Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012

HEp-2 Cells Classification via Fusion of Morphological and Textural Features

Ilias Theodorakopoulos, Dimitris Kastaniotis, George Economou and Spiros Fotopoulos
Electronics Laboratory
Department of Physics
University of Patras
Patras, Greece
{iltheodorako, dkastaniotis}@upatras.gr , {economou, spiros}@physics.upatras.gr

Abstract— **Autoimmune diseases are proven to be connected with the occurrence of autoantibodies in patient serum. Antinuclear autoantibodies (ANAs) identification can be accomplished in a laboratory using indirect immunofluorescence (IIF) imaging. ANAs are characterized by specific "visual" patterns on a humane epithelial cell line (HEp-2). The identification stage is usually done by trained and highly qualified physicians through visual inspection of slides using a fluorescence microscope. The presence of subjectivity in the identification process, the inter-observer variability, the increasing demand of highly specialized personnel, suggest that a realization of an automatic classification system is of great significance for the field of autoimmune diseases diagnosis. Moreover CAD systems can be used in a collaborative scheme in order to augment the physicians' capabilities. In this paper a system for automatic classification of staining patterns on single-cell fluorescence images is proposed. Our method utilizes morphological features extracted from a set of binary images derived via multi-level thresholding of fluorescence images. Furthermore, a modified version of Uniform Local Binary Patterns descriptor is incorporated in order to capture local textural information. The classification is performed using a non-linear SVM Classifier. The proposed method is evaluated using a publicly available dataset, recently released for the purposes of HEP-2 Cells classification competition at ICPR 2012, achieving up to 95.9% overall classification accuracy.**

Keywords-HEp-2 cells; fluorescence staining patterns; morphological features; rotation invariant LBPs; classification;

I. INTRODUCTION

Indirect immunofluorescence (IIF) imaging is the recommended laboratory technique to detect autoantibodies in patient serum, which have been confirmed to be in connection with the occurrence of autoimmune diseases such as systemic autoimmune rheumatic diseases, primary biliary cirrhosis and dermatomyositis. Antinuclear autoantibodies (ANAs) in patient serum are detected by a specific fluorescence pattern on a humane epithelial cell line (HEp-2). The fluorescence patterns are usually manually identified by trained and highly qualified physicians visually inspecting the slides using a fluorescence microscope.

Manual test evaluation though, commonly used in the laboratories, suffers from some important limitations. Firstly, the readings in IIF are subjected to inter-observer variability, limiting the reproducibility of the method used and inducing subjectivity to the results. Secondly, as ANA testing becomes more widespread used, the increasing demand for highly specialized personnel cannot always be satisfied. Additionally, the lack of standardization intensifies the limitations of human ability to detect and diagnose a disease during image interpretation, due to the non-systematic search patterns, the presence of noise and technical issues such as the photobleaching effect, which bleaches significantly the tissues in a few seconds. Lastly, the large amount of image data that is generated can make the detection of potential disease a really time-consuming process, which in turn increases the risk of oversight errors.

Automatic procedures aiming to determine the presence of autoantibodies in IIF, confronts with modern trends in other areas of medicine, offering a solution to all the above limitations and enables easier, faster and more reliable tests. Hence, an evident medical demand is the development of a Computer-Aided Diagnosis (CAD) system, which may support the physician's decision. CAD methods have definitely been proven effective in other contexts as they allow pre-selection of the cases to be examined, enabling the physician to focus his attention only on relevant cases, making it easier to carry out mass screening campaigns. Also, serve as decision fusion procedures, augmenting the physicians' capabilities and reducing errors.

In this context, several approaches have been proposed in the literature aiming to automate individual stages or the entire IIF diagnostic procedure. The typical flow of such procedures consists of five main steps namely image acquisition, image segmentation, mitosis detection, fluorescence intensity classification and staining pattern recognition. In [5] Hiemann et al. proposed a method for quality evaluation of fluorescence images based on o set of shape and textural parameters extracted from the images. Soda et al. in [15] propose an autofocus function that can deal with photobleaching effect during acquisition. Segmentation of fluorescence cells in indirect immunofluorescence images (IIF) was performed by Yu-Len et al. in [18] using the similarity-based watershed algorithm as a marker to prevent over-segmentation, and in [19] using an adaptive edged- based segmentation method. Automatic mitotic cells recognition was addressed by Foggia et

