CIMSA 2004 – IEEE International Conference on
Computational Intelligence for Measurement Systems and Applications
Boston, MA, USA, 14-16 July 2004

Morphological Classification of Blood Leucocytes by Microscope Images

Vincenzo Piuri, Fabio Scotti
University of Milan, Department of Information Technologies, via Bramante 65, 26013 Crema, Italy

Abstract *– The classification and the count of white blood cells in microscopy images allows the in vivo assessment of a wide range of important hematic pathologies (i.e., from presence of infections to leukemia). Nowadays, the morphological cell classification is typically made by experienced operators. Such a procedure presents undesirable drawbacks: slowness and it presents a not standardized accuracy since it depends on the operator's capabilities and tiredness. Only few attempts of partial/full automated systems based on image-processing systems are present in literature and they are still at prototype stage. This paper presents a methodology to achieve an automated detection and classification of leucocytes by microscope color images. The proposed system firstly individuates in the blood image the leucocytes from the others blood cells, then it extracts morphological indexes and finally it classifies the leucocytes by a neural classifier in Basophil, Eosinophil, Lymphocyte, Monocyte and Neutrophil.*

Keywords *– Automatic cell classification, morphological analysis, classifiers, leucocytes identification, image processing.*

I. INTRODUCTION

The microscope inspection of blood slides provides important qualitative and quantitative information concerning the presence of hematic pathologies. From decades this operation is performed by experienced operators, which basically perform two main analyses. The first is the qualitative study of the morphology of the cells and it gives information of degenerative and tumoral pathologies such as leukemia. The second approach is quantitative and it consists of differential counting the white blood's cells. Automated cell-counter systems (for example laser-based citometers [1]) are available on the market, but they are not image-based/morphological and they destroy the blood samples during the analysis [1, 2]. In addition, citometers do not allow direct morphological sub-classification of leucocytes, such as separating tumoral leucocytes from normal ones.

Fig. 1 shows a typical blood microscope image which has been digitalized by a CCD and acquired by a frame-grabber system [3]. Principal cells present in the blood are *red* blood cells, and the *white* cells (leucocytes). Leucocyte cells containing granules are called granulocytes (composed by *neutrophil*, *basophil*, *eosinophil*). Cells without granules are called agranulocytes (*lymphocyte* and *monocyte*). The percentage of leucocytes in human blood typically ranges between the following values: neutrophils 50-70%, eosinophils 1-5%, basophils 0-1%, monocytes 2-10%, lymphocytes 20-45% [4]. These cells provide the major defense against infections in the organism and their specific concentrations can help specialists to discriminate the presence or not of very important families of pathologies (i.e. the presence of mononucleosis, hepatitis diabetes, allergy, arthritis, anaemia, and many others) [4].

Unfortunately, the accuracy of cell classification and counting is strongly affected by individual operator's capabilities. In particular, the identification and the differential count of blood's cells is a time-consuming and repetitive task that can be influenced by the operator's accuracy and tiredness.

The diffusion of mass screening programs is continually pushing the need for fully-automated non-destructive systems for fast and accurate blood analysis. The proposed fully automated morphological classification system can be also indeed considered as a first step to the automatic detection/monitoring of blood pathologies such as various leukemia types inspecting the variation in *morphology* of white cells.

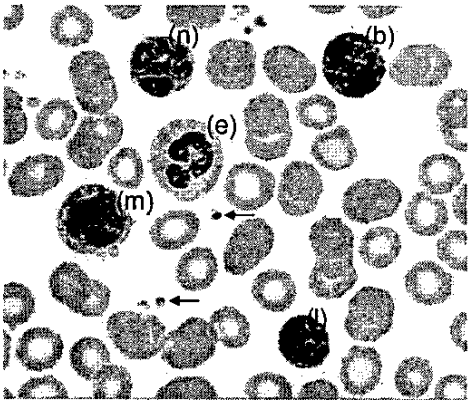

Fig. 1. Blood's white cells marked with colorant: basophil (b), eosinophil (e), lymphocyte (l), monocyte (m), and neutrophil (n). Arrows indicate platelets. Others elements are red cells.

