Automatic Recognition of Five Types of White Blood Cells in Peripheral Blood

Seyed Hamid Rezatofighi[1], Kosar Khaksari[1], Hamid Soltanian-Zadeh[1,2]

[1] Control and Intelligent Processing Center of Excellence, Department of Electrical and Computer Engineering, Faculty of Engineering, University of Tehran, Tehran 14395-515, Iran
h.tofighi@ece.ut.ac.ir, hszadeh@ut.ac.ir
[2] Image Analysis Lab., Department of Radiology, Henry Ford Hospital, Detroit, MI, USA
hamids@rad.hfh.edu

Abstract. An automatic system which is capable of recognizing white blood cells can assist hematologists in the diagnosis of many diseases. In this paper, we propose a new system based on image processing techniques in order to recognize five types of white blood cells in the peripheral blood. To segment nucleus and cytoplasm, a Gram-Schmidt orthogonalization method and a snake algorithm are applied, respectively. Moreover, three kinds of features are extracted from the segmented areas and two groups of textural features extracted by Local Binary Pattern (LBP) and co-occurrence matrix are evaluated. Best features are selected using a Sequential Forward Selection (SFS) algorithm and performances of two classifiers, ANN and SVM, are compared. In this application, the best result is obtained using LBP as the textural feature and SVM as the classifier. In sum, the results demonstrate that the methods are accurate and fast enough to execute in hematological laboratories.

Keywords: White blood cell, peripheral blood, segmentation, textural feature, feature selection, classification.

1 Introduction

Recognition and inspection of white blood cells in peripheral blood can assist hematologists in diagnosing many diseases such as AIDS, Leukemia, and blood cancer. Thus, this process is assumed as one of the most salient steps in hematological procedure. This analysis can be accomplished by automatic and manual approaches. Automatic methods usually examine white blood cells just quantitatively but not qualitatively, because they do not benefit from image processing techniques. Applying automatic systems juxtapose to image processing techniques may provide some qualitative evaluation and thus enhanced judgments. Furthermore, some of these tasks such as manually scrutinizing blood cells by experts are tedious and susceptible to error. Therefore, an automatic system based on image processing techniques can help the hematologists and expedite the trend.

Albeit not extensive, some methods are proposed in the literature for this purpose. Since segmentation is the most challenging step in white blood cells recognition procedure, improvement of nucleus and cytoplasm segmentation is the most widespread effort in many researches. For example, in [1], [2], [3] and [4], the authors suggested several methods to segment nuclei of white blood cells via

techniques that can be categorized into color-based methods. These methods are simple but are not capable of segmenting the white blood cells nucleus accurately. In addition, cytoplasm is colorless in most cases. Thus, its boundary is not detectable and cannot be segmented by these methods. Methods based on imaging techniques generate superior results. For example, the method proposed in [5] obtained more acceptable results using multi-spectral imaging techniques. In this method, intensity of each pixel in different spectra is used to construct the feature vectors and a support vector machine (SVM) is used for classification and segmentation. In spite of efficacy of this method for segmenting white blood cells components, this system's implementation is costly and thus cannot be used widely at all laboratories. Cytoplasm and nucleus segmentation via mathematical and contour models is the third method and also the most important one. In this field, some methods such as region growing [6], watershed [7], parameteric active contour deformable models [8], and also combination of the watershed technique and a parametric deformable model [9] are introduced in the literature. These methods are more complex and require more processing time in comparison with the first group of methods. However, their advantage is subtle more accurate segmentation.

Since morphological and textural features are the features which are elicited from white blood cells by a hematologist, many papers such as [1], [10], [11] use feature extraction methods on the basis of these features. For classification, Bayes classifier [12], different types of artificial neural networks (ANNs) such as feed-forward back-propagation [13] and [14], local linear map [15], fuzzy cellular neural network [16] are often used in the literature.

