JOURNAL OF AUTOMATIC CONTROL, UNIVERSITY OF BELGRADE, VOL 16:29-32, 2006 ©

Automatic Classification of Leukocytes

Stanislav Mirčić and Nikola Jorgovanović

Abstract — **This paper presents a novel algorithm for the automatic compilation of differential blood count (DBC), which is based on the direct analysis of a blood smear image and artificial neural networks. The results of the algorithm testing show high sensitivity of the algorithm in leukocyte detection and classification accuracy of 86%. Also, the algorithm enables the detection of potentially falsely classified leukocytes and in that way, with the help of a hematological expert, enables additional increase in the DBC compilation quality.**

Keywords — **Differential blood count, artificial neural networks.**

I. INTRODUCTION

THERE are five types of leukocytes in human blood: neutrophils, basophils, eosinophils, lymphocytes and monocytes [1]. They differ in their morphology, nucleus texture, size, affinity for different physiological colours and immune functions. Different pathological states of an organism can affect the change in the number of a single type of leukocytes. For that reason, the information about the exact number of leukocytes in patient's blood or their presence in percentages is very valuable from the diagnostic point of view.

The method for establishing the percentage of the presence of a single type of a leukocyte in blood is called Differential blood count (DBC). DBC is a highly used diagnostic method in medecine. Manual method is rather time-consuming and it involves the finding and classification of 100 leukocytes on a colured blood smear with the help of the microscope. In this case after the classification the numerical value of the leukocytes represents the percentage of the leukocyte presence in blood. The manual method, besides being time-consuming, has other drawbacks having in mind that often two haematological experts are engaged in the making of one DBC. Also, the quality of the DBC making in some cases depends largely on the experience and knowledge of the expert. These drawbacks of the manual method create high motivation for the making of a reliable automatic procedure for leukocyte classification.

First automatic methods for the DBC making made leukocyte classifications based solely on the nucleus and cytoplasm colour of the leukocyte in a coloured blood smear [2]. A significant progress in the performances of a device for the automatic DBC was made with the appearance of the Electrical Resistance Method [3], [4]. Because blood cells are a resistor that do not conduct electricity, the changes in electrical resistance occur when the blood cells pass through a detection apparatus (a minute hole) in a liquid through which electricity is passed. Counting the number of thesechanges indicates the blood cell count. As greater changes in electrical resistance occur when large blood cells pass through the apparatus, it is possible to distinguish blood cell types according to the magnitude of resistance. Also, a great progress has been made with the introduction of the Optical Flow Cytometry method which is the base of the majority of current commercial devices [5],[6],[7]. This method is based on the analysis of the scattered light which is obtained by the laser illumination of the cells or the analysis of light which appears when the fluorescent chemicals in or attached to the cells are excited.

Alongside the above mentioned methods for the automatic DBC, the methods for the leukocyte classification have been developed in a way similar to a haematological expert, based on blood smear image and with the help of up-to-date pattern recognition techniques [8-11]. Having in mind that these methods are based directly on the blood smear image it is possible to combine this method with the conventional manual method in order to obtain best results. Hence, in cases when haematological expert is not sure during the leukocyte classification process it is possible to seek „advice“ from the automatic classifier. It is even more useful to use the automatic DBC for a routine production of DBC and only in exceptional cases the automatic classifier may require haematologist's expertise. This of course implies the capability of the automatic classifier to detect cases in which the leukocyte classification could possibly be bad. Also, these types of the device enable blood smear images memorising for the purposes of further analyses or the comparisons of the results in different stages of treatment.

Because of their good performances in the field of pattern classification, NNs are frequently applied to the problem of the automatic classification of leukocytes based on image analysis. Even though there has been a significant progress in the performances of this kind of algorithms, the accuracy of classification of best classifiers has rarely exceeded 90%. This fact leaves a space for further developments of such algorithms.

The algorithm which will be presented in the following text represents a new algorithm for the computer classification of leukocytes based on digital image processing and neural networks. This algorithm has good performances of an autonomous classification and also enables the detection of the cases in which the leukocyte classification could possibly be bad. In that way it enables further expertise of a haematologist in special cases. Thus the algorithm provides quick creation of a best quality DBC.