In this paper we focus on the problem of identification and classification of white blood cells by microscope images. The proposed system firstly individuates in the blood image the leucocytes from the others blood cells, secondly it extracts morphological indexes and finally it classifies the leucocytes by a neural classifier (figure 2). Section 2 explains how to select white cells form the blood components in an image, section 3 describes how to perform a morphological analysis of white cells in order to extract a suitable set of indexes. These indexes are used to discriminate the 5 types of leucocytes by a classifier. Section 4 describes the design of

the classifier, how classification accuracy can be tested and how a proper classifier can be chosen from a set of candidates of different families (Nearest Neighbor classifier, Radial Basis function neural networks, feed-forward neural networks).

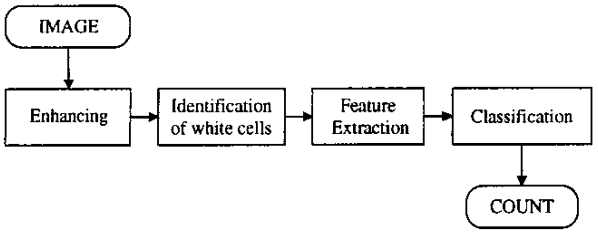

Fig. 2. The structure of modules composing the cells classification system.

II. IMAGE ENHACING AND IDENTIFICATION OF WHITE CELLS

The first processing step of the architecture enhances the input image and it selects the white cells present into the image by separating them from others blood's components (red cells and platelets).

Typical acquisition systems for optical microscopes produce very good quality color images: only low *quantization noise* is typically present in images and cells are represented with proper definition (cells have a diameter at least of 30 pixels). Occasionally, low contrast image can be acquired but, typically, background illumination tends to be uniform and adapt to subsequent image processing phases. These are our working hypotheses. If input images do not satisfy such hypotheses, suitable low-pass and band-pass filters should be envisioned to reduce the noise. Other filters can be adopted to reduce *defocusing* problems [5] and/or to reduce *non uniform backgrounds* [6]. Low contrast image can be effectively recovered by adaptive contrast stretching filters [5].

All the modules of the presented system work on gray-level images. This design choice, which apparently can be thought as a leak of information, indeed is driven by the consideration that the colorant used to mark the white cells can greatly chance the chromatic characteristics of cells from one experiment to another. It can be related to many uncontrollable experimental parameters such as operator capabilities to prepare the slides.

Proposed pre-processing steps are based on three very well-established assumptions [4]:

a) the colorant used in the preparation of the blood tends to concentrate only in white cells, in particular in their nuclei that are typically center-positioned (as shown in figure 1 and 3, white cells are the darker elements in images);

b) red cells are thinner in their center than in the edge, hence their nuclei appear more pale than the border (leucocytes are in the opposite situation);

c) platelets are much smaller than white and red cells.

Our approach to the identification of leucocytes in the blood image is based on a *adaptive prefiltering* and *segmentation*. The pre-filter uses hypotheses a), b) and c) to enhance only the leucocytes. It separates the others blood components with respect to the gray level intensities in order to achieve e very accurate leukocyte segmentation. The processing steps are as follows:

#1. *Contrast stretching*. Firstly, the input image is pre-filtered by a contrast stretching filter in order to enlighten *only* the leucocyte's nuclei (from figure 3.A to figure 3.B) [5]. As described, they tend to be the darker areas of the gray-level image (hypotheses a).

#2. *Opening*. Since the nuclei of leucocytes are a connected circular-shaped area of dark pixels and other blood components are not (hypothesis b and c), it is possible to enhance the nuclei and reduce other components by a morphological filter based on the *opening* operator with *structured element* [6,7]. This operator, if applied to gray level images, tends to enhance pixels which belong to areas of similar intensity *and* larger then the structured element. Conversely it tends to reduce the intensity of small groups of pixels smaller than the structured element.

