Pattern Recognition ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

ANA HEp-2 cells image classification using number, size, shape and localization of targeted cell regions

Gennady V. Ponomarev [a], Vladimir L. Arlazarov [b], Mikhail S. Gelfand [a,c], Marat D. Kazanov [a,*]

[a] A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Science, Russia
[b] Institute for Systems Analysis, Russia Academy of Science, Russia
[c] M.V. Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Russia

ARTICLE INFO

Keywords:
Immunofluorescent images
HEp-2 cells
Antinuclear antibodies
Image classification

ABSTRACT

The ANA HEp-2 medical test is a powerful tool in autoimmune disease diagnostics. The last step of this test, the interpretation of immunofluorescent images by trained experts, represents a potential source of errors and could theoretically be replaced by automated methods. Here we present a fully automatic method for recognition of types of immunofluorescent images produced by the ANA HEp-2 medical test. The proposed method makes use of the difference in number, size, shape and localization of cell regions that are targeted by the antinuclear antibodies – the humoral components of immune system that bind human antigens as a result of the immune system malfunction. The method extracts morphological properties of stained cell regions using a combination of thresholding-based and thresholding-less approaches and applies a conventional machine-learning algorithm for image classification.

1. Introduction

Many human autoimmune diseases are caused by a specific failure of immune system, which leads to production of its soluble components, antibodies, directed against self cells or tissues [1]. The antibodies of this type are referred to as autoantibodies [2]. Antinuclear antibodies (ANAs) [3] are a specific class of autoantibodies that attack different proteins and DNA inside the cell nucleus. The specificity of antinuclear antibodies varies depending on the type of an autoimmune disease and results in different localization of antibodies inside the nucleus, which can be exploited for the differentiation of a diagnosis. In clinical practice, the golden standard among laboratory tests detecting a presence and localization of ANAs is the indirect immunofluorescence (IIF) method [4]. This method produces patterns of stained cells by incubating the patient's serum with a certain cell line (HEp-2) and fluorescently labeled anti-human immunoglobulins [5–7]. The obtained stained cells are visualized using a fluorescence microscope and are analyzed by the trained personnel for establishing a diagnosis.

There are six predefined classes to which stained cells may be assigned (Fig. 1a) depending on the localization of antinuclear antibodies inside a nucleus. Each image class is associated with a single or a set of autoimmune diseases [8]. While immunofluorescence is known as a highly sensitive method for detecting antinuclear antibodies [9], some studies raise a concern that the interpretation of results depends on the personnel qualification level and many other factors [10,11]. This necessitates automation of the interpretation of ANA HEp-2 cell images. Several computer-based methods for classification of cell images [12–15] as well as preprocessing segmentation methods [16] were proposed in the literature, but they have not been benchmarked.

Recent appearance of a publicly available dataset of ANA HEp-2 cell images [17] and the associated image classification contest [8] can make a profound impact on the progress in this field. The method described in this paper was presented for the first time at that contest and here we describe it in complete detail, including some recent improvements.

2. Methods

The main idea of our approach is to utilize for automatic classification the morphological properties of stained-cell regions or domains as they obviously serve as the basis for the classification of HEp-2 cell immunofluorescent images by medical personnel. More specifically, we exploited the number, size, localization and shape of the stained-cell domains. The general scheme of the proposed method is as follows: (i) conversion of the original color

* Corresponding author. Tel.: +7 9166186778; fax: +7 4956500579.
E-mail addresses: krkdl@mail.ru (G.V. Ponomarev), mkazanov@gmail.com (M.D. Kazanov).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.09.027