Stanislav Mirčić is with the Faculty of Technical Science, University of Novi Sad, Serbia (e-mail: s_mircic@yahoo.com)

Nikola Jorgovanović is with the Faculty of Technical Science, University of Novi Sad, Serbia (e-mail: nikolaj@uns.ns.ac.yu)

II. METHODOLOGY

In order to successfully classify one leukocyte cell many morphological parameters of the cell must be taken into account. The most important features are: size of the cell (between 9 and 25μm), size of the nucleus, shape of the nucleus, the presence of granules in cytoplasm, the colour of the nucleus and of cytoplasm, the structure i.e. the texture of the nucleus etc [1]. Thus the automatic classifier must be able to evaluate quantitatively each of the cell features through adequate leukocyte image analyses and based on such numeric values set it must perform classification.

In the process of gathering the blood images for the purposes of this paper, many smears taken from different healthy adult persons and coloured with May-Grunwald-Giemsa cytochemicall stain were used. The images were obtained with the help of a microscope with an in-built CCD camera with the resolution of 320 x 240 pixels. Before processing, all the images were converted from RGB format to Gray Level images. An example of the image can be seen on Figure 1.

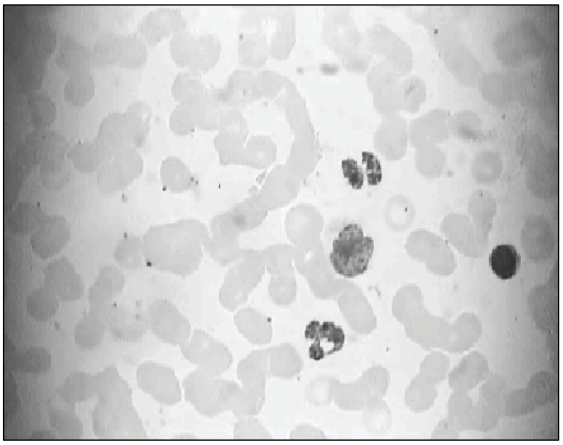
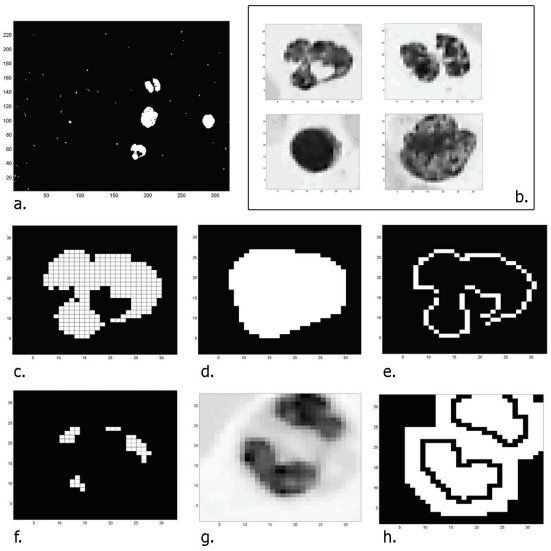


Figure 1. *One image of a blood preparation converted into gray level image in which four leukocytes (the cells of darker shades) and many erythrocytes which are all over the field of vision (the cells of lighter shades) can be observed.*

In the first phase of the image processing the correction of a non-uniform light (darkening on the edges of the image) and the elimination of stains which resulted from the uncleanness on the lens are made. This is obtained using a correction image. The correction image is formed by the averaging of the light intensity for every pixel in the image over the set of around 200 different blood smear images. Those averaged values are later used to correct pixel values of the new image in order to obtain a uniform light.

In the next processing phase a binary image is formed in a way that all pixels that are darker than a certain shade of grey are assigned the value of 1 and the lighter ones the value of 0. The shade of grey which is used as a threshold in forming the binary image is chosen in a way that in areas where the nuclei of the leukocytes are the binary image is assigned the value of 1 and in other parts of the image it is assigned the value of 0 (see Fig. 2a). The obtained binary image is then filtered with 2-D digital FIR filter with the Gaussian kernel of the size of 32x32 pixels. Such filtered binary image has local maximums in the areas of the nuclei of the leukocytes. By separation from the initial image of the area which is sized 32x32 pixels around the positions of the maximums in the filtered binary image, separated images of the leukocytes are obtained (see Fig. 2b).

