



Computational method for unsupervised segmentation of lymphoma histological images based on fuzzy 3-partition entropy and genetic algorithm



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

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

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

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

ARTICLE INFO

Article history:

Received 18 September 2016

Revised 28 January 2017

Accepted 23 March 2017

Available online 23 March 2017

Keywords:

Nuclear segmentation

Histological images

Lymphoma

Fuzzy 3-partition

Genetic algorithm

Valley-emphasis

ABSTRACT

Non-Hodgkin lymphoma is the most common cancer of the lymphatic system and should be considered as a group of several closely related cancers, which can show differences in their growth patterns, their impact on the body and how they are treated. The diagnosis of the different types of neoplasia is made by a specialist through the analysis of histological images. However, these analyses are complex and the same case can lead to different understandings among pathologists, due to the exhaustive analysis of decisions, the time required and the presence of complex histological features. In this context, computational algorithms can be applied as tools to aid specialists through the application of segmentation methods to identify regions of interest that are essential for lymphomas diagnosis. In this paper, an unsupervised method for segmentation of nuclear components of neoplastic cells is proposed to analyze histological images of lymphoma stained with hematoxylin-eosin. The proposed method is based on the association among histogram equalization, Gaussian filter, fuzzy 3-partition entropy, genetic algorithm, morphological techniques and the valley-emphasis method in order to analyze neoplastic nuclear components, improve the contrast and illumination conditions, remove noise, split overlapping cells and refine contours. The results were evaluated through comparisons with those provided by a specialist and techniques available in the literature considering the metrics of accuracy, sensitivity, specificity and variation of information. The mean value of accuracy for the proposed method was 81.48%. Although the method obtained sensitivity rates between 41% and 51%, the accuracy values showed relevance when compared to those provided by other studies. Therefore, the novelties presented here may already encourage new studies with a more comprehensive overview of lymphoma segmentation.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Lymphoma is a type of malignant disease that develops in cellular components called lymphocytes (Orlov et al., 2010). These cells represent one of the highest white blood cell populations responsible for the immunological defense of the body (Gartner & Hiatt, 2003). Lymphomas are divided into Hodgkin lymphomas (HL) and non-Hodgkin lymphomas (NHL), in accordance with

their combinations of morphological, genetic and clinical features (Mauriño & Siqueira, 2011).

The 2016 report of the National Cancer Institute of Brazil estimates almost 11,000 new cases of NHL (INCA, 2016). Chronic lymphocytic leukemia (CLL), follicular lymphoma (FL) and mantle cell lymphoma (MCL) belong to the NHL class, which corresponds to 85% of the lymphomas (Lowry & Linch, 2013). The American Cancer Society estimates for 2017 that about 72,240 new cases will be diagnosed and about 20,140 people will die from this cancer in the United States (ACS, 2017). Thus, nowadays there is a high demand for diagnoses, where its analysis and detection remain a challenge for pathologists.

Tissue samples stained with hematoxylin-eosin (H&E) have been used by pathologists for analysis and identification of NHL

* Corresponding author.

E-mail addresses: tosta.thaina@gmail.com (T.A. Azevedo Tosta), paulo.faria@ufu.br (P.R. Faria), neves.leandro@gmail.com (L. Alves Neves), marcelo.zanchetta@gmail.com (M.Z.d. Nascimento).

cancer structures. These procedures are essential for disease monitoring and more efficient definitions for treatments (Orlov et al., 2010). However, visual evaluation is a complex task due to the significant time involved, its subjectivity and variability between pathologists (Oger, Belhomme, & Gurcan, 2012; Sertel, Lozanski, Shana'ah, & Gurcan, 2010b).

Histological samples can be analyzed by computational techniques and this procedure has provided advances in the support for diagnosis and prognosis of lymphomas. The computational strategies can improve the accuracy and efficiency of the detection of cells linked to NHL cancers (Belkacem-Boussaid, Samsi, Lozanski, & Gurcan, 2011; Sertel, Catalyurek, Lozanski, Shana'ah, & Gurcan, 2010a) and patterns recognition (Orlov et al., 2010).

The segmentation of NHL structures is a crucial task in many clinical applications and the subsequent stages, including feature extraction and classification, which all rely heavily on the quality of this process. In this stage, techniques are applied in order to recognize the presence, distribution, size and morphological features useful for diagnosis (Haggerty, Wang, Dickinson, O'Malley, & Martin, 2014). However, such a task is complex due to features variations, mainly when distinguishing nuclear regions (Irshad, Veillard, Roux, & Racoceanu, 2014).

In this context, this paper presents an unsupervised segmentation method to aid pathologists in the identification of neoplastic nuclei of CLL, FL and MCL histological images. The proposed algorithm was divided into steps of preprocessing, segmentation and post-processing. In the preprocessing step, the histogram equalization and Gaussian filter were applied to the RGB color model channels. A technique based on thresholding was developed as a result of the combination between genetic algorithm (GA) and fuzzy 3-partition entropy method. Finally, the valley-emphasis technique and morphological operations of dilation and opening were applied in the post-processing step. The proposed method was tested on a public dataset comprised of 12 images of CLL, 62 of FL and 99 of MCL, which were obtained with magnification of 20 \times . The metrics of accuracy, sensitivity, specificity and variation of information were applied for quantitative evaluations. The performance of the proposed algorithm was compared to the results provided by the mean-shift technique (Comaniciu & Meer, 2002) and the approaches proposed by de Oliveira et al. (2013); Phoulady, Goldgof, Hall, and Mouton (2016); Vahadane and Sethi (2013); Wienert et al. (2012) and Paramanandam et al. (2016).

1.1. Related works

Several studies dedicated to the segmentation of NHL histological images are presented in the literature.

In the case of CLL images, the studies of Mohammed, Far, Naugler, and Mohamed (2013a, 2013b) presented nuclear, cellular and cytoplasmic segmentation methods of normal and neoplastic lymphocytes. In Mohammed, Far, Naugler, and Mohamed (2013b), the authors employed Otsu thresholding, canny edge detector, morphological operations and removal of 1% of local minima of watershed to reduce over and under segmentation errors. Further, the authors presented in Mohammed et al. (2013a), a segmentation method based on pixel classification using support vector machine (SVM) and K-means to reduce the feature set.

For MCL images, Yang, Tuzel, Meer, and Foran (2008) developed a segmentation method of overlapping cells using L_2 estimation (L_2E), gradient vector flow (GVF) and the CIE LUV color model for contour extraction. High curvature points were identified by a literature proposal. The canny edge detector was applied to detect inner edges, which correspond to candidate lines for separating the cells. These lines were analyzed through the Dijkstra algorithm to determine the best segmentation of the overlapping cells.

Studies related to the detection of CLL and MCL lesions are restricted to blood related images with different magnifications, such as 60 \times (Yang et al., 2008) and 100 \times (Mohammed et al., 2013b).

To segment follicular regions in FL images, different approaches were proposed, such as the methods of active contour model (Arora & Banerjee, 2013; Belkacem-Boussaid, Prescott, Lozanski, & Gurcan, 2010), region-based thresholding using curve evolution (Belkacem-Boussaid et al., 2011), thresholding based on mean brightness value (Zorman et al., 2007) and Otsu algorithm (Oger et al., 2012). To deal with significant color variations in the regions of interest (ROIs), Arora and Banerjee (2013) and Belkacem-Boussaid et al. (2010) applied a local energy function and active contour model to identify follicles from H&E stained tissue sections. In addition to the follicular regions segmentation, Arora and Banerjee (2013) also investigated the classification of the grades of FL, but the segmentation step was not evaluated. The authors of Belkacem-Boussaid et al. (2010) applied pre and post-processing for segmenting FL images. However, a limitation was noted in this approach consisting of the merging of follicles in the segmentation, which demands a new strategy for the separation of overlapping follicles.

The algorithm of Belkacem-Boussaid et al. (2011) stands out in the evaluation metrics of their different steps. For instance, the metrics of signal to noise ratio and texture contrast were applied to define the more adequate color channel. The preprocessing step was evaluated by the Haralick homogeneity metric, whereas overlapping follicles were analyzed by a concavity index. In Zorman et al. (2007), the authors used a pixel classification approach for follicles segmentation. The mean brightness values were considered as the threshold value in a pre-segmentation step. Different from the above described techniques, the method in Oger et al. (2012) proposed a segmentation of follicular regions on IHC images with registration of the identified regions on H&E images. However, the conformity metric does not reach satisfactory results due to identification of false positive regions.

Algorithms based on the Otsu thresholding method (Dimitropoulos, Michail, Koletsas, Kostopoulos, & Grammalidis, 2014; Michail et al., 2014), k-means (Oztan, Kong, Gurcan, & Yener, 2012; Sertel et al., 2009) and mean-shift clustering (Sertel et al., 2010a) were applied to detect centroblasts on FL images. After nuclear segmentation, Dimitropoulos et al. (2014) also investigated the extraction of morphological and textural features. Nucleoli detection and cytoplasm histogram analysis were used by applying the SVM classifier. This approach was also employed by Michail et al. (2014), however, intensity features were then classified by the linear discriminant analysis classifier. Furthermore, these studies applied empirical threshold values in the segmentation step for removal of red blood cells.

For FL grading, Oztan et al. (2012) presented a method for analyzing the more discriminant features among the cellular regions on FL images. For this purpose, features of graphs constructed from K-means segmentation results, information of intensity, texture and MBIR representations were combined, thus reaching the best results with the SVM classifier. The centroblast detection was also explored by the studies of Sertel et al. (2009) and Sertel et al. (2010a). However, Sertel et al. (2010a) considered the mean-shift method so as not to have to manually define the number of clusters, as demanded by the K-means method used by Sertel et al. (2009).

Considering segmentation methods for CLL, FL and MCL histological images, Table 1 summarizes the strengths and weaknesses of these techniques proposed in the literature.

A majority of the studies in Table 1 presents methods for detection and segmentation of FL due to its high incidence rate, which represents the second most common B-cell lymphoma

Table 1

Strengths and weaknesses of studies related to segmentation of the histological images stained with H&E.

Ref. and lesion images	Segmentation Method	Strengths	Weaknesses
Arora and Banerjee (2013) FL images.	Active contour model.	Robust segmentation to color variations.	Few training images for FL grading.
Belkacem-Boussaid et al. (2010) FL images.	Active contour model.	New filter for removal of noise and enhancement of follicle contours and robust method to low magnification images.	Manual selection of seed points of active contour model, unsatisfactory performance on low contrast images, limitation to identify nearby or small follicles individually.
Belkacem-Boussaid et al. (2011) FL images.	Region-based segmentation using curve evolution.	Use of pathological and biological features to define seed points and removal of false positives, new proposal for concavity detection.	Empirical definition of initial size of the curve, training with three cases to empirically define parameters, few images to the method evaluation, unsatisfactory performance for application on histological slides with low quality in their preparation and coloring processes.
Dimitropoulos et al. (2014) FL images.	Thresholding and Otsu.	Use of pathological properties to propose the hybrid classifier.	Empirical thresholding to remove red blood cells.
Dimitropoulos et al. (2016) FL images.	K-means and graph cut.	New evaluation criteria for splitting of elliptical components and overlapping nuclei.	Use of IHC and H&E images, training image set smaller than the test set.
Kong et al. (2011a) FL images.	Thresholding, efficient local Fourier transform, K-means and K-nn.	Proposal of a new color model.	Limitation to individually identify nuclei.
Kong et al. (2011b) FL images.	Thresholding, efficient local Fourier transform, K-means and K-nn.	Proposal of a new color model and a new method for splitting overlapping nuclei.	Parameter definition on five training images.
Luo et al. (2006) Blood images of MCL, Hairy Cell Leukemia and Plasma Cell Leukemia.	Otsu and watershed.	Use of feature invariant to rotation, translation and scale.	Manual definition of a threshold to feature selection.
Michail et al. (2014) FL images.	Thresholding and Otsu.	Use of pathological criteria.	Empirical definition of a threshold value for removal of red blood cells.
Mohammed et al. (2013b) Blood cell images.	Otsu, canny edge detector and subtraction between cellular and nuclear segmentations.	Removal of 1% of local minima to deal with over and under segmentation errors.	Cells located at image center, illumination conditions given as uniforms for Otsu thresholding.
Mohammed et al. (2013a) Blood cell images.	Otsu, K-means, SVM and subtraction between cellular and nuclear segmentations.	Over and under segmentation reduced by SVM.	Cells located at image center, few training images, limitation in cytoplasm segmentation and poor performance in overlapping lymphocytes.
Oger et al. (2012) FL images.	Otsu and intersection between R and B binary masks.	Use of stain properties for color channel selection.	Few images for the method evaluation and use of IHC images.
Oztan et al. (2012) FL images.	K-means.	Definition of high level descriptor of cytological components.	Empirical definition of relationship between graph vertices, limitation to application on overlapping cells.
Sertel et al. (2008a) FL images.	K-means.	Use of stain properties in feature extraction of follicles and of morphological and pathological features for centroblast detection.	Use of IHC and H&E images, with manual points definition for mapping between them.
Sertel et al. (2008b) FL images.	Thresholding and K-means.	Use of pathological criteria for definition of cytoplasm features.	Empirical thresholding to remove red blood cells and image background.
Sertel et al. (2009) Whole-slide images of FL.	Thresholding and K-means.	Use of the biological and pathological features for FL grading.	Training image set smaller than the test set, empirical definition of threshold value for removal of red blood cells and image background, use of texture feature on 40 × magnification images.
Sertel et al. (2010a) FL images.	Mean-shift.	Use of biological criteria for definition of detected features.	Definition of probability functions of centroblasts using a small training set.
Sertel et al. (2010b) FL images.	Gaussian mixture modeling with parameters estimation using expectation maximization.	Parameter estimation of fast radial symmetry transform adequate to different cell sizes, automation for parameters definition.	Texture extraction on 40 × magnification images.
Yang et al. (2008) Cases of MCL, CLL, FL and other lymphoma types.	L_2E and GVF algorithms, detection of high curvature points, canny edge detector and concave vertex graph.	High performance in a graph application, better quantitative results than watershed, use of pathological criteria for segmentation.	Empirical definition of parameters for high curvature points detection.
Zorman et al. (2007) FL images.	Thresholding.	New transform for follicle definition.	Empirical definition of color channel without contrast analysis.

(Canellos, Lister, & Young, 2006). For this reason, a method for segmentation of nuclear structures of CLL and MCL abnormalities is still a major challenge in pathology.

The studies on Table 1 present some limitations addressed by this work. For instance, different from the methods of Belkacem-Boussaid et al. (2010) and Sertel et al. (2008a), the proposed method in this study does not require any user interaction.

A common difficulty in lymphoma image processing is the segmentation of overlapping cells, as indicated by the studies of Belkacem-Boussaid et al. (2010), Kong, Belkacem-Boussaid, and Gurcan (2011a), Mohammed et al. (2013b) and Oztan et al. (2012). This difficulty was processed during the post-processing step, considering the valley-emphasis strategy. Moreover, the threshold values of Dimitropoulos et al. (2014),

Luo, Celenk, and Bejai (2006), Sertel et al. (2009) and Sertel et al. (2008b) were empirically defined and can present disadvantages for practical applications.

The methods described by Dimitropoulos, Barmpoutis, Koletsis, Kostopoulos, and Grammalidis (2016), Sertel et al. (2008a) and Oger et al. (2012) for segmentation of lymphoma used histological samples stained with IHC and H&E. Using these approaches, different types of images are required for the investigation and segmentation of an abnormality.

