proposed method is applied to images with 20× magnification obtained from lymph node biopsies, without similar studies in the literature.

This article presents a nuclear segmentation method of CLL neoplastic cells that overcomes relevant limitations of [9,11]. In [9], some histological sections were not used for containing lymphocytes too close to red blood cells, which negatively affected their system execution. In real clinical image databases, as the one used in this proposed study, images with that characteristic are present, which limits the application of [9] in medical practices. Moreover, both [9,11] are devoted to segment normal and neoplastic structures of blood images, but not of images obtained from lymph node biopsies, as in this proposal.

The present article identifies nuclear regions of neoplastic cells with a segmentation method applied to lymph node images with great differences of contrast and illumination. This method successfully handled them without any user's intervention, in contrast with [11], in which they needed supervised markings by a specialist in cellular segmentation. Regarding histological images with these differences, the application of Otsu's method as proposed in [9] is inadequate [15]. This fact raises the need of another segmentation technique to be applied to CLL images, or even histological images in general. For this purpose, evolutionary algorithms are still little explored but represent an attractive proposal since they can improve both robustness and processing time [16]. Combined with genetic algorithm (GA) technique, the fuzzy method contributes to the system performance, since it satisfactorily handles data with noise and also uncertainties of attributing each pixel to one of the analysed image structures [17].

This method consists of three steps: preprocessing, segmentation and post-processing. In the first one, deconvolution was applied to better represent nuclear regions, and its result was submitted to a contrast enhancement and noise removal by means of histogram equalization and mean filter. In the segmentation, ROIs were obtained by using fuzzy 2-partition entropy method combined with GA and wavelet horizontal component extraction. In

the post-processing, false positive regions were removed by means of the valley-emphasis method, and small deformations were corrected through morphological operations. We have simulated this methodology with a public image database [18], which was also manually segmented by a specialist. Its performance was evaluated by comparing it with the techniques of Vahadane and Sethi [19], Wienert et al. [20] and de Oliveira et al. [21].

The main contributions of this paper are summarized as follows:

- A new approach was investigated for unsupervised segmentation composed of the GA optimization technique and the fuzzy theory for CLL images from lymph node biopsies.
- Application of the proposed method on low magnification images, hence they offer less discriminant information of neoplastic nuclei. Besides, these images are from a public domain database, which have great variations of contrast and illumination, typical of real clinical environments.
- Investigation of wavelet transform detail sub-bands in the definition of thresholding value for segmentation of CLL images structures.
- Evaluation and comparison of smoothing and sharpening filters to improve leukemia histological images in the preprocessing step, and the post-processing step development for refinements in accordance with the morphological configuration of the regions manually segmented by a pathologist.

This paper is organized as follows: Section 2 describes the used image database and the theoretical concepts of this proposal. In Section 3, the manual segmentation performed by a specialist was compared with the unsupervised results, which were also compared with other methods of the literature. This section also presents the quantitative evaluation of filters and wavelet sub-bands applied in the preprocessing and segmentation steps. Finally, conclusion and future works are presented in Section 4.

2. Materials and methods

2.1. Database

In this study, CLL histological sections were digitized using a white light microscope Zeiss Axioskop with objective lens of 20× magnification and colour camera charge-coupled device (CCD) AxioCam MR5. These sections were stained with hematoxylin–eosin (H&E) digitized in the RGB model (24 bits of quantification, 8 bits for each colour), available for download [18].

From this database, 12 images were randomly chosen. Each image presents approximately two thousand cells, which is a quantity compatible with other studies related to segmentation of cellular and histological images [22–24]. ROIs were manually segmented by a specialist and automatically by this unsupervised segmentation technique. These images had great differences of contrast and illumination, as exemplified by the sub-images depicted in Fig. 1.

2.2. Proposed algorithm

In this section, the proposed method to segment nuclear regions of CLL neoplastic cells is detailed. Fig. 2 shows the fluxogram of this algorithm, implemented using MATLAB® language and the deconvolution plugin of ImageJ software [25]. Experiments were performed on a 1.7 GHz processor ultrabook (Acer M5-481T-6417) with 6 GB RAM.

2.2.1. Preprocessing

Firstly, the images were submitted to deconvolution [26]. This process enables the identification of structures distribution in

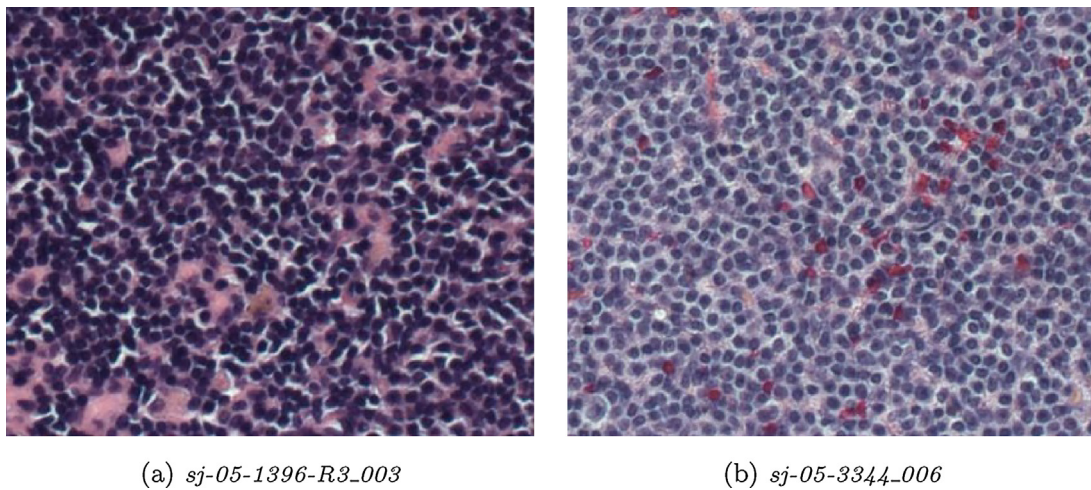


Fig. 1. Sub-images identified by case that were used to validate this proposed method, considering differences of contrast and illumination. (For interpretation of the references to color in text, the reader is referred to the web version of the article.).

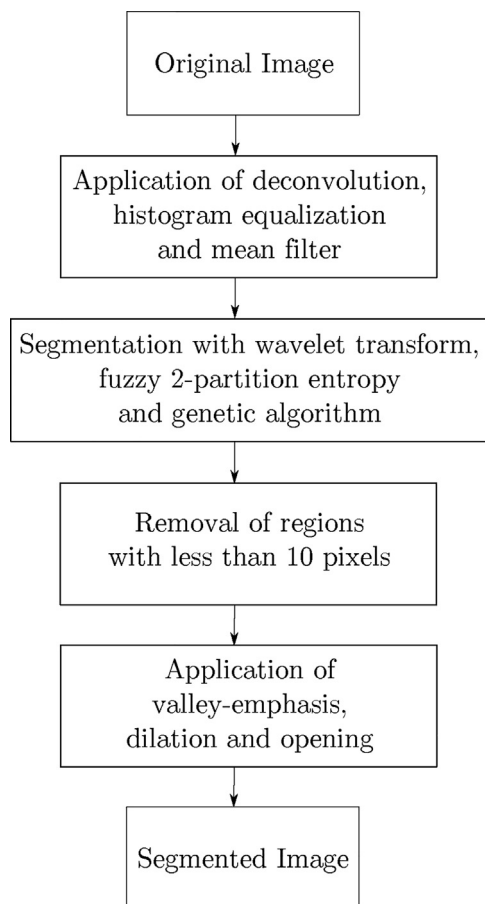


Fig. 2. Segmentation algorithm of CLL neoplastic cells.