Figure 2. *Transformations applied on a blood preparation image. a.) Binary image of the Fig. 1. (The pixels with the value of 1 are white and the ones with the value of 0 are black). b.) Separated leukocytes from the Fig. 1. c.) Binary image of the separated leukocyte image. d.) The smallest convex area which comprises all the pixels of the leukocyte nucleus from the Fig. 2.c. e.) the approximation of the leukocyte nucleus circumference from the Fig. 2.c. f.) The result of the double erosion of the binary image 2.c. g.) Separated image of an eosiphil granulocyte. h.) The white area comprises pixels in the nucleus and cytoplasm area of the cell from the Fig. 2.g.*



In every separated leukocyte image it is necessary to analyse quantitatively certain morphological features. The first characteristic to be analysed is the nucleus area. For the nucleus area measurement we took the number of pixels equal to 1 in the binary image of the leukocyte (Fig. 2.c.) divided by the total number of pixels in the image. The next characteristic to be analysed is the solidity which is obtained when the nucleus area, expressed in pixels and obtained from the binary image, is divided by the area of the smallest convex hull which comprises all the pixels of the nucleus (see Fig. 2.d.). This value is relevant to the classification since it represents the measurement of the nucleus fullness which is with some leukocytes such is lymphocyte (see Fig. 2.b in the left bottom corner) considerable and with some such is neutrophil granulocyte (Fig. 2.b top left corner) less considerable because this type of leukocyte often has its nucleus divided into more segments. The next measurement which also treats the complexity of the nucleus structure is circularity. It is obtained when the nucleus area is divided by the square of the nucleus circumference. The value obtained in such a way has a maximum for circular shapes of the nucleus and as the complexity of the nucleus increases (e.g. segmented nucleus) the value decreases. For the approximation of the nucleus circumference we took the number of pixels in the leukocytes image at which the absolute value of the light gradient exceeds a certain value which had been previously established. Having in mind that nucleus boundaries are characterized with a sudden change from the darker to a lighter shade of grey, the light gradient is large on these places, so in that way almost all the pixels placed on the nucleus circumference are separated (Fig. 2.e.).

The differences in the form of material out of which a nucleus is made are often described in medical literature.

For example, 'reticular', 'lumpy', 'granular' or 'smooth' nuclei are mentioned. Therefore, it is very important for the automatic classification that the information about the texture is quantified. In order to get quality information about the nucleus texture, it was measured only in the nucleus inside or in the inside of the segments of the nucleus. In that way the influence of the nucleus boundary texture on the final texture measurement was minimized, having in mind that all leukocytes had approximately the same form of the nucleus boundary. The pixels in whose surroundings the nucleus texture was measured were obtained in a way that the erode transformation of the binary image of leukocytes was done two times (Fig. 2.f.) [12]. Around every pixel obtained in such a way (Fig. 4c) an area of 6 x 6 pixels was separated from the initial grey level image and texture was measured within it. When measuring the texture the first thing to do was to calculate the approximation of the derivative (the difference of successive column elements) above every column of the matrix sized 6 x 6 pixels and then the average value of the derivative standard deviation above the whole matrix was calculated. Then the average texture value for all chosen pixels was calculated. This texture measure takes large values for 'rough' textures where the derivative value changes in an extensive range and small values for 'smooth' textures where the derivative is basically a constant one.

May-Grunwald-Giemsa colour which is used for colouring blood smears comprises two components. The first is an alkali dark blue component which attaches itself to the chromatin of the leukocytes which have the affinity for alkali colours. Such a leukocyte is e.g. basophil granulocyte. The second is an acid component of reddish colour which attaches itself to the chromatin of the cells which have the affinity for acid colours, e.g. in granules of eosinophil granulocytes. Because of these different affinities of the cells it is favourable to take into account the colour of the nucleus and of the cytoplasm of the cell when classifying leukocytes.