Some methods demand a training step, as the studies of Dimitropoulos et al. (2016), Mohammed et al. (2013b) and Sertel et al. (2010a). In these studies, the training step was composed of a small number of images, and this can be a limiting factor in the representation of ROIs features. Moreover, the methods of Belkacem-Boussaid et al. (2011) and Oger et al. (2012) employed only a few images for the testing step. The works of Belkacem-Boussaid et al. (2011) and Kong, Gurcan, and Belkacem-Boussaid (2011b) considered, respectively, only three and five images for defining the parameters of the algorithm, which can be inadequate for application in image datasets. In the approaches of Mohammed et al. (2013a, 2013b), only images with ROIs in their central regions were investigated. Besides, Mohammed et al. (2013a) assumed in their proposed approach uniform illumination of the images for Otsu application.

In their segmentation methods, Sertel et al. (2009, 2010b) used texture features obtained from images with magnification of $40\times$. Through this condition, these methods cannot reach high quality results in lower magnifications, since high magnifications can lead to more details of these features (Gurcan et al., 2009). Belkacem-Boussaid et al. (2010, 2011) indicated limitations for applications using low contrast images and low quality histological sample preparation. These conditions can be found in images available on public domain datasets. Moreover, the related studies used private image datasets for the evaluation of their methods. Researchers emphasize the importance of using public image datasets in order to demonstrate robustness of new approaches (Fuchs & Buhmann, 2011; Kothari, Phan, Stokes, & Wang, 2013; McCann, Ozolek, Castro, Parvin, & Kovacevic, 2015; Tafavogh, Catchpoole, & Kennedy, 2014).

1.2. Contributions of this work

In this work, an unsupervised computational algorithm is proposed to segment nuclei from neoplastic cells located on CLL, FL and MCL histological images. The main contributions are summarized as follows:

- A segmentation algorithm based on intrinsic features from nuclei of H&E histological lymphoma images with $20\times$ magnification, which allows for less details to perform segmentation;
- Application of techniques for contrast and illumination enhancements, allowing its application to be used on images with different conditions;
- A novel unsupervised method of automatic threshold selection for noise removal (non-neoplastic regions) at the segmentation step, based on GA associated to fuzzy 3-partition entropy technique;
- The evaluation of valley-emphasis thresholding and morphological operations in order to split identified cells and enhance the representativeness of nuclear shapes at the post-processing step;
- A performance evaluation using a public domain dataset, which is represented by variations commonly found in clinical practice;
- Contribution to state of the art of lymphoma images processing methods for segmentation of neoplastic cells in CLL and MCL

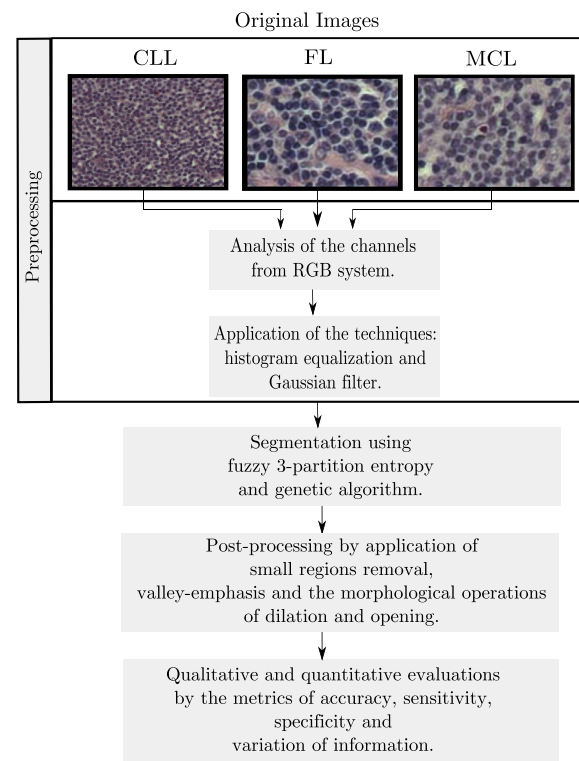


Fig. 1. Schematic illustration of the proposed method for unsupervised segmentation of nuclear neoplastic structures in histological images of lymphoma.

images, and new schemes for identifying centrocytes and centroblasts in cellular structures from FL.

This paper is organized as follows: the techniques and the dataset are described in Section 2. Section 3 discusses the obtained results in each step and presents the comparative evaluation of the proposed algorithm and other methods from the literature. Finally, the conclusion is presented in Section 4.

2. Materials and methods

In this section, the segmentation algorithm, the public domain dataset and the metrics applied to analyze the experiments are presented in details.

2.1. Method overview

The proposed method consists of three steps: preprocessing, segmentation and post-processing. In Fig. 1, are shown the steps developed to analyze the nuclear components of lymphoma images.

The algorithms were developed using MATLAB® language and the experiments were performed on an 1.7 GHz processor ultra-book (Acer M5-481T-6417) with 6GB RAM. The deconvolution plugin from ImageJ (2016) was also considered to analyze the images from the deconvolution process.

2.2. Dataset

The lymphoma cases considered in this study were obtained from a Zeiss AxioScope microscope with $20\times$ objective lens and an AxioCam MR5 CCD color camera. All images were obtained under the same equipments configurations, objective lens, camera and light source. These histological samples were stained with H&E and digitized using the RGB color model with 24 bits of quantization,

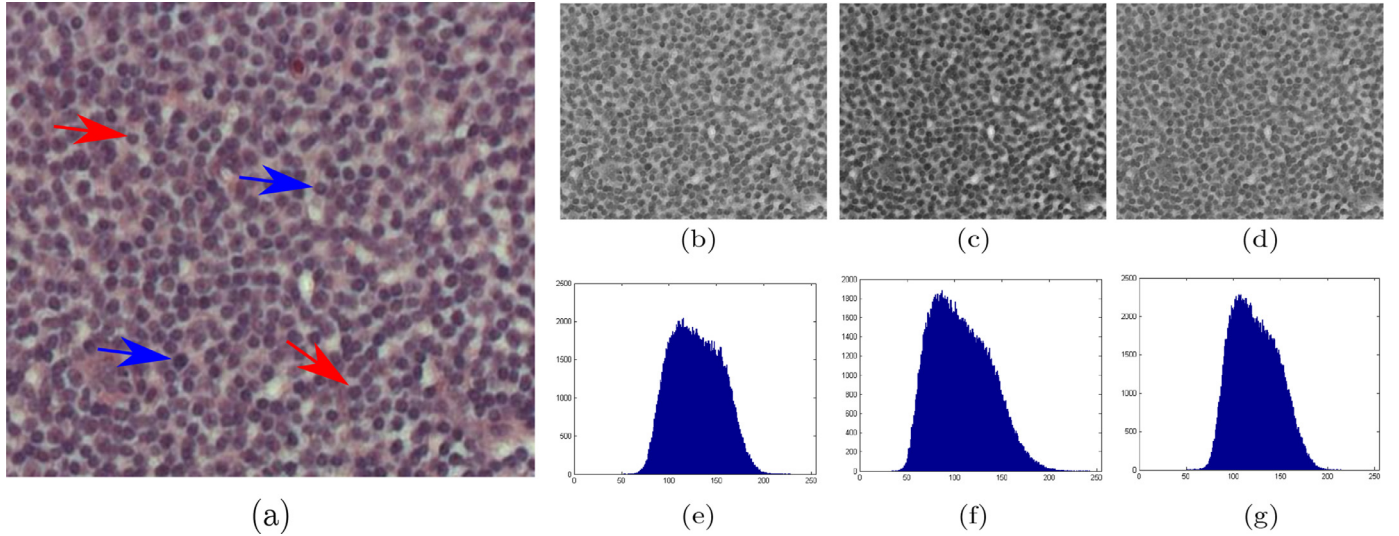


Fig. 2. Color channels separation from RGB color model of a subimage of the case *sj-05-5269-R10_002* from CLL lesion (a), with neoplastic nuclei indicated by red arrows and normal nuclei indicated by blue arrows, which exemplifies the contrast differences of the R (b), G (c) and B (d) channels, and their respective histograms illustrated by (e), (f) and (g). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

available for download (Shamir, Orlov, Eckley, Macura, & Goldberg, 2008).

The described images dataset was evaluated applying manual segmentation of subsets of each class. The image dataset considered in this work is composed of 12, 62 and 99 images of CLL, FL and MCL, respectively. Each case contains almost 2000 cells, similar amount as in studies of segmentation of cytological and histological images (Dimitropoulos et al., 2016; Gençtav, Aksoy, & Önder, 2012; Wang, Hu, Li, Liu, & Zhu, 2016). The lesions were manually marked by a specialist and automatically analyzed by the proposed method.

2.3. Preprocessing

The main purpose of the preprocessing step is to improve image representation for subsequent stages through contrast-enhancement and/or noise reduction (Gonzalez & Woods, 2000). There are different preprocessing techniques: the decomposition into each component of color models and the image normalization to standardize the image colors (Hoffman, Kothari, & Wang, 2014). In this work, the RGB, HSV, LAB, LUV, YCbCr and YIQ color models were applied for decomposition into their components. The deconvolution process (Ruifrok & Johnston, 2001) was also used to separate the two stains components of hematoxylin and eosin into different images. The entropy metric was used to choose the appropriate channel (see Section 2.6).

Fig. 2(b), (c) and (d), respectively, shows differences in contrast among the R, G and B channels of a CLL image (Fig. 2(a)). Histograms of the R, G and B color channels are illustrated by the Fig. 2(e), (f) and (g), respectively. As can be observed in Fig. 2(f), the G channel intensity level distribution is more uniform than the distribution of the R and B channels, indicating more contrast and, consequently, a better representation of the image (Gonzalez & Woods, 2000). This procedure was also applied to analyze the other lesions, leading to the choice of the B and R channels for FL and MCL images, respectively.

The histogram equalization technique was applied in order to normalize the color distribution across slides with different staining and illumination conditions (Jothi & Rajam, 2016). Equalization is a method for contrast-enhancement using histogram information. This method redistributes the intensity levels so that

the image histogram can have a uniform distribution (Gonzalez & Woods, 2000).

Given an image with $n = M \times N$ pixels, characterized by discrete values for the gray levels r , equalization uses a function of cumulative probability distribution as its transformation function, which is expressed by:

$$g_r = T(f_r) = \sum_{i=0}^r p_f(f_i) = \sum_{i=0}^r \frac{m_i}{n}, \quad (1)$$

where, m_i represents the frequency of gray level i and $p_f(f_i)$ is the probability of i -th gray level.

Then, a Gaussian filter was applied to reduce noise and smooth the lymphoma images. This technique consists in a convolution process that uses a mask characterized by its size and elements distribution, resulting in the sum of products between their elements and the intensity values of the image. Elements distribution of the mask is defined by a 2-dimensional Gaussian function, as given by:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (2)$$

where, x and y represent the image pixels coordinates and the standard deviation is expressed by σ .

In this paper, the parameter σ was investigated with assignments in the range $[0.5, 5.0]$, and the different sizes of the mask were explored applying 3×3 , 5×5 and 7×7 pixels.

Fig. 3 shows that the amount of smoothing and noise reduction is proportional to the mask parameters. Fig. 3(a), (e), (i) and (m) shows the R channel of the same image for a better visualization and comparison of results along the lines of this representation. Fig. 3(b), (c) and (d) illustrates the results applying masks with sizes 3×3 , 5×5 and 7×7 pixels, respectively, and the variable σ assigned to 0.5. Fig. 3(f), (g) and (h) are the obtained results with value 2 of σ and masks with sizes 3×3 , 5×5 and 7×7 pixels, respectively. Fig. 3(j)–(l) and (n)–(p) represents the obtained images of the investigation of parameter σ with values 3.5 and 5, respectively.

It is noticeable that the mask size can change image sharpness (see Fig. 3(f) and (h)). The sharpness level can hinder the distinction between nuclei and cytoplasm regions on the segmentation step (see Fig. 3(g), (h), (k), (l), (o) and (p)). In this step, a mask with size of 3×3 pixels and sigma 2, represented by Fig. 3(f), was

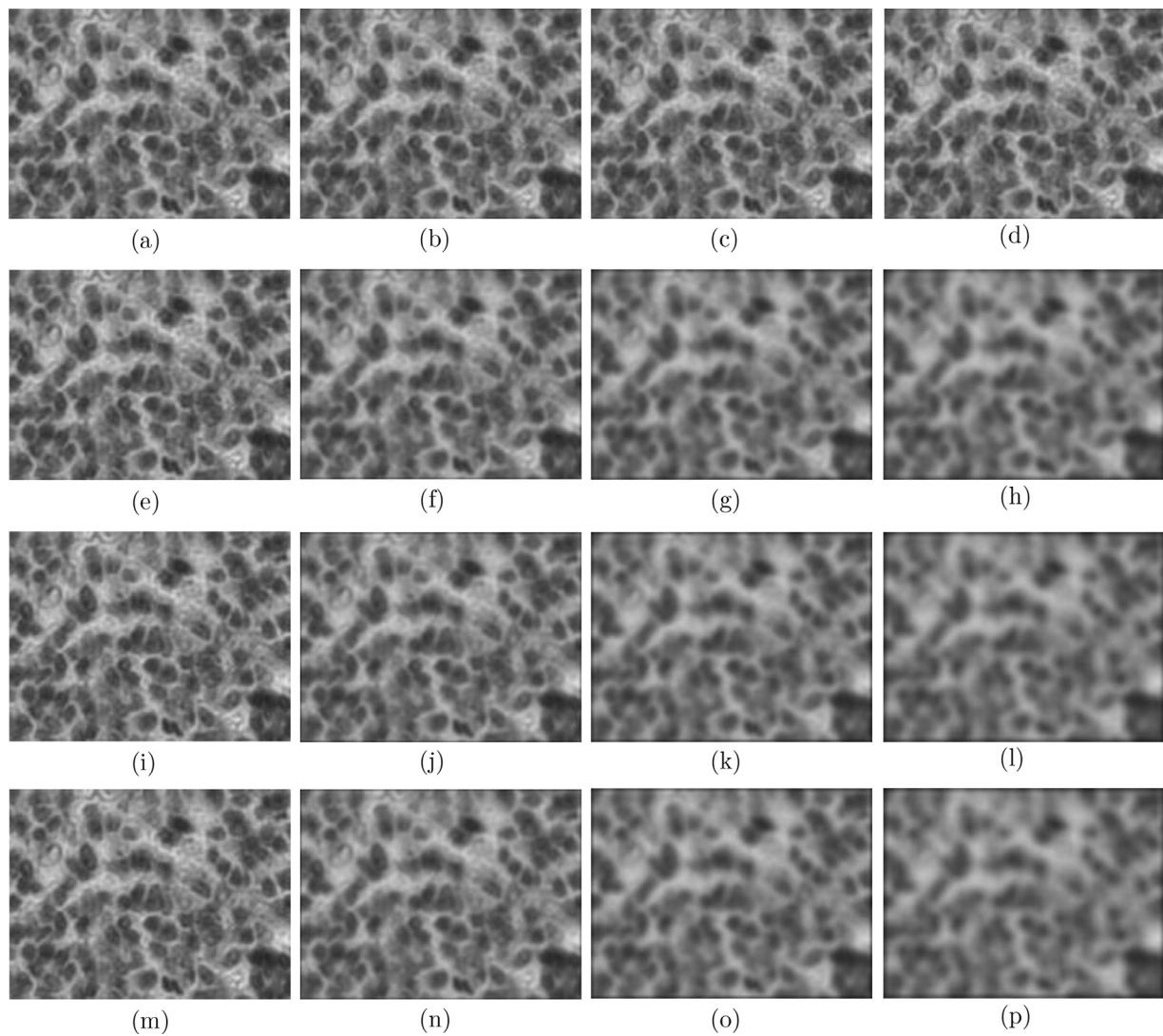


Fig. 3. Example of the use of Gaussian filter on a subimage of the case sj-04-4525-R4_001 from MCL class: (a), (e), (i), (m) R color channel from the original image, application with mask sizes of 3×3 ((b), (f), (j), (n)), 5×5 ((c), (g), (k), (o)) and 7×7 ((d), (h), (l), (p)) pixels, and assignments of sigma to 0.5 ((b)–(d)), 2 ((f)–(h)), 3.5 ((j)–(l)) and 5 ((n)–(p)).

applied since it yields better distinctions between nuclear regions and their surrounding areas.