In this paper, our purpose is to design a new system based on image possessing methods to classify five major groups of white blood cells in peripheral blood. Therefore, at first, segmentation of white blood cells nuclei is carried out via Gram-Schmidt method. Then, distinguishing basophils from the other samples are performed using features extracted from nucleus areas. As cytoplasm edge is unobservable, a snake algorithm is used after some preprocessing procedures in order to segment the cytoplasm. The features elicited from the nucleus and cytoplasm areas in both steps are categorized into color, morphological, and textural features. Two groups of textural features attained by the Local Binary Pattern (LBP) and the co-occurrence matrix are evaluated. The feature selection step is adjoined to this process for ameliorating the classifier performance and expediting the program trend. Finally, the performance of two different classifiers, SVM and ANN, when using different sets of features is compared.

The rest of the paper is organized as follows. In Section 2, we propose a pertinent system for recognition of five types of white blood cells. The experimental results are presented and discussed in Section 3. Finally, Section 4 is appropriated to presentation of the conclusions.

2   System Architecture

Designing an automatic system to recognize five types of white blood cells in a hematological image of the peripheral blood is the main purpose of this work. It is necessary to design a block diagram based on this type of dataset. Fig. 1 illustrates the block diagram of our proposed system.

As shown in this figure, the method has three major phases whose details are explained in the next sections.
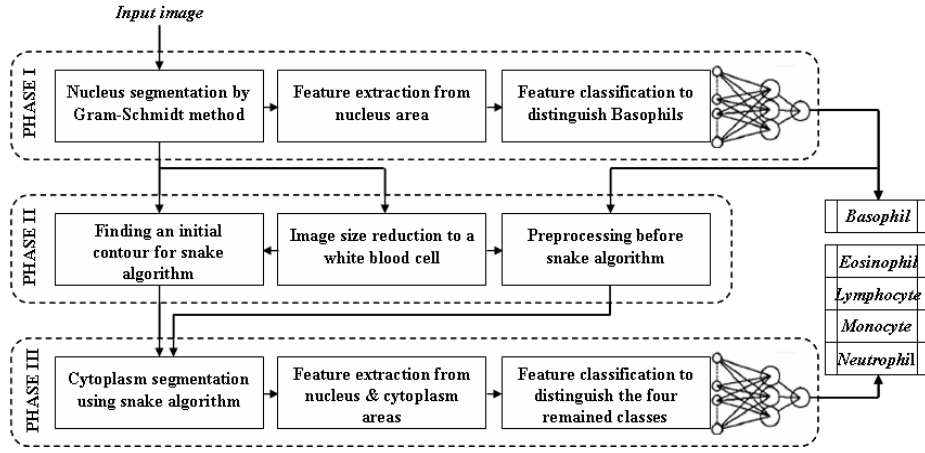
Fig. 1. The block diagram of the proposed system.

2.1 Phase I

Since in most of the samples, the boundary between nucleus and cytoplasm of basophils cannot be distinguished visually; these cells should not involve in segmentation of cytoplasm's step. Therefore, they should be recognized from the other samples in this Phase.

Segmentation of nucleus by Gram-Schmidt method: In this method, pixel intensities of the RGB components of the color image of the dataset are considered as 3D vectors. Then, v_1 as a desired vector is obtained by averaging the 3D vectors of the nucleus area in some samples. The v_2 and v_3 as undesired vectors are defined from the areas that are similar to the nucleus but are not the nucleus area. Using the Gram-Schmidt orthogonalization method proposed in [17] and v_1, v_2 and v_3, a weighting vector w is attained whose inner product with the pixel vectors results a composite image with higher intensity in the nucleus area compared to other areas (Fig. 2(a)). Next, by choosing an appropriate threshold based on the histogram information, we segment the image. The final result of the nucleus segmentation is shown in Fig. 2(b).
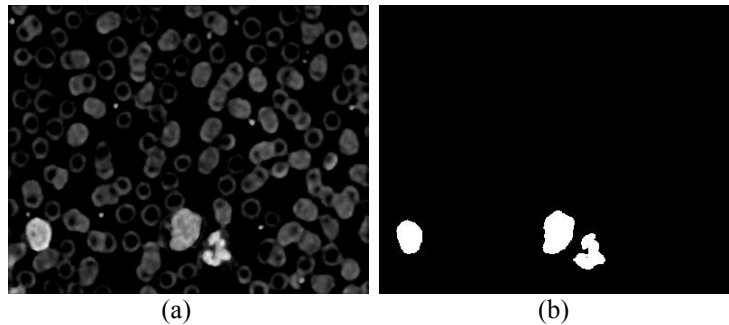