which the H&E dye got concentrated. The hematoxylin component stains acid parts of the cell, as the nucleus, in blue shades, whereas the eosin concentrates in alkaline, as the cytoplasm. The light applied to the tissue sample can be measured by the Lambert-Beer law with the absorption factor of hematoxylin and eosin. Hence, light transmission and brightness levels of *R*, *G* and *B* channels will correspond to dye components concentration in a nonlinear way. Since this relation is nonlinear, image intensity levels can-

not be straightforwardly used to separate those components, and so it is necessary to define the optical densities of each colour channel. Optical density is used for establishing a linear relation with concentrations of dyes, which finally determines the relative absorption of multiple components in the histological sample. The separation between hematoxylin and eosin is achieved through an orthonormal transformation of the optical densities, followed by its multiplication with RGB intensity values. Fig. 4 exemplifies the application of this technique to a CLL image. The H&E-stained image is presented by Fig. 4(a), and its deconvolution into each component is also showed by Fig. 4(b) and (c). Since hematoxylin stains acid regions, nuclear regions are identified by this component [6,27]. After deconvolution, histogram equalization technique was applied in order to adjust contrast. This technique redistributes image intensity levels to obtain a histogram of uniform distribution through a cumulative distribution function [28].

Histological images complexity is influenced by various factors, such as histological structures overlapping, blurred cell borders and nuclei corrupted by noise [29]. These noise can be caused by variations of physical and electronic processes of the digitization system and variability of light sources [30]. External interference and sensor configuration can also influence the quality of the resulting image, generating the so-called impulse noise [31]. CCD cameras, as the one used for obtaining the used CLL images, can cause noise by the influence of their light levels and sensor temperature [32]. These imaging systems can cause three types of noise: photon, dark and read noise [33]. The first two are types of shot noise with a Poisson distribution [33,34] and the read noise follows a Gaussian distribution [35]. This distribution can also be used to model noise generated by electronic components of image acquisition systems [36].

The tissue preparation can also introduce noise by resulting in white spots in the cellular structures and presenting craquelures, a typical characteristic of old paintings [37], as presented by Fig. 3. These are the most probable noise in histopathological images, called “crack” in [37]. Beside the noise introduced in its digitization step, the CLL images can also be corrupted by this noise, which present an intensity inhomogeneity in the histological regions.

In order to identify the most appropriate filter for minimizing possible degradation in the type of histological images investigated in this study, different filters were tested after the deconvolution and the histogram equalization. Thus, noise treatment and image smoothing were evaluated with the quantitative metric of entropy (*E*). The minimization of entropy indicates high homogene-

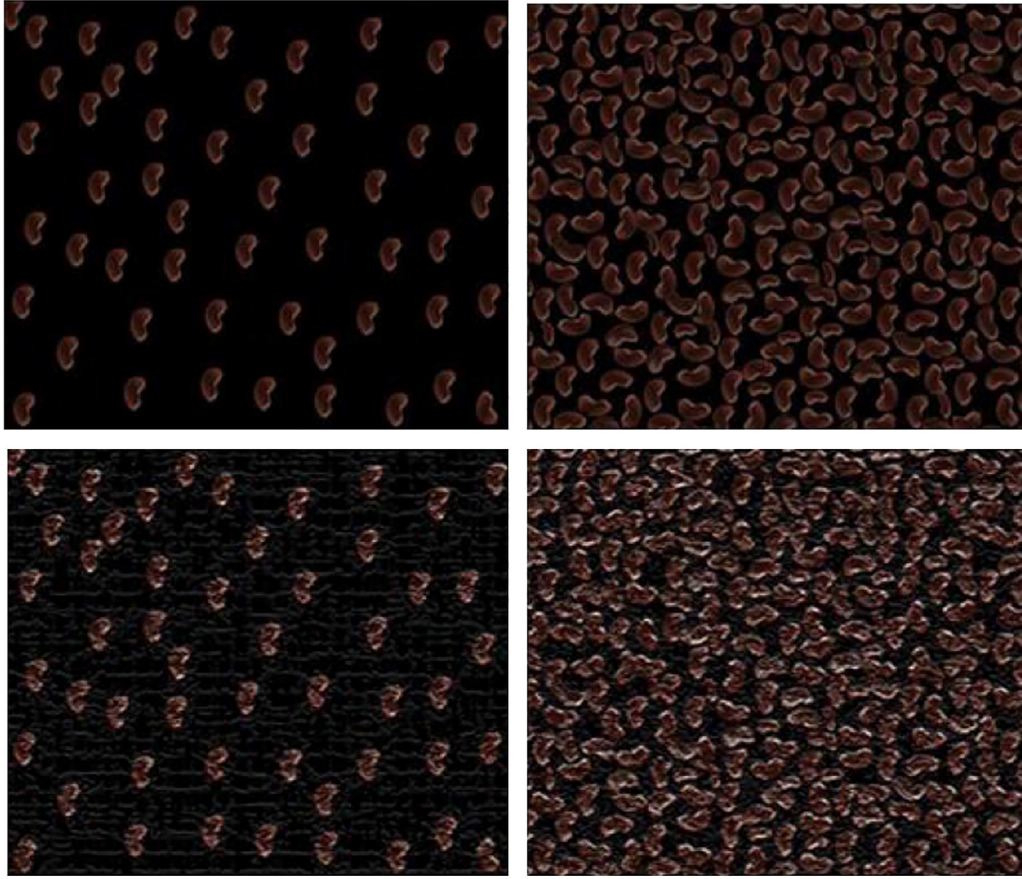


Fig. 3. An example of crack noise with the original images in the first row and the corrupted images in the second row. (For interpretation of the references to color in text, the reader is referred to the web version of the article.).

Figure extracted from [37] © 2010 Elsevier.

ity, which identifies the least noisy preprocessed image [38]. This metric is given by:

$$E = \sum_{k=1}^n p_k \cdot \log_2(p_k), \quad p_k = \frac{h_k}{\sum_{k=1}^n h_k}, \quad (1)$$

where h is the equalized image histogram and n is the total amount of image grey levels.

The evaluated filters were: the median [39], mean [15], Gaussian [40], wavelet-based [41] filters and the unsharp masking method [42]. The median filter is efficient to remove impulse noise but it is less effective to remove Gaussian and random noise [31,43]. The mean filter is usually applied to remove Gaussian noise [43] and the Gaussian filter is able to provide less blurring in random noise reduction [44]. The wavelet-based filter was described with good performance to reduce effects of speckle and Gaussian noise [45,46]. The unsharp masking was also evaluated for being effective in sharpening edges and less noise-sensitive in this process [47].

The technique that reached the minimum of entropy was the mean filter (see details in Table 1). This filter consists of replacing the intensity level of each pixel with the average of its neighbours [36]. In this method, the number of neighbouring pixels is given by a kernel of size 3×3 , defined by empirical tests. Fig. 4(d) exemplifies the output of this step.

2.2.2. Segmentation

Segmentation is one of the most complex and essential tasks in digital image processing [28]. In the proposed method this task was accomplished with thresholding, which is one of the most applied

methods of image segmentation since it is effective for defining disjoint regions and connected boundaries [48]. Recently, some intelligent optimization algorithms have been studied to determine threshold values [49,50]. Based on that, a segmentation technique was developed as a combination of fuzzy 2-partition entropy and GA [51] to separate CLL images into ROI and background.