Let us now discuss how the structured element of the opening filter can be adaptively designed. This step is important since it allows correct subsequent leukocyte segmentation. In this paper we propose to choose a structured element that mimics the average leukocyte cell, in particular a *circular pattern* of diameter **D**, where **D** is the average nucleus diameter. **D** can be given in input by the operator during the set-up of the system or automatically processed by images.

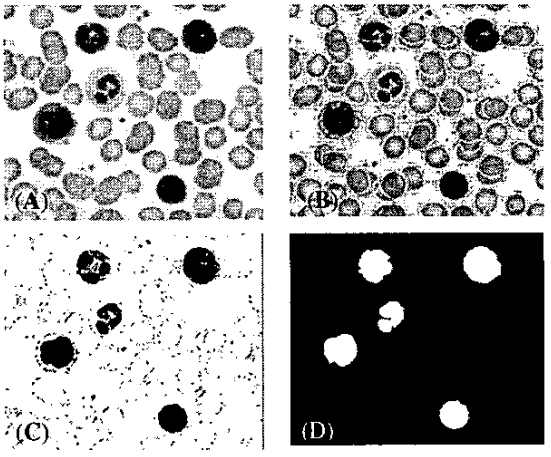

Fig. 3. White cells identification process: intermediate steps. (A): Gray-level converted image. (B): Contrast-stretched image. (C): opening of image B. (D): Segmented image of C.

Automatic detection of diameter of cells present into the image is typically performed by binarizing the image and counting the connected areas of image elements (mostly composed of red cells and leucocytes), and processing their average **A**. Considering elements as circular-shaped, the

average cell radius is hence $\mathbf{D'} = 2\sqrt{A/\pi}$. Since leucocytes are typical bigger than the average cell diameter, parameter D can hence be considered as $\mathbf{D} = \mathbf{D'}(1+\alpha)$ where $\alpha = 0.2$ for example. Other methods can be found in [6].

These operations are important in order to perfectly match only the leucocytes with segmentation: over-segmentation [8] will produce on overestimation of the real number of leucocytes in the image, immediately worsening the accuracy of the overall system.

Figure 3.C shows the effect of the opening operator with its structured element tuned on cell diameter **D'**. After this operation, the nuclei are properly enlightened and hence the segmentation can correctly locate and label each leucocyte.

Last step of this stage consists of the partition of the original image (A) in squared sub-images (*areas of interest*). Each sub-image must be properly located and it must entirely contain the leucocyte (within a portion of the background). The center of the sub-images will be located in the centroid position of the segmented nuclei. The dimension of the squared sub-image has been fixed to 3**D**.

III. AUTOMATIC MORPHOLOGICAL ANALYSIS

The previously described chain locates each leucocyte in a sub-image (figure4.a). It is now possible to extract morphological characteristic of each leucocyte in two subsequent steps: firstly processing the membrane and finally the nucleus.

A. Membrane detection

Let us now discuss the feature extraction phase of the external membrane of leucocytes. Figure 4 shows the proposed steps:

#3. *Contrast stretching*. A contrast stretching filter has been applied to the area of interest (figure 4.a) in order to enhance the contrast between the background (white) and the gray internal of the membrane (figure 4.b);

#4. *Edge detection*. This step reconstructs the borders of the membrane. Canny-based filters [6] have to be preferred for their intrinsic capacity to ensure in output a continuous edge. The output of edge-detection filtering is a binarized image (figure 4.c).

#5. *Dilation*. The morphological operator called dilation has been used [6] to better connect separated points of the perimeter and to make the perimeter of the cell as a *connected item* (thicker more than one pixel, [7,9]) as shown in figure 4.d.

#6. *Filling*. Last step consists of filling internal holes of the biggest connected elements of the processed image [6,9]. The biggest connected element in the region of interest is indeed the membrane perimeter (figure 4.d). The result of the processing is plotted in figure 4.e.

