Transactions on
Case-Based Reasoning
Vol.2, No 1 (2008) 29 - 40
© ISSN: 1867-366X (Journal),
ISBN: 978-3-940501-08-0,
IBaI Publishing ISSN 1864-9734

ibai Publishing
www.ibai-publishing.org

Diagnosis of Lymphatic Tumors by Case-Based Reasoning on Microscopic Images

Sara Colantonio[1], Petra Perner[2], and Ovidio Salvetti[1]

[1] Institute of Information Science and Technologies, Italian National Research Council,
Via G. Moruzzi 1, 56124 Pisa, Italy
{Name.Surname}@isti.cnr.it
[2] Institute of Computer Vision and Applied Computer Sciences - IBaI
pperner@ibai-institut.de
http://www.ibai-institut.de/

Abstract. In this paper, a novel method for diagnosing lymphatic tissue tumors is presented. Microscopic specimen images are analyzed for extracting and characterizing malignant cells. A case-based reasoning approach is followed for classifying morphologic and densitometric cell features so as to provide a final diagnosis.

Keywords: Case-based reasoning, Image Analysis, Lymphatic Tissue Tumor

1 Introduction

Automated systems for the early diagnosis of lymphatic tumors, based on the morphological analysis of blood cells in microscopic specimen images [1], have been the pursuit of many research activities and projects in the last years. Pathologists usually make diagnosis by analyzing the morphologic and densitometric characteristics of specimen cells for distinguishing the malignant from benign nature of a tumor [2, 3]. The same procedure is also followed for distinguishing among different stages and forms of a tumor: i.e., *Aggressive Lymphoid Tumors* (de novo large and mixed cell lymphomas and transformed chronic lymphocytic leukemia - ALT), *Indolent Chronic Lymphocytic Leukemia* (ICLL) and *Reactive Lymphoid Hyperplasia* (RLH).

Unfortunately, the preparation of microscopic samples is a complex process, which often results in the partial destruction of molecular and cellular structures. Professional intuition and complex routine performance are needed to yield meaningful results. Computerized applications devoted to the automated analysis of photomicrographic images of lymphatic tissue imprints have recently appeared as a viable solution to simplify the experts' analysis and diagnosis tasks. Some attempts to cope with these problems have been made, even though resulted in only prototypical applications [4, 5, 6]. However, the variation of the appearance of the different types of cells is very large so that generalization based classifiers such as Artificial Neural Networks, Decision Trees or rule-based classifiers are not appropriate for this application. They need a long learning phase and having knowledge acquisition from the expert until they perform well. Therefore, a classifier based on case-based reasoning seems to be the better choice for an automatic diagnosis system since it is based on cases that the expert can easier describe and it can learn on different levels from a set of cases until the best system performance is achieved.

In this paper, a novel method is presented for recognizing malignant cells in microscopic cell images and, thus, providing a final diagnosis. Recognition is performed by reasoning on similar cases already processed and diagnosed, which are stored in a *base of cases*. The method is the current result of a research activity that started some years ago at CNR-ISTI in cooperation with the Russian Academy of Science and has coped with different tasks of the microscopic cell image processing [7, 8]. The cooperation continued with IBaI [11], that is studying case-based reasoning approaches for microscopic cell image processing.

In the Section 2, diagnosis of lymphatic tumors based on blood samples and the material are described. Section 3 presents the case-based reasoning approach. The image segmentation is briefly described in Subsection 3.1. The feature extraction is presented in Subsection 3.2. The case-based classifier is presented in Subsection 3.3. Results on the performance of the system are given in Section 4. Finally, we summarize our work in Section 5.

2 The Application - Diagnosis of Lymphatic Tumors based on Blood Samples

All cancer staging systems seek to identify clinical and pathological features that can predict outcome or guide therapy. Non-invasive methods for the early detection of disseminating disease are, hence, of great interest.