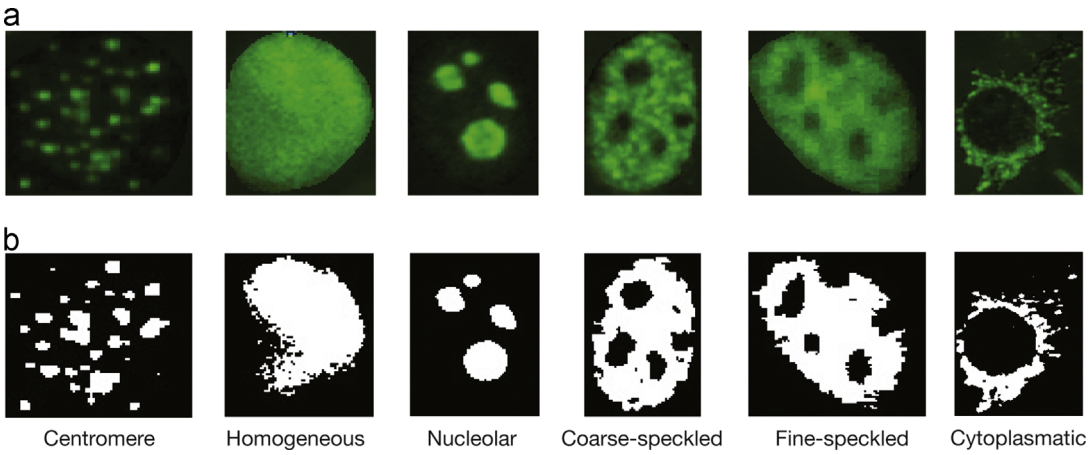

Fig. 1. Images of ANA HEp-2 stained cells. (a) Six classes of stained-cells images defined in the course of ICPR 2012 ANA HEp-2 Cells Classification Contest and their binary versions and (b) obtained using Otsu thresholding.

Table 1
The full list of the image features used for classification and the estimates of their discrimination abilities calculated by F-test.

No.	Image feature	F-test *p*-value	F-test rank
1	Object number	4.07E−89	7
2	Average object area	3.43E−101	5
3	Object area variance	7.75E −12	17
4	Maximum object area	1.66E−123	4
5	Hole number	8.18E−58	11
6	Average hole area	2.73E−43	15
7	Hkole area variance	4.98E−44	14
8	Maximum hole area	2.19E−59	10
9	Distance to cell boundaries	3.58E−75	9
10	Longest concave arc length	5.94E−51	13
11	LIV number	2.63E−82	8
12	Average LIV area	3.83E−20	16
13	Average LIV depth	5.37E−55	12
14	Big LIV number	1.99E−96	6
15	LIP number	3.97E−141	1
16	Average LIP area	5.74E−125	3
17	Average LIP height	1.52E−134	2

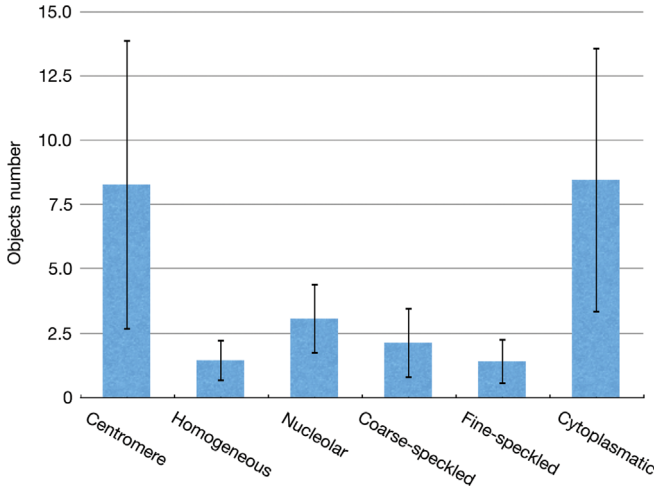

Fig. 2. Distribution of the object number for ANA HEp-2 image classes.

image into a grayscale image; (ii) image thresholding using the Otsu binarization method (Fig. 1b) [18]; (iii) extraction of image features; (iv) image classification using a SVM classifier [19]. The Otsu binarization and SVM methods were applied in their original form without any modifications. Below we provide a detailed description of the cell image features that are used for classification. We logically divided all features into two groups based on the extraction approach: the first group comprises features extracted from the binary image obtained from the original image using thresholding, while the second group of features are the ones extracted from the original image. The full list of the image features along with the estimates of their discrimination abilities, calculated independently for each feature using the F-test, is presented in Table 1. We will further refer to any connected groups of pixels, which were assigned as foreground pixels after the binarization, as objects.

2.1. *Feature group I: features extracted from the binary image*

Object number: The number of stained regions inside the cell is clearly different among several ANA HEp-2 cell image classes. This is a direct consequence of varying biological nature of the antibodies' targets. Thus, the largest number of stained objects usually can be found for the Centromere class of images. The antibodies' targets of this class are the chromosome centromere proteins, so the object number usually corresponds to the normal chromosome number (46 for a human cell). The second largest object number usually corresponds to the Nucleolar class of images where the antibodies target the nucleoli. The number of nucleoli in a human cell generally varies from one to ten depending on a cell phase [20]. The images of the Cytoplasmatic class have a diffuse staining throughout the cytoplasm of a cell, which after binarization often splits into several objects. Thus, contrary to other classes, images of the Cytoplasmatic class have a larger variance of the number of objects in an image. The Homogeneous and both speckled classes have, as a rule, an entirely stained nucleus, which results in formation of a single object after binarization. The distribution of object number across the image classes observed for images of the ICPR HEp-2 Cells Classification Contest training dataset [8] is shown in Fig. 2.