Another way to improve the histological image quality is based on color normalization. An investigation of the method proposed by Macenko et al. (2009) is presented in Section 3.1.1.

2.4. Segmentation

The proposal of an unsupervised segmentation method is one of the most difficult tasks in digital image processing. Many techniques can be applied in this step, such as thresholding, region growing, graphs and watershed (Gonzalez & Woods, 2000). Thresholding is one of the main methods considered on histological images as an efficient tool for segmentation of different structures (Meijering, 2012; Oswal, Belle, Diegelmann, & Najarian, 2013; Smochinã, Hergehelegiu, & Manta, 2011). Thus, the approach chosen here was based on multilevel thresholding to distinguish nucleus, cytoplasm and background. The definition of the threshold values was obtained by GA and fuzzy 3-partition entropy technique (Yin, Zhao, Wang, & Gong, 2014) in order to obtain an unsupervised method. There are other optimization algorithms that can be adopted to find the multilevel thresholding. Therefore, the evolutionary methods of artificial bee colony (ABC) (Bose & Mali,

2016; Heris, 2015a), cuckoo search (CS) (Bhandari, Kumar, & Singh, 2015; Yang, 2009), differential evolution (DE) (Cuevas, Zaldívar, & Perez-Cisneros, 2016; Heris, 2015b), particle swarm optimization (PSO) (Biswas, 2014; Remamany, Chelliah, Chandrasekaran, & Subramanian, 2015) and wind driven optimization (WDO) (Bayraktar, 2013; Bayraktar, Komurcu, & Werner, 2010) were also investigated to this problem (see Section 3.2.1). Also, the cytoplasm areas were considered during the segmentation step in order to obtain better results of nuclear contours. Fig. 4 shows the main steps of GA used to calculate the threshold values (Paulinas & Ušinskas, 2015).

Firstly, it is necessary to define the GA initial population composed of a set of individuals. Then, the histogram was calculated and normalized using the preprocessed image. In this stage, the population size was empirically chosen as 60 individuals, in which each individual was represented as a chromosome encoded as six values related to the normalized histogram intensity levels.

The fuzzy 3-partition entropy technique was applied to define the values of each individual. The S and Z functions, represented by Eqs. (3) and (4), were considered to quantify the membership degree of gray levels (k) to each investigated region. In this work, two pairs of these functions were necessary for the segmentation of the three analyzed classes (nucleus, cytoplasm and background)

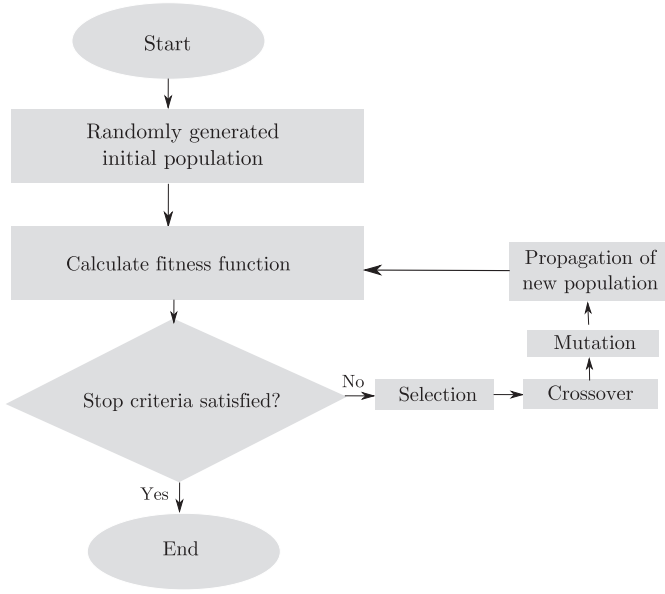


Fig. 4. Flowchart of GA algorithm for threshold values definition (adapted from Kaushik, Singh, Singhal, and Singh, 2013 © 2013 Citeseer).

(Yin et al., 2014).

$$S(k, u, v, w) = \begin{cases} 1, & k \leq u \\ 1 - \frac{(k-u)^2}{(w-u) \cdot (v-u)}, & u < k \leq v \\ \frac{(w-u) \cdot (k-v)^2}{(k-w)^2}, & v < k \leq w \\ 0, & k > w \end{cases} \quad (3)$$

$$Z(k, u, v, w) = 1 - S(k, u, v, w). \quad (4)$$

The functions given by S and Z were considered as membership functions, which classify the pixels into three fuzzy sets, referred as the three analyzed classes (M_n related to the membership degree to nuclear regions, M_c to cytoplasm regions and M_b to background). These functions are represented by Eqs. (5)–(7):

$$M_n(k) = S(k, u_1, v_1, w_1), \quad (5)$$

$$M_c(k) = \begin{cases} Z(k, u_1, v_1, w_1), & k \leq w_1, \\ S(k, u_2, v_2, w_2), & k > w_1, \end{cases} \quad (6)$$

$$M_b(k) = Z(k, u_2, v_2, w_2), \quad (7)$$

where, k represents the brightness levels of the image, in this study, $0 \leq k \leq 255$, and u_1, v_1, w_1, u_2, v_2 and w_2 , where $0 \leq u_1 < v_1 < w_1 < u_2 < v_2 < w_2 \leq 255$, are parameters that determine the distribution of membership degrees of each intensity level, as shown in Fig. 5.

In Fig. 5 the pixels that contain intensity levels between the parameter u_1 and the intersection point of M_n and M_c curves have a higher membership degree with the dark set, which is a feature of lymphoma image nuclear regions. The values of pixels contained in the interval defined by intersection points of the curves are associated to cytoplasm regions and values held in the interval between the intersection point of M_c and M_b curves and the parameter w_2 are associated to background and other irrelevant information.

The values that represent each individual were obtained by the membership functions parameters and analyzed considering the entropy function, represented in Eq. (8):

$$H(u_1, v_1, w_1, u_2, v_2, w_2) = -P_n \cdot \log(P_n) - P_c \cdot \log(P_c) - P_b \cdot \log(P_b), \quad (8)$$

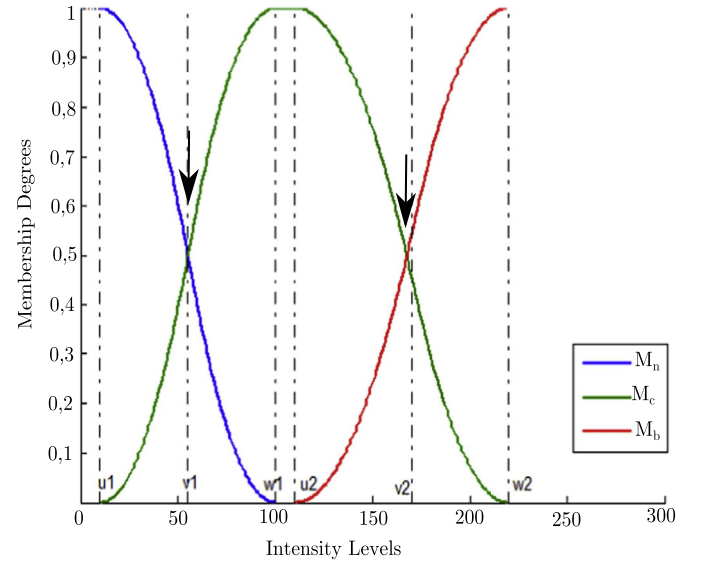


Fig. 5. Representative graph of membership functions used by fuzzy 3-partition entropy method with indication of threshold values by black arrows (adapted from Yin et al., 2014 © 2014 Elsevier).

where, the probability value (P) of each structure was obtained by:

$$P_n = \sum_{k=0}^{255} h(k) \cdot M_n(k), \quad (9)$$

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

$$P_b = \sum_{k=0}^{255} h(k) \cdot M_b(k), \quad (11)$$

where, $h(\cdot)$ represents the normalized histogram of the image.

In this step, the H value must be maximized, indicating a greater amount of extracted information and allowing the definition of the best combination of parameters ($u_1, v_1, w_1, u_2, v_2, w_2$). The threshold values are represented by black arrows in the intersection points of the M_n, M_c and M_b functions in Fig. 5. Such values can also be defined considering the membership degrees, where both of them correspond to their assignment to 0.5, as can be seen on the Y-axis in Fig. 5 (Yin et al., 2014).

The selection step must be applied to select individuals considered as adequate solutions. These individuals are maintained in the next generation through the elitism process, and they are used by the crossover and mutation steps (Paulinas & Ušinskas, 2015). In this step, 30% of the population was used, so 18 from 60 individuals that compose each population were considered in each subsequent generation. This value was chosen to reach population diversity, leading to a higher amount of new individuals and, consequently, new solutions.

The crossover operation must be applied to the population in order to generate new individuals. This stage will be repeated until the population size is complete. In this study, the crossover probability (Jianli & Baoqi, 2009) was defined with the value 0.65. This decision was based on empirical tests of the crossover probability defined between 0.5 and 1.0.

The value of crossover point (CP) (Jianli & Baoqi, 2009) was randomly selected among the parameters u_1, v_1, w_1, u_2, v_2 and w_2 for obtaining two descendants. Fig. 6 illustrates an example of crossover, in which two individuals were used to obtain two descendants based on the CP assignment to value 2.

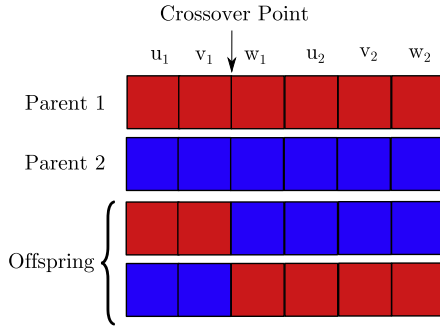


Fig. 6. Example of crossover performed on two individuals for obtaining their offspring.

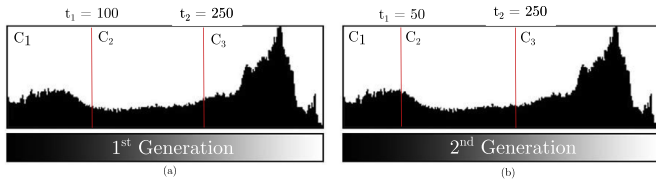


Fig. 7. Example of application of the Hammouche et al. (2008) algorithm for definition of stop criteria over two generations ((a) and (b)) using a histogram representation.

The mutation was applied to a subset after the crossover step. In this study, mutation was applied with a 0.01 probability to generate similar solutions by new attribute values (Lad, Agrawal, & Pandya, 2014).

Finally, the method proposed by Hammouche, Diaf, and Siarry (2008) was applied as GA stop criteria. In this work, the execution was interrupted when the average of intensity levels of the analyzed structures remains the same over two consecutive generations. Fig. 7 represents an example of a sufficient condition of stop criteria. In Fig. 7(a), the result obtained in a generation is presented and Fig. 7(b) illustrates the result obtained in the subsequent generation. The average values between the classes 1 and 2 (t_1), and 2 and 3 (t_2) are also presented. Thus, considering that the average between the classes 2 and 3 was the same in both generations (250), the method is finished.

2.5. Post-processing

To reduce the number of possible false positive regions in the segmentation step, it is necessary to use techniques capable to refine the segmented structures. Initially, the area was computed for each segmented object. Regions with areas smaller than 10 pixels were removed since empirical tests show that these regions represent small noise.

The segmentation stage was also incapable to remove intra-nuclear regions of the lymphoma images. These regions are characterized by intensity levels that are brighter than the nuclei that compose them. Therefore, the valley-emphasis technique (Ng, 2006) was applied on regions identified by the segmentation step mapped on the preprocessed images.

This technique considers the smallest probability of occurrence (p_T) of the intensity levels to determine a threshold value. Besides, the variance between classes is maximized, according to the second term in Eq. (12), as defined by Otsu (1979):

$$T = \text{Max}\{(1 - p_T) \cdot (\omega_1(T)\mu_1^2(T) + \omega_2(T)\mu_2^2(T))\}, \quad (12)$$

where, ω represents the occurrence probability of the considered classes (nuclei and intra-nuclear regions) and μ corresponds to the average of intensity levels of each class.

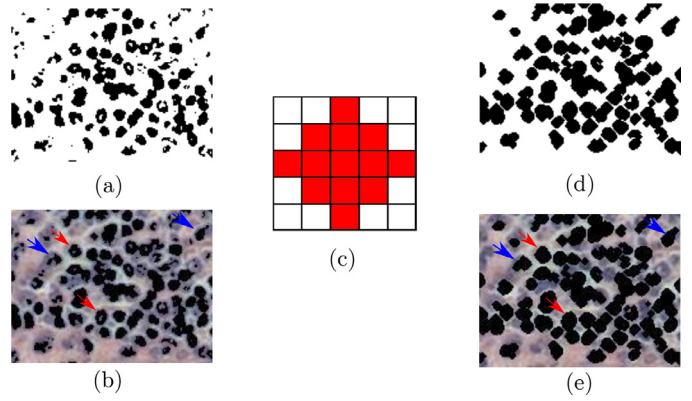


Fig. 8. Example of binary dilation with the increase of objects area, indicated by red arrows, and filling holes, identified by blue arrows: binary image (a), its correspondent regions identified on the original image (b), the result of this operation represented by a binary image (d) and the original image (e) using a structuring element with disk distribution with radius 2 (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

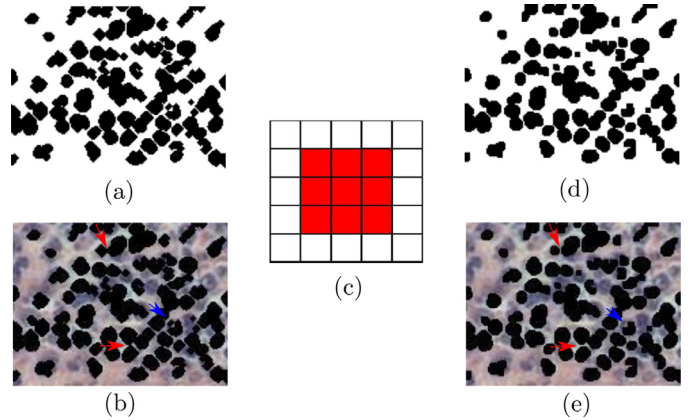


Fig. 9. Example of opening operation that allows to eliminate isolated pixels identified by red and blue arrows: binary image (a), original image (b) and the result of opening represented by a binary image (d) and its mapping on the original image (e) obtained using a structuring element with size of 3×3 with a square distribution (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, operations of dilation and opening were applied on the obtained image from the valley-emphasis method. The dilation operation allows filling holes and increasing the objects areas (Gonzalez & Woods, 2000; Pedrini & Schwartz, 2007), as indicated by red and blue arrows, respectively, in the Fig. 8(b) and (e). Fig. 8 illustrates the execution of this operation with a structuring element with disk distribution and radius 2 (Fig. 8(c)). Fig. 8(a) presents the binary image, in which segmented regions are represented by black pixels, and Fig. 8(b) presents its mapping on the original image. The obtained binary results are shown in Fig. 8(d) and the identification in relation to original image is presented in Fig. 8(e).

The opening operation was applied to eliminate isolated pixels and to merge small neighboring regions. Fig. 9 illustrates this operation using a binary image (Fig. 9(a)) and its correspondent identification on the original image (Fig. 9(b)). In this example, a 3×3 pixels structuring element with square distribution was used (Fig. 9(c)), obtaining objects represented by a binary image (Fig. 9(d)) and the original image (Fig. 9(e)). Red and blue arrows indicate differences between regions before and after its application.

