International Journal of Laboratory Hematology
The Official journal of the International Society for Laboratory Hematology

ORIGINAL ARTICLE INTERNATIONAL JOURNAL OF LABORATORY HEMATOLOGY

Automatic classification of atypical lymphoid B cells using digital blood image processing

S. ALFÉREZ*, A. MERINO†, L. E. MUJICA*, M. RUIZ*, L. BIGORRA†, J. RODELLAR*

*Universitat Politècnica de Catalunya, Barcelona, Spain
†Servei d'Hemoterapia-Hemostasia, Centre de Diagnóstic Biomédic, Hospital Clínic de Barcelona, Barcelona, Spain

Correspondence:
Anna Merino, Department of Hemotherapy-Hemostasis, Hospital Clinic, Villarroel 170, 08036 Barcelona, Spain.
Tel.: +34 932275400 2309;
Fax: +34 932279374;
E-mail: amerino@clinic.ub.es

The authors have nothing to disclose.

doi:10.1111/ijlh.12175

Received 2 September 2013; accepted for publication 25 October 2013

Keywords
Atypical lymphoid cells, hematological cytology, digital image processing, automatic cell classification, peripheral blood, morphological analysis

SUMMARY

Introduction: There are automated systems for digital peripheral blood (PB) cell analysis, but they operate most effectively in non-pathological blood samples. The objective of this work was to design a methodology to improve the automatic classification of abnormal lymphoid cells.
Methods: We analyzed 340 digital images of individual lymphoid cells from PB films obtained in the CellaVision DM96:150 chronic lymphocytic leukemia (CLL) cells, 100 hairy cell leukemia (HCL) cells, and 90 normal lymphocytes (N). We implemented the *Watershed Transformation* to segment the nucleus, the cytoplasm, and the peripheral cell region. We extracted 44 features and then the clustering Fuzzy C-Means (FCM) was applied in two steps for the lymphocyte classification.
Results: The images were automatically clustered in three groups, one of them with 98% of the HCL cells. The set of the remaining cells was clustered again using FCM and texture features. The two new groups contained 83.3% of the N cells and 71.3% of the CLL cells, respectively.
Conclusion: The approach has been able to automatically classify with high precision three types of lymphoid cells. The addition of more descriptors and other classification techniques will allow extending the classification to other classes of atypical lymphoid cells.

INTRODUCTION

Peripheral blood (PB) is an organic fluid easily accessible, and its study is the initial analytical step in the diagnosis of most of the hematological and nonhematological diseases [1]. Frequently, the blood smear provides the primary or the only evidence of a specific diagnosis, remaining an important diagnostic tool even in the age of molecular analysis [2]. Morphological evaluation of leukemia and lymphoma cells is essential for their diagnosis and classification. In the World Health Organization (WHO) classification, atypical cell morphology, along with immunophenotype and genetic changes,

remains essential in defining lymphoid neoplasms [3].

Despite the significant improvements during the last years in hematology analyzers, no significant progress has been made in terms of automatic classification of atypical PB cells. These devices are limited to identifying normally circulating leukocytes and flagging abnormal cells, without being able to classify the abnormal leukocytes.

The close collaboration between cytologists, mathematicians, and engineers over the last few years has made possible the development of automatic methodologies for digital image processing of normal blood cells. Some equipments are able to preclassify cells in different categories by applying neural networks, extracting a large number of measurements and parameters that describe the most significant cell morphological characteristics [4]. These systems, when integrated in the daily routine, represent an interesting technological advance as they are able to preclassify most of the normal blood cells in PB [5].

Atypical lymphoid cells are the most difficult pathological cells to classify using morphology features only [6] so that few studies of automatic classification of these cells with satisfactory results have been published. In most of the previous studies, the lymphoid cell classification has been addressed with pattern recognition systems to separate the cells into categories [7–10]. Nevertheless, the image processing techniques used in some of the papers are not useful for the current digital images, because the present acquisition technology is based on charge-coupled device sensors [7].