Object size: The cell regions targeted by antibodies in the ANA HEp-2 test differ not only in numbers but also by size. The smallest ones among the domains targeted by antinuclear antibodies are the centromeric regions of chromosomes stained in the respective class of images. The cell nucleoli that are stained in the Nucleoar class images are slightly larger than centromeres. The largest objects' sizes may be attributed to the entire stained cell nuclei of the Homogeneous and both speckled classes. At that, the two latter have a slightly reduced size in comparison with the whole

Please cite this article as: G.V. Ponomarev, et al., ANA HEp-2 cells image classification using number, size, shape and localization of targeted cell regions, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.09.027

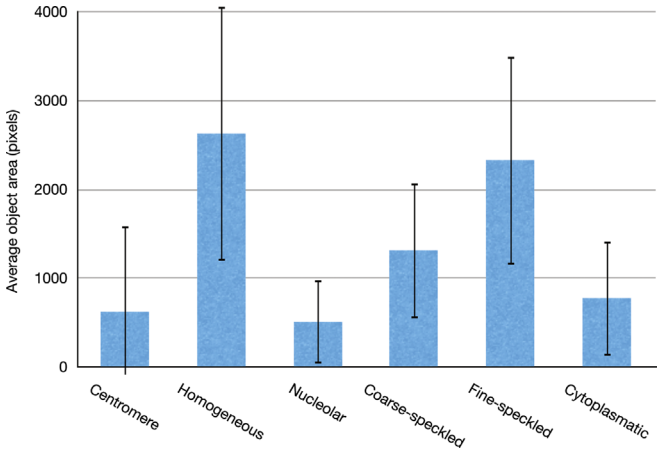

Fig. 3. Distribution of the average object area for ANA HEp-2 image classes.

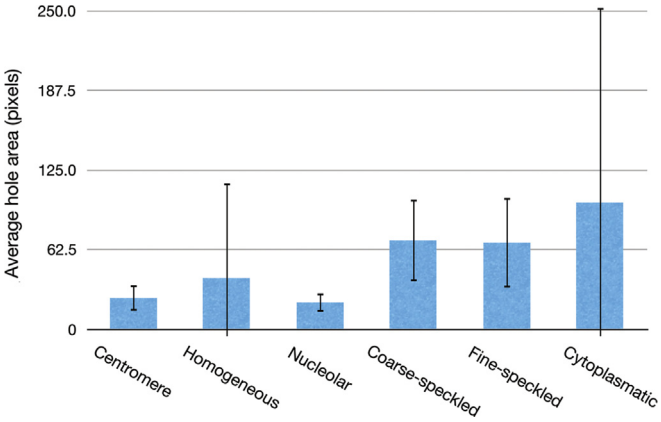

Fig. 4. Distribution of the average hole area for ANA HEp-2 image classes.

nucleus due to the exclusion of negatively stained nucleoli. Size of objects in the Cytoplasmatic class images highly depends on the inhomogeneity of the stained cytosol area, which cause splitting of the entire stained cytosol into several objects after binarization. The particular features related to the size of stained objects, which we used in our method, were the average object area, variance of the object area and maximum of the object area. The distribution of the mean object area across the image classes is presented in Fig. 3.

Holes inside objects: A prominent feature that is peculiar for two speckled classes is the negatively stained nucleoli located inside a positively stained nucleus. The nucleoli appear in the speckled classes of images as readily observable round dark holes inside a bright nucleus circle. In contrast, other image classes do not possess any similar structures except for rare artifacts. To capture a presence of holes inside objects we introduce the following features: the number of holes, as well as mean, variance and maximum of the hole area. The distribution of the mean hole area is presented in Fig. 4.