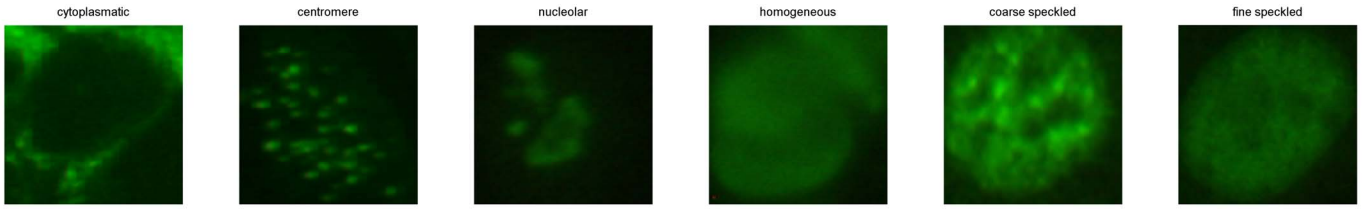

Figure 1. Examples of fluorescence staining patterns

al. [1] combining morphological features along with textural features extracted using LBPs. In [13] fluorescence intensity classification is performed using statistical features and the incorporation of decision rules along with three specialized classifiers (experts) trained on positive, negative and intermediate samples respectively.

The problem of staining pattern recognition has been approached by Parner et al. [9] using automatic thresholding via Otsu's algorithm in order to segment the individual cells, followed by texture model estimation obtained by various realizations of compact random sets. In [17] textural and statistic features are incorporated along with Self-Organizing Map classification scheme, in order to Classify HEp-2 cell Images. In the same context, Soda et al. [14] use an aggregation of binary classifiers operating on statistical and spectral textural features, also introducing a reliability measure of the final classification. Finally, evaluation results for IIF CAD systems integrating autofocus, intensity classification and staining pattern classification procedures have been reported in [4, 5, 11].

In the current work we present a method for HEp-2 cells classification based on fluorescence staining patterns, utilizing tools from two well-established fields of image processing i.e. image morphology and texture analysis. More specifically, morphological features extracted from binary images, derived via application of multiple thresholding operations on the fluorescence cell images, along with textural features extracted using a modified version of Local Binary Patterns, are simultaneously incorporated in order to achieve enhanced classification performance. The rest of this paper is organized as follows: In Section 2 the adopted taxonomy of fluorescence patterns is described. A detailed description of the incorporated features is given in section 3. In Section 4 the evaluation procedure is outlined and the corresponding experimental results are provided. Finally, conclusions are drawn in section 5.

II. TAXONOMY

Positive HEp-2 samples may reveal different patterns of fluorescent staining that are relevant to diagnostic purposes. Although more than thirty different nuclear and cytoplasm patterns could be identified [16], the taxonomy adopted in the current work is the one followed by [12, 1] where patterns are classified into one of the five following groups:

- Homogeneous: diffuse staining of the interphase nuclei and staining of the chromatin of mitotic cells. Numerous of these sera were shown to be positive for antibodies against double-stranded DNA and/or histones.
- Speckled: a fine or coarse granular nuclear staining of the interphase cell nuclei. This basic pattern insludes fine to coarse nuclear staining with slight or without nucleolar staining and without staining of metaphase cell chromatin.
- Nucleolar staining: large coarse speckled staining within the nucleus, less than six in number per cell
- Cytoplasmic: fine fluorescent fibres running the length of the cell; it is frequently associated with other autoantibodies to give a mixed pattern;
- Centromere: several discrete speckles (~ 40−60) distributed throughout the interphase nuclei and characteristically found in the condensed nuclear chromatin during mitosis as a bar of closely associated speckles.

Examples of fluorescence patterns from each of the above groups are shown in figure 1.

III. FEATURES EXCTRACTION

As shown in the literature [1, 17, 13] morphological and textural features can provide a very efficient description of the characteristics of fluorescence images, in terms of discrimination between different groups of patterns. In this context, we chose to utilize both types of features, in order to enhance the classification performance of the proposed system. To this purpose, two sets of features (morphological and textural) are extracted from each fluorescence image, the union of which constitutes the representative vector of the image.