Morphological distinction between various types of lymphoid cells requires experience and skill, and, moreover, objective values do not exist to define cytological variables. CLL cells are typically small lymphocytes with clumped chromatin and scant cytoplasm. HCL cells are larger than normal lymphocytes (N), and they have abundant weakly basophilic cytoplasm with irregular 'hairy' margins. This study presents a new methodology for lymphocyte recognition to allow the 'automatic classification' of abnormal lymphoid cells circulating in PB in some B lymphoid neoplasms, such us chronic lymphocytic leukemia (CLL) and hairy cell leukemia (HCL) cells.

MATERIAL AND METHODS

Blood sample preparation and digital image acquisition

Samples from patients with CLL and HCL were included in the study. The diagnoses were established by clinical and morphological findings as well as characteristic immunophenotype of the lymphoid cells. Specifically, CLL cells had the phenotype CD5+, CD19+, CD23+, CD25+, weak CD20+, CD10−, FMC7−, and dim surface immunoglobulin (sIg) expression. All the patients with HCL had lymphoid cells with the phenotype CD11c+, CD25+, FMC7+, CD103+, and CD123+.

Blood samples were obtained from the routine workload of the Core Laboratory of the Hospital Clínic of Barcelona. Venous blood was collected into tubes containing K_3EDTA as anticoagulant. The samples were analyzed by a cell counter Advia 2120 (Siemens Healthcare Diagnosis, Deerfield, IL, USA), and PB films were automatically stained with May–Grünwald–Giemsa in the SP1000i (Sysmex, Japan, Kobe) within 4 h of blood collection.

The quality of the smears and cell morphology was assessed by hematologists prior to the image study. We selected 340 lymphoid cell images from PB films, where 90 images were lymphocytes from healthy patients, 100 were lymphoid cells from patients with HCL, and 150 were lymphoid cells from patients with CLL. Each individual cell image had a resolution of 367 × 360 pixels, and they were obtained by the CellaVision DM96 system (Lund, Sweden).

Novel method for lymphocyte classification

In this study, we developed a novel method for lymphocyte recognition based on 3 steps: (i) color segmentation, (ii) feature extraction, and (iii) classification. They are shortly described in the remaining of this section.

Color segmentation

A digital blood image is composed of a finite number of pixels. Each one has a particular location and color value, which can be represented in several spectral components or color spaces: RGB, HSV, Lab, among others [11]. The goal of the segmentation procedure is

to separate lymphoid cells captured by microscope from other objects in the image [12–14].

In this work, lymphoid cell segmentation was obtained using the watershed transformation (WT), which was applied only on the gradient of the green component from RGB color space [15]. As a final result, three different regions of the cell were identified (segmented): the cytoplasm, the nucleus, and the peripheral zone around the cell.

Feature extraction

The objective of this stage is to obtain information about the objects in the image under analysis. A number of 44 features were used in this work, which are related, respectively, to geometry (10), texture (30), basophilia intensity (3), and cytoplasm external profile (1). They are summarized as follows:

Geometric features

These features are quantitative geometric interpretations of the cell and nucleus shapes. For each cell and nucleus, we calculated the following: '*areas*', '*diameters*', '*perimeters*', and '*conic eccentricities*'. Then, '*nucleus/cytoplasm ratio*' was calculated by dividing the respective areas. The '*nucleus eccentricity relative to the cell center*' was calculated as the distance between the center of the cell and the nucleus [9].

Texture features

Several statistical measures were used to describe the texture of the cytoplasm and nucleus regions [16]. The *skewness* measures the asymmetry of the shape; the *kurtosis*, the relative flatness; the *energy*, the uniformity; and the *entropy*, the variability. In addition, the *mean* and the *standard deviation* were calculated. Other second-order statistical features were considered the following: *contrast*, *homogeneity*, *correlation*, *energy*, *entropy*, and *difference variance* [17].

Granulometrical features of the nucleus

The granulometry estimates the size distribution of the bright and dark spots on the image. The granulometrical curve places the information from the dark spots on the left (negative co-ordinates) and the information from the bright spots on the right (positive co-ordinates) [18]. From the granulometrical curve of the lymphoid cell, we calculated four features: *mean*, *standard deviation*, *skewness*, and *kurtosis*, to discriminate the different types of nuclear texture and improve chromatin description.

Basophilia features of the cytoplasm. Cytoplasmic basophilia can be estimated by color analysis. The *Lab* color space is characterized by its approximation to human perception. Therefore, the *means* of the intensities for each color component are appropriate to represent the basophilia degree of the cytoplasm [9].

