Applied Soft Computing 64 (2018) 49–58

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

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

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

[a] Federal University of ABC, Center of Mathematics, Computer Science and Cognition, Av. dos Estados, 5001, 09210-580 Santo André, São Paulo, Brazil
[b] Federal University of Uberlândia, Department of Histology and Morphology, Institute of Biomedical Science, Av. Amazonas, S/N, 38405-320 Uberlândia, Minas Gerais, Brazil
[c] São Paulo State University (UNESP), Department of Computer Science and Statistics, R. Cristóvão Colombo, 2265, 15054-000 São José do Rio Preto, São Paulo, Brazil
[d] Federal University of Uberlândia, Faculty of Computer Science, Av. João Naves de Ávila, 2121, 38400-902 Uberlândia, Minas Gerais, Brazil

ARTICLE INFO

Article history:
Received 22 October 2016
Received in revised form
20 September 2017
Accepted 22 November 2017
Available online 5 December 2017

Keywords:
Nuclei segmentation
Genetic algorithm
Wavelet transform
H&E-stained histological images
Chronic lymphocytic leukemia

ABSTRACT

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

1. Introduction

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

The diagnosis of this neoplasia requires histological sections analysis of either lymph nodes or blood. Such analysis is performed by a specialist in identifying cancer cells. This is crucial for the patient, since only a correct diagnosis can make it possible an adequate medical follow-up, a right identification of the illness degree, and also the best guidance to efficient treatments [4]. However, analyses of specialists are partly subjective, highly time-consuming and also influenced by the fact that leukemia has many variations. This all makes diagnosing a complex task [5].

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

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

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

https://doi.org/10.1016/j.asoc.2017.11.039
1568-4946/

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

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

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

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

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

This method consists of three steps: preprocessing, segmentation and post-processing. In the first one, deconvolution was applied to better represent nuclear regions, and its result was submitted to a contrast enhancement and noise removal by means of histogram equalization and mean filter. In the segmentation, ROIs were obtained by using fuzzy 2-partition entropy method combined with GA and wavelet horizontal component extraction. In the post-processing, false positive regions were removed by means of the valley-emphasis method, and small deformations were corrected through morphological operations. We have simulated this methodology with a public image database [18], which was also manually segmented by a specialist. Its performance was evaluated by comparing it with the techniques of Vahadane and Sethi [19], Wienert et al. [20] and de Oliveira et al. [21].

The main contributions of this paper are summarized as follows:

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

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

2. Materials and methods

2.1. Database

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

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

2.2. Proposed algorithm

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

2.2.1. Preprocessing

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