A. Morphological Features

Every fluorescence pattern group is characterized by unique optical properties originating from the nature of the depicted cells, as described in section 2. In order to express appropriately these characteristics, as numerical values forming a representative vector, a novel set of morphological features is incorporated. The proposed features set is a variant of the 1D Boolean texture models proposed in [2] and especially the two dimensional extension proposed by Perner et al. in [9].

Initially, each cell image is converted to grayscale and a median filter is applied, using a 3x3 pixels window, in order to eliminate isolated extremities of the intensity values. Note that the filtered image is used as means to calculate a set of thresholding levels and no feature is extracted from it. To this purpose, the minimum and maximum intensity values of the

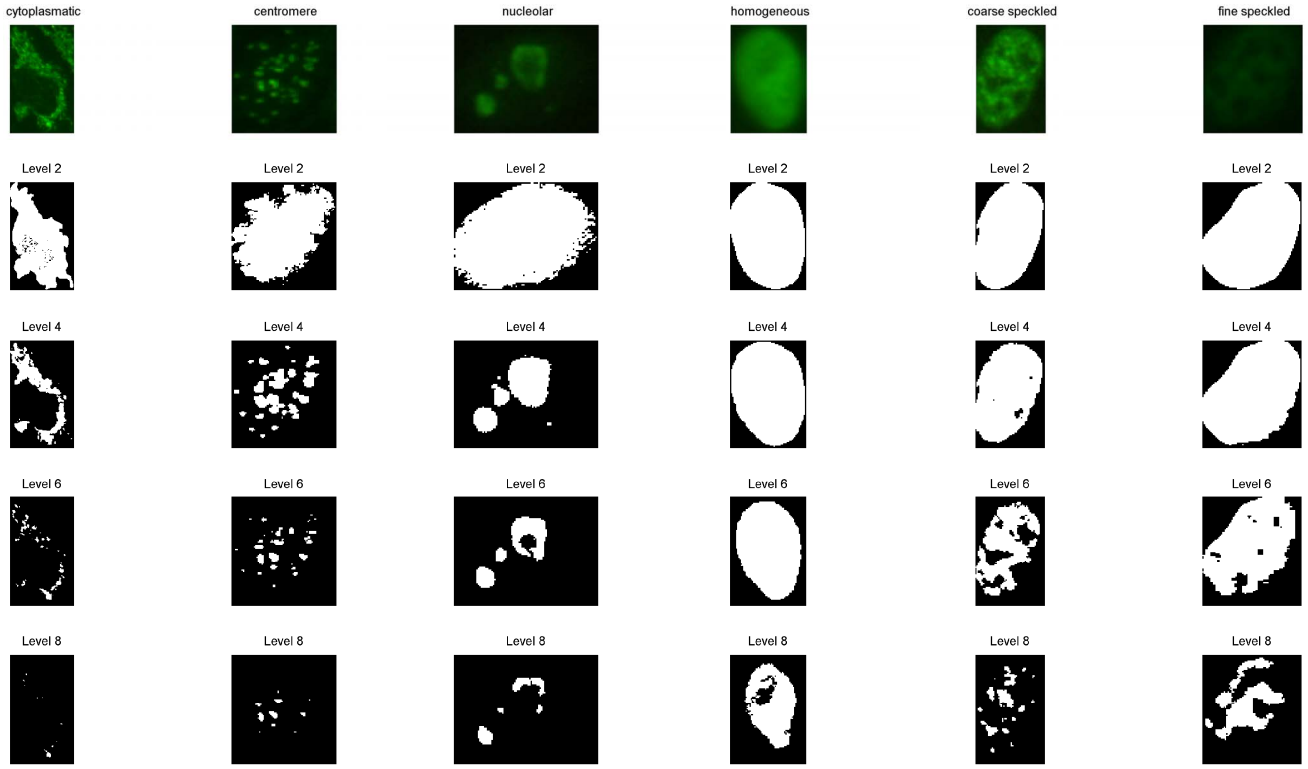

Figure 2. Examples of the binary images resulting from multiple level thresholding for various staining patterns.

filtered image are considered, and 9 equally spaced values in the range of intensity extremes are computed.