Cytoplasmic profile feature. In this study, we propose a novel method to characterize the cytoplasmic profile. It estimates the projections of the cytoplasm using the peripheral region around the cell segmented by WT. This feature is obtained using thresholding segmentation to the green component and counting the pixels of this region.

Classification

In this work, the main objective of the classification step was to obtain an automatic clustering using the features extracted from each image in order to analyze how they can provide relevant information for the detection of normal, HCL, or CLL lymphoid cells.

All features were stored in a data matrix, which was used as the input data for the classification. The unsupervised classification methodology Fuzzy c-mean (FCM) was applied. Similar input data were grouped in each cluster with certain membership degree [19]. Finally, the maximum membership value was considered to select the cluster for each lymphoid cell.

RESULTS

In the first step, WT was effective in separating the nucleus of the cell. Besides, it allowed segmenting more regions, specifically the outer profile of the cytoplasm, which is crucial to extract the useful information to discriminate different types of lymphocytes. More-

over, its computational cost was low. Figure 1 shows the images corresponding to the different stages that we obtained applying the WT segmentation to the lymphoid cells. The lymphoid original cell stained with MGG is shown in Figure 1a. The WT was applied only on the gradient of the green component from RGB color space (Figure 1b). As the gradient highlights the edges (high intensity changes) of the objects, some external and internal markers were included as minimum values over the gradient image to improve the delimitation of the different regions as shown in Figure 1c. Thereby, the over-segmentation was avoided, and only the entire lymphoid cell was separated (the darkest region on Figure 1d). Once the entire lymphoid cell was separated, new markers were imposed (Figure 1e), and the WT was applied again to segment the nucleus (Figure 1f). Afterward, mathematical morphology operations were performed to improve the quality of the regions from the nucleus and cytoplasm. Finally, 3 different regions of the cell were identified: the cytoplasm, the nucleus, and the peripheral zone around the cell (Figure 1g,h).

Corresponding to the second step (feature extraction), Figure 2 shows an example of N, HCL, and CLL lymphoid cell images (Figure 2a) and their granulometrical curves. Figure 2b shows how these curves discriminate the types of nuclear texture in the different lymphoid cells, improving chromatin description. To obtain information from each curve, four features were calculated as follows: *mean, standard deviation, skewness,* and *kurtosis*.

Figure 3 displays an example of cytoplasmic profile feature extraction obtained in one of the hairy cell images. After the segmentation of the cell (Figure 3a), the peripheral zone around the cell was selected (Figure 3b). The histogram representation of this region showed an intermediate lobe that contained most of these 'hairy' projections (Figure 3c). Then, the presence or absence of these projections was determined (Figure 3d). Finally, this area was quantitated. The