(a)   (b)

Fig. 2. (a) Resulting image with higher intensity in nucleus area relative to the other areas after applying the Gram-Schmidt orthogonalization method in a hematological image. (b) Final result of nucleus segmentation after thresholding.

Feature extraction from nucleus area and classification to distinguish basophils: As previously stated, basophils should be recognized and separated from the other types of white blood cells in this step. Therefore, some morphological features such as nucleus area and perimeter, number of the separated parts of nucleus, mean and variance of the nucleus boundaries and roundness criterion of nucleus are extracted from the segmented area. Color features are the other features extracted as a normalized vector of averaged nucleus color. To extract textural features, the co-occurrence matrix and the local binary pattern are applied and the results are compared.

Co-occurrence Matrix: The co-occurrence matrix is constructed on the basis of gray levels with the distances d and angles φ. In fact, this matrix describes the second order probabilistic features. Fourteen features are extracted from the co-occurrence matrix that explain contrast, homogeneity, entropy and others that properly represent the image textural properties. To make the features rotation invariant, 4 matrices are usually computed at 4 angles and the average of these matrices specifies the 14 features [18].

Local Binary Patterns (LBP): Local Binary Pattern (LBP) is another feature for texture processing. Because LBP analyzes textures in different radii, it can be supposed as a multi-resolution textural feature. Two features are usually extracted for each radius. The first one is LBP^{riu2} which represents the structure of texture and the other one is *VAR* which depicts changes in the gray levels [19].

2.2 Phase II

In this Phase, the main purpose is to prepare the image for the snake algorithm in order to segment the cytoplasm. To this end, image size reduction, preprocessing before the snake algorithm, and finding an initial contour for the snake algorithm are applied. These are explained below.

Image size reduction: Using the nucleus area segmented in the previous phase, we find the center of each nucleus and fit an appropriate window around to get sub images with a complete white blood cell. This trend makes the segmentation process easier. In our research, a 141*141 window is used.

Finding an initial contour for snake algorithm: To find an initial contour for the snake algorithm, the morphological dilation operation is first applied to the segmented nucleus region. The structuring element for the dilation operation is a square with an adaptive size based on the nucleus size. In this step, the boundary of this region is used as the initial contour. Fig. 3 shows this procedure.

|		
(a)	(b)	(c)
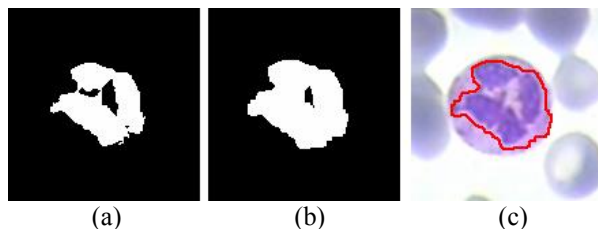
Fig. 3. (a) A segmented nucleus, (b) The image after dilation operation, (c) Initial contour.

The adaptive parameter, size of the structuring element, is because of the differences in the white blood cells sizes. To have a congruous initial contour next to

the cytoplasm edge, the contour of a small nuclei should be processed with the smaller dilation operator than a large nuclei.