Object localization: Localization of various targets inside a cell seems to be an important classification feature. While the stained centromeric regions of chromosomes and nucleoli tend to be uniformly spread throughout the entire nucleus with a gap between them and the nucleus boundary, stained cytoplasm is located outside of the nucleus and surrounds it being immediately adjacent to the cell boundaries. Similar to the cytoplasmic case, the concentration of fluorescent antibodies of the two speckled classes is higher in the boundaries of the image, i.e. near the nuclear membrane, due to negatively stained nucleoli. Thus, we introduce a feature that describes whether the antibodies' targets are located closer to the preliminary segmented boundaries of a nucleus (or a cell in the Cytoplasmatic case), or closer to the center of image. For that, we calculate the average of the shortest distances from each object's pixel to the image boundaries, i.e nucleus or cell boundaries

$$L = \frac{\sum_{i,j} d_{i,j}}{N},$$

where i is the object index, j is the pixel index of the ith object, $d_{i,j}$ is the shortest distance from the jth pixel of the ith object to the image boundaries and N is the total number of pixels. The distribution of this feature among image classes is shown in Fig. 5.

Object shape: The objects' outer contours represent an approximate circle for the majority of image classes. The exception is the Cytoplasmatic class where the stained cytoplasmic content that surrounds the nucleus as a rule splits into several irregularly shaped objects. Such objects often have an extended concave part of the contour, which is adjacent to the nuclear membrane (Fig. 1). This property can be useful for recognition of the Cytoplasmatic class images (Fig. 6). The respective parameter is the length of the longest continuous concave segment of the object's contour. To assign a single value of this parameter to each particular image we introduced a corresponding image feature as the maximum value of this parameter among all image objects

$$C = \max_i \{c_i\},$$

where c_i is the longest continuous concave segment of contour of the ith object.

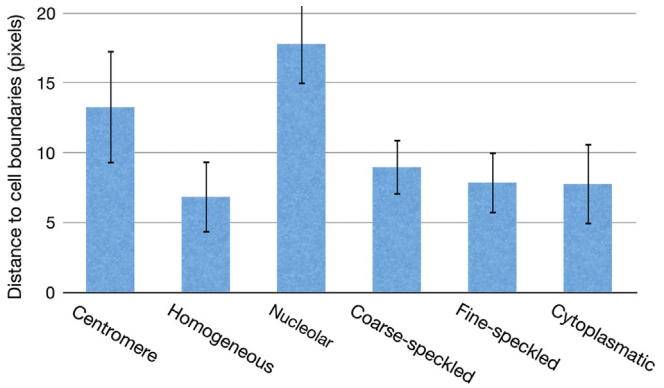

Fig. 5. Distribution of the mean distance from objects to image boundaries for ANA HEp-2 image classes.

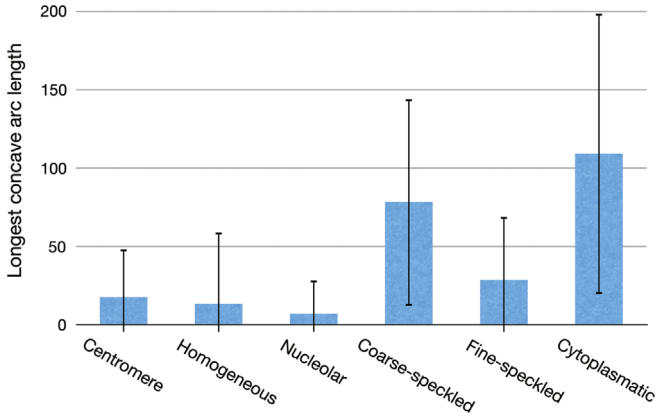

Fig. 6. Distribution of the longest concave arc length for ANA HEp-2 image classes.

Please cite this article as: G.V. Ponomarev, et al., ANA HEp-2 cells image classification using number, size, shape and localization of targeted cell regions, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.09.027

2.2. Feature group II: features extracted from the original image

Local intensity valleys: The main drawback of the hole detection procedure described above, which implies the recognition of holes from the binary image obtained using a single threshold, was occasional loss of holes located on other intensity levels. To overcome this problem we developed an approach for detection of all local intensity valleys (LIV) on the original image, which can be visually perceived as the holes in objects. For this purpose we organized a loop that iterates through the image intensity levels with a small step s and performs a binarization procedure at each iteration. This allowed us to track the object holes throughout the intensity levels gathering information about its maximum size, and also to detect the levels at which each hole appears and degrades. The knowledge about the depth of the intensity valley, i.e. the value of difference between the levels of hole appearance l_a and degradation l_d, can help to distinguish spurious valleys due to noise, which usually have a small depth, from the real ones, which as a rule is of significant depth.