In particular, the diagnosis and classification of lymphoid tumors is becoming increasingly complex as well as fundamental. Current classification schemes incorporate morphologic features, immunophenotype, molecular genetics, and clinical data to specifically categorize leukemias into various subtypes. Although sophisticated methodologies are frequently used to detect characteristic features conferring diagnostic, prognostic, or therapeutic implications, a thorough microscopic examination remains essential to the pathologic evaluation. Such a procedure requires that blood cells be sampled in different loci and by various modalities (e.g., aspiration, washing, culture, smear or scraping). Different staining procedures are

used for fixing the sample and highlighting a specific cell type. For instance, for the diagnosis of different types of leukaemia, e.g. myelogenous or lymphocitic, bone marrow smears or lymphatic nodes aspirations are respectively employed, the former contains mostly granulocytes and red cells, while the latter mostly lymphocytes.

By far, the analysis under magnification is mainly performed qualitatively by cytologists for differentiating cells or identifying their physio-pathological conditions. Though human expertise is hard, if not impossible (at least at the moment), to replicate in computerized applications, automated cell image analysis tools may support and improve cytologists' work in several ways. Actually, current advanced microscopes can easily acquire high resolution microscopic cell images from which a considerable number of significant measurements, i.e. *features*, can be extracted. Automating the feature extraction process yields an objective, quantitative, detailed and reproducible evaluation of cell morpho-functional characteristics and allows the analysis of a large quantity of images. Moreover, by suitably classifying these features, an automated diagnosis can be obtained in a low-cost, standard-accurate system.

Due to the high variability of cell types and their appearance into a single specimen, adaptive classification methods reveal to be more suitable for recognizing the different types of diseases. In particular, case-based reasoning represents, above all, a viable solution since it avoids an explicit elicitation of experts' knowledge, is capable of coping with the large variety of possible cases, and can produce as output a prototype image which can be shown to cytologists for a visual comparison and explanation of the diagnosis results.

According to such considerations, a case-based reasoning method has been devised for the diagnosis of different types of lymphatic tumours via the analysis of microscopic cell images.

Footprints of lymphoid organs were obtained within 20-30 mins after biopsy, fixed in methanol and stained by Romanovsky-Giemsa stain. Digital images were obtained using a custom video-microscopic system based on Leica DMRB microscope, equipped with Kodak DC-290 camera and Planapo oil immersion objectives x63 or x100. The equivalent area of one pixel was 0.0036 or 0.0064 μm^2. The specimen pictures were stored as RGB-images in 24-bit TIFF format.

A base of cases of 1500 photos of specimens of 34 patients was obtained by considering 22 cases of ALT, 10 cases of ICLL and 2 cases of RLH. This way, a large variety of possible cell conditions was stored.

Unfortunately, experts have a good idea about the appearance of healthy and tumor cells, but they usually rely on their professional intuition when different cell conditions concur in the same sample. Therefore, our base of cases was submitted for examination to experts who marked pathological nuclei in the image and isolated them for being classified, considering a number of healthy nuclei as well.

3 The Case-based Reasoning Approach

Experts usually diagnose lymphoid tumors by evaluating a number of characteristics of sampled blood cells. Two examples of microscopic images showing an ALT and an ICLL cases are shown in Fig. 1. The main diagnostic criteria are as follows [7]:

- size of nuclei and density of different lymphoid cells in the specimen,
- shape of nuclei (round, elliptical, folded),
- presence of invaginations,
- textural characteristics of chromatin (condensed/dispersed, chromatin fibril diameter, presence of granules of condensed chromatin and size of the granules), and
- presence or absence of nucleoli.

In many cases, these can be not so sharply true and experts try to employ their experience and expertise to make a final diagnosis.

The main idea behind an automated diagnosis application is to translate these criteria into a number of features automatically computed from microscopic cell images. Usually, these features are processed by a classifier able to interpret image content and hence provide the final diagnosis.