Preprocessing before snake algorithm: Due to high accumulation of the red blood cells, they may touch the cytoplasm of the white blood cells. Thus, the boundary between the cytoplasm and the red blood cells may not be distinguished when the color image is changed into gray-scale. To solve this problem, the image is enhanced by color histogram equalization. Next, the enhanced image is transferred into the Hue-Saturation-Intensity (HSI) space. The final image is attained by extracting the saturation plate from the HSI image. Based on this idea, we have a gray-scale image that has good discrimination between the boundaries of the cytoplasm and the red blood cells. Then, the image is smoothed by a Gaussian kernel to eliminate the cytoplasm cavities and inhomogeneities and ameliorates the image for the snake algorithm (Fig. 4).

(a) (b) (c) (d)
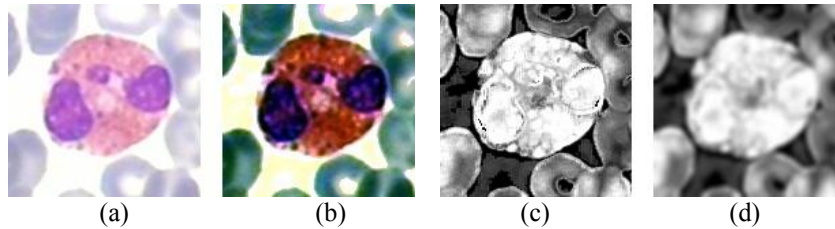
Fig. 4. (a) A sample image. (b) The image after Histogram Equalization. (c) The image after extracting the saturated plate. (d) The smoothed image by a Gaussian kernel.

2.3 Phase III

The main aim of this phase is to recognize the four remaining classes of the white blood cells. For this purpose, after preprocessing and initial contour detection, the snake algorithm is applied to segment the cytoplasm. After segmentation, textural and morphological features are extracted from both of the nucleus and the cytoplasm and the four cell types are classified.

Cytoplasm segmentation using snake algorithm: The snake algorithm used in this paper is the algorithm proposed in [20].It starts from the obtained initial contour and its parameters are set as $\alpha = 2, \beta = 5, \gamma = 0.7, \rho = 0.4$. The snake algorithm ends when no snake points move to new positions for four consecutive iterations.

Feature extraction from both nucleus and cytoplasm areas, and classifying the four remaining classes: in this step, features are extracted from the cytoplasm area in combination with the features extracted from the nucleus area in phase I. Retrospectively, these features are categorized into three groups of morphological, textural, and color features. The morphological features are cytoplasm area and whole cell body perimeter, mean and variance of the cytoplasm boundaries, roundness of the whole cell and the ratio between the cytoplasm and nucleus areas. Textural features are also extracted from the cytoplasm area by a co-occurrence matrix and the local binary pattern and their results compared. At the end, a normalized vector of the average cytoplasm color is extracted as color features.

3 Experimental Results

The proposed method was evaluated by 251 blood smear slide images acquired by a light microscope from stained peripheral blood using the Digital Camera-Sony-Model No. SSC-DC50AP with magnification of 100. The images contain 720*576 pixels and were classified by a hematologist into the normal leukocytes: basophil, eosinophil, lymphocyte, monocyte, and neutrophil. Also, the areas related to the nuclei were manually distinguished by an expert.

3.1 Segmentation Results

In order to quantitatively evaluate the results of the nucleus and cytoplasm segmentations, the following similarity measure is defined.

$$T_s = 100 \times \frac{A_{program} \cap A_{expert}}{\max(A_{program}, A_{expert})} \tag{1}$$

where $A_{program}$ is the segmented area by the algorithm and A_{expert} is the segmented area by an expert. When these two areas are the same, T_s is 100. In Table 1, the resulting measures for each kind of the white blood cells and their overall segmentation are presented.

Table 1. Similarity measures (Ts) for the segmentation of different types of white blood cells.