At this stage, the correctness of previously performed steps can be controlled. It is important to note that we expect a cell perimeter slightly around the value $\pi\mathbf{D}$ previously processed. If the obtained value of the perimeter is much smaller than this, wrong membrane detection can be occurred. In this case, the optional dilation filter (filtering step #3) should be applied or re-applied. If the perimeter values do not reaches the correct expected value, the region of interest should be discarded in order to avoid further errors in the feature extraction phases. If the processed perimeter value is much bigger than expected values, it could mean that two superimposed cells have been considered as one. Hence, the analysis of the region of interest can be repeated without the dilation filter, which probably joined two partially superimposed cells. Again, if the perimeter values do not return to acceptable values, the region of interest must be discarded. In our analysis we consider as correct values of the perimeter the values that range between 100% and 250% of $\pi\mathbf{D}$.

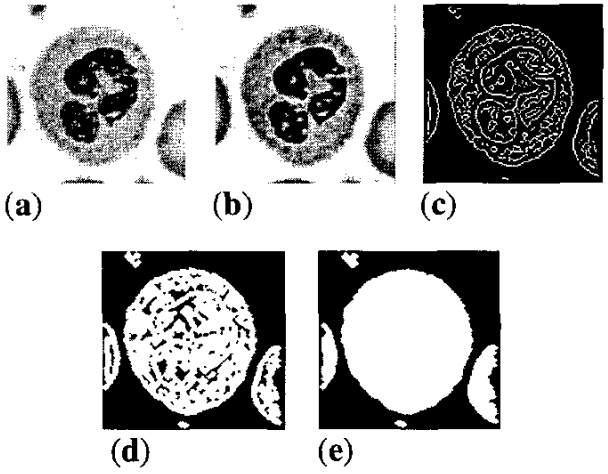

Fig. 4. From area of interest (a) to feature extraction of external membrane leucocyte's characteristics. Phases: (b) contanst stretching, (c) Canny's edge detection, (d) dilation operator, (e) internal holes filled.

Previous steps transformed the leucocyte's portion in the region of interest into a fully connect area in a black and white image (figure 4.e). At this stage, the system can easily processe the membrane parameters such as *area*, *perimeter* and geometrical *momentums* of the external membrane [6,9,10]. A complete description of features extracted from this sub-image is given in next sections.

B. Nucleus detection

The nucleus feature extraction exploits the results of the membrane detection phase. Since by definition, for not-broken cells, the nucleus is internal to the membrane, it is useful to reload the contrast-stretched image (figure 4.b) and process the following filtering operation:

#7. *Cropping.* In this step we crop the external portion of the membrane (figure 5.a). This step ensures a more robust extraction of the nucleus features since it excludes undesired image elements producing images similar to the one plotted in figure 5.b [6,11].

#8. *Contrast stretching.* Contrast stretching filtering is applied to better separate nucleus from cytoplasm (figure 5.b).

#9. *Minimum intensity homogenization.* This step applies a 4x4 low-pass morphological filter that ensures a better homogenization of the nucleus's color (only areas darker than a fixed threshold are uniformly averaged in their gray-level by a 4x4 matrix). Figure 5.c show only the area above the fixed threshold. In [6,10] different methods are proposed to automatically select the threshold.

#10. Step #4, #5 and #6 are then repeated. Images d), e) and f) in figure 5 are the corresponding examples.

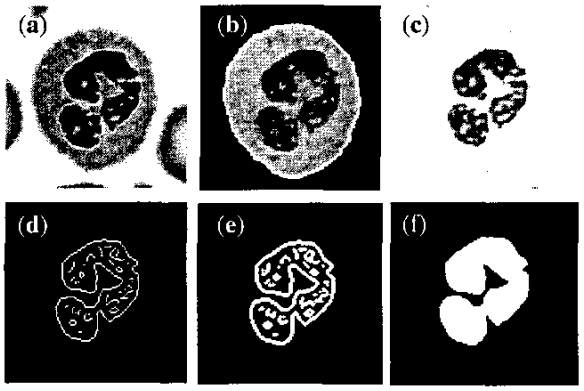

Fig. 5. From area of interest (a) to nucleus' feature extraction. Phases: (b) contrast stretching of cropped membrane, (c) low-pass filtering of only darker areas, (d) Canny based edge detection (e) dilation operator, (e) internal holes filled.