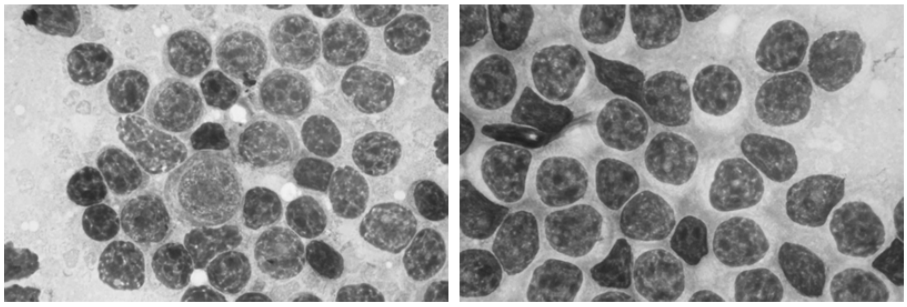

Fig. 1. Two sample images, acquired as photos of lymphoid tissue specimens stained according to the Romanovsky-Giemsa technique, which correspond to (left) an Aggressive Lymphatic Tumor (ALT), and (right) an Indolent Chronic Lymphatic Leukemia (ICLL).

For aiding experts' activities, the case based approach to image interpretation is particularly suitable. It mainly consists in the construction of a base of known (i.e. classified) cases, which is used for classifying a new case by selecting the most similar one contained in the archive as output. When dealing with images, this results in the provision of a final diagnosis, i.e. the one associated to the retrieved case, but also in the image itself, which can be shown to experts for additional comparison and explanation.

Such an approach has been followed for defining the method proposed in this paper. A base of photomicrographic images of lymphatic tissue imprints was created

to select and describe diagnostically important features of lymphocyte images. Each image is processed according to a multi-step procedure for extracting the cells it contains (*Segmentation* step), characterizing them according to a features-based representation (*Features Extraction* step) and, hence, interpreting its content by reasoning on the case base (*Image Interpretation* step). The result of this procedure, sketched in Fig. 2, is the image of the most similar case to the one at hand and its associated diagnosis.

In the following sections, each step of the procedure is described in detail.

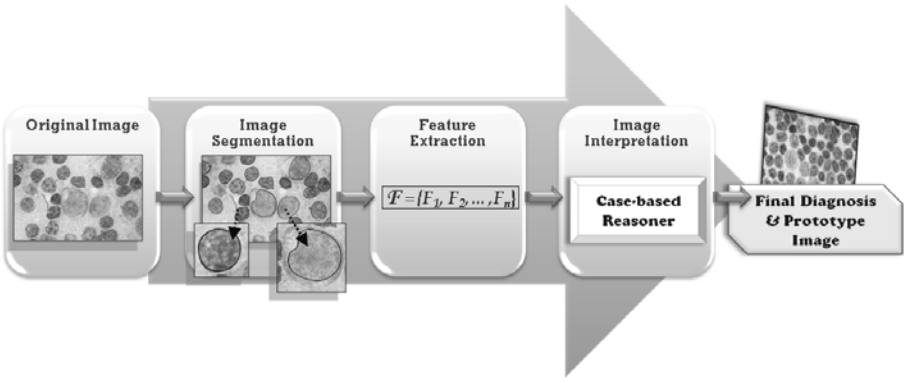

Fig. 2. The multi-step procedure for the interpretation of microscopic images and lymphatic tumor diagnosis.

3.1 Image Segmentation

The first step applied to microscopic images consists in isolating the cells with their cytoplasma and the nuclei from the background. Digital images are acquired from stained specimens of lymphoid tissue. The final aim is to extract the main cell structures (i.e., nucleus and cytoplasm). Unfortunately, staining artifacts, cells that touch each other and physiological characteristics of cell nuclei (e.g., the chromatin distribution) make the segmentation process a particularly difficult task.

A two-step approach was defined for coping with these difficulties as described in [8, 9]. A coarse-to-fine classification is applied for firstly detecting cells and their structures, and, then, refining the segmentation. A complete description of the method can be found in [8, 9], herein a brief summary of its functioning is reported as well as examples of results.