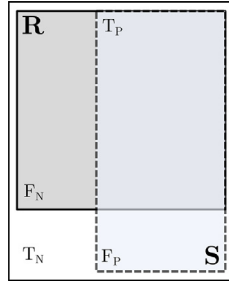


Fig. 10. Relation between regions identified by a specialist (R) and computational techniques (S) (adapted from Oger et al., 2012 © 2012 Elsevier).

2.6. Evaluation methods

In the preprocessing step, a metric based on entropy was applied to define the appropriate color model and filters. This metric was applied considering that its low value indicates high homogeneity and, consequently, absence of noise on the images, which could represent problems for the segmentation (Tsai, Lee, & Matsuyama, 2008). In addition, minimum entropy also indicates effective performance of filters capable to handle illumination differences, which can also represent complications in the segmentation (Leong, Brady, & McGee, 2003). Thus, the low value of entropy corresponds to the desired condition for the results from preprocessing stage, calculated by Eqs. (13) and (14) (Belkacem-Boussaid et al., 2010):

$$E = \sum_{i=1}^N p_i \cdot \log_2(p_i), \quad (13)$$

$$p = \frac{H_t(I(p, q))}{\sum_{t=1}^N H_t(I(p, q))}, \quad (14)$$

where, H represents the image histogram, N is the total number of intensity levels, t is the index that defines the considered intensity level, and i is the index that demonstrates the histogram variation.

In the segmentation stage, there are several metrics that can be used to evaluate quality of the segmented image (Estrada & Jepson, 2009; Ghose et al., 2012). In this study, four metrics were considered: accuracy, sensitivity, specificity (Insana, Meyers, & Grossman, 2000) and variation of information (Wu, Zhao, Luo, & Shi, 2015).

The evaluation of a segmentation method can be performed by calculating the overlapping regions of the segmented image and the regions of a reference image demarcated by a specialist (gold-standard). Then, the following parameters were considered: true positive (T_P), corresponding to the amount of correctly detected pixels, true negative (T_N) that represent the amount of correctly undetected pixels, false positive (F_P) that is related to the number of incorrectly detected pixels, and the false negative (F_N) that indicates incorrectly undetected pixels (Chang et al., 2014). Fig. 10 illustrates the relation between these concepts, gold-standard regions (R) and objects identified by unsupervised segmentation techniques (S).

The metrics of sensitivity (Se), specificity (Sp) and accuracy (Ac) are defined in Eqs. (15), (16) and (17), respectively. Sensitivity defines the amount of manually detected pixels that were also correctly segmented by the proposed algorithm. Specificity is determined to quantify the percentage of correctly defined true negatives (Oger et al., 2012). Accuracy corresponds to the quantification of proposed segmentation hit rate to the manual segmentation (Byrd, Zeng, & Chouikha, 2007).

$$Se = \frac{T_P}{T_P + F_N}, \quad (15)$$

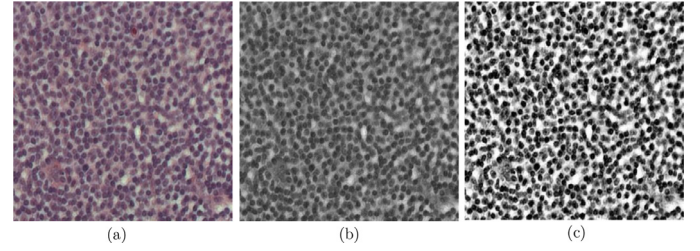


Fig. 11. Example of application of the preprocessing methods on a CLL image: (a) original image, (b) extracted G channel and (c) application of histogram equalization and Gaussian filter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

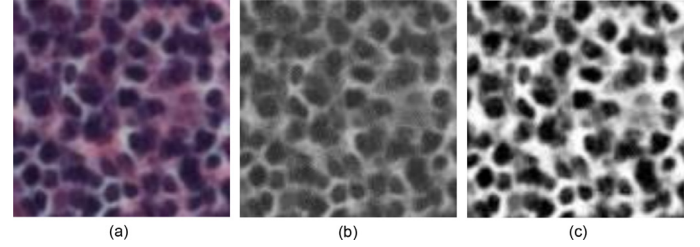


Fig. 12. Example of application of histogram equalization and Gaussian filter on B channel from a FL image: (a) original image, (b) extracted B channel and (c) application of preprocessing techniques. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$Sp = \frac{T_N}{T_N + F_P}, \quad (16)$$

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

The metric of variation of information (Vol) is defined to quantify the distance between R and S , as demonstrated by Eq. (18), where $H(X)$ represents the entropy of X and $I(X, Y)$ corresponds to mutual information between X and Y .

$$Vol(R, S) = H(R) + H(S) - 2I(R, S) \quad (18)$$

It should be noted that low values of Vol are required to express more similarities between unsupervised and manual segmentations (Wu et al., 2015). Eqs. (15)–(17) indicate that high values obtained from these calculations represent better results of the unsupervised segmentation.

3. Results and discussion

The results of each step and their limitations are presented in details in this section. Furthermore, comparative results are discussed applying quantitative and qualitative analyzes.

3.1. Preprocessing

Fig. 11(a), (b) and (c) shows a CLL class original image, the extraction of its G color channel and the preprocessed image, respectively. Results of preprocessing step of a FL image are presented by Fig. 12(a)–(c). A MCL image is illustrated by Fig. 13(a), as well as the extraction of its R channel and the result of this step by Fig. 13(b) and (c), respectively.

Table 2 presents entropy values obtained by the application of histogram equalization and Gaussian filter on each considered color channel. Minimum entropy values, represented in bold, were obtained by G , B and R channels for CLL, FL and MCL lesions, respectively. This metric was used in this study, since its minimum value indicates a good performance measure related to preprocessing filters for dealing with illumination differences and noise, as presented by Belkacem-Boussaid et al. (2010).

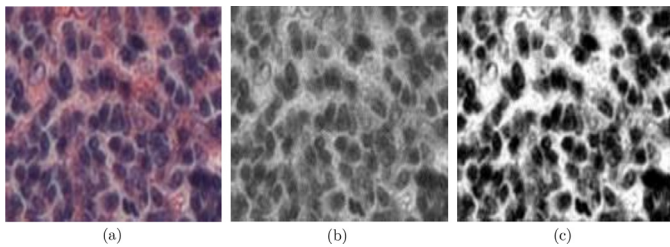


Fig. 13. Example of preprocessing application on a MCL image: (a) original image, (b) extracted R channel and (c) result of application of preprocessing techniques that reached minimum entropy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Entropy values obtained from application of the preprocessing methods on each channel from the color models of RGB, HSV, LAB, LUV, YCbCr, YIQ, hematoxylin and eosin.

Channels	CLL	FL	MCL
R (RGB)	−7.7922	−7.7889	−7.9830
G (RGB)	−7.9900	−7.7889	−7.4814
B (RGB)	−7.7914	−8.0021	−7.4822
H (HSV)	−7.7155	−7.7505	−7.7969
S (HSV)	−7.7847	−7.7981	−7.7753
V (HSV)	−7.7910	−7.7889	−7.7816
L (LAB)	−7.4909	−7.7954	−7.7823
a (LAB)	−7.5604	−7.8108	−7.4377
b (LAB)	−7.4346	−7.8081	−7.4539
L (LUV)	−7.0308	−7.6281	−7.0504
U (LUV)	−6.8558	−7.5759	−6.8782
V (LUV)	−7.4842	−7.6906	−7.6759
Y (YCbCr)	−7.7702	−7.7949	−7.7813
Cb (YCbCr)	−7.3676	−7.5534	−7.3814
Cr (YCbCr)	−7.2570	−7.5603	−7.2217
Y (YIQ)	−7.6907	−7.7949	−7.6817
I (YIQ)	−5.1639	−6.5765	−5.9798
Q (YIQ)	−7.4339	−7.6344	−7.2159
Hematoxylin	−7.4469	−7.7884	−7.6084
Eosin	−7.4848	−7.3846	−7.5996

3.1.1. Normalization analysis

The normalization method of Khan (2013); Macenko et al. (2009) was applied to compare the obtained results from the preprocessing stage. Normalization techniques allow the removal of color inconsistencies and it can be considered an important processing step for histopathological images (Li & Platanotis, 2015). In this method, it is necessary to choose a reference image so its colors can be mapped on the processed images. In this regard, the lymphoma images were analyzed in the search for the image with the highest contrast between the background and the ROIs. In this application, Fig. 14 illustrates the chosen image and its histogram. The sparse distribution of the histogram (Fig. 14(b)) indicates a high contrast of this image (Gonzalez & Woods, 2000).

Original images histograms were non-uniform. This condition makes the application of the proposed segmentation inadequate since this color distribution indicates low contrast, making the identification of ROIs more difficult. In this distribution, the threshold values defined by the proposed segmentation are not capable of identifying the neoplastic nuclei. Even after normalization, the images had non-uniform histograms. The displacement of histogram peaks to the left, darkening pixels that were brighter, still indicates low contrasts. To demonstrate this limitation, Fig. 15 illustrates a CLL image after the proposed preprocessing step (Fig. 15(a)) and after the normalization with conversion to grayscale (Fig. 15(c)). Their histograms are also shown, where it is noted that the normalized histogram presents a non-uniform distribution, and consequently a lower contrast.

Through the entropy metric used for the preprocessing evaluation, the normalization technique also presented higher results than the proposed preprocessing step. CLL, FL and MCL lesions presented quantitative results of −7.0409, −7.1282 and −7.0442, respectively. These values are higher than the minimum values presented by Table 2, indicating poorer performance for illumination differences and noise presence.

3.2. Segmentation

Figs. 16(a), 17(a) and 18(a) present original images of CLL, FL and MCL, respectively. Figs. 16(b), 17(b) and 18(b) present images resulting from preprocessing step of each class, and Figs. 16(c), 17(c) and 18(c) illustrate the histograms of preprocessed images, as well as the threshold values obtained by the segmentation step, indicated by red lines. Figs. 16(d), 17(d) and 18(d) illustrate segmentation results by binary images. Figs. 16(e), 17(e) and 18(e) present the original images with regions identified by the segmentation. Finally, Figs. 16(f), 17(f) and 18(f) show in details some regions with correspondents highlighted for this step analysis.

Figs. 16(f), 17(f) and 18(f) present irregular contours of identified regions, which is in contrast to the expected results for this application. Thus, the post-processing step was necessary to correct the resulting contours. Furthermore, the minimum size of true positive regions were determined in order to remove the false positive regions, which were associated to small noises. Nuclear segmentation containing two identified nuclei as one object was also observed, as represented by some regions with red contours in the Figs. 16(f), 17(f) and 18(f).

3.2.1. Analysis of evolutionary algorithms

The segmentation results obtained by the optimization methods of ABC, CS, DE, GA, PSO and WDO are presented on Table 3.

Considering the CLL lesion, the best results of accuracy, sensitivity, specificity and variation of information metrics were 81.06% (GA), 47% (ABC, CS, DE, PSO and WDO), 89% (GA) and 1.11 (ABC), respectively. As noted on Table 3, the highest value of sensitivity was shared among the vast majority of the used methods, with exception of the GA. However, the best specificity result was obtained by this method, which directly influenced its accuracy, which stood out from among the others.

When the FL and MCL were considered, the best results of accuracy, specificity and variation of information of FL and MCL lesions were obtained by the GA. The best sensitivity (57%) of FL lesion was shared among ABC, DE and PSO. In the MCL lesion, sensitivity reached the highest value (50%) in the DE technique. However, again, the GA achieved the lowest sensitivity and the highest specificity.

The presented results indicate the poor performance of GA in the identification of the gold-standard regions, represented by the sensitivity metric. The results shown in Table 3 can also indicate that DE provided the best values for the three groups. Considering these values, DE can obtain good threshold values by the use of only one best-proposed solution over the steps of selection, crossover and mutation. Meanwhile, the GA uses a set of best solutions in the steps of selection and crossover, exploring more possible solutions at the cost of more variations in its results. Through a consideration of the best results presented across most of the metrics of the analyzed lesions, the GA technique is considered the best optimization method for this application.

The other methods have presented intermediate results due to their randomness. ABC uses a random approach in the proposal of new solutions by the employed bees. In CS method, the proposed solutions, represented by bird nests, also use randomness

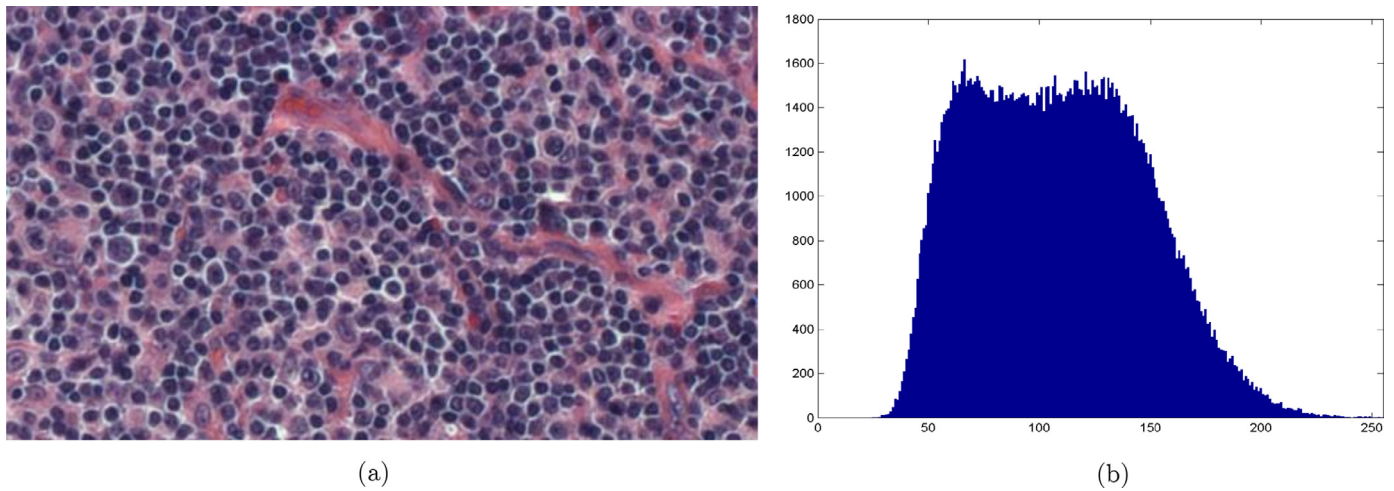


Fig. 14. Reference image for normalization analysis in the preprocessing step (a) and its histogram (b).

Table 3

Average results of segmentation through the metrics of accuracy, sensitivity, specificity and variation of information using the ABC, CS, DE, GA, PSO and WDO on lymphoma images.

Lesion	Optimization methods	Accuracy	sensitivity	Specificity	Variation of Information
CLL	ABC	79.92%	47%	86%	1.11
	CS	80.07%	47%	87%	1.20
	DE	79.94%	47%	86%	1.21
	GA	81.06%	41%	89%	1.14
	PSO	80.02%	47%	86%	1.20
	WDO	79.88%	47%	86%	1.21
FL	ABC	81.44%	57%	85%	1.11
	CS	81.69%	56%	85%	1.10
	DE	81.47%	57%	85%	1.11
	GA	82.83%	51%	87%	1.04
	PSO	81.70%	57%	85%	1.11
	WDO	81.52%	56%	85%	1.13
MCL	ABC	78.40%	47%	84%	1.20
	CS	79.56%	42%	84%	1.15
	DE	79.29%	50%	80%	1.15
	GA	80.76%	42%	86%	1.09
	PSO	79.55%	46%	84%	1.15
	WDO	79.30%	47%	84%	1.16

independent of the best solution of each generation. In this approach, if a nest with a cuckoo egg defined by a random parameter is discovered, this nest is removed, even if it is the best solution of the current generation. Although it uses the best local and global solutions in its exploration step, PSO presents the influence between the updates of particle position and its velocity. Considering this dependency relationship, PSO convergence becomes limited by controlling the exploration ability of this method (Shi & Eberhart, 1998). The influence between variables is also noted in the WDO method, given by the updates of the position and velocity of a particle, which also limits the exploration of new solutions.