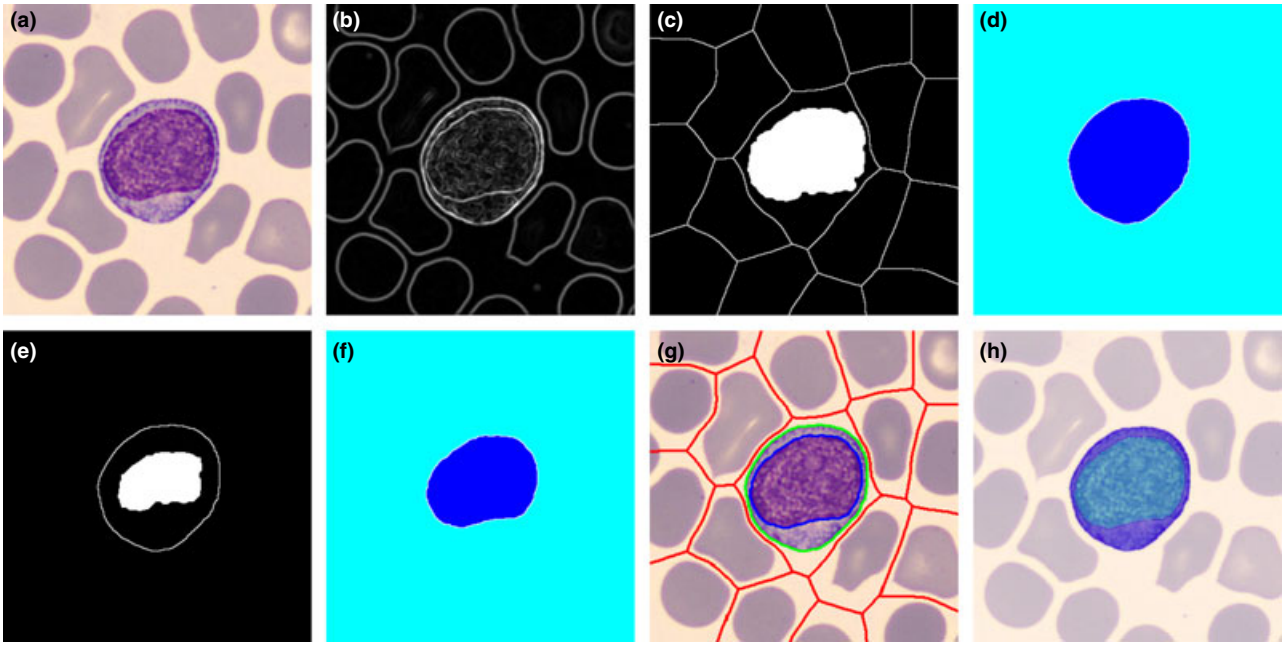

Figure 1. Different stages of the watershed segmentation (WT): The original cell (a) was processed to obtain the external and internal markers (c). The WT was calculated on the gradient of green component (b). The markers limit the WT to segment the cell (d). Once the lymphoid cell was separated, its edges are used as the new external marker and the thinned mask of the nucleus as the new internal marker (e) in the WT to segment the nucleus (f). Finally, the watershed lines (g) showed the regions of interest: the nucleus, the cytoplasm (h), and the peripheral zone around the cell.

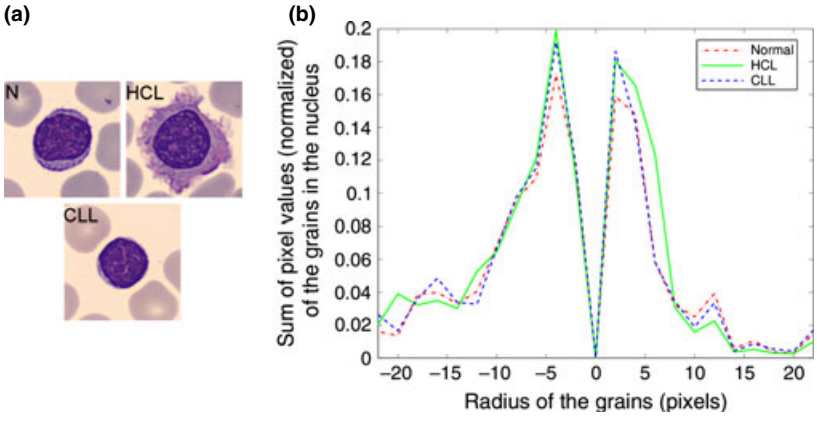

Figure 2. Normal (N), HCL and CLL lymphocytes (a), and their corresponding granulometric curves (b), which places the information from the dark spots on the left (negative co-ordinates) and the information from the bright spots on the right (positive co-ordinates).