In order to get information about the colour of the inside of the nucleus and cytoplasm of a cell and not about the colour of the surrounding cells or boundaries of a nucleus from an image of a leukocyte, it is necessary to previously separate the pixels which are in the area of interest. This can be achieved by applying an adequate combination of dilation and erosion on the binary leukocyte image (Figures 2.g. and 2.h.). Further, 256 shades of grey in grey level image are divided into 15 intervals. Pixels which according to their colour belong to an interval and at the same time belong to the chosen pixels from the transformed binary image are recounted for every interval. In this way we obtain a certain histogram for 15 different intervals of grey.

Numeric histogram values together with the values of morphological characteristics (area, relative area, circularity and texture) create a classification vector which comprises 19 values. That vector is then submitted to a trained artificial neural network which classifies leukocytes based on it.

In this paper a feed-forward NN with 19 input neurons, 5 output neurons and two hidden layers with 120 and 70 neurons was used. The final network topology was chosen after a series of testing of different topologies as a topology with a minimal classification error in test data set. The need for a topology with two hidden layers is logical considering the type of the problem to which the NN was applied. Considering that leukocytes from two different classes can often be visually similar, an overlap in the input space of characteristics is very probable. Further, some leukocytes from the same class can vary greatly in their morphological characteristics. Such is with lymphocytes which appear in two shapes; as common lymphocytes sized 8 - 10μm and large ones sized approximately 15μm. Such considerably different morphological characteristics imply the existence of two or more separate areas in the input space which belong to the same class of leukocytes. Such a classification problem requires a richer NN structure which is provided by the network with two hidden layers that can separate even most complex class areas in the input space.

Because of the large number of weights and neurons that the network had, there was a risk of the overfitting of training data which could have negative effects on generalizing network performances. For that reason the early stopping method was used during the network training. The method would stop the training if in 5 consecutive iterations the error in vector test set rose. The network was trained in batch mode with back-propagation algorithm.

For transfer functions of all the neurons sigmoid functions were used. Because of that, after input vector processing by the NN we chose a class whose output neuron had the greatest activation for the class to which the input vector belonged. Even though it is recommended to use Hard-Limit transfer functions for neurons in output layer for classification problems, for the transfer functions in the output layer the said sigmoid functions were used. This was done because the neuron with the sigmoid transfer function could take values from continuous interval between 0 and 1 at the output. The value can be interpreted as a degree of the input vector belonging to the class that the output neuron represents (of course, this should be taken with a reservation having in mind that the value at the NN output is not the probability of the input vector belonging to some type but rather some non-linear function of this probability). This could be useful in potential future applications of this algorithm in commercial device in a way that in cases when more NN outputs have high values (or when none of the outputs have values higher than 0) leukocyte image which is being classified is memorized and shown to an expert who can classify disputable images. In that way in cases when NN is 'not certain' to which class the input vector belongs it can leave the leukocyte classification to a medical specialist and it would mean the device performances improvement with the minimal additional effort of an expert.

The NN was trained with the resilient modification of the basic back propagation algorithm [13]. When the weight of the network changes, this modification of the trained algorithm uses the sign of the gradient for determining the course of change. It establishes the numeric value of the change with a special algorithm.

The whole algorithm was made and tested in MATLAB® 6.0 programme package.

III. RESULTS

For the purposes of the algorithm testing and in cooperation with Haematology Clinic of the Institute for Internal Diseases in Novi Sad, around 500 images of normal blood smears from various adult persons were collected. In around 300 images there were one or more leukocyte cells. All images were interpreted after the collection and all the leukocytes were classified by an expert. When recording the preparations we chose the images which comprised one of the following four types of leukocytes: neutrophil granulocyte, eosinophil granulocyte, lymphocyte and monocyte. basophil granulocyte was not included in this algorithm testing having in mind that it rarely appears in the blood of a normal person, only one per 200 leukocyte cells. Therefore, to obtain a sufficient number of images of this

leukocyte it is necessary to inspect and classify a huge number of leukocytes (around 20000) which was not envisaged for this testing. When collecting snapshots, special attention was paid in order to have fairly the same number of leukocytes of all the four above mentioned types in the obtained set.