The introduced procedure can be summarized as follows:

- the LIV area is $HS = \max_i(HS^i)$, where HS^i is the area of the hole, which corresponds to the local intensity valley detected in the binary image generated by thresholding at the intensity level i, $i \in [l_a, l_d]$.
- the LIV depth is $HD = l_d - l_a$;
- an image feature describing the number of LIVs having significant area values is $HN_T = |\{H_i\}|,\ i : HS_i > T_{LIV\_big}$, where T_{LIV_big} is the threshold on the LIV area, HS_i is the area of the ith hole, $|\{H_i\}|$ is the number of LIVs satisfying the condition.

Local intensity peaks: Similarly to the detection of local intensity valleys corresponding to holes in objects, we applied a procedure that detects local intensity peaks (LIP), which correspond to fluorescently labeled centromeric regions in Centromere images and small bright spots inside the cell nucleus in Coarse-speckled images. The introduced procedure iterates through the intensity levels and tracks LIP formation starting from the brightest level of image. To detect only small area LIPs and to distinguish them from noise objects we defined two empirical thresholds: the first one, T_{LIP_area}, was established to define the maximum possible LIP area and the second one, T_{LIP_height}, to define the minimum contrast of actual LIPs, i.e. the minimal allowed difference between levels of LIP appearance and degradation. We defined the level of LIP appearance as the intensity level at which any new object with area less than T_{LIP_area} has been detected in the binary image obtained by thresholding for a current iteration. Accordingly, the level of degradation of the LIP was defined as the intensity level at which the area of the object exceeded the T_{LIP_area} threshold. The calculated LIP contrast, or the LIP height, which we defined as a difference between levels of LIP appearance and degradation, allowed us to differentiate actual centromeric regions from noise. As a result, we introduced two new image features. the LIP number and the average LIP height, which corresponded to the number of detected small-area local intensity peaks in the image and to the average contrast of the detected LIPs relative to the background, respectively.

2.3. Optimization and performance evaluation

The evaluation of performance of a particular version of the method was performed by 10-fold cross-validation on the training set. We used Accuracy metric as the main criterium for the estimation of the method efficiency. For deeper understanding of the method classification quality we constructed the method confusion matrix representing numbers of misclassification cases between each pair of image classes. Analysis of the confusion matrix led us to the introduction of new features aiming to discriminate poorly separated classes. We also confirmed the discrimination power of the proposed features using the F-test (Table 1). Calculated F-test p-values demonstrated high, but uneven classification ability of the image features. It was expected because some features like the object number are different for the majority of classes, while other features like the arc length or hole number are specific for individual classes.

Our method has several parameters, which were optimized using images from the training set. Most parameters were introduced to distinguish real objects from noise. The complete list of parameters is as follows: minimal object area (T_{obj_area}), minimal hole area (T_{hole_area}), maximum LIP height (T_{LIP_height}), maximum LIP area (T_{LIP_area}), minimal big LIV area (T_{LIV_big}). The value of the minimal object size was derived from the analysis of the Centromere images containing the smallest stained regions among all classes. Given knowledge of the number of chromosomes in the cell it was relatively easy to distinguish centromere regions from noise objects and to derive the value of T_{obj_area}. The value of the minimal hole area T_{hole_area} was set to the minimal object area. Values of other three parameters, which were involved in the calculation of thresholding-less image features, were optimized based on the properties of objects obtained on the training set using thresholding-based image features only.

To make the method universal with regard to the image resolution, we calculated all image features in relative units. Thus, all features that describe areas, e.g. object and hole areas, were normalized by the cell area. All one-dimensional features were normalized by the constant calculated as the diameter of a circle whose area equals to the cell area.

3. Results

The key step of the ANA HEp-2 medical test is the interpretation of obtained stained patterns of HEp-2 cells for establishing a correct diagnosis. The ANA HEp-2 test produces diverse staining patterns due to staining of different cell regions or domains. These domains, nucleoli, nucleus, cytosol and chromosomes, differ in size, shape, number and localization inside the cell. This allows interpreters to distinguish staining patterns characteristic for different autoimmune deceases. The main idea of the proposed approach is to segment the stained regions in the image using a binarization technique and then to catch their structural properties as extracted features.