Initially, the input image was submitted to a wavelet transform. Discrete wavelet transforms enable signals separation and also their precise local description [52]. These transforms are commonly used to simultaneously define features in the domains of time and frequency. Given a 2D signal, as an image, rows and columns are decomposed into approximation (low–low), horizontal (low–high), vertical (high–low) and diagonal (high–high) sub-bands [53]. The approximation sub-band highlights the image illumination, and the other sub-bands represent intensities of edges [54]. Fig. 5 illustrates the wavelet decomposition process of a 2D signal [52].

Decomposition through wavelet transform enables different image frequency levels analyses. High frequency components enhance boundaries and other fine details of image objects, such as the nuclear regions we are interested in [55,56]. Daubechies 5, Symlet 8 and Bi-orthogonal 3.7 wavelet families were analysed and the last one gave the best results, as presented in Section 3. Fig. 6 shows a decomposition example over a preprocessed CLL image (Fig. 6(a)), and also the approximation (Fig. 6(b)), horizontal (Fig. 6(c)), vertical (Fig. 6(d)) and diagonal (Fig. 6(e)) sub-bands.

In this step, an investigation was performed with the proposal to analyse the differences of wavelet details sub-bands of healthy and neoplastic nuclei. Since the approximation sub-band contains only low-frequency components and it corresponds to a low reso-

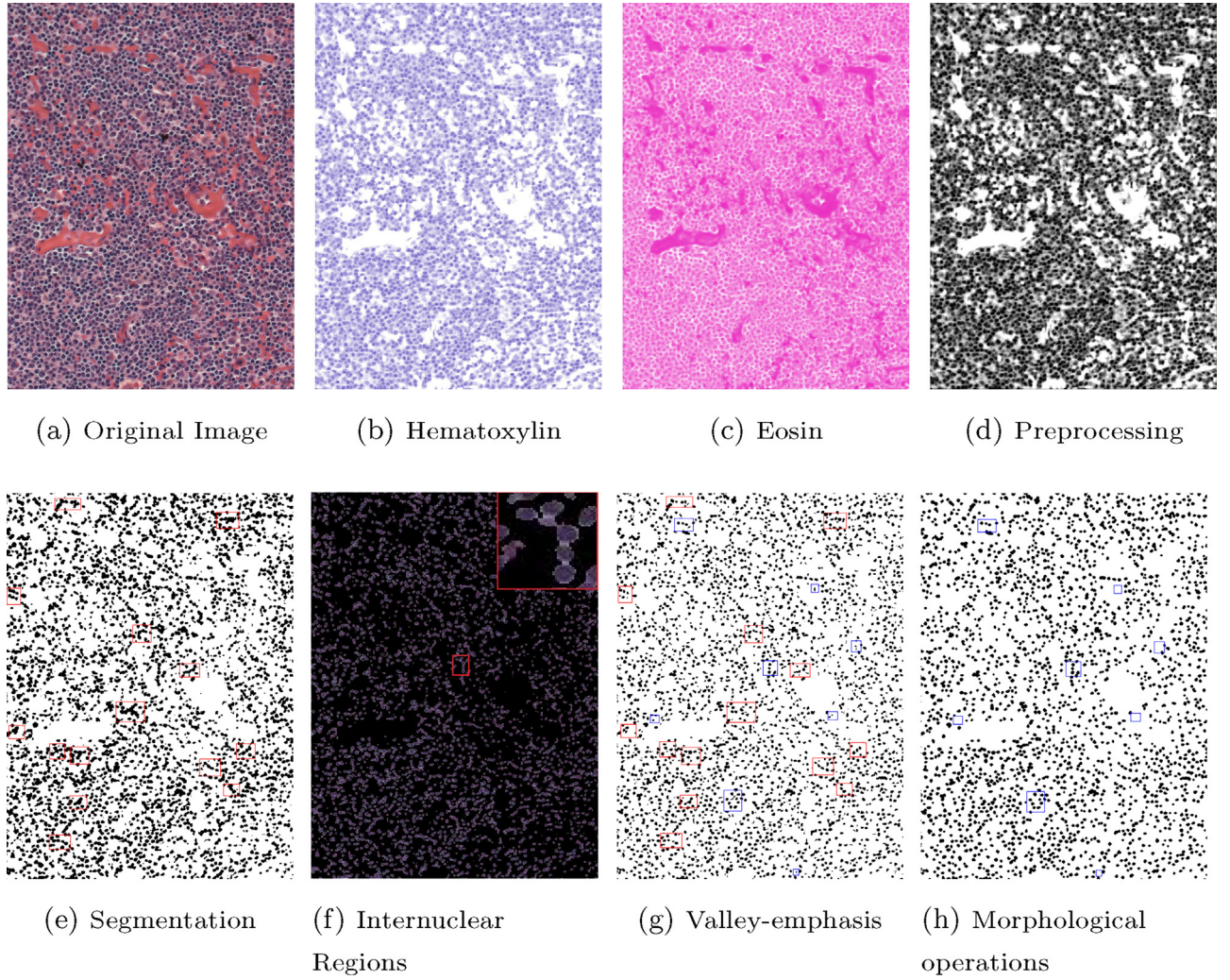


Fig. 4. Application of the proposed method to a CLL sub-image of *sj-03-852-R2.012* case: (a) original image, (b) hematoxylin component, (c) eosin component, (d) preprocessing result, (e) the result of segmentation step with different nuclei identified as one object in rectangular red contours, (f) segmentation mask applied to the original image with a zoomed internuclear region, (g) result of valley-emphasis application and (f) the output of morphological operations. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

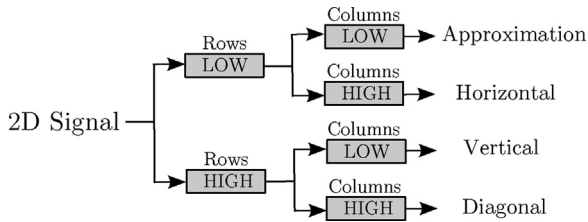


Fig. 5. Decomposition of a 2D signal by wavelet transform. Figure adapted from [52] © 2015 IEEE.

lution of the signal representation, it does not allow the analysis of signal details in the high frequencies [52,56]. So this sub-band was not considered in this analysis. For this purpose, the CLL images were manually segmented to identify some healthy nuclei, as illustrated by Fig. 7, where these nuclei are identified by white pixels and the neoplastic ones by black pixels. All the details sub-bands were normalized in the interval $[0, 255]$ for this analysis. For each image, six normalized histograms were obtained, corresponding to the healthy and the neoplastic nuclei represented in the horizontal, vertical and diagonal details sub-bands. The maximization of the histograms distance metric [57], described in the Section 2.2.4, was used to analyse the details sub-bands. Through this metric, the

horizontal sub-band has reached the highest distance, presenting the most discriminant coefficients and being used in the segmentation. The evaluation of each detail sub-band for the whole image database is shown in Section 3 (Tables 2–5).

In order to apply the GA, 150 random triples (u, v, w) were taken, in which $0 \leq u < v < w \leq 255$. These triples are the so-called *individuals* in GA general theory. The number 150 was defined empirically, and the variables u, v, w were used to determine the threshold value. In order to define the variables of each triple, the fuzzy 2-partition entropy technique was applied [58]. For the identification of the two regions (neoplastic nuclei and background), association degrees were quantified by the functions S and Z of each brightness level k , given by Eqs. (2) and (3), respectively.