Beside leukocytes, different kinds of dirtiness which after the smear colouring have a colour that is similar to the leukocytes colour can often be found in blood smear. In some cases they can be 'falsely' separated and processed in the phase of separation and analysis of the leukocyte image. For that reason the NN with 4 neurons in the output layer which correspond to the leukocyte types was supplemented with an additional output neuron which is to be active when an analysed image contains some dirt instead of a leukocyte. In this way the errors made in the leukocyte separation phase were 'filtered' by the NN.

One half of the blood smear images was used for the NN training while the other half was used for the classification performances testing. Special attention was paid to having the approximately same number of cells from all types in both of the groups.

The percentage of the accurately classified leukocytes with the chosen NN topology was 86% from the total number of test set leukocytes (there were around 200 leukocytes in the test set). The obtained device sensitivity in leukocytes detecting was 99% while the percentage of dirtiness classified as leukocytes was around 4% from the total number of false positive detected leukocytes.

This proved to be a fairly good result having in mind that the classification accuracy of commercial classifiers rarely exceeds 90% [8]. Since the training set is rather small in comparison with the dimensionality of the space in which the NN does the classification (classifying vector has 19 dimensions), it can be expected that with a larger training set the classifier could obtain even better results.

IV. CONCLUSION

This paper presents the algorithm for the automatic classification of leukocytes using digital image processing and artificial NNs. Based on the classification testing we showed that the algorithm accurately classifies up to 86% of the test set leukocytes. Having in mind this result, the algorithm can serve as a good starting basis for the creation of a commercial automatic leukocytes classifier based on blood smear analisis image. Also, the paper presents the way in which the potentially falsely classified leukocyte images can be separated and presented to a haematological expert for further analysis in order to obtain better performances in the creation of DBC.

REFERENCES

[1] Gyton, A.,*Medicinska fiziologija*, Medicinska knjiga, Beograd - Zagreb, 1985.

[2] Tycko DH, Anbalagan S, Liu HC, Ornstein L, "Automatic leukocyte classification using cytochemically stained smears," J Histochem Cytochem. 1976 Jan;24(1):178-94.

[3] Rikiya Tanabea, Seiichi Hatab and Akira Shimokohbeb, "MEMS complete blood count sensors designed to reduce noise from electrolysis gas.", Microelectronic Engineering, Volume 83, Issues 4-9, April-September 2006, Pages 1646-1650.

[4] Loos H, Blok-Schut B, Kipp B, van Doorn R, Meerhof L., "Size distribution, electronic recognition, and counting of human blood monocytes.", Blood. 1976 Nov;48(5):743-53

[5] de Jonge R, Brouwer R, van Rijn M, van Acker BA, Otten HJ, Lindemans J., "Automated analysis of pleural fluid total and differential leukocyte counts with the Sysmex XE-2100." Clin Chem Lab Med. 2006;44(11):1367-71

[6] Miller RE., "Hematology: automated white blood cell differential counting flow-analysis." Clin Lab Med. 1981 Mar;1(1):127-50.

[7] Ducrest S, Meier F, Tschopp C, Pavlovic R, Dahinden CA., "Flowcytometric analysis of basophil counts in human blood and inaccuracy of hematology analyzers.", Allergy. 2005 Nov;60(11):1446-50.

[8] Swolin B, Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks--evaluation of DiffMaster Octavia, Clinical and Laboratory Haematology, 2003 Jun;25(3):139-47

[9] Sinha, N.; Ramakrishnan, A.G., "Automation of differential blood count",TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, Volume 2, Issue , 15-17 Oct. 2003 Page(s): 547 - 551 Vol.2

[10] Theera-Umpon, N.; Gader, P.D., "System-level training of neural networks for counting white bloodcells", Systems, Man and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, Volume 32, Issue 1, Feb 2002 Page(s):48 – 53

[11] Piuri, V. Scotti, F., "Morphological classification of blood leucocytes by microscope images.", Computational Intelligence for Measurement Systems and Applications, 2004. CIMSA. 2004 IEEE International Conference, 14-16 July 2004, Pages 103- 108.

[12] E.R.Davies, Machine Vision: Theory, Algorithms, Practicalities, Academic press, San Diego, 1997.

[13] Riedmiller, M., and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," Proceedings of the IEEE International Conference on Neural Networks, 1993