The classification quality of the proposed method was evaluated using the MIVIA HEp-2 Images Dataset that is a publicly available dataset of HEp-2 cell images [17]. The evaluation consisted of two experiments. In the first experiment the whole dataset was divided into the training and test sets in exactly the same way as it was done at the ICPR 2012 HEp-2 Cells Classification Contest [8]. These two datasets were constructed from the original images of 28 ANA HEp-2 test slides each of which contained several dozens of HEp-2 stained cells. All HEp-2 cells were preliminary segmented and presented separately along with segmentation masks. We will further refer to the images of presegmented cells simply as cells and to the images of ANA HEp-2 test slides simply as images. The overall accuracy, i.e. the percentages of correctly classified cells, for training and test sets were 95.56% (689 out of 721 cells) and 70.57% (518 out of 734 cells), respectively. The confusion matrices corresponding to the cell classification results are presented in Tables 2 and 3, respectively. Based on cell classification we further assigned class labels to the images as the most frequent class of cells in the image. The overall accuracies of image classification for training and test sets

Please cite this article as: G.V. Ponomarev, et al., ANA HEp-2 cells image classification using number, size, shape and localization of targeted cell regions, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.09.027

Table 2
The cell-level confusion matrix obtained for the MIVIA training set. Overall accuracy is 95.56%.

Cell class	Centromere (%)	Homogeneous (%)	Nucleolar (%)	Coarse speckled (%)	Fine speckled (%)	Cytoplasmatic (%)
Centromere	98.56	0.96	0.48	0.00	0.00	0.00
Homogeneous	0.00	94.00	2.67	0.67	2.67	0.00
Nucleolar	0.00	0.00	100.00	0.00	0.00	0.00
Coarse speckled	0.92	0.00	0.00	99.08	0.00	0.00
Fine speckled	0.00	17.02	0.00	3.19	79.79	0.00
Cytoplasmatic	0.00	0.00	0.00	0.00	0.00	100.00

Table 3
The cell-level confusion matrix obtained for the MIVIA test set. Overall accuracy is 70.57%.

Cell class	Centromere (%)	Homogeneous (%)	Nucleolar (%)	Coarse speckled (%)	Fine speckled (%)	Cytoplasmatic (%)
Centromere	98.56	0.96	0.48	0.00	0.00	0.00
Homogeneous	0.00	94.00	2.67	0.67	2.67	0.00
Nucleolar	0.00	0.00	100.00	0.00	0.00	0.00
Coarse speckled	0.92	0.00	0.00	99.08	0.00	0.00
Fine speckled	0.00	17.02	0.00	3.19	79.79	0.00
Cytoplasmatic	0.00	0.00	0.00	0.00	0.00	100.00

Table 4
The image-level confusion matrix obtained for the MIVIA training set. Overall accuracy is 100%.

Image class	Centromere (%)	Homogeneous (%)	Nucleolar (%)	Coarse speckled (%)	Fine speckled (%)	Cytoplasmatic (%)
Centromere	100.00	0.00	0.00	0.00	0.00	0.00
Homogeneous	0.00	100.00	0.00	0.00	0.00	0.00
Nucleolar	0.00	0.00	100.00	0.00	0.00	0.00
Coarse speckled	0.00	0.00	0.00	100.00	0.00	0.00
Fine speckled	0.00	0.00	0.00	0.00	100.00	0.00
Cytoplasmatic	0.00	0.00	0.00	0.00	0.00	100.00

Table 5
The image-level confusion matrix obtained for the MIVIA test set. Overall accuracy is 100%.

Image class	Centromere (%)	Homogeneous (%)	Nucleolar (%)	Coarse speckled (%)	Fine speckled (%)	Cytoplasmatic (%)
Centromere	100.00	0.00	0.00	0.00	0.00	0.00
Homogeneous	0.00	100.00	0.00	0.00	0.00	0.00
Nucleolar	0.00	0.00	100.00	0.00	0.00	0.00
Coarse speckled	0.00	0.00	0.00	100.00	0.00	0.00
Fine speckled	0.00	0.00	0.00	0.00	100.00	0.00
Cytoplasmatic	0.00	0.00	0.00	0.00	0.00	100.00

were both 100%. The confusion matrices of image-level classification are presented in Tables 4 and 5, respectively.

In the second experiment we applied leave-one-out technique on the image level. For the all 28 ANA HEp-2 test slide images the same numbers of runs were performed where in each run the method was trained on cells belonging to 27 images and then tested on cells extracted from the remaining image. The summary of the classification results for all images including the percentage of correctly classified cells, as well as the cell-level and image-level classification confusion matrices are presented in Tables 6–8.