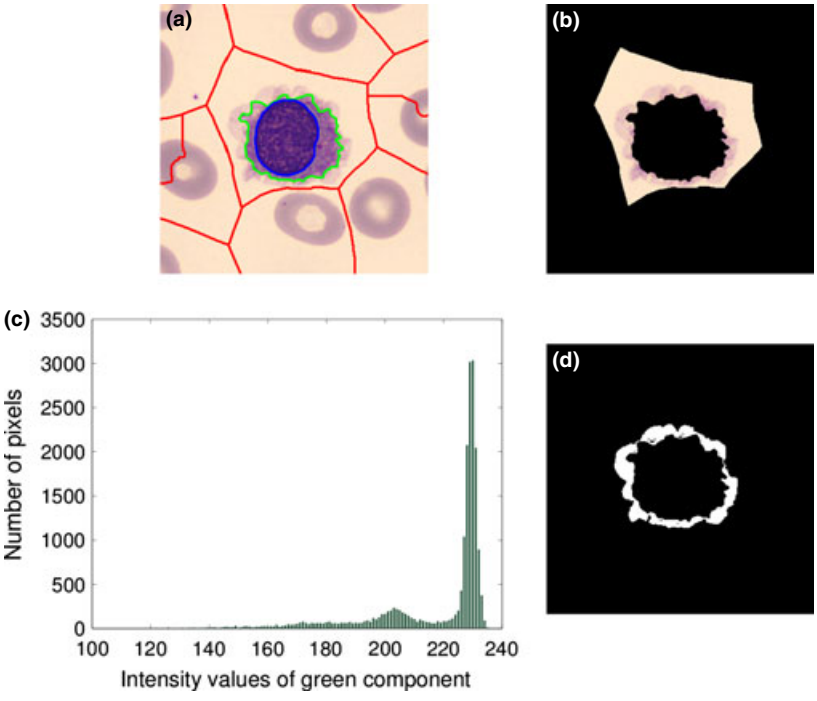

Figure 3. Stages to calculate the HCL cell cytoplasmic feature. After the cell segmentation (a), the peripheral zone around the cell was selected (b). The histogram representation of this region showed an intermediate lobe that contained most of the 'hairy' projections (c). Then, the presence of these projections was determined (d). Finally, this area was quantitated.

novel cytoplasmic profile feature proposed in this work was decisive for the detection of the hairy cells. Figure 4 shows the characteristic cytoplasmic profile feature for all the cells. HCL cells showed very high values of this feature compared with CLL and N lymphoid cells.

Afterward, for the classification step, the 44 features of the 340 available cells were used to create the data matrix. It was automatically clustered into three groups using FCM producing three membership functions. As every cell pertains to one of the three groups with different degrees of membership, we used the criterion that each lymphoid cell belongs to the group with the highest membership value. The left part of Table 1 gives a summary of the whole data obtained in the first FCM classification step. This shows an excellent classification on the group 3 because it included 98% of the HCL cells. However, the groups 1 and 2 contained 75.6% of normal lymphoid cells and 62.7% of CLL cells, respectively. Figure 5 contains 3 plots corresponding to each group. The horizontal axis represents each individual cell, while the vertical axis gives its membership value. These 3 values represent the probability of belonging to each