	Basophil	Eosinophil	Lymphocyte	Monocyte	Neutrophil	Overall
Nucleus	94.7%	90.81%	88.86%	96.7%	94.05%	**93.02%**
Cytoplasm	-	95.55%	93.05%	81.23%	97.25%	**91.79%**
Average	**94.7%**	**93.22%**	**90.01%**	**91.23%**	**96.23%**	**93.09%**

According to Table I, it can be inferred that the accuracy result of nucleus segmentation for the lymphocyte class is lower than the other classes. The main reason is that the color of the cytoplasm is analogous to the color of the nucleus in many of the lymphocytes samples, especially young ones. Therefore, the segmentation error for this type of the white blood cells is larger than the others. Also, since in most of the cases, the vitreous cytoplasm of monocytes is colorless, even the deformable model with a congruous preprocessing is unable to find the cytoplasm boundaries precisely. Therefore, the accuracy result of the cytoplasm segmentation is worse than those of the other classes.

3.2 Classification Results

In this paper, classification is performed in two sections, discriminating the basophils from the other types of white blood cells in phase I and recognizing the remaining classes in phase III.

To appraise the performance of the classifiers and the result of our proposed algorithms in recognizing the white blood cells, an accuracy criterion is used.

Experimental result for basophil classification: To compare the performance of the textural features, two groups of features, extracted from the nucleus area, are created. These two groups include similar morphological and color features but they are

different in textural features. To classify these features, at first, feature dimension is reduced by a SFS algorithm [21] and the results for ANN [22] and SVM [23] are compared by means of a total accuracy criterion. In this research, a Multi-Layer Perceptron (MLP) [22] ANN is used. Fig. 5(a) and (b) illustrate the results related to the local binary pattern and the co-occurrence matrix, respectively.

Some points are construed from these figures. The first point is that reduction in dimension of features aggravates the classification as expected. The second one is that the overall accuracy does not have considerable escalation after 15 features for both ANN and SVM. This occurs because after selecting 15 features, the differentiation between the features of each class is at a maximum and increasing the features dimension perplexes the classifier. The third one is that the ANN classifier has more fluctuation in the overall accuracy in comparison with the SVM. These changes are because the MLP is not trained well as a result of trapping in a local minimum. The other point is that the ANN and SVM classifiers have similar performances in most of the feature dimensions. But due to the stability of the SVM in training, it may be preferred. The last point is that the groups of the features attained from the co-occurrence matrices have generally better performance in comparison with the groups of features obtained from LBP. However, the time required for calculating the first group of features is significantly higher than the second one.
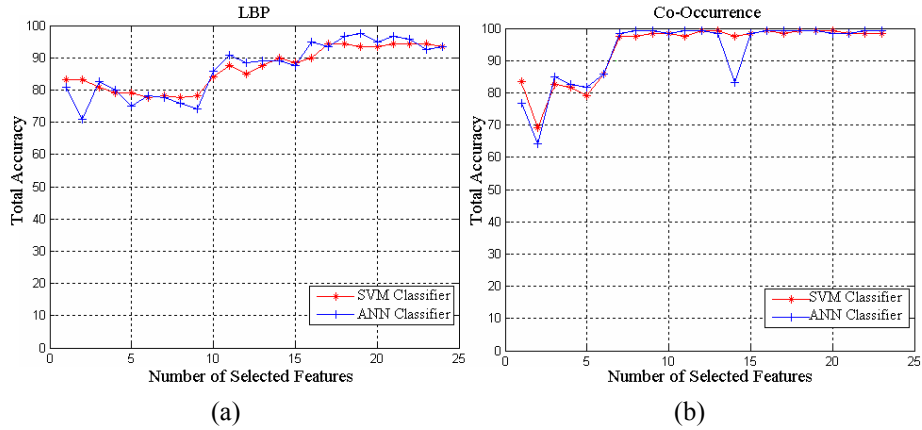
Fig. 5. The overall accuracy results for the ANN and SVM using: (a) LBP (b) Co-occurrence matrix as their textural features, to distinguish basophils.