4. Conclusions

This paper presents an automatic method for classification of immunofluorescent cell images obtained in the ANA HEp-2 medical test. The proposed method exploits the biological nature of cell regions or domains targeted by fluorescently labeled antibodies by assigning obtained immunofluorescent images into six predefined classes for establishing a correct diagnosis. The method captures the structural content of a stained cell regarding the fluorescently labeled domains, particularly, the domain number, size, shape and localization. The obtained image features are used for classification by means of a conventional machine-learning algorithm. At the time of writing our method demonstrated the highest classification quality (95.56% and 70.57% accuracies on the training and test set, respectively) among the methods presented at the ICPR 2012 HEp-2 Cell Classification contest. Most of these methods are texture-based ones. Contrary to texture-based methods, which usually use a large number of features, our method uses a small number of features that should decrease a computation time and facilitate further improvement of the method. Moreover, our method is independent of the cell image resolution. Overall, the accuracy of the best methods on the contest's test set lies in the vicinity of 70 percent, which leaves space for further research in this field.

Please cite this article as: G.V. Ponomarev, et al., ANA HEp-2 cells image classification using number, size, shape and localization of targeted cell regions, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.09.027

Table 6
The results of cells classification on 28 slides obtained for the leave-one-out experiment.

Image ID	True class	Centromere		Homogeneous		Nucleolar		Coarse speckled		Fine speckled		Cytoplasmatic	
1	Homogeneous	0	0.00%	49	80.33%	5	8.20%	1	1.64%	6	9.84%	0	0.00%
2	Fine speckled	1	2.08%	8	16.67%	1	2.08	22	45.83%	16	33.33%	0	0.00%
3	Centromere	86	96.63%	0	0.00%	2	2.25%	1	1.12%	0	0.00%	0	0.00%
4	Nucleolar	4	6.06%	16	24.24%	27	40.91%	0	0.00%	12	18.18%	7	10.61%
5	Homogeneous	4	8.51%	27	57.45%	7	14.89%	0	0.00%	8	17.02%	1	2.13%
6	Coarse speckled	4	5.88%	0	0.00%	0	0.00%	63	92.65%	1	1.47%	0	0.00%
7	Centromere	47	8.93%	0	0.00%	8	14.29%	1	1.79%	0	0.00%	0	0.00%
8	Nucleolar	45	80.36%	0	0.00%	11	19.64%	0	0.00%	0	0.00%	0	0.00%
9	Fine speckled	0	0.00%	28	50.87%	0	0.00%	12	26.09%	6	13.04%	0	0.00%
10	Coarse speckled	6	18.18%	0	0.00%	0	0.00%	12	36.36%	8	24.24%	7	21.21%
11	Coarse speckled	1	2.44%	0	0.00%	0	0.00%	5	12.20%	35	85.37%	0	0.00%
12	Coarse speckled	2	4.06%	1	2.04%	0	0.00%	37	75.51%	0	0.00%	9	18.37
13	Centromere	38	82.61%	0	0.00%	4	8.70%	2	4.35%	2	4.35%	0	0.00%
14	Centromere	5	7.94%	49	77.78%	2	3.17%	0	0.00%	5	7.94%	2	3.17%
15	Fine speckled	5	7.94%	10	15.87%	2	3.17%	22	34.92%	14	22.22%	10	15.87%
16	Centromere	35	92.11%	0	0.00%	3	7.89%	0	0.00%	0	0.00%	0	0.00%
17	Coarse speckled	0	0.00%	0	0.00%	4	21.05%	9	43.37%	6	31.58%	0	0.00%
18	Homogeneous	0	0.00%	7	16.67%	1	2.38%	0	0.00%	34	80.95%	0	0.00%
19	Centromere	48	73.85%	0	0.00%	16	24.62%	1	1.54%	0	0.00%	0	0.00%
20	Nucleolar	39	84.76%	1	2.17%	6	13.04%	0	0.00%	0	0.00%	0	0.00%
21	Homogeneous	4	6.56%	45	73.77%	4	6.56%	0	0.00%	8	13.11%	0	0.00%
22	Homogeneous	2	1.68%	65	54.62%	9	7.58%	23	19.33%	9	7.56%	11	8.24%
23	Fine speckled	1	1.96%	45	88.24%	0	0.00%	4	7.84%	1	1.96%	0	0.00%
24	Nucleolar	3	4.11%	12	16.44%	56	76.71%	2	2.74%	0	0.00%	0	0.00%
25	Cytoplasmatic	10	41.67%	0	0.00%	2	8.33%	1	4.17%	0	0.00%	11	45.83%
26	Cytoplasmatic	3	8.82%	0	0.00%	0	0.00%	1	2.94%	0	0.00%	30	88.24%
27	Cytoplasmatic	0	0.00%	1	2.63%	0	0.00%	6	15.79%	0	0.00%	31	81.58%
28	Cytoplasmatic	0	0.00%	0	0.00%	1	7.69%	2	15.38%	0	0.00%	10	76.92%

Table 7
The cell-level confusion matrix obtained for the leave-one-out experiment. Overall accuracy is 54.77%.