Again it is possible to control the exactness of the performed steps by controlling area's value of the extracted nucleus (similarly to the controls described after step #4 concerning the external membrane). Of course the nucleus' area must have a value minor of the area of the external membrane. In addition, we fixed that the value must be above a fixed threshold (20% of A in our study) since we assume that a classifiable leukocyte must have a nucleus. Extracted cells with values outside these ranges must be discarded from the classification/counting.

In leukocyte classification, an important index used by experts is the *count of the nucleus's lobes.* For example Neutrophils can have up to 5 lobes, Eosinophils typically 2 but rarely up to 3-4, Basophil almost 2-3, Lymphocyte always 1 (not considering tumor deformed Lymphocyte), etc. The extraction of this value has been automatically achieved with *iterative erosion filtering* of the binarized nucleus (figure 5.e).

Figure 6 shows the variation of the numbers of objects in the binarized-and-filled image of the nucleus through iterative erosion filtering. The correct number of lobes can be easily extracted by counting the number of connected objects with area larger than a fixed threshold (equals to 5 pixels in our experiments) in every single iteration, and extracting the maximum obtained value. In figure 6, the maximum value of separated objects is obtained in iteration number 9-12 and it is equal to 4, hence the nucleus has 4 lobes.

C. Features extracted

Let us now resume the most important indexes that can be extracted from sub-images and their biological relevance in the leucocytes classification. Firstly, we separately evaluated for the binarized membrane and the nucleus the following 8+8 indexes. We used the standard procedure present in the Matlab Image Processing Toolbox [6]: Area, Perimeter, Convex Area, Solidity, Major Axis Length, Orientation, Filled Area, Eccentricity. In addition we processed the ratio between the cell and nucleus areas, the nucleus' "*rectangularity*" (ratio between the perimeter of the tightest bounding rectangle and the nucleus perimeter), the cell "*circularity*" (ratio between the perimeter of the tightest bounding circle and the cell perimeter), the number of lobes, and finally the solidity, area and mean gray-level intensity of the cytoplasm. Total number of extracted features is 23.

The choice of indexes has been driven by suggestions of hematopathologist present in the literature [5]. In fact, experts classify cells by qualitatively evaluating the same cells features.

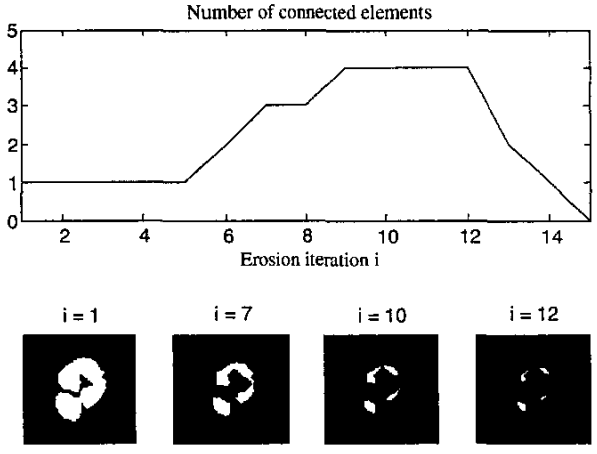

Fig. 6. Extracting the number of nucleus' lobes. Correct number of lobes corresponds to maximum number of connect objects (with area > 5) during iterative erosion. 4 lobes are present and detected.

IV. CLASSIFIER CREATION AND TEST

The system has been tested using sample-images extracted from an accredited image repository [3]. Our dataset consists of 113 images that contain 134 expert-labeled leucocytes. The test of the overall systems has been divided in three phases.

In the first phase we tested the first three modules of the system (*enhancing*, *identification* and *feature extraction*). Input images have been processed by modules and we obtained 134 sub-images with single correctly positioned leucocytes. The system correctly produced for each leukocyte its vector of extracted features.