According to the above conclusion, selecting 15 features and utilizing SVM is the best way to have optimal performance in classification results and speed. To compare these two groups of features, Tables 2 and 3 show confusion matrices, the accuracy, and the overall accuracy for these two groups when 15 features are selected and SVM is used as the classifier, respectively.

Table 2. Confusion matrix, Accuracy, and Overall Accuracy for 15 LBP features and SVM classifier.

	Recognized Basophil	Recognized Non-Basophil	Accuracy
Basophil	50	0	100%
Non-Basophil	23	150	86.71%
Overall Accuracy			89.69%

Table 3. Confusion matrix, Accuracy, and Overall Accuracy for 15 Co-Occurrence features and SVM classifier.

	Recognized Basophil	Recognized Non-Basophil	Accuracy
Basophil	49	1	98%
Non-Basophil	5	168	97.11%
Overall Accuracy			98.64%

According to these tables, the results of classification with the features of the co-occurrence matrix are superior to those of the LBP. However, considering a trade-off between accuracy and processing time, LBP may be preferred. The ratio between the times required for feature extraction using the co-occurrence and LBP methods is 20 to one.

Experimental result for classifying the four remaining groups of white blood cells: Similar to the previous section, the performance of the two groups of features whose textural features are extracted by the LBP and the co-occurrence matrix are compared whilst these features are obtained from both of the nucleus and cytoplasm areas. For a second time, the SFS algorithm is applied in order to select the best features in a prespecified dimension and the results are compared for ANN and SVM by the accuracy criterion. Fig. 6(a) delineates the overall accuracy of the SVM and ANN for the first group of features and Fig. 6(b) demarcates the result for another group of features.

(a)   (b)
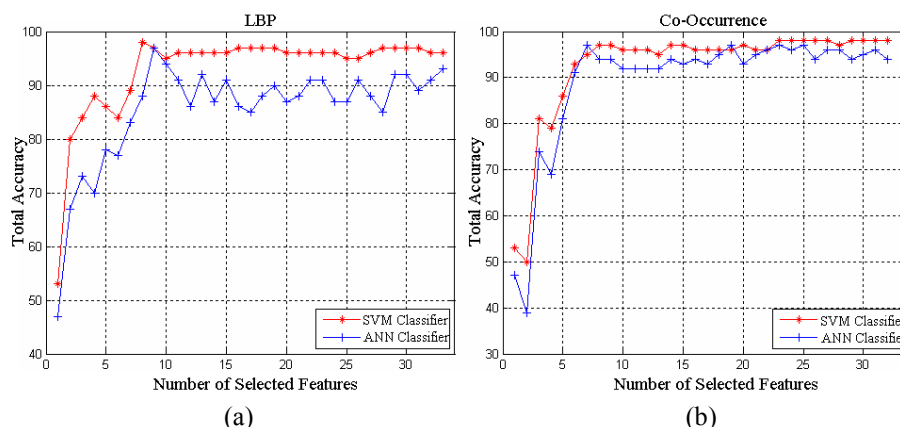
Fig. 6. The overall accuracy results of the ANN and SVM using: (a) LBP; (b) Co-occurrence matrix as their textural features, in order to distinguish the 4 remaining classes.

According to the above figures, some points related to the performance of the classifiers are analogous to the conclusion discussed in the previous section. For instance, reduction in dimension of features exacerbates the classifier performance. Furthermore, the oscillation of the overall accuracy of the ANN classifier is considerably more than that of the SVM. Other points inferred from the figures are: after 10 features, curves do not have significant escalation for both of the ANN and SVM, the SVM classifiers have superior performance in this case in comparison with the ANN. In addition, it is obvious that the groups of features obtained from the co-occurrence matrices have generally superior performance again.