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

$$Z(k, u, v, w) = 1 - S(k, u, v, w). \quad (3)$$

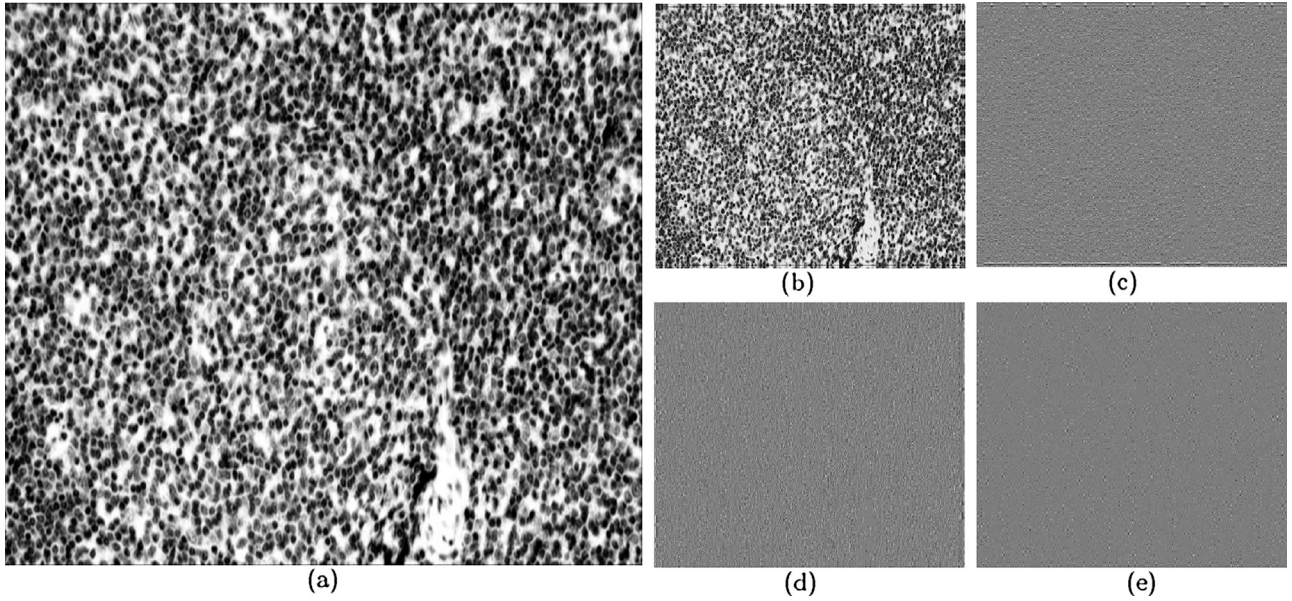


Fig. 6. Decomposition with Bi-orthogonal wavelet transform applied to a CLL sub-image: (a) original image, (b) approximation, (c) horizontal, (d) vertical and (e) diagonal sub-bands.

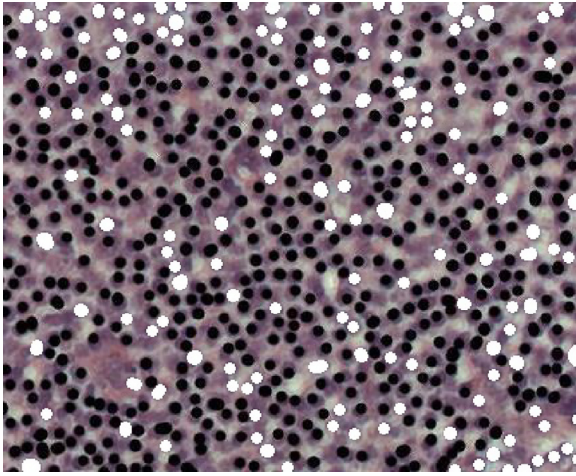


Fig. 7. To identify the most discriminant sub-band between healthy (white pixels) and neoplastic (black pixels) nuclei, these regions were manually segmented and used to calculate their histograms distance. (For interpretation of the references to color in text, the reader is referred to the web version of the article.).

The variables u , v and w were used to obtain the membership degrees M . These quantify how close each level k is to the groups that will be identified, M_n to nuclear regions and M_b to background. The membership degrees are computed by Eqs. (4) and (5).

$$M_n(k) = Z(k, u, v, w), \quad (4)$$

$$M_b(k) = S(k, u, v, w). \quad (5)$$

The parameters u , v and w determine the distributions of the association degrees. The functions M_n and M_b intersect at a point that represents the threshold value to be applied to the wavelet horizontal sub-band of the preprocessed image. Intensity levels between zero and this intersection point have a greater association degree with the nuclear regions, while levels between this point and 255 have a greater degree with the background.

Values determined by each triple were submitted to an evaluation through entropy calculation, given by:

$$H(u, v, w) = -P_n \cdot \log(P_n) - P_b \cdot \log(P_b), \quad (6)$$

where the probabilities P are:

$$P_n = \sum_{k=0}^{255} h(k) \cdot M_n(k), \quad (7)$$

$$P_b = \sum_{k=0}^{255} h(k) \cdot M_b(k), \quad (8)$$

and $h(\cdot)$ represents the normalized histogram of the wavelet horizontal component, P_n is the nuclear probability and P_b corresponds to the background probability.

The value H must be maximized, which means extracting the greatest possible quantity of information. This will define the best combination of parameters (u , v , w) that determines the most efficient separation between the analysed regions.

According to GA workflow a selection step is included, represented by the elitist selection in this method. Namely, only the triples with the greatest values of entropy were kept in next generations [59]. This approach keeps 50% of the population in each generation. Subsequently, the crossover step was used to combine the selected triples [60]. For that, the parameters of crossover probability and crossover point were defined. Crossover probability was empirically assigned to 0.65, which was applied to determine the selected triples during the combination process. The crossover point was randomly generated between 1 and 3 to define the combination of the parameters of the triples. Finally, the mutation step was applied, using a mutation probability parameter, which leads to a populational diversity by changing the parameters of some triples [61]. In the proposed method this probability was empirically defined as 0.01. The GA stop criteria was defined by Hammouche et al. method [62]. This technique interrupts GA execution when average intensity values of both ROI and background remains the same over two consecutive generations.

The output of this step is exemplified by Fig. 4(e), where we notice that corrections must be applied as a post-processing.

2.2.3. Post-processing

In order to eliminate small false positive objects, regions with area below 10 pixels were removed. This value corresponds to the smallest true positive area manually identified by the spe-