3.3. Post-processing

Fig. 19 illustrates the application of valley-emphasis on a case of each lesion. Fig. 19(a), (b) and (c) shows a CLL segmented image, a region in details and the corresponding area obtained by applying valley-emphasis method, respectively. Fig. 19(d)–(f) illustrate a case of FL lesion containing a segmented image, a segmented area in details and the same region after application of this technique. Fig. 19(g)–(i) corresponds to a segmented image from a MCL case, one of its regions and the result of application of this step on this area.

Fig. 19(c), (f) and (i) indicates that some cells were characterized by holes and irregular contours (marked by red arrows). Thus,

morphological operations of dilation and opening were applied and Fig. 20 illustrates an example of each group after the application of these methods. Fig. 20(a), (d) and (g) presents segmented images with application of valley-emphasis method on the CLL, FL and MCL images, respectively. Fig. 20(b), (e) and (h) illustrates regions from CLL, FL and MCL cases in details, respectively, with irregular regions marked by red arrows. Finally, Fig. 20(c), (f) and (i) presents the results of morphological operations marked by red arrows indicating the obtained corrections.

3.4. Comparative analysis of results

It is important to note that there were not found any computational techniques that had been developed for the analysis of CLL and MCL groups that contemplate histological images from lymph nodes biopsy. Considering segmentation methods for FL images, there are no studies exclusively related to the segmentation of both its characteristic structures, centrocytes and centroblasts, as proposed by this study. Therefore, some studies related to segmentation of nuclear structures in histological images stained with H&E were used for comparing results. The mean-shift method (Comaniciu & Meer, 2002) was used due to its wide application by works proposed for segmentation of histological images, such as Xing and Yang (2013) and Sertel et al. (2010a). The techniques proposed by Vahadane and Sethi (2013),

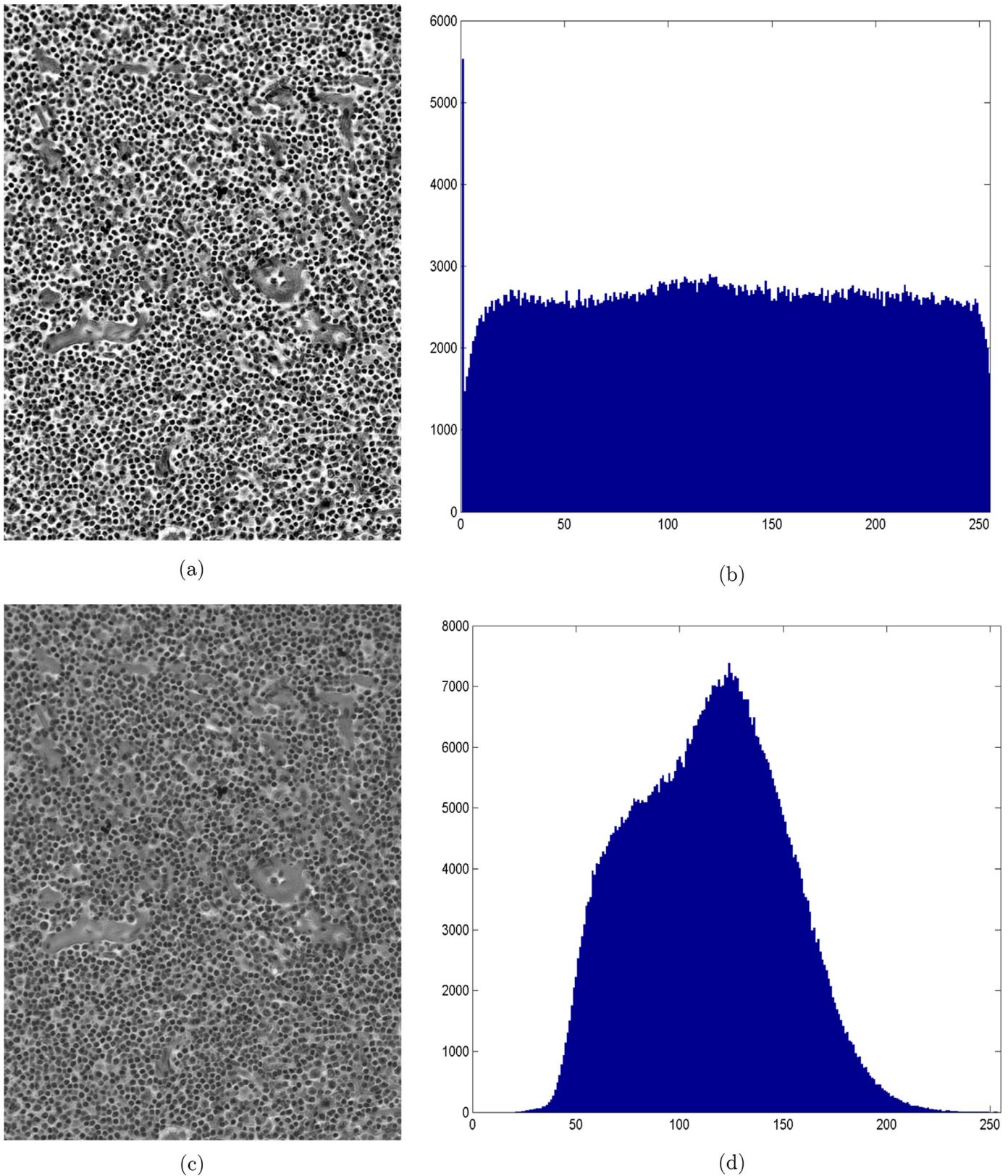


Fig. 15. A CLL image after preprocessing step (a), its histogram (b), and the same image after normalization of Macenko et al. (2009) (c), and its histogram (d), indicating a lower contrast.

Wienert et al. (2012), de Oliveira et al. (2013), Phoulady et al. (2016) and Paramanandam et al. (2016) were also applied for presenting nuclear segmentation methods of histological images.

Figs. 21(a), 22(a) and 23(a) present original images with 20 × magnification of CLL, FL and MCL classes, respectively. Refer-

ence images manually segmented by the specialist are represented by Figs. 21(b), 22(b) and 23(b). Figs. 21(c), 22(c) and 23(c) illustrate results obtained by application of the proposed segmentation method on CLL, FL and MCL images, respectively.

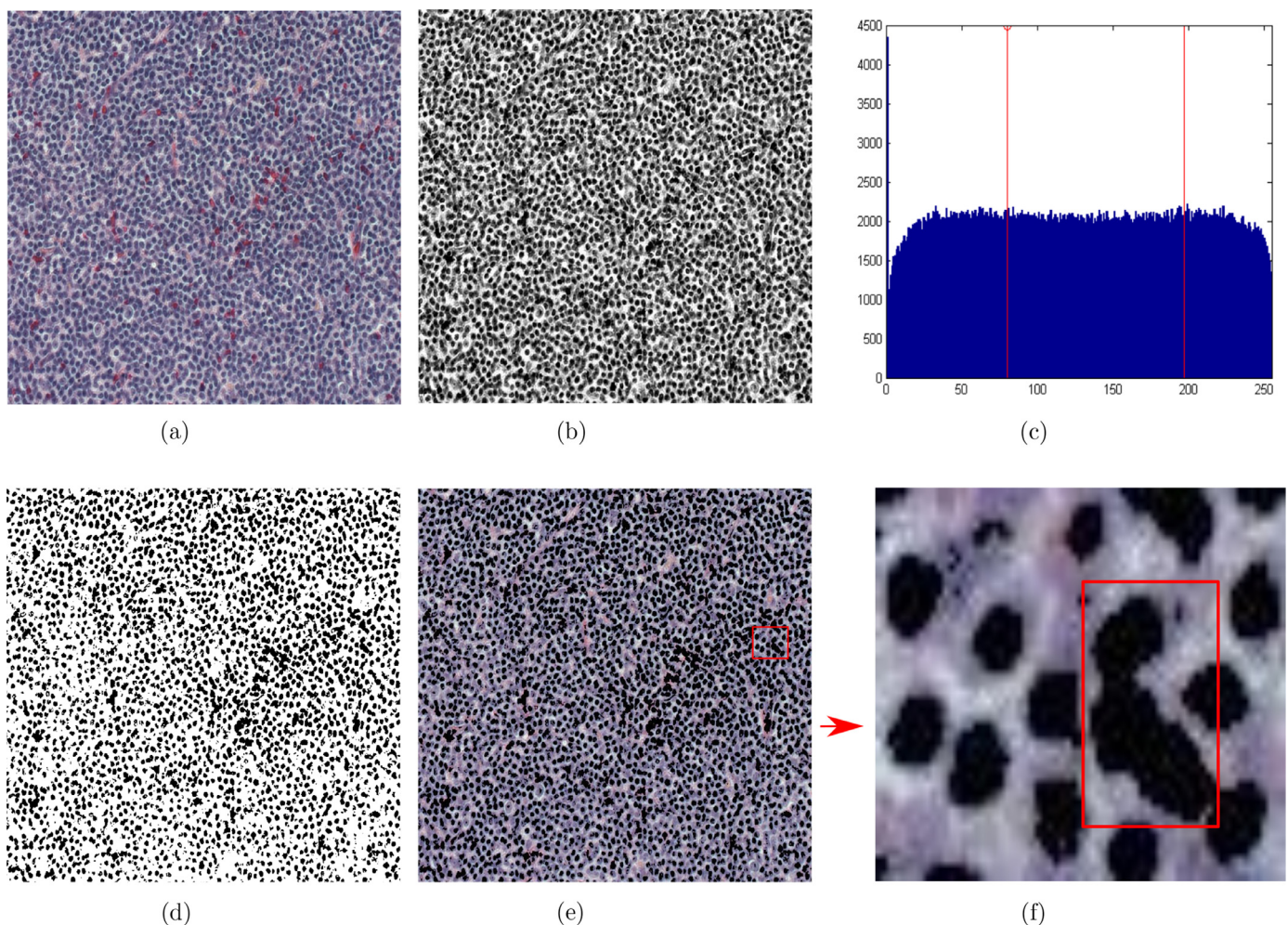


Fig. 16. Example of segmentation application on a subimage of sj-05-3344_006 case from CLL class: (a) original image, (b) image resulting from preprocessing step, (c) histogram of preprocessed image with threshold values of 80 and 197, indicated by red lines, (d) binary image that were obtained from the segmentation, (e) mapping of regions identified in segmentation on the original image and (f) the regions in details for result analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figs. 21(d), 22(d) and 23(d) show the results obtained by the mean-shift technique (Comaniciu & Meer, 2002). Figs. 21(e), 22(e) and 23(e) present the results after application of the method found in Vahadane and Sethi (2013). Figs. 21(f), 22(f) and 23(f) illustrate the results of the Wienert et al. (2012) technique. Figs. 21(g), 22(g) and 23(g) show the results of de Oliveira et al. (2013) algorithm. Finally, Figs. 21(h), 22(h) and 23(h), and Figs. 21(i), 22(i) and 23(i) present the results of Phoulady et al. (2016) and Paramanandam et al. (2016), respectively.

In comparison to manual segmentation (Figs. 21(b), 22(b) and 23(b)), the results provided by the proposed method (Figs. 21(c), 22(c) and 23(c)) show the identified nuclei and false positive regions (indicated by red arrows). These results (Figs. 21(c), 22(c) and 23(c)) show a larger number of nuclei detected (false positive) to the marking performed by the specialist. However, its spatial distribution was similar to the specialist segmentation. The segmentation result of FL (Fig. 22(c)) indicates that the proposed method enabled the identification of a significant quantity of true positive regions, as indicated by the blue arrows, corresponding to centrocytes and centroblasts.

Results of mean-shift method are presented by Figs. 21(d), 22(d) and 23(d). The results show false positive regions (indicated by red arrows). It is also noted that regions identified by this method have presented larger areas than the regions marked

by the specialist, which represents an overlapping problem. The amount of false positive regions is still expressive in results of Wienert et al. (2012) (Figs. 21(f), 22(f) and 23(f)). The false negative regions were shown to be of a larger quantity in this method (indicated by red arrows).

Figs. 21(e), 22(e) and 23(e) show the results obtained by Vahadane and Sethi (2013) technique on the different lymphoma classes. One observes that many regions were identified with several nuclei (marked by rectangular contours in red color), which result in a possible incorrect analysis of nuclear regions.

The method of de Oliveira et al. (2013) (Figs. 21(g), 22(g) and 23(g)) presented overlapping in all classes. Using its own pre and post-processing, this proposal did not enable the separation of identified regions containing more than one nucleus. This limitation can result in erroneous analyses in possible steps of feature extraction and classification.

The results of Phoulady et al. (2016) (Figs. 21(h), 22(h) and 23(h)) also present a considerable overlapping rate, mainly in the CLL lesion, where identified regions have larger areas than those manually segmented. Although it has consistent nuclear morphology, this method segmented false positive regions, indicated by red arrows, and false negative nuclei, indicated by blue arrows in Fig. 22(h). A common limitation of this method among the lesions was the identification of two nuclei as one object, indicated by red rectangles.

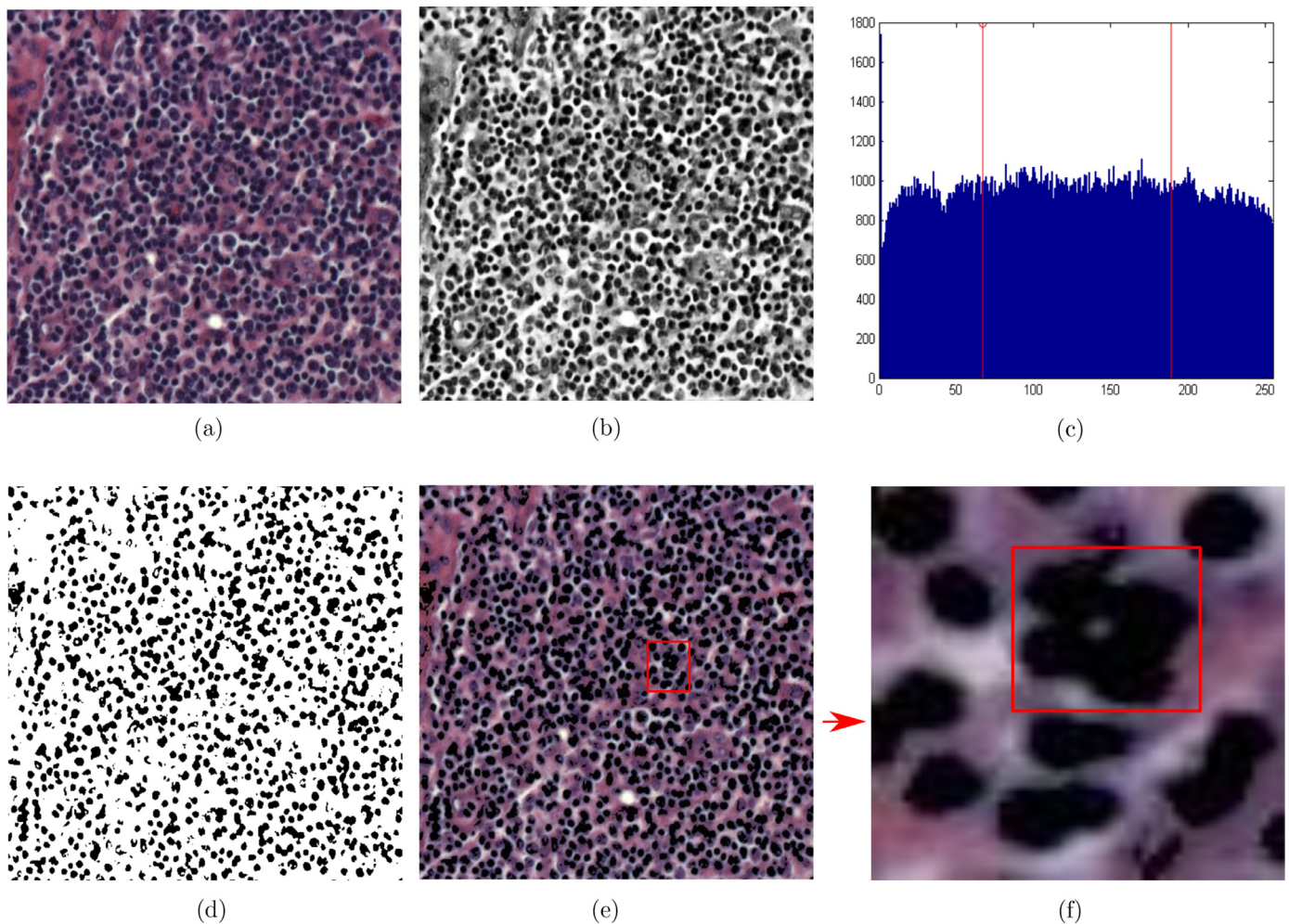


Fig. 17. Example of segmentation application on a subimage of *sj-05-6124-R3_006* case from FL class: (a) original image, (b) image resulting from preprocessing, (c) histogram of the previous image with thresholds 67 and 189, indicated by red lines, (d) binary image from the segmentation, (e) identification of regions obtained from segmentation on the original image and (f) the regions for result analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

Quantitative results obtained by segmentation methods applied on CLL lesion.