The method consists in applying to each image a two-stage procedure as follows:

1. *Cell dislocation detection*: a color space transformation is firstly applied to images for converting them from RGB to HSV space. Then, homogeneous image regions are detected via a cluster analysis, by applying the *fuzzy c-means* algorithm.

2. *Cells contours extraction*: each homogeneous region is further processed by computing a feature vector for each pixel. These are then classified by a dedicated Artificial Neural Network (ANN), which refines the segmentation by distinguishing among different kinds of structures (e.g. cell nucleus, cytoplasm, background, artifacts). As a result, each pixel gets the label of the cell part it belongs to according to the classification.

Two original images and the automatic labeled objects in these images are shown in Fig.3 and Fig. 4.

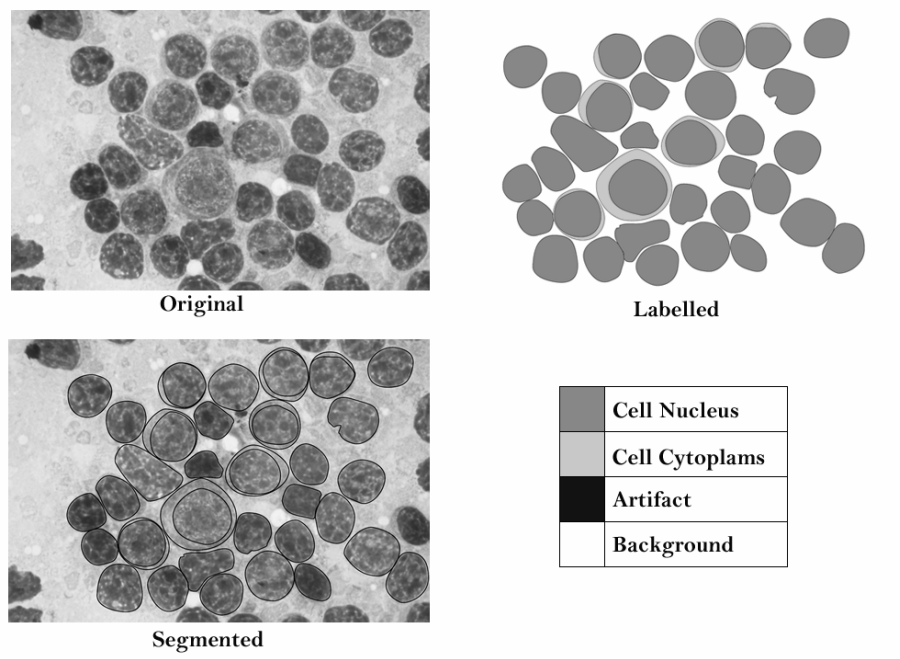

Fig. 3. The result of the segmentation step for an ALT microscopic image.

3.2 Automatic Cell Characterization by Image Features

Once segmented, cell structures, in particular cell nuclei, are characterized by computing a number of morphological and densitometric features (see table I). A good representative description was selected to encode the diagnostic criteria applied by hematologists. Precisely, morphological features, such as size and roundness, were computed for assessing nucleus shape, while densitometric and textural features were considered for characterizing the chromatin and nucleoli distribution. Moreover, a scale-space approach followed by the identification of the iso-intensity curves was employed for identifying the *critical points* (points with vanishing gradient [10]) and, hence, detecting and describing the peak of chromatin distribution [5].