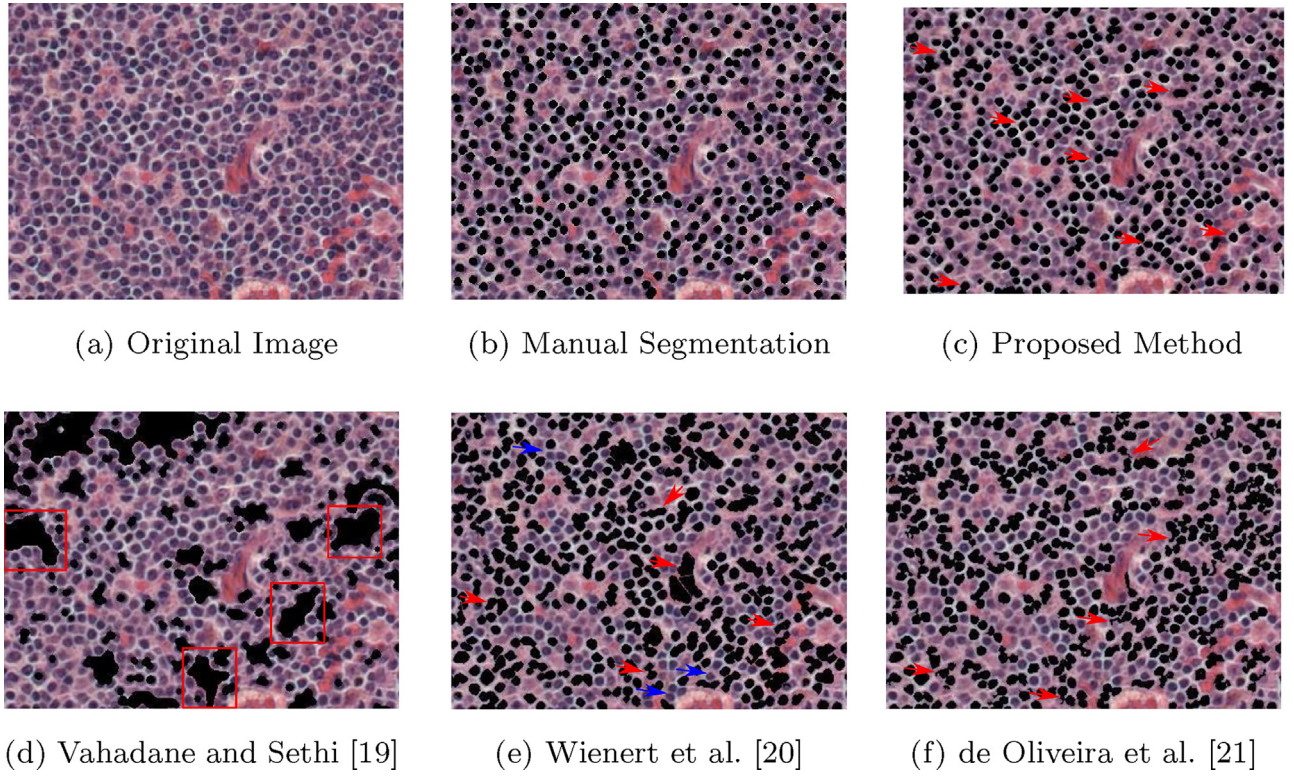


Fig. 8. Comparison of results obtained from applying unsupervised segmentation techniques to a histological section diagnosed as CLL, case sj-03-852-R2.012. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

cialist. The segmentation step was also unable to identify some nuclei individually, revealing internuclear regions. Such regions were characterized by brighter intensity levels than the ROIs, as illustrated by the zoomed area in Fig. 4(f). In order to correct this limitation, the valley-emphasis method [63] was applied to regions characterized by area greater than 50 pixels, corresponding to the average areas identified by the specialist. This method defines a threshold value given by the valley regions of the preprocessed image histogram. Its pixels were then separated into two classes, $C_1 = \{0, 1, \dots, t\}$ and $C_2 = \{t+1, t+2, \dots, 255\}$, where t is the threshold value. In this application, classes C_1 and C_2 represent neoplastic nuclei and internuclear regions, respectively. The probabilities of these classes are:

$$\omega_1(t) = \sum_{i=0}^t p_i \quad \text{and} \quad \omega_2(t) = \sum_{i=t+1}^{255} p_i, \quad (9)$$

where p_i is the occurrence probability of grey level i . Their averages are given by:

$$\mu_1(t) = \sum_{i=0}^t ip_i / \omega_1(t) \quad \text{and} \quad \mu_2(t) = \sum_{i=t+1}^{255} ip_i / \omega_2(t). \quad (10)$$

Through Eqs. (9) and (10) the valley-emphasis method defines the threshold value t using Eq. (11). Maximizing $(1 - p_t)$ corresponds to placing the threshold value in the histogram valley, while the second term is the interclass variance that must also be maximized, as in Otsu's method [64].

$$t = \max_{0 \leq t \leq 255} \{(1 - p_t) \cdot (\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t))\}. \quad (11)$$

The result of this application is shown by Fig. 4(g), where regions contoured by red rectangles can be compared with their corresponding regions in the segmentation step, represented by Fig. 4(e).

After applying valley-emphasis, some noise, connections between objects and irregular contours still remained, as indicated by blue rectangles in Fig. 4(g). Therefore, the morphological operation of opening was applied, since this operation can enlarge the space between objects, smooth contours and remove small noise. Dilation operation was also used to restore object areas, making the automatically identified regions compatible with the ones analysed by the specialist [28,36]. In this study the structural elements of these operations had been empirically defined with disk and square distributions, respectively. The first one is characterized by a radius of two pixels, and the second one by a side length of two pixels. The result of these operations is depicted in Fig. 4(h), with a better representation of nuclear contours.

2.2.4. Evaluation metrics

The evaluation of the wavelet details sub-bands was performed by the histograms distance metric (χ^2) [57]. This metric was analysed to compare the coefficients of the horizontal, vertical and diagonal details sub-bands of healthy and neoplastic nuclei in CLL images. The larger this result, the more discriminant the detail sub-band is:

$$\chi^2(H, N) = \frac{1}{2} \sum \frac{(H_i - N_i)^2}{(H_i + N_i)}, \quad (12)$$

where, H_i and N_i represent the i th bin of the healthy and neoplastic histograms, respectively.

In order to evaluate the obtained results, the metrics of accuracy (AC) and dice index (DI) [65] were used by comparing automatic and manual segmentations. AC is given by Eq. (13), which quantifies pixels correctly segmented by the unsupervised method [66].

$$AC = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, \quad (13)$$

where T_p represents the amount of true positive pixels, and T_N , F_p and F_N correspond to true negative, false positive and false negative pixels, respectively. DI is defined by the ratio between intersection and union of manual (A) and unsupervised (B) segmentations:

$$DI(B, A) = \frac{2|B \cap A|}{|B| + |A|}. \quad (14)$$

Eqs. (13) and (14) indicate that the higher the values of these metrics, the better the unsupervised segmentation.

3. Results and discussion

The filters evaluation is presented by Table 1, which shows the obtained results with the entropy metric after the preprocessing step. The mean filter achieved the lowest value in this metric, as indicated in bold. According to [67] the lower this metric, the better the filter at treating noise in images.

Table 1

Entropy results of the filters median, mean, Gaussian, wavelet-based and of the unsharp masking method to the CLL images after deconvolution and histogram equalization.

Median filter	Mean filter	Gaussian filter	Wavelet-based filter	Unsharp masking method
−5.9256	−7.9520	−7.9469	−7.9284	−7.8074

Table 2 shows the results of histograms distance metric using the details sub-bands of the wavelet functions (Daubechies 5, Symlet 8 and Bi-orthogonal 3.7). This metric shows that the larger its value, the better is the separation between healthy and neoplastic nuclei. Through this metric, the horizontal sub-band has reached the highest distance (in bold), presenting the most discriminant coefficients and being used in the segmentation.

Table 2

Quantitative results of histograms distance of healthy and neoplastic nuclei using Daubechies, Symlet and Bi-orthogonal wavelets.

Sub-band	Daubechies wavelet	Symlet wavelet	Bi-orthogonal wavelet
Diagonal	4.0743×10^{-5}	3.8892×10^{-5}	3.4474×10^{-5}
Horizontal	8.5090×10^{-5}	8.8958×10^{-5}	8.7001×10^{-5}
Vertical	6.6749×10^{-5}	6.9075×10^{-5}	6.8991×10^{-5}

The results of accuracy and dice index metrics were also evaluated to define the best wavelet function, since the differences of the horizontal sub-band were very small among these functions. The results of these metrics are presented in Tables 3–5. The subtle differences in the histogram distance of the horizontal sub-band lead to variations of accuracy and dice index of the wavelet functions, which defined the Bi-orthogonal 3.7 wavelet as the one used to search for the threshold value, since it reached the highest results in both metrics (in bold).