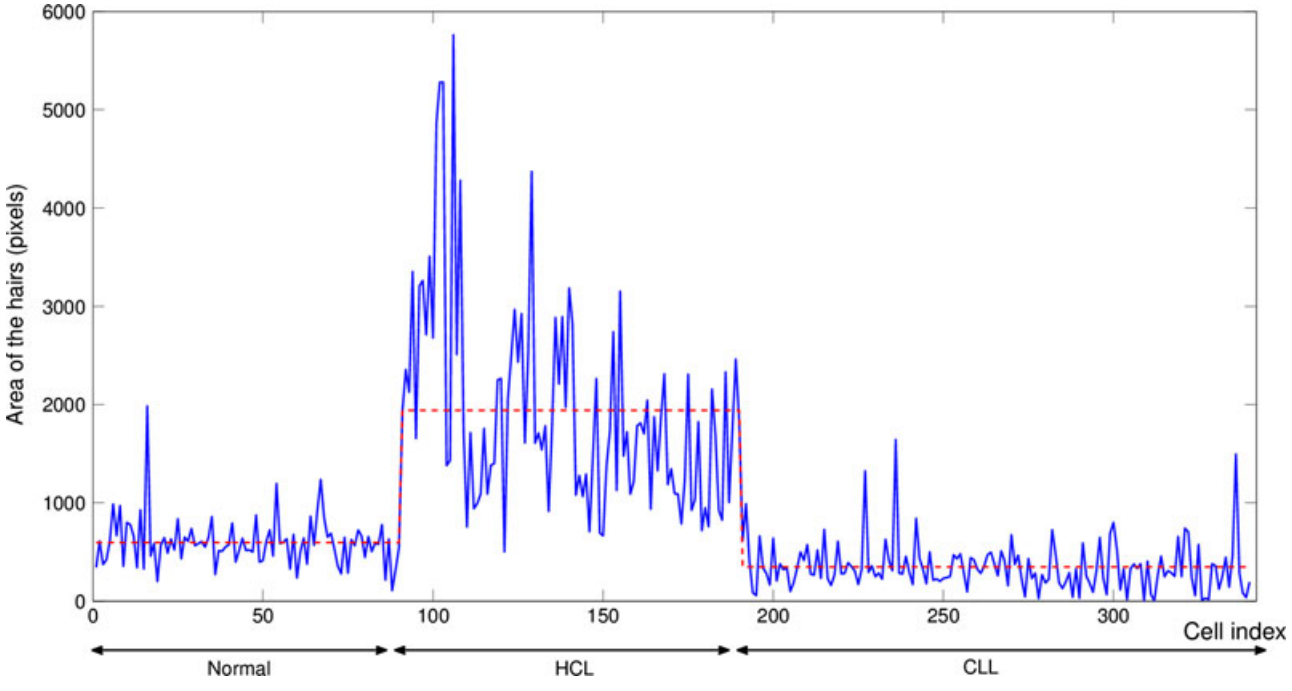

Figure 4. Cytoplasmic profile feature in Normal (N), HCL, and CLL lymphoid cells. HCL cells showed very high values of this feature compared with CLL and N lymphoid cells.

Table 1. Two-step classification process. First, 3 different classes of data were obtained. Each group has cells of the three types, that is, the group 3 has 98% of the HCL cells. A second Fuzzy C-means (FCM) was applied using the texture features only. It resulted in two new groups with 83.3% of N lymphoid cells and 71.3% of CLL cells, respectively

	FCM step 1			FCM step 2	
Type	Group 1, %	Group 2, %	Group 3, %	New group 1, %	New group 2, %
N	75.6	20	4.4	**83.3**	12.2%
HCL	1	1	**98**	2	0%
CLL	30	62.7	7.3	21.3	**71.3%**

N, normal lymphoid cells; HCL, hairy cell leukemia cells; CLL, chronic lymphocytic leukemia cells.
The bold values are made to highlight the best percentages in the results.

group, their sum being equal to 1. In every plot, the data set was sorted in this way: The first 90 images belong to N lymphocytes, the following 100 to HCL cells and the last 150 to CLL cells. From Figure 5, it was clear to assure that both the normal cells and the CLL cells do not belong to group 3 due to their low membership values. It was also clear that HCL cells had high probability of belonging to group 3. On the contrary, from Figure 5, it was difficult to infer to which group (1 or 2) belong the normal and the CLL cells due to the high variance of their membership values in groups 1 and 2.

To improve the classification, once the HCL group was identified, the set of the remaining cells was clustered again using FCM. In this new clustering process, only two types were considered (N and CLL cells),

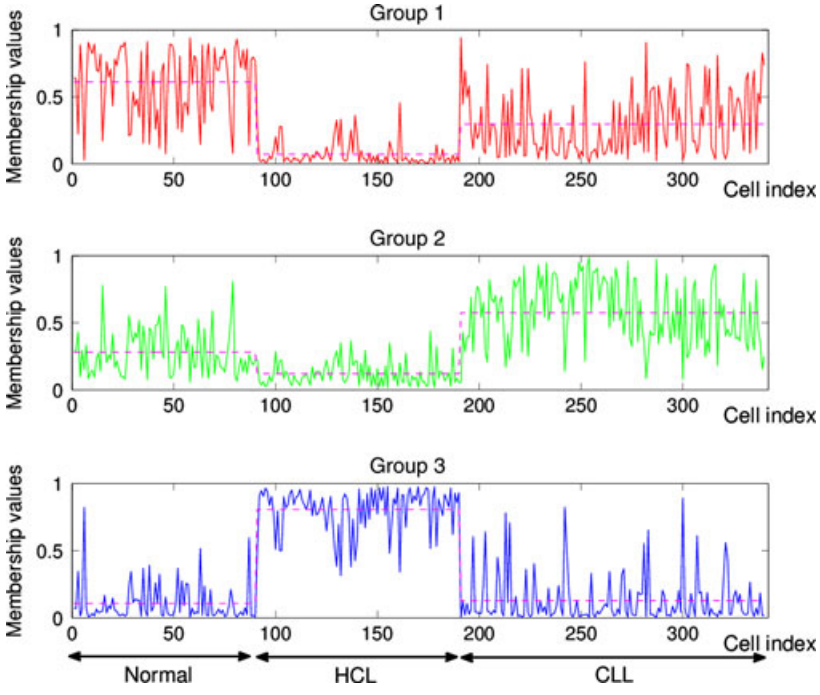

Figure 5. Membership function of each type of cell: normal lymphocytes, hairy cells (HCL), and lymphoid cells from chronic lymphocytic leukemia (CLL). The horizontal axis represents the cells, while the vertical axis represents the probability of belonging to each group. The horizontal line for each type of cell represents the mean of their membership values in each group.