In conclusion, considering both of the classification accuracy and processing time, SVM classifiers and feature dimension of 10 can be considered as optimal.

Tables 4 and 5 illustrate the confusion matrix, accuracy, and overall accuracy when 10 features are selected and SVM is used as a classifier with the local binary pattern and the co-occurrence matrix, respectively. According to these tables, in this case, the classification accuracy for the two defined groups of features seems to be equal. However, noting this fact that calculation of the features extracted by the co-occurrence matrix is noticeably more computational, using LBP for the textural features is proposed for this phase again.

Table 4. Confusion matrix, Accuracy, and Overall Accuracy for 10 LBP features and SVM classifier.

	Recognized Eosinophil	Recognized Lymphocyte	Recognized Monocyte	Recognized Neutrophil	Accuracy
Eosinophil	19	0	0	0	100%
Lymphocyte	0	27	2	0	93.1%
Monocyte	0	0	23	1	95.83%
Neutrophil	1	0	0	27	96.43%
Overall Accuracy					96%

Table 5. Confusion matrix, Accuracy, and Overall Accuracy for 10 the co-occurrence features and SVM classifier.

	Recognized Eosinophil	Recognized Lymphocyte	Recognized Monocyte	Recognized Neutrophil	Accuracy
Eosinophil	18	0	0	1	94.74%
Lymphocyte	0	27	2	0	93.1%
Monocyte	0	0	23	1	95.83%
Neutrophil	0	0	0	28	100%
Overall Accuracy					96%

4 Conclusion

In this paper, we proposed a system in order to recognize five groups of white blood cell in the peripheral blood. The proposed system has a reasonable processing time and is sufficiently accurate. The overall segmentation result of 93% and classification accuracies of 90% and 96% in phases I and III verify the accuracy of the system. Regarding the processing time, the program requires 10 seconds for analyzing a single white blood cell on a Pentium-4 PC, running at 3.2 GHz, with 1 GB of RAM and MATLAB. Hence, differential counting of 100 white blood cells lasts about 16 minutes. As a comparison, an expert requires almost 15 minutes to carry out this process. Thus, this program can be used in the hematological laboratories.

Notwithstanding mentioned advantages, the proposed method may need initial calibration at the start point of the program when new datasets with different characteristics are introduced to the system. This is due to the required alignment of the initial vectors in the Gram-Schmidt method. As a future work, an algorithm can be

designed to align the three preceded vectors in the Gram-Schmidt method automatically. In addition, it is cogent to add a new class for the white blood cells that do not belong to those five classes. This is due to the fact that sometimes other cells called Blast appear in the peripheral blood. This cell type is more frequently found in the abnormal blood samples.