Table 3

Quantitative results of the proposed method using each sub-band of the Daubechies wavelet.

Sub-band	Accuracy (%)	Dice Index (%)
Diagonal	80.08 ± 2.10	87.99 ± 1.54
Horizontal	80.13 ± 2.14	88.01 ± 1.61
Vertical	80.36 ± 2.11	88.18 ± 1.91

Studies of segmentation of CLL histological images from lymph node biopsies stained with H&E were not found in the specialized literature. For this reason, computational methods of nuclear segmentation of H&E images were considered for comparison of results. These methods were applied to the image database described in Section 2.1, as well as the proposed algorithm, for a

Table 4

AC and DI results of the proposed method applied to each sub-band of the Symlet wavelet.

Sub-band	Accuracy (%)	Dice Index (%)
Diagonal	80.77 ± 2.28	88.52 ± 1.62
Horizontal	79.70 ± 1.99	87.66 ± 1.49
Vertical	80.91 ± 2.30	88.61 ± 1.63

Table 5

Quantitative results of the proposed method application to each sub-band of the Bi-orthogonal wavelet.

Sub-band	Accuracy (%)	Dice Index (%)
Diagonal	80.25 ± 2.21	88.11 ± 1.56
Horizontal	81.43 ± 2.59	89.01 ± 1.80
Vertical	80.22 ± 2.14	88.08 ± 1.57

fair comparison. In order to avoid bias, the implementation of these methods was supplied by their very authors. Fig. 8(a)–(c) show an original image, its manual segmentation and the result obtained by this method, respectively. The techniques of Vahadane and Sethi [19], Wienert et al. [20], and de Oliveira et al. [21] were chosen for comparison with results illustrated by Fig. 8(d)–(f), respectively.

Vahadane and Sethi's method [19] segments nuclei in histological images using the watershed algorithm. In their preprocessing step the R channel was extracted, with subsequent application of Gaussian filter and morphological operations of erosion and dilation. For the segmentation, watershed, Otsu's and radial symmetry transform techniques were employed. Finally, the morphological operations were used again in order to eliminate false positive regions.

Wienert et al. [20] developed an approach based on contours to segment cell nuclei in histological images. The preprocessing detects contours by means of Sobel method. Afterwards, a distance transform was used to remove small objects that do not belong to the boundaries, and also to separate identified objects. Colour information extracted by deconvolution, and removal of regions with less than 50 pixels were used to identify the nuclear regions.

Considering de Oliveira's method [21], it proposes a nuclear segmentation algorithm in prostate histological images. The preprocessing used the R channel combined with a contrast enhancement technique. For nuclear segmentation, an optimal threshold value was defined by the minimum-cross-entropy thresholding technique. The post-processing used connected component labelling and also area of identified objects in order to eliminate false positive regions.

In Fig. 8(c) the red arrows indicate false positive regions identified by the proposed method, since they do not fit the specialist's markings (Fig. 8(b)). Vahadane and Sethi's method [19] presented some inconsistency by grouping different nuclei into single regions, as indicated by rectangular red contours in Fig. 8(d). Considering Wienert et al. [20], it presented an expressive number of false positive (indicated by red arrows) and also many more false negative regions (suggested by blue arrows) than our proposed method. Finally, de Oliveira et al. [21] identified objects by segmenting regions of quite greater area than the specialist's, leading to a higher overlapping rate, as indicated by red arrows in Fig. 8(f).

Table 6 presents the quantitative results of this method using the metrics of accuracy and dice index, which were 81.43 ± 2.59 and 89.01 ± 1.80 , respectively. This table also includes the results of Vahadane and Sethi [19], Wienert et al. [20] and de Oliveira et al. [21].

Regarding the accuracy, the highest value was given by this proposed method. The second best result was given by Vahadane and Sethi [19], namely $77.07 + 5.40\% = 82.47\%$. But they can guarantee only $77.07 - 5.40\% = 71.67\%$, which is below the

81.43 – 2.59% = 78.84% of this method. In this sense, the proposed method ensures the highest rate of accuracy in Table 6. The same conclusion holds for dice index. Namely, this method ensures at least 87.21% (89.01 – 1.80%) of coincidence with the manual segmentation. This is the highest dice index result in Table 6, followed by Wienert et al. [20] with 85.21%.

Table 6
Quantitative results of the proposed method and others'.

Techniques	Accuracy (%)	Dice Index (%)
Proposed method	81.43 ± 2.59	89.01 ± 1.80
Vahadane and Sethi [19]	77.07 ± 5.41	86.39 ± 3.79
Wienert et al. [20]	78.53 ± 2.02	86.82 ± 1.61
de Oliveira et al. [21]	70.60 ± 2.60	80.18 ± 2.56

Combining accuracy and dice index analysis, the proposed method offers a relevant contribution to the improvement of computer aided diagnosis. Although specifically related to CLL neoplastic cells, this present study can be a supplementary tool for both computer-aided detection (CADE) and computer-aided diagnosis (CADx) of such cells. The same assertion holds for the other methods in Table 6, since their strategies could be combined with this proposal. For instance, considering that Wienert et al. [20] segmented quite more false negative regions than this method, this fact can indicate a clue about reducing false positive rate.

The proposed combination of techniques achieved relevant results for this application. In the preprocessing step, deconvolution technique was allied with histogram equalization and mean filter. This method showed to be effective for achieving the most appropriate representation of nuclear regions in histological images. The good performance of this proposal was also due to the wavelet transform, which showed to be an efficient tool for image analysis since it captures high frequency that corresponds to the ROIs boundaries. GA evaluation through Shannon's entropy (Eq. (6)) was efficient for quantifying information on the segmented structures. The post-processing improved the results by increasing their consistence with the manual segmentation, particularly in the case of ROI morphological aspects. Hence, the post-processing is necessary for extraction of features and classification used in CAD systems, even though these are not implemented in this method yet. The main limitation of this proposal is the quantity of false positive regions still present in the obtained results.

4. Conclusion and future works

One of the essential tasks in the diagnosis of different types of cancer is the identification of neoplastic cell nuclei in histological images. When manually accomplished, this task becomes tedious, costly and subjective. Therefore, digital image processing techniques are useful for helping pathologists to identify these structures of interest in an objective way. This present study proposed an unsupervised method to segment neoplastic nuclear regions in CLL images. In the steps of preprocessing, segmentation and post-processing, the combined techniques showed relevant results at separating ROIs by considering colour features and dismissing information about shape and spatial distribution.

The combination between the GA method and the fuzzy 2-partition entropy reached a good performance in the segmentation of the neoplastic nuclei, even on the complex images used for validation. With low magnification and variations of contrast and illumination, these images allowed to demonstrate the robustness of the proposed method. The filters evaluation in the preprocessing step was important to obtain images with a better quality for the segmentation. The quantitative wavelet sub-bands analyses also contributed to the segmentation once they allowed to define

the most discriminate sub-band between healthy and neoplastic nuclei. The stop criteria used in the GA method made this proposal an unsupervised method, with no user intervention for this purpose. Finally, the post-processing enhanced the morphological features of the neoplastic nuclei, making them in accordance with the evaluation of a pathologist.