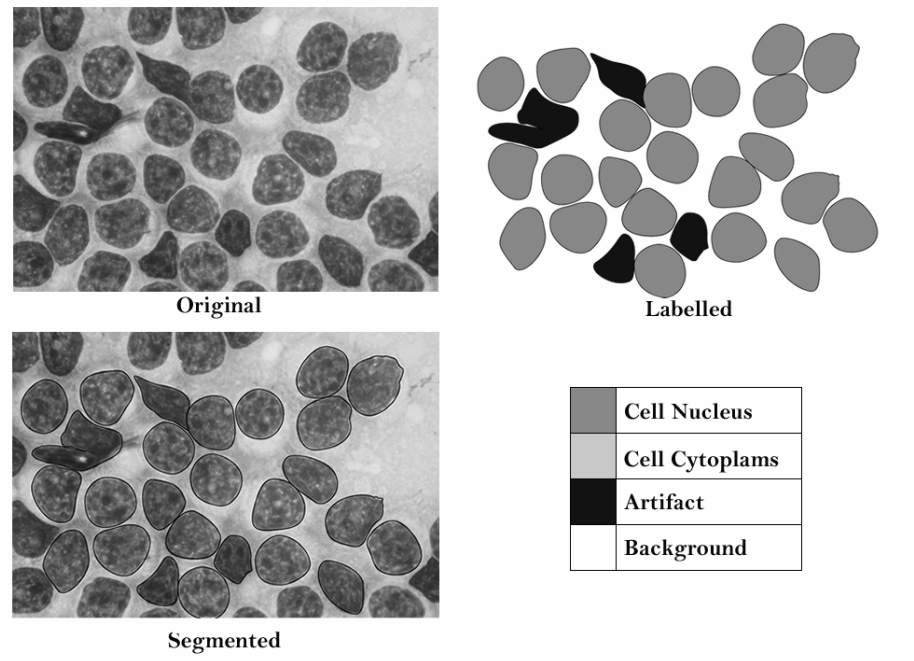

Fig. 4. The result of the segmentation step for an ICLL microscopic image

Fourier features were used for describing the nuclei textural properties (see table I). Precisely, the Fast Fourier Transform was applied to compute the spectrum of squared images obtained by placing each extracted nucleus against the black background [11]. For each $2N \times 2N$ gray-level image, a function $f(r)$ was computed as the sum of the matrix elements located on the semi-circle with the radius r centered in the middle of the image, with $r=0,1,\ldots,N$. While a function $f(\theta)$ were obtained as the sum of the matrix elements located on the line segment of the length N originating from the middle of a matrix and forming the angle θ with the horizontal axis (anticlockwise), $\theta \in [0,\pi]$.

Table I summarizes type and definition of the features considered.

3.3 The Case-based Classifier

A case-based reasoner classifies a sample according to the cases in a case base by selecting the most similar case as output. The cases consist of the feature-based descriptive model introduced in the previous section.

A proper similarity measure is necessary to retrieve the most similar case, but in most applications there is no a-priori knowledge available that suggests the right similarity measure. The method of choice to select the proper similarity measure is therefore to apply a subset of the numerous similarity measures known from statistics

to the problem and to select the one that performs best according to a quality measure such as, for example, the classification accuracy. The other choice is to automatically

Table I. Features considered, divided by class

Feature Class	Feature
Morphological	Area, Circularity
Densitometric	Mean, Variance, Skewness, Kurtosis of the histogram distribution for each RGB components
Scale Space	Number of critical points
Textural	1. For $f(r)$: Mean, Variance, Skewness, Kurtosis, Number of local maxima, Sum of local maxima, Value of lobal maximum, Number of local minima, Sum of local minima, Value of global minimum **2. For $f(\theta)$: Number of local maxima, Sum of local maxima, Value of global maximum, Number of local minima, Sum of local minima, Value of global minimum**

build the similarity metric by learning the right attributes and attribute weights. While constructing a case base for the case-based classifier, selecting some prototypical examples can be useful for reducing the number of examples used for classification. This results in better generalization and a more noise tolerant classifier. Therefore, a function to assess a collection of prototypes and identify redundancy is useful. Moreover, an important variable in a case-based reasoner is the value used to determine the number of closest cases and the final class label, i.e. the value of the *k* nearest cases.