Subsequently, a set of binary images $I^T(x,y)$ are constructed via application of thresholding operation to the initial (grayscale) image $I(x,y)$, utilizing the above set of 9 values T as threshold values. Significant amount of information is carried by the resulting sequence of binary images regarding the spatial distribution of intensities on the depicted staining pattern, expressed in the form of patterns of homogenous regions of Boolean "True" value, namely objects, with differentiating properties along the various threshold levels. Examples of the resulting binary images for various staining patterns are illustrated in Figure 2. In order to quantify this information, Connected Component Analysis is performed in each binary image $I^T(x,y)$, and the following set of morphological features is computed for each of them:

1) Number of detected objects: N^T.

2) Density in Binary Image T:

$$D^T = \frac{\sum_{i=1}^{N^T} A_i^T}{A_{cell}} \tag{1}$$

where A_i^T is the area of the ith object of the thresholded image at level T, and A_{cell} is the area of the cell in the initial image.

3) Mean Objects' Solidity:

$$S^T = \frac{1}{N^T} \sum_{i=0}^{i=N^T} S_i^T \tag{2}$$

where S_i^T is the solidity of the ith object of the thresholded image at level T and $S_i^T = A_i^T \big/ convex\_hull\{A_i^T\}$.

It has to be noted that objects of size less than 1% of the mean objects' size of each binary image, are considered as noise and ignored during the calculation of the above features.

Finally, the complexity of the cell's contour is considered as an additional feature. We chose to quantify the complexity as the difference between the cell's contour and the perimeter of the equivalent circle (a circle with the same area as the cell):

$$CC = P_{cell} - P_{EqCirc} = P_{cell} - 2\sqrt{\pi A_{cell}}\,, \tag{3}$$

where P_{cell} and P_{EqCirc} are the cell's and the equivalent circle's diameter, respectively.

B. Textural Features

While the above morphological features are able of capturing structural characteristics of the depicted staining patterns, the local textural information is considered to be of equal importance regarding the discrimination between the various patterns [1], as also evidenced by the experimental

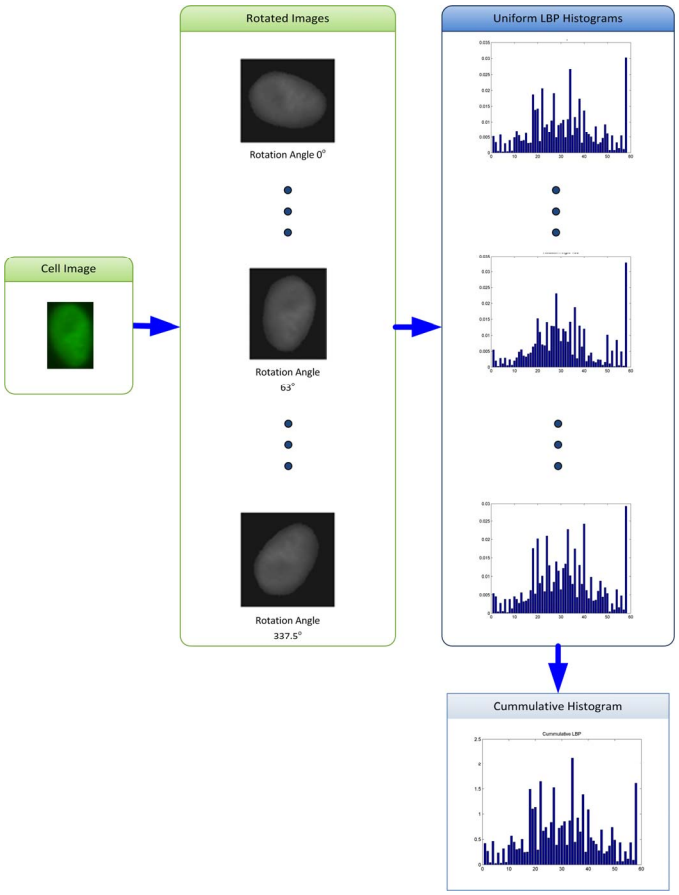

Figure 3. Overview of the proposed variation of Local Binary Patterns Descriptor.

results shown in section IV. In order to exploit this type of information, we have incorporated a variation of the LBP textural descriptor.

Local Binary Patterns were originally proposed by Ojala et al [7] as a method for summarizing the local structure of an image. In the original LBPs the use of a 3x3 non-parametric Kernel was proposed, presenting high discriminative ability on texture classification problems. From this 3x3 window a binary word is extracted as follows: The center pixel is used as a threshold and intensity comparison with the 8 neighboring pixels is performed. Each pixel having intensity value greater than threshold is assigned the binary value 1, otherwise the binary value zero.

