Expert Systems With Applications 81 (2017) 223–243

Contents lists available at ScienceDirect

Expert Systems With Applications

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

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.

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).

http://dx.doi.org/10.1016/j.eswa.2017.03.051
0957-4174/

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, Shanaah, & 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×. 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 Dijsktra 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 × (Yang et al., 2008) and 100 × (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, Koletsa, 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