Case-based reasoning was implemented with ProtoClass [13], whose features include options for *k-value selection* for the closest cases, *prototype selection*, *feature-subset selection*, *feature weighting*. A schema of the classification process is sketched in Fig. 5: it consists in a *design phase* (*Learning Unit*) and the *normal classification phase* (*Classification Unit*). The classification phase starts after we have evaluated the

classifier and determined the right features, feature weights, the number of prototypes and the cases.

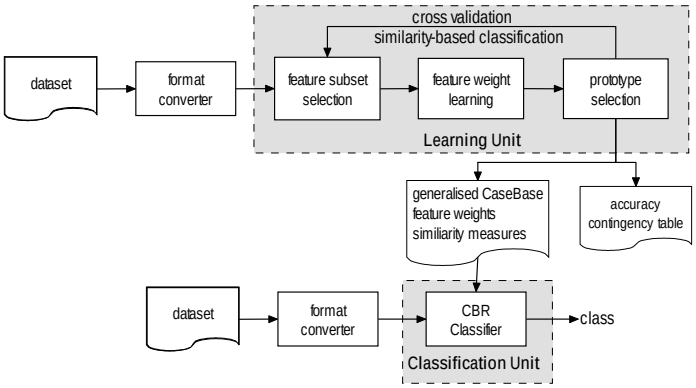

Fig. 5 Case-based Classifier

4 Results

The method proposed has been experimented on the dataset described in Table II.

Table II. Description of the base of cases. The *Normal* (Norm) cases correspond to healthy image nuclei

	Patients	Images	Nuclei
ALT	22	968	1760
ICLL	10	444	660
RHL	2	88	140
Norm	-	-	500
Tot	34	1500	3060

ProtoClass has been applied to the feature vectors extracted from images. For the first tests, only the selection of the most effective *k*-value has been investigated. Cross validation has been adopted for evaluating the reasoner performance according to three different measures: the *correctness* or *accuracy p* (as in eq. 1, the number of correctly classified samples according to the number of samples); the *class specific quality* p_{ki} (as in eq. 2 the number of correctly classified samples for one class *i* to all samples of class *i*); and the *classification quality* p_{ti} (as in eq. 3, the number of correctly classified samples of class *i* to the number of correctly and falsely classified samples into class *i*).

$$p = \frac{\sum_{i=1}^{m} c_{ii}}{\sum_{i=1}^{m}\sum_{j=1}^{m} cij} \tag{1}$$

$$p_{ki} = \frac{c_{ii}}{\sum_{j=1}^{m} c_{ji}} \tag{2}$$

$$p_{ti} = \frac{c_{ii}}{\sum_{j=1}^{m} cij} \tag{3}$$

where c_{ij} is the number of samples of class j assigned to class i (i.e., c_{ii} is the number of correctly classified samples), and m is the number of classes.

Results of ProtoClass application for the three measures are reported in the contingency matrix shown in Table III, assessing their amount for three different values of k, i.e. of the closest cases. The best performance is achieved with a k-value of seven.

5 Conclusion and Discussion

In this paper, a novel method for the diagnosis and differentiation of lymphatic tissue tumors has been introduced and evaluated. It relies on the application of case-based reasoning on microscopic cell images. Actually, among different adaptive classification methods, case-based reasoning appeared the most suitable solution since it avoids an explicit elicitation of experts' knowledge, is capable of coping with the large variety of possible cases, and can produce as output a prototype image that can be shown to cytologists for a visual comparison and explanation of the diagnosis results.

The method is based on a multi-step procedure that consists in the extraction of the relevant cell structures, their characterization by means of a vector of features selected according to diagnostic criteria followed by experts, and in the application of a case-based reasoner. The tool ***ProtoClass*** has been employed for the last classification step.

The performance of the classification has been evaluated according to different measures in order to better assess the functioning of the reasoner. The first tests have dealt only with performance evaluation according to different value of k closest cases. This has highlighted that best results are obtained with the higher value of k (i.e., k=7).