Afterwards, the 8 binary values are considered as a binary representation of a decimal number, and a histogram of all the corresponding values computed across the image is calculated.

While the 3x3 kernel, which is a radius 1 LBP operator, efficiently describes the local spatial textural information using 2^8=256 bin histogram as a descriptor, later in [10] it was observed that the largest proportion of information was contained in a small subset of the initial 256 local binary patterns. This subset is the Uniform Binary Patterns, where the number of transitions from 1 to 0 or backwards is limited to a maximum of two. This results in drastically reducing the overall descriptor dimensions, while maintaining the necessary information.

Another way of capturing useful spatial information is the incorporation of LBP operators with radius greater than one [8]. This leads into a multiscale texture exploration.

While uniformity and scalability are very important properties of LBP operators, discrimination can be further enhanced by taking into account the local structure of the image during the local pattern extraction procedure (improved LBPs [6]). There, instead of comparing the neighboring pixels of radius N {1, 2} with the central pixels' intensity value, boundary points are compared with the mean intensity value of all neighboring pixels.

Having considered the above, an application can be further benefit by incorporating a rotation invariant extension of the LBP operator [8]. The drawback here is the crudeness of the angular space quantization (especially in the radius 1 LBP operator) that can be somehow over passed by using a radius 2 LBP operator, leading to a 22.5 degrees angular quantization. While this is a very efficient way to capture some notion of rotation invariance, it lacks in the expressiveness of global orientation. In this context, in [3] rotation invariance is achieved using a global "principal" orientation of texture images. Although this approach establishes an appropriate alignment of the processed images prior to the application of LBP operator, often leads to suboptimal results due the miscalculation of principal orientation.

In this work, we propose an alternative approach for the extraction of rotation invariant uniform LPBs, in order to express local textural characteristics of Hep-2 fluorescence images. Given an image, a set of 80 rotated images (4.5 degrees intervals in the range 0o - 360o) are computed using bilinear interpolation. The radius 1 LBP operator is then applied in each image, and the histogram of the uniform patterns is computed. The resulted histograms, after normalization, are accumulated into one final descriptor. An overview of the procedure is illustrated in Figure 3.

The $9\times3+1=28$ morphological features along with the 58-bin LBPs histogram are concatenated in order to form the 86-dimension representative vector of the cell's image. The classification is performed in the 86-dimensional feature space using non-linear Support Vector Machines Classifier with Gaussian RBF kernel.

IV. Experimental Evaluation

In order to evaluate the performance of the proposed scheme for fluorescence staining pattern classification, we used the only publicly available dataset provided by the HEp-2 cell classification contest, hosted by the ICPR 2012 conference [20, 1]. The dataset consists of 721 single-cell fluorescence images, and constitutes the training subset of the contest (the test set has not been released yet). HEp-2 images were acquired by means of a fluorescence microscope (40-fold magnification) coupled with a 50Wmercury vapor lamp and with a digital camera utilizing a CCD with square pixel of 6.45 μm. The initial images have a resolution of 1388x1038 pixels, a color depth of 24 bits and they are stored in BMP format.

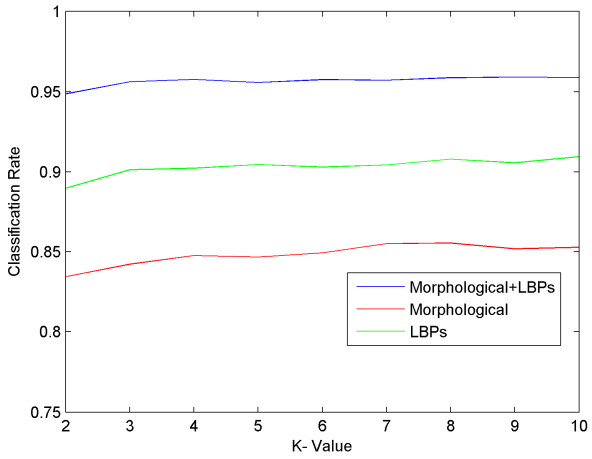

Figure 4. Classification performance of the various feature sets for variable k.