Techniques	Accuracy	Sensitivity	Specificity	Variation of Information
Proposed Method	81.06%	41%	89%	1.14
Mean-shift (Comaniciu & Meer, 2002)	79.74%	48%	86%	1.21
Vahadane and Sethi (2013)	77.07%	21%	88%	1.16
Wienert et al. (2012)	78.53%	47%	84%	1.25
de Oliveira et al. (2013)	70.60%	67%	71%	1.43
Phoulady et al. (2016)	71.97%	70%	72%	3.77
Paramanandam et al. (2016)	81.85%	4%	96%	0.42

The segmentation of Paramanandam et al. (2016) presented a poor performance across all lesions. The parameter of this algorithm corresponding to the width of a typical region of interest was set to 4, which was shown to be an adequate value for application on the images used. Some false negative regions are observed in Figs. 21(i), 22(i) and 23(i), as well as some false positive regions, indicated by red arrows. Although the irregular contours, some true positive regions were segmented, as indicated by blue arrows in Fig. 22(i). Fig. 23(i) highlights, using red rectangles, segmented regions that cover both true positive and false positive regions.

In order to evaluate the proposed method and the aforementioned techniques, Tables 4–6 show their quantitative per-

formance. These results were obtained by the metrics of accuracy, sensitivity, specificity and variation of information of the proposed method, mean-shift and techniques of Vahadane and Sethi (2013), Wienert et al. (2012), de Oliveira et al. (2013), Phoulady et al. (2016) and Paramanandam et al. (2016).

Evaluation of accuracy metric, which quantifies how close unsupervised segmentation is from manual segmentation, indicates that the proposed method reached average results of 81.06%, 82.83% and 80.76% for the classes CLL, FL and MCL, respectively. This result is higher than the values obtained by the methods used for comparison in FL class, but for the CLL and MCL classes, Paramanandam et al. (2016) reached the highest accuracy results. However, this quantitative metric may be insufficient for

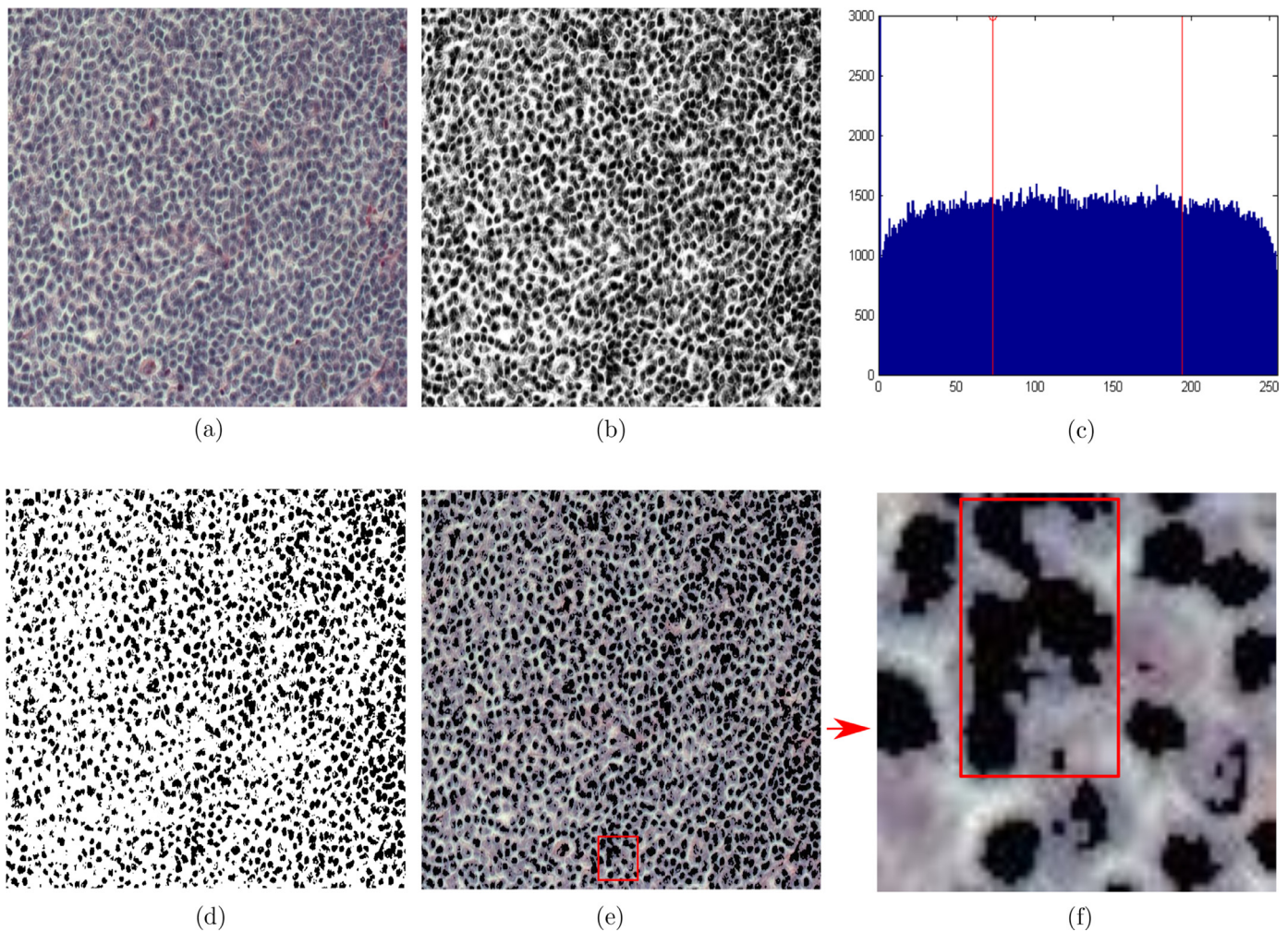


Fig. 18. Example of application of segmentation step using a subimage of sj-05-768_013 case from MCL class: (a) original image, (b) image resulting from preprocessing step, (c) preprocessed image histogram with threshold values 73 and 194, indicated by red lines, (d) binary image from the segmentation, (e) identification of regions obtained from segmentation on the original image and (f) the regions for result analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5

Performance of the segmentation methods for FL images.

Techniques	Accuracy	Sensitivity	Specificity	Variation of Information
Proposed Method	82.83%	51%	87%	1.04
Mean-shift (Comaniciu & Meer, 2002)	81.62%	58%	84%	1.10
Vahadane and Sethi (2013)	79.38%	25%	86%	1.08
Wienert et al. (2012)	81.05%	53%	85%	1.11
de Oliveira et al. (2013)	70.47%	69%	70%	1.33
Phoulady et al. (2016)	74.56%	70%	75%	3.42
Paramanandam et al. (2016)	82.45%	10%	92%	0.46

Table 6

Evaluation of the application of the segmentation methods on MCL lesion.

Techniques	Accuracy	Sensitivity	Specificity	Variation of Information
Proposed Method	80.76%	42%	86%	1.09
Mean-shift (Comaniciu & Meer, 2002)	79.50%	48%	83%	1.15
Vahadane and Sethi (2013)	80.67%	17%	89%	0.96
Wienert et al. (2012)	79.95%	41%	85%	1.12
de Oliveira et al. (2013)	67.10%	68%	66%	1.37
Phoulady et al. (2016)	73.35%	59%	75%	3.36
Paramanandam et al. (2016)	82.24%	9%	91%	0.45

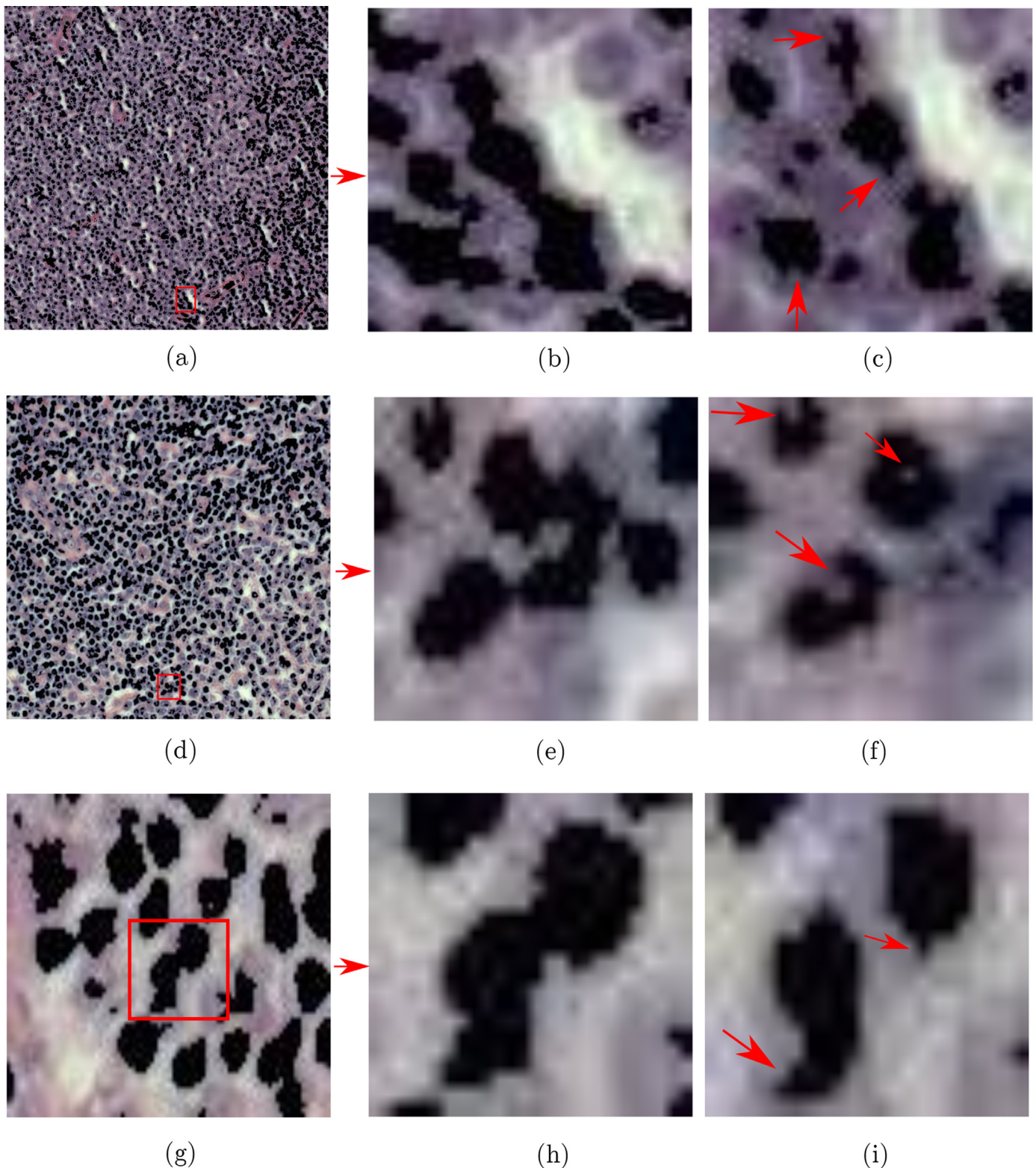


Fig. 19. Example of application of valley-emphasis segmentation method for nuclei splitting: (a), (d) and (g) represent segmented images of CLL, FL and MCL, respectively, (b), (e) and (h) illustrate zoomed regions of the correspondents segmented images, (c), (f) and (i) present their regions results from this process. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

the analysis of segmentation techniques. Considering the results of Paramanandam et al. (2016), segmentation errors occurred in the nuclei identification, as observed in Figs. 21(i), 22(i) and 23(i).

For the sensitivity and specificity metrics, the proposed method presented low values in the first metric and high values in the sec-

ond one. The method of mean-shift and Wienert et al. (2012) presented similar results. The lowest sensitivities were reached by Vahadane and Sethi (2013) and Paramanandam et al. (2016). In combination with their high specificity values, the accuracy metric of these methods are higher than or very close

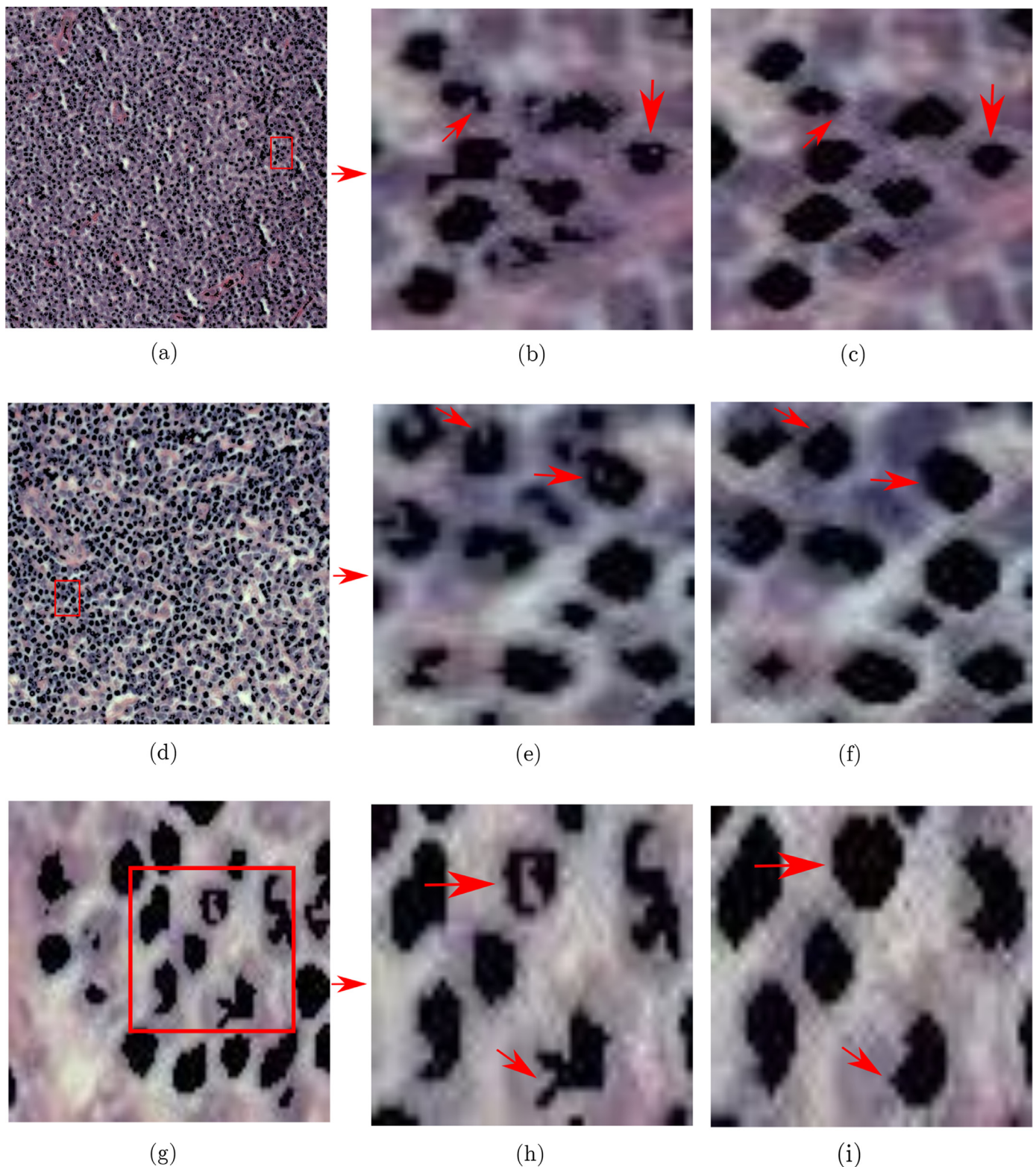


Fig. 20. Example of morphological operations application for correction of segmented cells: (a), (d) and (g) illustrate segmented images of CLL, FL and MCL, respectively, (b), (e) and (h) present their segmented regions highlighted and zoomed, (c), (f) and (i) present their results after this processing. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

to the proposed method, even with their very poor qualitative results. The results with the highest sensitivity were obtained by [de Oliveira et al. \(2013\)](#) and [Paramanandam et al. \(2016\)](#), but their specificities were the lowest when compared to other methods, indicating higher overlapping rates.