Table III. Contingency matrix and the values of the three evaluation measures

		ALT			ICLL			RHL			Norm		
k	1	3	7	1	3	7	1	3	7	1	3	7
ALT	1620	1676	1688	68	43	49	10	9	5	52	32	18
ICLL	37	25	17	583	603	628	19	15	6	31	17	9
RHL	3	1	0	7	6	2	102	118	129	18	15	9
Norm	7	5	1	15	12	6	27	21	19	451	462	474
P_1	97.18%	98.18%	98.94%	86.63%	90.81%	91.68%	64.56%	72.39%	81.13%	81.70%	87.83%	92.94%
P_M	92.05%	95.23%	95.91%	88.33%	91.36%	95.15%	72.86%	84.29%	92.14%	90.20%	92.40%	94.80%
P	0.9007	0.934	0.9539									

Future activities will consist in the evaluation of other options such as feature subset selection and prototype selection. The first option would be useful to highlight which are the relevant features for the diagnosis.

Acknowledgments.
This work was partially supported by the European Community, under the Sixth Framework Programme, Information Society Technology – ICT for Health, within the STREP project HEARTFAID (IST-2005-027107), 2006-2009.

References

1. Matutes E, Polliack A. Morphological and immunophenotypic features of chronic lymphocytic leukemia. Rev Clin Exp Hematol. 2000 Mar;4(1):22-47.
2. Churakova Zh.V., Gurevich I.B., Jernova I.A., et al.: Selection of Diagnostically Valuable Features for Morphological Analysis of Blood Cells, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, Vol. 13. (2003) 2 381-383
3. Jaffe E.S, Harris N.L., Stein H., et al.: Pathology and Genetics of Tumors of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press (2001)
4. Foran Dj. et al.: Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy, IEEE Trans. Inf. Technol. Biomed., pp. 265-73, December 2000
5. Gurevich I.B., Khilkov A., Murashov D.: A Technique for Extraction of Diagnostic Data from Cytological Specimens. In Proc. of CIARP 2003, LNCS 2905, pp. 274–281, 2003.
6. Piuri V., Scotti F.: Morphological Classification of Blood Leucocytes by Microscope Images. Proc. International Symposium on Computational Intelligence for Measurement Systems and Applications, Boston, MD, USA, July 2004
7. Ablameyko S., Di Bona S., Gurevich I., et al.: Towards Automated Analysis of Cytological and Histological Specimen Images, AITTH'2005 (Minsk), vol. 1, pp. 27–31, 2005
8. Colantonio S., Salvetti O., Gurevich I.B.: A Two-Step Approach for Automatic Microscopic Image Segmentation Using Fuzzy Clustering and Neural Discrimination. Pattern Recognition and Image Analysis, 2007, Vol. 17, No. 3, pp. 428–437. © Pleiades Publishing, Ltd., 2007
9. Colantonio S., Gurevich I.B., Salvetti O.: Automatic fuzzy-neural based segmentation of microscopic cell images. Int. Journal of Signal and Imaging Systems Engineering, Vol. 1:1, Inderscience, 2008
10. Florack, L.M.J., Kuijper, A.: The topological structure of scale-space images. Journal of Mathematical Imaging and Vision, Vol. 12, N1 (2000) 65–80
11. Gurevich I.B., Harazishvili D., Jernova I., Khilkov A., Nefyodov A., Vorobjev I. Information Technology for the Morphological Analysis of the Lymphoid Cell Nuclei. J. Bigun and T. Gustavsson (Eds.): SCIA 2003, LNCS 2749, pp. 541−548, Springer-Verlag Berlin Heidelberg 2003
12. Little S., Salvetti O., Perner P.. Evaluating a Case-Based Classifier for Biomedical Applications 21st IEEE International Symposium on Computer-Based Medical Systems 17-19 June 2008 pp. 584 – 586, 2008
13. Perner P., Prototype-Based Classification, Applied Intelligence 28: 238-246, 2008