In the second phase, experts associated the correct classification of each extracted leukocyte. The resulting dataset (a 23x134 matrix of features and 134 classifications vector) has been used to test the last module of the system: the classifier.

The capability of selected features in separating the 5 classes can be qualitatively evaluated by plotting the classes with respect to the three most relevant features: cell area, nucleus area and gray intensity of the cytoplasm (figure 7). The most relevant features have been found by applying a feature selection technique called *forward selection* based on nearest neighbor classifier evaluated with Leave One Out method [13].

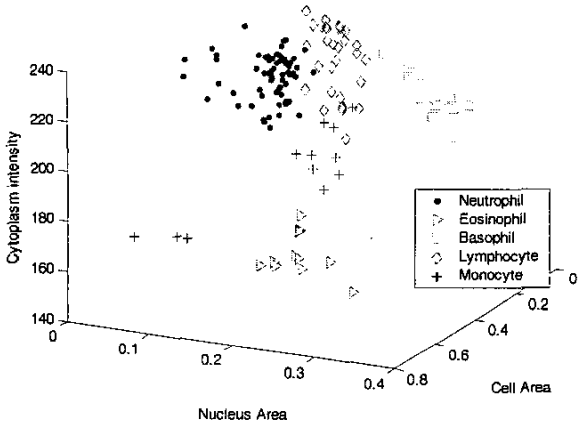

Fig. 7. Separation of classes using the three most relevant features: cell and nucleus area (normalized), and gray intensity of cytoplasm.

In the third phase we created, tested and compared different families of classifiers [14]. Since at this stage of the research the dimension of the dataset is limited to 134 samples, the evaluation of the system accuracy has been performed by N-fold cross-validation technique where N was chosen equals to 10 [13]. The first classifier family we considered is the nearest neighbor classifiers (kNN). Different kNN's have been created ranging the number of nearest neighbors k from 1 to 15 and considering the Euclidean, cubic and Manhattan norm [13]. Feed-forward neural networks with log-sigmoidal activation function (FF-NN) and with two hidden layers have been created by ranging the number of the hidden units from 2 to 50. We used the Levemberg Marquard training method with Bayesian regularization present in the Matlab Neural Network Toolbox [6]. Also Radial Basis Function neural networks have been tested with different number of centers (from 1 to 120) [13].

In our experiments, we considered also a parallel classifier built with feed-forward neural networks (figure 8) [15]. In this classifier family we have one FF-NN for each class. Each neural network is separately trained to recognize if the input sample belongs to the own class or not. The outputs of the neural networks range from 1 to 0 depending on if the sample belongs or not to the class. The final decision module selects the network which has the higher value in output and it associates the final classification to that class [16]. For example, if an unknown sample corresponding to a Basophil cell is put in input, the output of the network associated with this class (labeled with "b" in figure 8) should be the highest. Hence, the decision module will produce in output the class "Basophil". The neural networks required to build the parallel classifier have been trained and selected with the same method previously described for FF-NN.

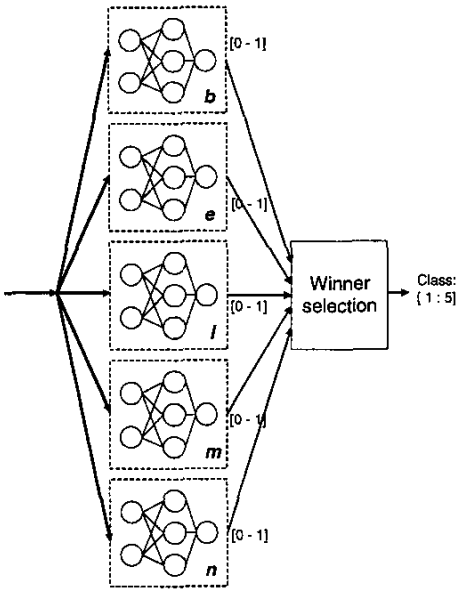

Fig. 8. Parallel structure of neural networks. Five neural networks have in input the same input vector and classify one single class (b,e,l,m,n). Outputs range from 0 to 1. A decision module produces the final classification (b,e,l,m,n).