Our experimental results indicate that the proposed technique achieves a good performance, with accuracy and dice index values above 80% and nearly 90%, respectively. In order to draw comparisons, other studies related to nuclear segmentation of H&E-stained histological images were applied to validate the proposed method. Compared with these studies, our method presents a better representation of nuclear contours and achieves lower overlapping rates for the images of the used database.

In future works, other features of neoplastic nuclei will be explored, such as texture and topology, in order to remove the false positive regions. Moreover, semantic features will also be studied to explore the identified regions. Neoplastic cells can be detected with these features since they give a computational representation of some criteria used by pathologists to diagnose CLL. We also intend to investigate feature extraction and classification of CLL images and other lesion groups, and different evaluation metrics of GA individuals [68,69]. Furthermore, we intend to use more images in future evaluations of the proposed method.

Acknowledgements

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

References

- [1] K.R. Rai, P. Jain, Chronic lymphocytic leukemia (CLL) – then and now, *Am. J. Hematol.* 91 (3) (2016) 330–340.
- [2] J. Maly, J.S. Blachly, Chronic lymphocytic leukemia: exploiting vulnerabilities with targeted agents, *Curr. Hematol. Malig. Rep.* 11 (1) (2016) 52–60.
- [3] M.d.A. Martins, F.J. Carrilho, V.A.F. Alves, E.A.D. Castilho, G.G. Cerri, C.L. Wen, *Clínica Médica*, vol. 3, 1st ed., 2011.
- [4] N.V. Orlov, W.W. Chen, D.M. Eckley, T.J. Macura, L. Shamir, E.S. Jaffe, I.G. Goldberg, Automatic classification of lymphoma images with transform-based global features, *IEEE Trans. Inf. Technol. Biomed. Eng.* 14 (4) (2010) 1003–1013.
- [5] O. Sertel, G. Lozanski, A. Shana'ah, M.N. Gurcan, Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation, *IEEE Trans. Biomed. Eng.* 57 (10) (2010) 2613–2616.
- [6] L.P. Gartner, J.L. Hiatt, *Tratado de Histologia em Cores*, 2nd ed., Guanabara Koogan, 2003.
- [7] H. Irshad, A. Veillard, L. Roux, D. Racocanu, Methods for nuclei detection, segmentation, and classification in digital histopathology: a review – current status and future potential, *IEEE Rev. Biomed. Eng.* 7 (2014) 97–114.
- [8] S. Ong, X. Jin, R. Sinniah, et al., Image analysis of tissue sections, *Comput. Biol. Med.* 26 (3) (1996) 269–279.
- [9] E.A. Mohammed, B.H. Far, C. Naugler, M.M.A. Mohamed, Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding, *Canadian Conference on Electrical and Computer Engineering*, IEEE (2013) 1–5.
- [10] K. Jiang, Q. Liao, Y. Xiong, A novel white blood cell segmentation scheme based on feature space clustering, *Soft Comput.* 10 (1) (2006) 12–19.
- [11] E.A. Mohammed, B.H. Far, C. Naugler, M.M.A. Mohamed, Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation, *International Conference on e-Health Networking, Applications and Services* (2013) 622–626.
- [12] N. Guo, L. Zeng, Q. Wu, A method based on multispectral imaging technique for white blood cell segmentation, *Comput. Biol. Med.* 37 (1) (2007) 70–76.
- [13] X. Liu, G. Feng, Kernel bisecting k-means clustering for SVM training sample reduction, *International Conference on Pattern Recognition*, IEEE (2008) 1–4.
- [14] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener, Histopathological image analysis: a review, *IEEE Rev. Biomed. Eng.* 2 (2009) 147–171.
- [15] J.M. Haggerty, X.N. Wang, A. Dickinson, C.J. O'Malley, E.B. Martin, Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin, *BMC Med. Imaging* 14 (1) (2014) 7.
- [16] S. Sarkar, S. Das, S.S. Chaudhuri, Hyper-spectral image segmentation using Rényi entropy based multi-level thresholding aided with differential evolution, *Expert Syst. Appl.* 50 (2016) 120–129.