Variation of information enables the quantification of distances between unsupervised and manual segmentations. The proposed algorithm presented relevant results for all classes, but the lowest values were obtained by [Paramanandam et al. \(2016\)](#). However, the performance of this algorithm in qualitative evaluation does not

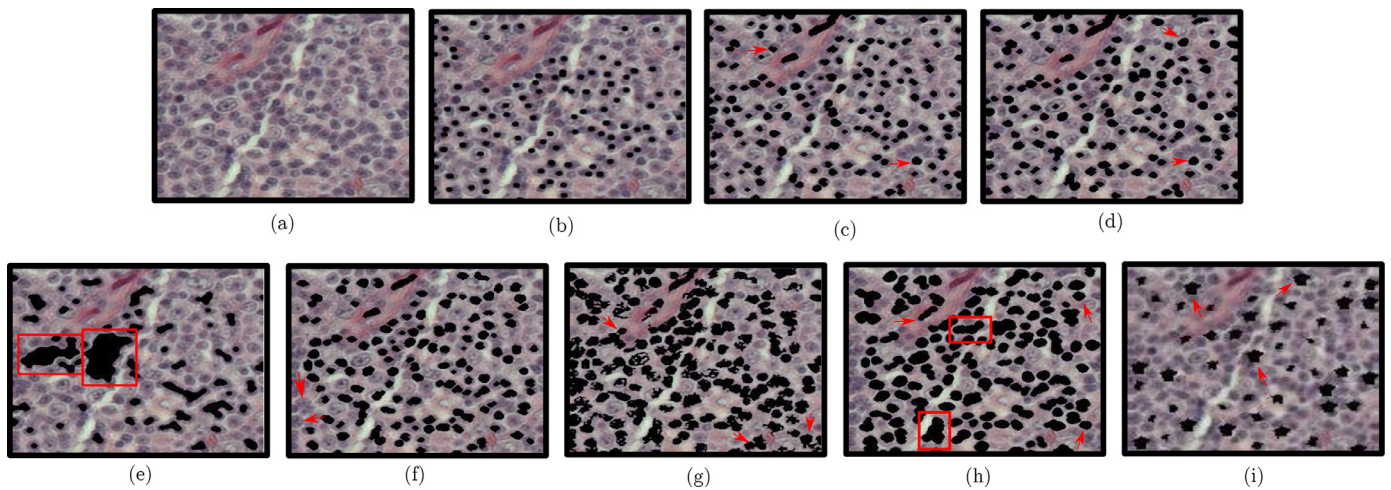


Fig. 21. Comparison of segmentation techniques on a CLL sample: (a) original image, (b) image segmented by the specialist, (c) proposed method, (d) mean-shift, (e) technique of Vahadane and Sethi (2013), (f) method of Wienert et al. (2012), (g) algorithm of de Oliveira et al. (2013), (h) proposal of Phoulady et al. (2016) and (i) study of Paramanandam et al. (2016). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

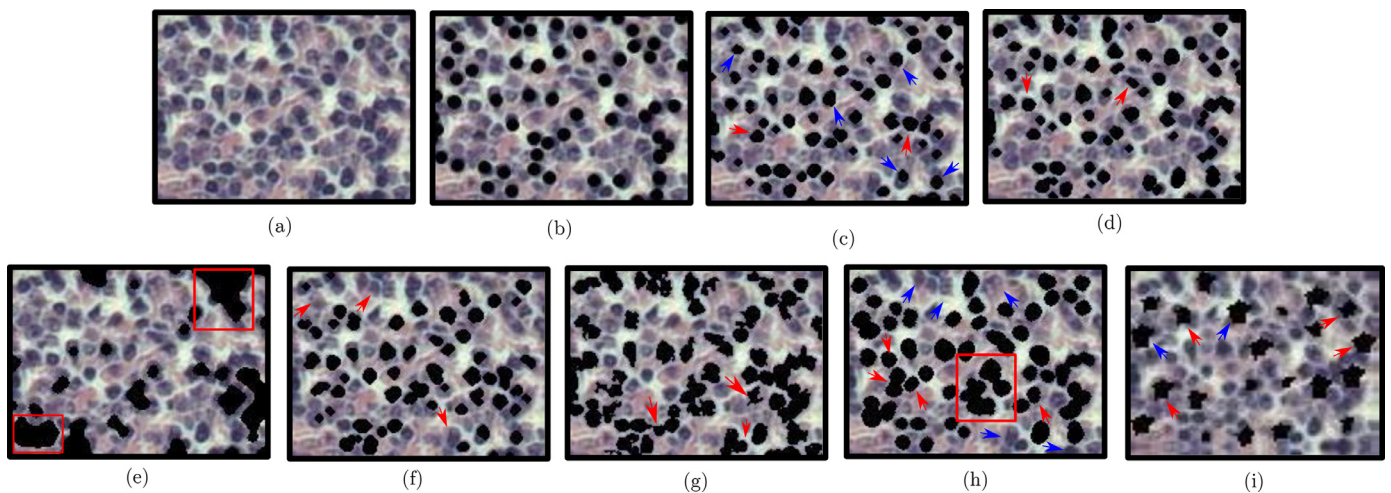


Fig. 22. Results of segmentation techniques for FL image: (a) original image, (b) manually segmented image, (c) proposed method, (d) mean-shift, (e) study of Vahadane and Sethi (2013), (f) method of Wienert et al. (2012), (g) proposal of de Oliveira et al. (2013), (h) algorithm of Phoulady et al. (2016) and (i) technique of Paramanandam et al. (2016). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

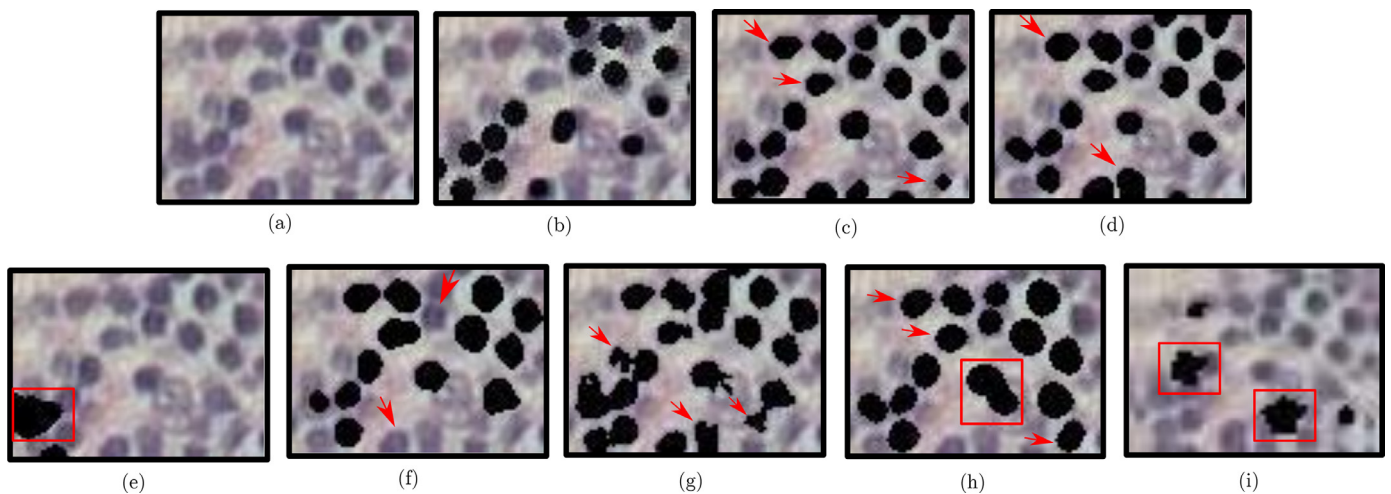


Fig. 23. Segmentation MCL image with different segmentation techniques: (a) original image, (b) image segmented by the specialist, (c) proposed algorithm, (d) mean-shift, (e) algorithm of Vahadane and Sethi (2013), (f) study of Wienert et al. (2012), (g) technique of de Oliveira et al. (2013), (h) method of Phoulady et al. (2016) and (i) algorithm of Paramanandam et al. (2016). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

represent this condition, since the results showed incorrect identification of nuclei regions in the lymphoma images.

Through a consideration of the obtained results, one notes the good performance of the proposed method for splitting ROIs. The same is not observed in the other techniques, a consequence of a high overlapping rate, as indicated by the results of [de Oliveira et al. \(2013\)](#). The results of [Phoulady et al. \(2016\)](#) indicate that the proposed method using the valley-emphasis technique was effective, since this study also used only intensity information for this purpose. In contrast to the large amount of false negative regions identified by [Vahadane and Sethi \(2013\)](#) and [Paramanandam et al. \(2016\)](#), the proposed algorithm and the techniques of mean-shift and [Wienert et al. \(2012\)](#) presented a larger amount of true positive identifications. However, mean-shift and the method of [Wienert et al. \(2012\)](#) presented larger quantities of false negative regions, as indicated by the subtle differences in the specificity metric and by the qualitative results. Thus, even though the proposed method presents a high false positive rate, it is still capable of identifying more true positive regions.

3.4.1. Complexity analysis

A complexity analysis was performed to verify the behavior of the used methods. The proposed approach is composed of three stages: preprocessing, segmentation and post-processing. The asymptotic behavior was estimated at $O(n^2)$, considering that the segmentation step requires the most significant processing time with operations of selection, crossover and mutation that are required by GA.

The approaches of [Wienert et al. \(2012\)](#), [Vahadane and Sethi \(2013\)](#), [de Oliveira et al. \(2013\)](#), [Phoulady et al. \(2016\)](#) and [Paramanandam et al. \(2016\)](#) were based on different criteria and did not provide detailed discussions with focus in asymptotic behavior. This can be a limitation for comparisons with our method. However, this study considered a general analysis and these methods suggest asymptotic behaviors of $O(n^2)$, for instance, of [Vahadane and Sethi \(2013\)](#) and [de Oliveira et al. \(2013\)](#). Thus, the proposed method presents behavior similar with the estimated costs for some relevant methods that are available in literature.

4. Conclusion

This study presented an unsupervised segmentation method of nuclei from neoplastic cells for histological images of CLL, FL and MCL, stained with H&E. Qualitative evaluation of the proposed method reached relevant results for images with $20 \times$ magnification. The proposed method was also quantitatively evaluated considering images manually segmented by a specialist. The methods of mean-shift, [Vahadane and Sethi \(2013\)](#), [Wienert et al. \(2012\)](#), [de Oliveira et al. \(2013\)](#), [Phoulady et al. \(2016\)](#) and [Paramanandam et al. \(2016\)](#) were applied on images of a public domain dataset for comparison purposes.

The proposed method presents robust results with great representativeness of nuclear contours. In comparison to other methods, the described technique presents more consistent contours with the specialist segmentation, as represented by the [Figs. 21–23](#), with a low overlapping rate. These results were obtained considering the processing stages, which enabled the identification of boundary regions more accurately. In addition, the proposed preprocessing presented better results than the normalization. Through its application, the normalization has not reached a satisfactory performance for the enhancement of illumination differences and noise treatment, as quantitatively indicated by the entropy metric.

The GA was also important for defining the adequate threshold values considering different solutions among a large set of parameter combinations of the fuzzy 3-partition method. In comparison

with bio-inspired methods, the GA reached the best overall quantitative results. Exploring a set of the best parameters, the search space became wider, while being directed toward the definition of the best possible threshold values. The DE method also presented a good performance among the evolutionary algorithms analyzed, but it uses only one best solution in its iterations, which can lead to fast convergence with poor exploration of the search space. The low usage of randomness, different from ABC and CS methods, along with solution updates independent from other variables, in contrast to PSO and WDO, also point to the methodology used by GA.

Through qualitative and quantitative analyses, the combination of methods used in the different steps of this study was able to reach more effective results than the compared techniques. Even with noise amplification by histogram equalization, the preprocessing step was successful in allowing contrast enhancements. Besides, the Gaussian filter application makes the reduction of possible noise effects in the later stages. Due to segmentation limitations, the post-processing was crucial for producing a greater definition of contours, as well as reaching the splitting of the nuclei. The valley-emphasis application was performed via intensity analysis of the merged regions, where the suitability of this method was noted for the regions color distribution. Since the histograms of these regions were close to an unimodal distribution, the best threshold value should be located in the valley regions, which this method reached. In the search for coherent morphological properties, the operations used were also defined through analyses of segmented region contours.

Considering its application on images with great contrast and illumination differences, the proposed method provided relevant results for the investigated groups by applying intrinsic features of nuclei, disregarding their spatial distributions and shapes. Recently, there have been a limited amount of studies related to segmentation of CLL and MCL lesions. Thus, another important aspect is that the proposed method contributes through a new strategy for segmentation of these lesions. In regards to histological images of FL, there are many papers dedicated to the detection and segmentation of this tissue. However, this method demonstrated relevant results since it enabled the identification of both centrocytes and centroblasts.

One limitation of the proposed method is the amount of false positive regions presented by the obtained results. The investigated segmentation methods presented similar behavior when the metrics of sensitivity and specificity are analyzed. Future works should consider optimization of the proposed method and the development of a diagnosis support system using the described steps. In the first proposal, validation can be performed using images segmented by different specialists. This method can also be evaluated for application on different public domain lymphoma images, to verify its performance. Besides, new approaches of the GA method should be explored for its computational cost could be reduced. For this purpose, crossover and mutation steps will be improved eliminating randomness from their implementations. In the second proposal, features of identified regions could be explored, such as investigations of frequency-based information, for removal of false positive objects, thus constituting a detection step.