Table 1 reports the results of the best classifier for each family that as been found during the validation test.

Notably the parallel classification system achieves the lowest mean miss-classification error (with lowest standard deviation). Furthermore, the corresponding processing time is

not the highest with respect to other tested classifiers. That information is very important to design the final systems when enlarged dataset will be considered. In fact, the computation complexity of the kNN classifier is almost proportional to the training dataset size. On the contrary feed-forward neural networks' complexity is related to the present number of neurons which tends to be much lower than the number of training samples.

TABLE 1. Accuracy of the best classifiers (23 features)

Type	Mean error	Standard deviation	Execution time	Note
KNN	0,14	0,091	1254 μs	K=1, Euclidean norm
FF-NN	0,18	0,10	309 μs	[20,1] Neurons
RBF	0,17	0,11	1986 μs	100 centers
Parall. FF-NN	0.08	0.09	852 μs	Composed By 5 FF-NN

V. CONCLUSIONS

This paper presented a methodology to achieve a fully automated detection and classification of leucocytes by microscope color images identifying the following classes: Basophil, Eosinophil, Lymphocyte, Monocyte and Neutrophil. Experiments show that the final classification module implemented by means of a parallel classifier composed by feed-forward neural classifiers achieves an accurate solution with minor computational complexity than traditional nearest neighbor classifier.

Results indicate that the morphological analysis of blood's white cells is achievable and it offers remarkable classification accuracy. Further studies will be focused on identification of tumor deformations in the cell morphology for a fully-automated diagnostic system.

REFERENCES

[1] Abbott Diagnostics Website, http://www.abbott.com/products/diagnostics.htm/

[2] Beckman Coulter Website, http://www.coulter.com/coulter/ Hematology/

[3] F. Cillesen, W. Der Meer, "Atlas of Blood Cell Differentiation", Elsevier Science B.V., Amsterdam, The Netherlands, 1998

[4] "Blood", Keith Breden Taylor and Julian B. Schorr, Colliers Encyclopaedia, Vol 4, 1978

[5] J.S. Lim "Two dimensional signal and image processing" Prentice Hall 1990

[6] R. C. Gonzalez, R. E. Woods, S.L. Eddins, "Digital Image Processing Using MATLAB",Pearson Prentice Hall Pearson Education, Inc., New Jersey, USA, 2004

[7] H. J. A. M. Heijmans (1994): Morphological Image Operators, Academic Press, New York.

[8] K. Catellman. Digital Image Processing , Prentice-Hall, NJ, 1996

[9] R. M. Haralick, S. R. Sternberg , X. Zhuang, "Image Analysis Using Mathematical Morphology", IEEE Transactions on Pattern Analysis and Machine Intelligence, 9 (4): 532-550, 1987.

[10] P. Kuosmanen and J. Astola (1995) "Soft Morphological Filtering", Journal of Mathematical Imaging and Vision, 5 (3): 231-262.

[11] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification and Scene Analysis, second ed., New York, John Wiley & Sons, 2000

[12] Andrea Biondi, Giuseppe Cimino, Rob Pieters, and Ching-Hon Pui , "Biological and therapeutic aspects of infant leukemia ",Blood, Vol. 96 No. 1 July, pp. 24-33, 2000:

[13] A.K. Jain, R.P.W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 4–37, 2000

[14] R. Duin, "A note on comparing classifiers", Pattern Recognition Letters, 1996, vol. 1, pp. 529–536.

[15] K.Woods, W. P. Kegelmeyer Jr., K.Bowyer, "Combination of Multiple Classifiers Using Local Accuracy Estimates", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 4, april 1997

[16] T.K. Ho, J.J. Hull, and S.N. Srihari, "Decision Combination in Multiple Classifier Systems", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, pp. 66–75, Jan. 1994.