- [17] G. Begelrnan, E. Gur, E. Rivlin, M. Rudzsky, Z. Zalevsky, Cell nuclei segmentation using fuzzy logic engine, *International Conference on Image Processing*, vol. 5, IEEE (2004) 2937–2940.
- [18] L. Shamir, N. Orlov, D.M. Eckley, T.J. Macura, I.G. Goldberg, IICBU 2008: a proposed benchmark suite for biological image analysis, *Med. Biol. Eng. Comput.* 46 (9) (2008) 943–947.
- [19] A. Vahadane, A. Sethi, Towards generalized nuclear segmentation in histological images, *International Conference Bioinformatics and Bioengineering*, IEEE (2013) 1–4.
- [20] S. Wienert, D. Heim, K. Saeger, A. Stenzinger, M. Beil, P. Hufnagel, M. Dietel, C. Denkert, F. Klauschen, Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach, *Sci. Rep.* 2 (2012) 503.
- [21] D.L.L. de Oliveira, M.Z. do Nascimento, L.A. Neves, M.F. de Godoy, P.F.F. de Arruda, D. de Santi Neto, Unsupervised segmentation method for cuboidal cell nuclei in histological prostate images based on minimum cross entropy, *Expert Syst. Appl.* 40 (18) (2013) 7331–7340.
- [22] A. Gençtav, S. Aksoy, S. Önder, Unsupervised segmentation and classification of cervical cell images, *Pattern Recognit.* 45 (12) (2012) 4151–4168.
- [23] K. Dimitropoulos, P. Barmoutis, T. Koletsa, I. Kostopoulos, N. Grammalidis, Automated detection and classification of nuclei in PAX5 and H&E-stained tissue sections of follicular lymphoma, *Signal Image Video Process.* 11 (1) (2016) 1–9.
- [24] P. Wang, X. Hu, Y. Li, Q. Liu, X. Zhu, Automatic cell nuclei segmentation and classification of breast cancer histopathology images, *Signal Process.* 122 (2016) 1–13.
- [25] ImageJ (Image Processing and Analysis in JAVA), 2016, URL <http://imagej.nih.gov/ij/index.html> (accessed 11.10.16).
- [26] A.C. Ruifrok, D.A. Johnston, Quantification of histochemical staining by color deconvolution, *Anal. Quant. Cytol. Histol. Int. Acad. Cytol. Am. Soc. Cytol.* 23 (4) (2001) 291–299.
- [27] J.C. Junqueira, J. Carneiro, *Histologia Básica*, 10th ed., Guanabara Koogan, 2004.
- [28] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 2nd ed., Prentice Hall Upper Saddle River, 2002.
- [29] L. He, L.R. Long, S. Antani, G.R. Thoma, Histology image analysis for carcinoma detection and grading, *Comput. Methods Programs Biomed.* 107 (3) (2012) 538–556.
- [30] J.C. Russ, M. Parry-Hill, M.W. Davidson, Reducing Noise, Olympus America, Inc. and The Florida State University, 2016, URL <http://micro.magnet.fsu.edu/primer/digitalimaging/russ/reducingnoise.html> (accessed 05.09.17).
- [31] K.R. Spring, J.C. Russ, M. Parry-Hill, M.W. Davidson, Median Filters for Digital Images, Olympus America, Inc. and The Florida State University, 2016, URL <https://micro.magnet.fsu.edu/primer/java/digitalimaging/processing/medianfilter/index.html> (accessed 14.09.17).
- [32] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 3rd ed., Prentice Hall Upper Saddle River, 2008.
- [33] T.J. Fellers, M.W. Davidson, CCD Noise Sources and Signal-To-Noise Ratio, Olympus America, Inc. and The Florida State University, 2015, URL <https://micro.magnet.fsu.edu/primer/digitalimaging/concepts/ccdsnr.html> (accessed 13.09.17).
- [34] P. Patidar, M. Gupta, S. Srivastava, A.K. Nagawat, Image de-noising by various filters for different noise, *Int. J. Comput. Appl.* 9 (4) (2010).
- [35] J. Joubert, Y. Sabharwal, D. Sharma, Digital camera technologies for scientific bio-imaging. Part 3. Noise and signal-to-noise ratios, *Microsc. Anal.* 1 (2011).
- [36] H. Pedrini, W.R. Schwartz, *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*, Thomson Learning, 2008.
- [37] O.S. Al-Kadi, Texture measures combination for improved meningioma classification of histopathological images, *Pattern Recognit.* 43 (6) (2010) 2043–2053.
- [38] K. Belkacem-Boussaid, J. Prescott, G. Lozanski, M.N. Gurcan, Segmentation of follicular regions on H&E slides using a matching filter and active contour model, in: *SPIE Medical Imaging*, vol. 7624, International Society for Optics and Photonics, 762436–1–762436–11, 2010.
- [39] K. Belkacem-Boussaid, S. Samsi, G. Lozanski, M.N. Gurcan, Automatic detection of follicular regions in H&E images using iterative shape index, *Comput. Med. Imaging Graph.* 35 (7) (2011) 592–602.
- [40] E. Michail, E.N. Kornaropoulos, K. Dimitropoulos, N. Grammalidis, T. Koletsa, I. Kostopoulos, Detection of centroblasts in H&E stained images of follicular lymphoma, *Signal Processing and Communications Applications Conference*, IEEE (2014) 2319–2322.
- [41] C.S. Anand, J.S. Sahambi, Wavelet domain non-linear filtering for MRI denoising, *Magn. Reson. Imaging* 28 (6) (2010) 842–861.
- [42] M. Tsiplakidou, M.G. Tsipouras, N. Giannakeas, A.T. Tzallas, P. Manousou, Automated detection of liver histopathological findings based on biopsy image processing, *Information* 8 (1) (2017) 36.
- [43] M. Lysaker, S. Osher, X. Tai, Noise removal using smoothed normals and surface fitting, *IEEE Trans. Image Process.* 13 (10) (2004) 1345–1357.
- [44] K.R. Spring, J.C. Russ, M. Parry-Hill, M.W. Davidson, Spatial Averaging, Olympus America, Inc. and The Florida State University, 2016, URL <https://micro.magnet.fsu.edu/primer/java/digitalimaging/processing/spatialaveraging/index.html> (accessed 14.09.17).
- [45] N. Attias, S. Gupta, Wavelet based techniques for speckle noise reduction in ultrasound images, *Int. J. Eng. Res. Appl.* 4 (2) (2014) 508–513.
- [46] M. Nasri, H. Nezamabadi-pour, Image denoising in the wavelet domain using a new adaptive thresholding function, *Neurocomputing* 72 (4) (2009) 1012–1025.
- [47] A.P. Sheppard, R.M. Sok, H. Averdunk, Techniques for image enhancement and segmentation of tomographic images of porous materials, *Phys. A: Stat. Mech. Appl.* 339 (1) (2004) 145–151.
- [48] K.R. Castleman, *Digital Image Processing*, Pearson Education Instruments, Inc., New Delhi, India, 2012.
- [49] W. Tao, J. Tian, J. Liu, Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm, *Pattern Recognit. Lett.* 24 (16) (2003) 3069–3078.
- [50] W. Tao, H. Jin, L. Liu, Object segmentation using ant colony optimization algorithm and fuzzy entropy, *Pattern Recognit. Lett.* 28 (7) (2007) 788–796.
- [51] S. Yin, X. Zhao, W. Wang, M. Gong, Efficient multilevel image segmentation through fuzzy entropy maximization and graph cut optimization, *Pattern Recognit.* 47 (9) (2014) 2894–2907.
- [52] C.B. Fiallos, M.G. Pérez, A. Conci, V.H. Andaluz, Automatic detection of injuries in mammograms using image analysis techniques, *International Conference on Systems, Signals and Image Processing*, IEEE (2015) 245–248.
- [53] S. Lahmiri, Image characterization by fractal descriptors in variational mode decomposition domain: application to brain magnetic resonance, *Phys. A: Stat. Mech. Appl.* 456 (2016) 235–243.
- [54] Q. Abbas, Segmentation of differential structures on computed tomography images for diagnosis lung-related diseases, *Biomed. Signal Process. Control* 33 (2017) 325–334.
- [55] D. Edwin, S. Hariharan, Wavelet based segmentation of liver tumors from CT images, *Int. J. Appl. Eng. Res.* 11 (6) (2016) 4607–4614.
- [56] E.F. Oliveira, A.G.C. Bianchi, L.d.S. Martins-Filho, R.F. Machado, Granulometric analysis based on the energy of wavelet transform coefficients, *Rev. Escola Minas* 63 (2) (2010) 347–354.
- [57] O. Pele, M. Werman, The quadratic-chi histogram distance family, in: *European Conference on Computer Vision*, Springer, 2010, pp. 749–762.
- [58] Y. Tang, W. Mu, Y. Zhang, X. Zhang, A fast recursive algorithm based on fuzzy 2-partition entropy approach for threshold selection, *Neurocomputing* 74 (2011) 3072–3078.
- [59] M. Paulinas, A. Ušinskas, A survey of genetic algorithms applications for image enhancement and segmentation, *Inf. Technol. Control* 36 (3) (2007) 278–284.
- [60] W. Lin, W. Lee, T. Hong, Adapting crossover and mutation rates in genetic algorithms, *J. Inf. Sci. Eng.* 19 (5) (2003) 889–903.
- [61] J. McCall, Genetic algorithms for modelling and optimisation, *J. Comput. Appl. Math.* 184 (1) (2005) 205–222.
- [62] K. Hammouche, M. Diaf, P. Siarry, A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation, *Comput. Vis. Image Understand.* 109 (2) (2008) 163–175.
- [63] H. Ng, Automatic thresholding for defect detection, *Pattern Recognit. Lett.* 27 (14) (2006) 1644–1649.
- [64] N. Otsu, A thresholding selection method from gray-level histograms, *Automatica* 11 (1975) 23–27.
- [65] G. Wu, X. Zhao, S. Luo, H. Shi, Histological image segmentation using fast mean shift clustering method, *Biomed. Eng. Online* 14 (1) (2015) 24.
- [66] K.A. Byrd, J. Zeng, M. Chouikha, A validation model for segmentation algorithms of digital mammography images, *J. Appl. Sci. Eng. Technol.* 1 (2007) 41–50.
- [67] D. Tsai, Y. Lee, E. Matsuyama, Information entropy measure for evaluation of image quality, *J. Digit. Imaging* 21 (3) (2008) 338–347.
- [68] R.A. Fisher, *The Genetical Theory of Natural Selection: A Complete Variorum Edition*, Oxford University Press, 1999.
- [69] Z. Abo-Eleneen, G. Abdel-Azim, A novel algorithm for image thresholding using non-parametric fisher information, in: *International Electronic Conference on Entropy and Its Applications*, vol. 1, Multidisciplinary Digital Publishing Institute, 2014, pp. 15.