and only the texture features were used. The right part of Table 1 gives a report of the results in this step: the percentage of normal lymphoid cells increased to 83.3% in the new group 1, and the percentage of CLL cells increased to 71.3% in the new group 2. In this case, this was clear enough to distinguish these two types of cells, because their membership values were quite different for a significant percentage of cells as observed in the right part of Table 1.

DISCUSSION

In this study, a group composed by normal and two types of atypical lymphoid cells (HCL and CLL) has been analyzed. HCL and CLL cells were selected in our work for their representative morphology and the large number of these cells that we found in the routine workload in our laboratory.

Cell morphology is subject to variability in slide making and staining procedures. To minimize this variability, the images used in this work were obtained in a standard and reproducible way using automatic staining and the CellaVision DM96 analyzer. The system scanned the slides identifying different types of white blood cells (WBC). It takes digital cell images and uses artificial neural networks to analyze them [4, 5]. The analyzer preclassifies WBC but is not able to separate the different abnormal lymphoid cells circulating in PB in some B-cell lymphoid neoplasms [20].

Because atypical lymphoid cells are the most difficult ones to be classified using only morphology features [6], in this work, a new methodology is proposed combining segmentation, feature extraction, and classification algorithms. We showed that this automated image-based methodology extracted granulometrical, basophilia, and cytoplasmic profile features in an objective and reproducible way. Then, this methodology could help to provide a new generation of automated systems to assist in the diagnosis through hematological cytology.

Our results showed that texture descriptors were the most relevant in CLL lymphoid cell discrimination. Moreover, nuclear characteristics are important features in morphological diagnosis. The nuclear staining pattern reflects chromatin organizations, and, in addition, the CLL cells typically contain clumped chromatin [21]. Therefore, it supplies a good descriptor.

In a previous work [9], granulometry was used to describe cytoplasmic profile feature. Although that work showed good segmentation and description results, it was not completed with further studies

toward the discrimination among different groups of similar diagnosis.

In our work, a novel cytoplasmic profile feature is proposed based on a simple thresholding of the peripheral zone around the cell. As we expected, this feature was crucial for the HCL cells detection, as in PB stained with MGG, they show a soft, blue–gray cytoplasm with hair-like cytoplasmic projections [22]. On the other hand, this feature could be used for the detection of another atypical lymphoid cells with cytoplasmic villous, such as the splenic marginal zone lymphoma.

Concerning to the classification process, 26 features (geometrical and second-order statistical features) were used in [8] to automate the classical microscopic diagnosis, obtaining good results in the classification of CLL cells but only with respect to the different 'normal' types of leukocytes from PB. In our work, 44 features are used adding other geometrical and second-order statistical features as well as basophilia, granulometrical, first-order statistical, and cytoplasmic profile features. In addition, we distinguish three types of lymphoid cells: normal, CLL, and HCL. It is relevant to remark that, up to our knowledge, hairy cells have never been automatically classified before.

One of the main contributions of this work is the segmentation methodology of the peripheral region of the cell, in addition to the nucleus and cytoplasm, which allows proposing a new cytoplasmic profile feature describing information about the villous. This enables to achieve an accurate classification of HCL cells.

In summary, the methodology presented in this paper has been able to discriminate between three groups of lymphoid cells with encouraging results. The overall goal is to combine medical, engineering, and mathematical backgrounds to provide more objective and reproducible estimation of the lymphoid cell morphology than the standard microscopy analysis. The addition of more descriptors and other classification techniques to the developed methodology will allow extending to other classes of atypical lymphoid cells. We are progressing with further work in this direction, in order that this methodology could be useful in clinical practice.