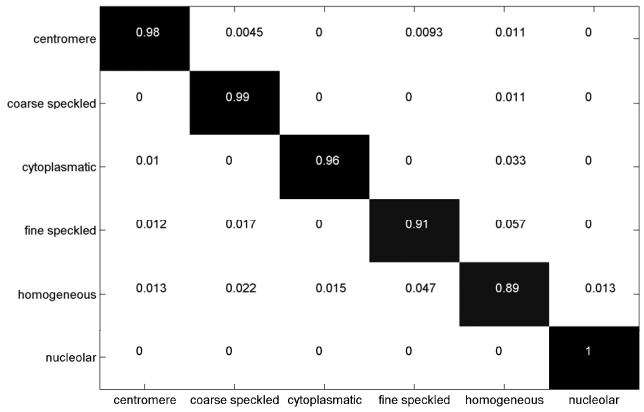

Figure 5. Confussion Matrix for 10-fold validation procedure using morphological and textural features' fusion. Each row is normalized to unit sum.

The images were manually processed by specialists in order to produce single-cell images and provide ground truth information. Firstly, a biomedical engineer manually segmented the cells by the use of a tablet PC. Subsequently, each single-cell image was verified and annotated by a medical doctor specialized in immunology. Thus, the dataset contains the cell images, the corresponding binary masks indicating the exact area of the cell, and the information regarding each cell's pattern and intensity.

The performance evaluation of the proposed classification scheme was performed using k-fold validation procedure, where the dataset is randomly partitioned into k equally sized subsets, one of which is used as the test set and the rest k-1 form the training set. The procedure is repeated k times with each of the k subsets used exactly once as the test set. The k results are then averaged to the final result. On each iteration, the non-linear SVM which is responsible for the final classification of the extracted features, is trained using the data provided by the corresponding training set.

In order to demonstrate the discriminative ability for the two incorporated groups of features separately, additional experiments were conducted using the same validation procedure except that only one group of features was used each time in order to represent the staining patterns. The obtained results for the morphological features, LBPs and the fusion of those are illustrated in figure 4 by means of overall classification accuracy. The experiments were conducted using k in the range of 2 to 10 so as to test the generalization of the classification scheme. For each experimental configuration (combination of k and features set) 10 runs of execution were performed (with different random initializations) and the overall results were averaged.

The highest overall classification accuracy of 95.9% is achieved by exploiting both the morphological and the textural features. The highest classification rate achieved by the morphological features is 85.3% and the corresponding rate for the LBPs is 91.0%. As inferred by the results illustrated in figure 4, both types of features can be considered successful on extracting the necessary information able to discriminate the different staining patterns up to some extent. The local textural information though, expressed via the incorporated LPBs descriptor, proves to be of greater importance in terms of discriminative ability, compared to the structural characteristics captured by the morphological features. However, fusing the two features sets into a large representative vector, can significantly benefit the classification accuracy of the proposed scheme. This indicates that the proposed morphological and textural features can extract complementary information from the cell images, regarding the illustrated staining pattern.

Finally, the confusion matrix for 10-fold validation using both morphological and textural features is illustrated in figure 5, summarizing the distribution of misclassifications along the various staining patterns. Each row of the confusion matrix is normalized to unit sum. It is evident that the most indistinguishable patterns are the "homogeneous" and "fine speckled", since the 4.7% of the homogeneous samples were classified as fine speckled, and the 5.7% of the fine speckled samples were classified as homogeneous. The reasoning behind this behavior is the fact that the optical characteristics that distinguish between the two patterns are inherently vague, as can been easily inferred by figure 2. In general though, the overall confusion percentages (off-diagonal elements) remain reasonably low.

V. CONCLUSIONS

In this paper we have presented an approach for automatic classification of staining patterns on fluorescence cell images. A set of morphological features, extracted from a set of binary images derived via multiple-level thresholding, were proposed in order to capture structural characteristics of the depicted staining patterns. Additionally, a variation of the well-known uniform LBPs descriptor is proposed in order to summarize the local textural information of each pattern. The two sets of features are fused into one representative vector, and a non-linear SVM is trained in order to classify the staining patterns. The proposed scheme was evaluated using the only publicly available dataset of fluorescence cell images, achieving 95.9% overall classification rate. Unfortunately, the dataset was recently released and there are no reported results in the literature to compare with.

ACKNOWLEDGMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

REFERENCES

[1] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Early experiences in mitotic cells recognition on HEp-2 slides," in Proceedings of the 2010 IEEE 23rd International Symposium on Computer-Based Medical Systems: IEEE Computer Society.