Acknowledgement

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

References

- ACS (2017). What are the key statistics about non-Hodgkin lymphoma? Online. Accessed 01.22.2017.

- Arora, B., & Banerjee, S. (2013). Computer assisted grading schema for follicular lymphoma based on level set formulation. In *Students conference on engineering and systems (SCES)* (pp. 1–6). IEEE.
- Bayraktar, Z. (2013). The wind driven optimization (wdo) algorithm. Online. Accessed 12.21.2016 <http://www.mathworks.com/matlabcentral/fileexchange/44865-the-wind-driven-optimization-wdo-algorithm>.
- Bayraktar, Z., Komurcu, M., & Werner, D. H. (2010). Wind driven optimization (wdo): A novel nature-inspired optimization algorithm and its application to electromagnetics. In *2010 IEEE antennas and propagation society international symposium* (pp. 1–4). IEEE.
- Belkacem-Boussaid, K., Prescott, J., Lozanski, G., & Gurcan, M. N. (2010). Segmentation of follicular regions on h&e slides using a matching filter and active contour model. *SPIE medical imaging*. International Society for Optics and Photonics.
- Belkacem-Boussaid, K., Samsi, S., Lozanski, G., & Gurcan, M. N. (2011). Automatic detection of follicular regions in h&e images using iterative shape index. *Computerized Medical Imaging and Graphics*, 35(7), 592–602.
- Bhandari, A. K., Kumar, A., & Singh, G. K. (2015). Tsallis entropy based multilevel thresholding for colored satellite image segmentation using evolutionary algorithms. *Expert Systems with Applications*, 42(22), 8707–8730.
- Biswas, P. (2014). Searching/tuning/optimizing by particle swarm optimization (psa) method. Online. Accessed 12.21.2016 <http://www.mathworks.com/matlabcentral/fileexchange/43541-particle-swarm-optimization-pso>.
- Bose, A., & Mali, K. (2016). Fuzzy-based artificial bee colony optimization for gray image segmentation. *Signal, Image and Video Processing*, 1–8.
- Byrd, K. A., Zeng, J., & Chouikha, M. (2007). A validation model for segmentation algorithms of digital mammography images. *Journal of Applied Science & Engineering Technology*, 1.
- Canellos, G. P., Lister, T. A., & Young, B. (2006). *The lymphomas* (2nd ed.).
- Chang, V., Saavedra, J. M., Castañeda, V., Sarabia, L., Hirschfeld, N., & Härtel, S. (2014). Gold-standard and improved framework for sperm head segmentation. *Computer Methods and Programs in Biomedicine*, 117(2), 225–237.
- Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.
- Cuevas, E., Zaldivar, D., & Perez-Cisneros, M. (2016). Image segmentation based on differential evolution optimization. In *Applications of evolutionary computation in image processing and pattern recognition* (pp. 9–22). Springer.
- de Oliveira, D. L. L., do Nascimento, M. Z., Neves, L. A., de Godoy, M. F., de Aruda, P. F. F., & de Santi Neto, D. (2013). Unsupervised segmentation method for cuboidal cell nuclei in histological prostate images based on minimum cross entropy. *Expert Systems with Applications*, 40(18), 7331–7340.
- Dimitropoulos, K., Barmoutis, P., Koletsa, T., Kostopoulos, I., & Grammalidis, N. (2016). Automated detection and classification of nuclei in pax5 and h&e-stained tissue sections of follicular lymphoma. *Signal, Image and Video Processing*, 1–9.
- Dimitropoulos, K., Michail, E., Koletsa, T., Kostopoulos, I., & Grammalidis, N. (2014). Using adaptive neuro-fuzzy inference systems for the detection of centroblasts in microscopic images of follicular lymphoma. *Signal, Image and Video Processing*, 8(1), 33–40.
- Estrada, F. J., & Jepson, A. D. (2009). Benchmarking image segmentation algorithms. *International Journal of Computer Vision*, 85(2), 167–181.
- Fuchs, T. J., & Buhmann, J. M. (2011). Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics*, 35(7), 515–530.
- Gartner, L. P., & Hiatt, J. L. (2003). *Tratado de Histologia em Cores* (2nd ed.).
- Gençtav, A., Aksoy, S., & Önder, S. (2012). Unsupervised segmentation and classification of cervical cell images. *Pattern Recognition*, 45(12), 4151–4168.
- Ghose, S., Oliver, A., Martí, R., Lladó, X., Vilanova, J. C., Freixenet, J., ... Meriaudeau, F. (2012). A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. *Computer Methods and Programs in Biomedicine*, 108(1), 262–287.
- Gonzalez, R. C., & Woods, R. E. (2000). *Processamento de Imagens Digitais*. Edgard Blucher.
- Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. *Biomedical Engineering, IEEE Reviews in*, 2, 147–171.
- Haggerty, J. M., Wang, X. N., Dickinson, A., O'Malley, C. J., & Martin, E. B. (2014). Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin. *BMC Medical Imaging*, 14(1).
- Hammouch, K., Diaf, M., & Siarry, P. (2008). A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation. *Computer Vision and Image Understanding*, 109(2), 163–175.
- Heris, S. M. K. (2015a). Implementation of artificial bee colony in matlab. Online. Accessed 12.21.2016 <http://www.yarpiz.com/>.
- Heris, S. M. K. (2015b). Implementation of differential evolution (de) in matlab. Online. Accessed 12.21.2016 <http://www.yarpiz.com/>.
- Hoffman, R. A., Kothari, S., & Wang, M. D. (2014). Comparison of normalization algorithms for cross-batch color segmentation of histopathological images. In *Annual international conference of the IEEE Engineering in medicine and biology society* (pp. 194–197). IEEE.
- ImageJ (2016). ImageJ (image processing and analysis in java). Online. Accessed 16.09.2016 <http://imagej.nih.gov/ij/index.html>.
- INCA (2016). Estimativa 2016 – incidência de câncer no Brasil. Online. Accessed 09.16.2016.
- Insana, M., Meyers, K., & Grossman, L. (2000). Handbook of medical imaging: Medical image processing and analysis.
- Irshad, H., Veillard, A., Roux, L., & Racoceanu, D. (2014). Methods for nuclei detection, segmentation, and classification in digital histopathology: A review – current status and future potential. *IEEE Reviews in Biomedical Engineering*, 7, 97–114.
- Jianli, L., & Baoqi, Z. (2009). The segmentation of skin cancer image based on genetic neural network. In *World congress on computer science and information engineering*: 5 (pp. 594–599). IEEE.
- Jothi, J. A. A., & Rajam, V. M. A. (2016). A survey on automated cancer diagnosis from histopathology images. *Artificial Intelligence Review*, 1–51.
- Kaushik, D., Singh, U., Singhal, P., & Singh, V. (2013). Medical image segmentation using genetic algorithm. *International Journal of Computer Applications*, 81(18).
- Khan, A. (2013). Stain normalisation toolbox. Online. Accessed 12.23.2016 <http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/software/sntoolbox/>.
- Kong, H., Belkacem-Boussaid, K., & Gurcan, M. (2011a). Cell nuclei segmentation for histopathological image analysis. *SPIE medical imaging*. 79622R–79622R, International Society for Optics and Photonics.
- Kong, H., Gurcan, M., & Belkacem-Boussaid, K. (2011b). Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting. *Medical Imaging, IEEE Transactions on*, 30(9), 1661–1677.
- Kothari, S., Phan, J. H., Stokes, T. H., & Wang, M. D. (2013). Pathology imaging informatics for quantitative analysis of whole-slide images. *Journal of the American Medical Informatics Association*, 20(6), 1099–1108.
- Lad, K., Agrawal, M., & Pandya, M. M. (2014). Survey on genetic algorithms & basic operators. *International Journal of Advanced Information Science and Technology (IJIAIST)*, 22(22), 44–48.
- Leong, F. W., Brady, M., & McGee, J. (2003). Correction of uneven illumination (vignetting) in digital microscopy images. *Journal of Clinical Pathology*, 56(8), 619–621.
- Li, X., & Plataniotis, K. N. (2015). A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics. *IEEE Transactions on Biomedical Engineering*, 62(7), 1862–1873.
- Lowry, L., & Linch, D. (2013). Non-Hodgkin's lymphoma.
- Luo, Y., Celenk, M., & Bejai, P. (2006). Discrimination of malignant lymphomas and leukemia using radon transform based-higher order spectra. *Medical imaging*. International Society for Optics and Photonics. 61445K–1–61445K–10.
- Mackenro, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., ... Thomas, N. E. (2009). A method for normalizing histology slides for quantitative analysis. In *ISBI*: 9 (pp. 1107–1110).
- Mauriño, B. B., & Siqueira, S. A. C. (2011). Classificação dos Linfomas.
- McCann, M. T., Ozolek, J. A., Castro, C. A., Parvin, B., & Kovacevic, J. (2015). Automated histology analysis: Opportunities for signal processing. *Signal Processing Magazine*, 32(1), 78–87.
- Meijering, E. (2012). Cell segmentation: 50 years down the road. *IEEE Signal Processing Magazine*, 29(5), 140–145.
- Michail, E., Kornaropoulos, E. N., Dimitropoulos, K., Grammalidis, N., Koletsa, T., & Kostopoulos, I. (2014). Detection of centroblasts in h&e stained images of follicular lymphoma. In *22nd signal processing and communications applications conference (SIU)* (pp. 2319–2322). IEEE.
- Mohammed, E. A., Far, B. H., Naugler, C., & Mohamed, M. M. A. (2013a). Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation. *15th international conference on e-health networking, applications and services*. IEEE.
- Mohammed, E. A., Far, B. H., Naugler, C., & Mohamed, M. M. A. (2013b). 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.
- Ng, H. (2006). Automatic thresholding for defect detection. *Pattern Recognition Letters*, 27(14), 1644–1649.
- Oger, M., Belhomme, P., & Gurcan, M. N. (2012). A general framework for the segmentation of follicular lymphoma virtual slides. *Computerized Medical Imaging and Graphics*, 36(6), 442–451.
- Orlov, N. V., Chen, W. W., Eckley, D. M., Macura, T. J., Shamir, L., Jaffe, E. S., & Goldberg, I. G. (2010). Automatic classification of lymphoma images with transform-based global features. *IEEE Transactions on Information Technology in Biomedicine*, 14(4), 1003–1013.
- Oswal, V., Belle, A., Diegelmann, R., & Najarian, K. (2013). An entropy-based automated cell nuclei segmentation and quantification: Application in analysis of wound healing process. *Computational and Mathematical Methods in Medicine*.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(62–66), 1.
- Oztan, B., Kong, H., Gurcan, M. N., & Yener, B. (2012). Follicular lymphoma grading using cell-graphs and multi-scale feature analysis. *SPIE medical imaging*: 8315.
- Paramanandam, M., O Byrne, M., Ghosh, B., Mammen, J. J., Manipadam, M. T., Thamburaj, R., & Pakrashi, V. (2016). Automated segmentation of nuclei in breast cancer histopathology images. *PLoS One*, 11(9), e0162053.
- Paulinas, M., & Ušinskis, A. (2015). A survey of genetic algorithms applications for image enhancement and segmentation. *Information Technology and Control*, 36(3).
- Pedrin, H., & Schwartz, W. R. (2007). *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. Thomson Learning.
- Phoulady, H. A., Goldof, D. B., Hall, L. O., & Mouton, P. R. (2016). Nucleus segmentation in histology images with hierarchical multilevel thresholding. *SPIE medical imaging*. 979111–979111, International Society for Optics and Photonics.

- Remamany, K. P., Chelliah, T., Chandrasekaran, K., & Subramanian, K. (2015). Brain tumor segmentation in mri images using integrated modified pso-fuzzy approach. *International Arab Journal of Information Technology (IAJIT)*, 12.
- Ruifrok, A. C., & Johnston, D. A. (2001). Quantification of histochemical staining by color deconvolution. *Analytical and Quantitative Cytology and Histology/The International Academy of Cytology [and] American Society of Cytology*, 23(4), 291–299.
- Sertel, O., Catalyurek, U. V., Lozanski, G., Shanaah, A., & Gurcan, M. N. (2010a). An image analysis approach for detecting malignant cells in digitized h&e-stained histology images of follicular lymphoma. In *20th international conference on pattern recognition (ICPR)* (pp. 273–276). IEEE.
- Sertel, O., Kong, J., Lozanski, G., Catalyurek, U., Saltz, J. H., & Gurcan, M. N. (2008a). Computerized microscopic image analysis of follicular lymphoma. *Medical imaging*: 6915. International Society for Optics and Photonics.
- Sertel, O., Kong, J., Lozanski, G., Catalyurek, U., Saltz, J. H., & Gurcan, M. N. (2009). Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading. *Journal of Signal Processing Systems*, 55(1–3), 169–183.
- Sertel, O., Kong, J., Lozanski, G., Shana'ah, A., Catalyurek, U., Saltz, J., & Gurcan, M. (2008b). 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.
- Sertel, O., Lozanski, G., Shana'ah, A., & Gurcan, M. N. (2010b). Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation. *IEEE Transactions on Biomedical Engineering*, 57(10), 2613–2616.
- Shamir, L., Orlov, N., Eckley, D. M., Macura, T. J., & Goldberg, I. G. (2008). Ilicbu 2008: A proposed benchmark suite for biological image analysis. *Medical & Biological Engineering & Computing*, 46(9), 943–947.
- Shi, Y., & Eberhart, R. C. (1998). Parameter selection in particle swarm optimization. In *International conference on evolutionary programming* (pp. 591–600). Springer.
- Smochinã, C., Herghelegiu, P., & Manta, V. (2011). Image processing techniques used in microscopic image segmentation. *Technical Report*. Gheorghe Asachi Technical University of Iași.
- Tafavogh, S., Catchpoole, D. R., & Kennedy, P. J. (2014). Cellular quantitative analysis of neuroblastoma tumor and splitting overlapping cells. *BMC Bioinformatics*, 15(1).
- Tsai, D., Lee, Y., & Matsuyama, E. (2008). Information entropy measure for evaluation of image quality. *Journal of Digital Imaging*, 21(3), 338–347.
- Vahadane, A., & Sethi, A. (2013). Towards generalized nuclear segmentation in histological images. In *13th international conference on bioinformatics and bioengineering (BIBE)* (pp. 1–4). IEEE.
- Wang, P., Hu, X., Li, Y., Liu, Q., & Zhu, X. (2016). Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Processing*, 122, 1–13.
- Wienert, S., Heim, D., Saeger, K., Stenzinger, A., Beil, M., Hufnagl, P., ... Klauschen, F. (2012). Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach. *Scientific Reports*, 2.
- Wu, G., Zhao, X., Luo, S., & Shi, H. (2015). Histological image segmentation using fast mean shift clustering method. *Biomedical Engineering Online*, 14(1), 24.
- Xing, F., & Yang, L. (2013). Robust selection-based sparse shape model for lung cancer image segmentation. In *Medical image computing and computer-assisted intervention (MICCAI)* (pp. 404–412). Springer.
- Yang, L., Tuzel, O., Meer, P., & Foran, D. J. (2008). Automatic image analysis of histopathology specimens using concave vertex graph. In *Medical image computing and computer-assisted intervention (MICCAI): 11* (pp. 833–841). Springer.
- Yang, X. (2009). Cuckoo search (cs) algorithm implementation. Online. Accessed 12.21.2016 -cs-algorithm <http://www.mathworks.com/matlabcentral/fileexchange/29809-cuckoo-search>.
- Yin, S., Zhao, X., Wang, W., & Gong, M. (2014). Efficient multilevel image segmentation through fuzzy entropy maximization and graph cut optimization. *Pattern Recognition*, 47(9), 2894–2907.
- Zorman, M., Kokol, P., Lenic, M., Rosa, J. L. S., Sigut, J. F., & Alay' on, S. (2007). 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.