References

1. Sabino, D.M.U., Costa, L.D.F., Rizzatti, E.G., Zago, M.A.: A Texture Approach to Leukocyte Recognition. Real-Time Imaging. 10 (4), 205--216 (2004)
2. Angulo, J., Flandrin, G.: Automated Detection of Working Area of Peripheral Blood Smears Using Mathematical Morphology, Anal. Cell. Pathol. 25, 37--49 (2003)
3. Wu, J., Zeng, P., Zhou, Y., Oliver, C.: A Novel Color Image Segmentation Method and Its Application to White Blood Cell Image Analysis. In: 8th International Conference on Signal Processing (ICSP06), vol. 2, pp. 245--248. IEEE Press, Beijng (2006)
4. Umpon, N.T.: Patch-based White Blood Cell Nucleus Segmentation Using Fuzzy Clustering. ECTI Trans. Electrical. Electronic. Communications. 3 (1), 5--10 (2005)
5. Guo, N., Zeng, L., Wu, Q.: A Method Based on Multi-Spectral Imaging Technique for White Blood Cell Segmentation. Comput. Biol. Med. 37(1), 70--76 (2006)
6. Chassery, J.M., Garbay, C.: An Iterative Segmentation Method Based on Contextual Color and Shape Criterion. IEEE Trans. Pattern Anal. Machine Intell. 6 (6), 795--800 (1984)
7. Jiang, K., Liao, Q.M., Dai, S.Y.: A Novel White Blood Cell Segmentation Scheme Using Scale-Space Filtering and Watershed Clustering. In: 2nd International Conference on Machine Learning and Cybernetics, vol. 5, pp. 2820--2825. (2003)
8. Zamani F., Safabakhsh, R.: An Unsupervised GVF Snake Approach for White Blood Cell Segmentation Based on Nucleus. In: 8th International Conference on Signal Processing (ICSP06), vol. 2. IEEE Press, Beijng (2006)
9. Park J., Keller, J.M.: Snakes on the Watershed. IEEE Trans. Pattern. Anal. Mach. Intell. 23 (10), 1201--1205 (2001)
10. Umpon N.T., Dhompongsa, S.: Morphological Granulometric Features of Nucleus in Automatic Bone Marrow White Blood Cell Classification. IEEE Trans. Inf. Technol. Biomed. 11 (3), 353--359 (2007)
11. Thiran J.P., Macq, B.: Morphological Feature Extraction for the Classification of Digital Images of Cancerous Tissues. IEEE Trans. Bio. Med. Eng. 43 (10), 1011--1020 (1996)
12. Hengen, H., Spoor, S.L., Pandit, M.C.: Analysis of Blood and Bone Marrow Smears Using Digital Image Processing Techniques. In: M. Sonka, J.M. Fitzpatrick, (Eds.), Proceedings of the SPIE, Medical Imaging 2002: Image Processing, vol. 4684, pp. 624--635. (2002)
13. Umpon N.T., Gader, P.D.: System-Level Training of Neural Networks for Counting White Blood Cells. IEEE Trans. Syst., Man, Cybern. 32 (1), 48--53 (2002)
14. Long, X., Cleveland, W.L., Yao, Y.L.: A New Preprocessing Approach for Cell Recognition. IEEE Trans. Inf. Technol. Biomed. 9 (3), 407--412 (2005)
15. Nattkemper, T.W., Ritter, H.J., Schubert, W.: A Neural Classifier Enabling High-Throughput Topological Analysis of Lymphocytes in Tissue Sections. IEEE Trans. Inf. Technol. Biomed. 5 (2), 138--149 (2001)
16. Shitong, W., Min, W.: A New Detection Algorithm (NDA) Based on Fuzzy Cellular Neural Networks for White Blood Cell Detection. IEEE Trans Inf. Technol. Biomed. 10 (1), 5--10 (2006)
17. Rezatofighi, S.H., Soltanian-Zadeh, H., Sharifian, R., Zoroofi, R.A.: A New Approach to White Blood Cell Nucleus Segmentation Based on Gram-Schmidt Orthogonalization," 2nd International Conference on Digital Image Processing (ICDIP09), pp. 107--111. IEEE Press, Thailand (2009)
18. Haralick, R., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man. Cyber., 3 (6), 610--621 (1973)

19. Ojala, T., PietikaÈ inen, M., MaÈenpa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Anal. Machine Intell. 24 (7), 971-987 (2002)
20. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. of Comput. Vision. 1 (4), 321--331 (1988)
21. Pudil, P., Novovičová, J., Kittler, J.: Floating Search Methods in Feature Selection. Pattern Recogn. Lett. 15 (11), 1119--1125 (1994)
22. Basheer I.A., Hajmeer, M.: Artificial Neural Networks: Fundamentals, Computing, Design, and Application. J. Microbiol. Meth. 43 (1), 3--31 (2000)
23. Burges, C.J.C.: A tutorial on Support Vector Machines for Pattern Recognition. Knowledge Discovery and Data Mining, 2 (2), 121--167 (1998)