ACKNOWLEDGEMENTS

S. Alférez acknowledges the support provided by the Universitat Politècnica de Catalunya · BarcelonaTech (UPC) under a PhD grant. L. Bigorra appreciates the support provided by the J.L Castaño foundation through a research grant.

REFERENCES

1. Houwen B. The differential cell count. Lab Hematol 2001;7:89–100.
2. Bain BJ. Diagnosis from the blood smear. N Engl J Med 2005;353:498–507.
3. Swerdlow SH, Campo E, Harris NL. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. France: IARC Press; 2008.
4. Ceelie H, Dinkelaar RB, van Gelder W. Examination of peripheral blood films using automated microscopy; evaluation of Diffmaster Octavia and Cellavision DM96. J Clin Pathol 2007;60:72–9.
5. Briggs C, Longair I, Slavik M, Thwaite K, Mills R, Thavaraja V, Foster A, Romanin D, Machin SJ. Can automated blood film analysis replace the manual differential? An evaluation of the CellaVision DM96 automated image analysis system. Int J Lab Hematol 2009;31:48–60.
6. Gutiérrez G, Merino A, Domingo A, Jou JM, Reverter JC. EQAS for peripheral blood morphology in Spain: a 6-year experience. Int J Lab Hematol 2008;30:460–6.
7. Bergmann M, Heyn H, Müller-Hermelink HK, Harms H, Aus HM. Automated recognition of cell images in high grade malignant lymphoma and reactive follicular hyperplasia. Anal Cell Pathol 1990;2:83–95.
8. Sabino DMU, Da Fontoura Costa L, Rizzatti EG, Zago MA. A texture approach to leukocyte recognition. Real-Time Imaging 2004;10:205–16.
9. Angulo J, Klossa J, Flandrin G. Ontology-based lymphocyte population description using mathematical morphology on color blood images. Cell Mol Biol 2006;52:3–16.
10. Tuzel O, Yang L, Meer P, Foran DJ. Classification of hematologic malignancies using texton signatures. Pattern Anal Appl 2007;10(4):277–90.
11. González RC, Woods RE. Digital Image Processing. Upper Saddle River, NJ: Prentice Hall; 2008.
12. Xu C, Prince JL. Snakes, shapes, and gradient vector flow. IEEE Trans Image Process 1998;7:359–69.
13. Sadeghian F, Seman Z, Ramli AR, Abdul Kahar BH, Saripan M-I. A framework for white blood cell segmentation in microscopic blood images using digital image processing. Biol Proced Online 2009;11:196–206.
14. Alférez E, Merino A, Mujica L, Rodellar J. Morphological features using digital image processing in lymphoid neoplasias [Abstract]. Int J Lab Hematol 2011;33(Suppl. 1):53.
15. Beucher S, Meyer F. The morphological approach to segmentation: the watershed transformation. In: Mathematical Morphology in Image Processing. Dougherty E (ed.). New York: Marcel Dekker; 1992: 433–81.

16. Materka A, Strzeleck M. Texture analysis methods - a review. Brussels: Technical University of Lodz, Institute of Electronics, COST B11 report; 1998. Available from: http://www.eletel.p.lodz.pl/programy/cost/pdf_1.pdf
17. Haralick RM, Shanmugan K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610–21.
18. Angulo J. A mathematical morphology approach to the analysis of the shape of cells. In: Progress in Industrial Mathematics at ECMI 2006. Bonilla LL. Moscoso M. Platero G. Vega JM. (eds). Leganes, Spain: Springer; 2006: 543–7.
19. Pal NR, Bezdek JC. On cluster validity for the fuzzy c-means model. IEEE Trans Fuzzy Syst 1995;3:370–9.
20. Merino A, Brugués R, Garcia R, Kinder M, Bedini JL, Escolar G. Comparative study of peripheral blood morphology by conventional microscopy and Cellavision DM96 in hematological and non hematological diseases [Abstract]. Int J Lab Hematol 2011;33(Suppl. 1):112.
21. Jahanmehr SH, Rogers M, Zheng J, Lai R, Wang C. Quantitation of cytological parameters of malignant lymphocytes using computerized image analysis. Int J Lab Hematol 2008;30:278–85.
22. Summers TA, Jaffe ES. Hairy cell leukemia diagnostic criteria and differential diagnosis. Leuk Lymphoma 2011;52:6–10.