[2] P. Garcia, M. Petrou, and S. Kamata, "Short communication: The use of Boolean model for texture analysis of grey images," Comput. Vis. Image Underst., vol. 74, pp. 227-235, 1999.

[3] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," Pattern Recognition, vol. 43, pp. 706-719.

[4] R. Hiemann, N. Hilger, J. Michel, J. Nitschke, A. BÖHm, U. Anderer, M. Weigert, and U. Sack, "Automatic Analysis of Immunofluorescence Patterns of HEp-2 Cells," Annals of the New York Academy of Sciences, vol. 1109, pp. 358-371, 2007.

[5] R. Hiemann, N. Hilger, U. Sack, and M. Weigert, "Objective quality evaluation of fluorescence images to optimize automatic image acquisition," Cytometry Part A, vol. 69A, pp. 182-184, 2006.

[6] J. Hongliang, L. Qingshan, L. Hanqing, and T. Xiaofeng, "Face detection using improved LBP under Bayesian framework," presented at Image and Graphics, 2004. Proceedings. Third International Conference on, 2004.

[7] T. Ojala, M. Pietikäinen and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognition, vol. 29, pp. 51-59, 1996.

[8] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, pp. 971-987, 2002.

[9] P. Perner, H. Perner, and B. Müller, "Mining knowledge for HEp-2 cell image classification," Artificial Intelligence in Medicine, vol. 26, pp. 161-173, 2002.

[10] M. Pietikäinen, T. Ojala, and Z. Xu, "Rotation-invariant texture classification using feature distributions," Pattern Recognition, vol. 33, pp. 43-52, 2000.

[11] A. Rigon, F. Buzzulini, P. Soda, L. Onofri, L. Arcarese, G. Iannello, and A. Afeltra, "Novel opportunities in automated classification of antinuclear antibodies on HEp-2 cells," Autoimmunity Reviews, vol. 10, pp. 647-652.

[12] U. Sack, S. Knoechner, H. Warschkau, U. Pigla, F. Emmrich, and M. Kamprad, "Computer-assisted classification of HEp-2 immunofluorescence patterns in autoimmune diagnostics," Autoimmunity Reviews, vol. 2, pp. 298-304, 2003.

[13] P. Soda, G. Iannello, and M. Vento, "A multiple expert system for classifying fluorescent intensity in antinuclear autoantibodies analysis," Pattern Anal. Appl., vol. 12, pp. 215-226, 2009.

[14] P. Soda and G. Iannello, "Aggregation of Classifiers for Staining Pattern Recognition in Antinuclear Autoantibodies Analysis," Information Technology in Biomedicine, IEEE Transactions on, vol. 13, pp. 322-329, 2009.

[15] P. Soda, A. Rigon, A. Afeltra, and G. lannello, "Automatic Acquisition of Immunofluorescence Images: Algorithms and Evaluation," presented at Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on, 2006.

[16] D. H. Solomon, A. J. Kavanaugh, P. H. Schur, and G. American College of Rheumatology Ad Hoc Committee on Immunologic Testing, "Evidence-based guidelines for the use of immunologic tests: Antinuclear antibody testing," Arthritis Care & Research, vol. 47, pp. 434-444, 2002.

[17] H. Yi-Chu, H. Tsu-Yi, C. Chin-Yuan, C. Wei-Ta, L. Yu-Chih, and H. Yu-Len, "HEp-2 cell images classification based on textural and statistic features using self-organizing map," in Proceedings of the 4th Asian conference on Intelligent Information and Database Systems - Volume Part II. Kaohsiung, Taiwan: Springer-Verlag.

[18] H. Yu-Len, C. Chia-Wei, H. Tsu-Yi, and J. Yu-Lang, "Outline Detection for the HEp-2 Cell in Indirect Immunofluorescence Images Using Watershed Segmentation," presented at Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC '08. IEEE International Conference on, 2008.

[19] H. Yu-Len, J. Yu-Lang, H. Tsu-Yi, and C. Chia-Wei, "Adaptive Automatic Segmentation of HEp-2 Cells in Indirect Immunofluorescence Images," in Proceedings of the 2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (sutc 2008): IEEE Computer Society, 2008.

[20] http://mivia.unisa.it/hep2contest/index.shtml