Cell class	Centromere	Homogeneous (%)	Nucleolar (%)	Coarse speckled (%)	Fine speckled (%)	Cytoplasmatic (%)
Centromere	72.55	13.73	9.80	1.40	1.96	0.56
Homogeneous	3.03	58.48	7.88	7.27	19.70	3.64
Nucleolar	37.76	12.03	41.49	0.83	4.98	2.90
Coarse speckled	6.19	0.48	1.90	60.00	23.81	7.62
Fine speckled	3.37	43.75	1.44	28.85	17.79	4.81
Cytoplasmatic	11.93	0.92	2.75	9.17	0.00	75.23

Table 8
Image-level confusion matrix obtained for the leave-one-out experiment. The overall accuracy is 71.43%.

Image class	Centromere (%)	Homogeneous (%)	Nucleolar (%)	Coarse speckled (%)	Fine speckled (%)	Cytoplasmatic (%)
Centromere	100.00	0.00	0.00	0.00	0.00	0.00
Homogeneous	0.00	80.00	0.00	0.00	20.00	0.00
Nucleolar	50.00	0.00	50.00	0.00	0.00	0.00
Coarse speckled	0.00	0.00	0.00	80.00	20.00	0.00
Fine speckled	0.00	50.00	0.00	50.00	0.00	0.00
Cytoplasmatic	0.00	0.00	0.00	0.00	0.00	100.00

Conflict of interest statement

None declared.

References

[1] B. Bolon, Cellular and molecular mechanisms of autoimmune disease, Toxicologic Pathology 40 (2) (2012) 216–229.
[2] A. Lleo, P. Invernizzi, B. Gao, M. Podda, M.E. Gershwin, Definition of human autoimmunity–autoantibodies versus autoimmune disease, Autoimmunity Reviews 9 (5) (2010) A259–266.
[3] E.M. Tan, Autoantibodies to nuclear antigens (ANA): their immunobiology and medicine, Advances in Immunology 33 (1982) 167–240.
[4] I.D. Odell, D. Cook, Immunofluorescence techniques, Journal of Investigative Dermatology 133 (1) (2013) e4.
[5] D.H. Solomon, A.J. Kavanaugh, P.H. Schur, Evidence-based guidelines for the use of immunologic tests: antinuclear antibody testing, Arthritis & Rheumatism 47 (4) (2002) 434–444.
[6] E.M. Tan, Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology, Advances in Immunology 44 (1989) 93–151.
[7] U. Sack, K. Conrad, E. Csernok, I. Frank, F. Hiepe, T. Krieger, A. Kromminga, P. von Landenberg, G. Messer, T. Witte, R. Mierau, Autoantibody detection using indirect immunofluorescence on HEp-2 cells, Annals of the New York Academy of Sciences 1173 (2009) 166–173.
[8] P. Foggia, G. Percannella, P. Soda, M. Vento, Benchmarking hep-2 cells classification methods, IEEE Transactions on Medical Imaging PP (99) (2013) 1.
[9] E. Ulvestad, Performance characteristics and clinical utility of a hybrid ELISA for detection of ANA, Acta Pathologica, Microbiologica et Immunologica Scandinavica 109 (3) (2001) 217–222.

Please cite this article as: G.V. Ponomarev, et al., ANA HEp-2 cells image classification using number, size, shape and localization of targeted cell regions, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.09.027

[10] D.H. Solomon, R.H. Shmerling, P.H. Schur, R. Lew, J. Fiskio, D.W. Bates, A computer based intervention to reduce unnecessary serologic testing, Journal of Rheumatology 26 (12) (1999) 2578–2584.
[11] R. Hiemann, T. Buttner, T. Krieger, D. Roggenbuck, U. Sack, K. Conrad, Challenges of automated screening and differentiation of non-organ specific autoantibodies on HEp-2 cells, Autoimmunity Reviews 9 (1) (2009) 17–22.
[12] P. Perner, H. Perner, B. Muller, Mining knowledge for HEp-2 cell image classification, Artificial Intelligence in Medicine 26 (1–2) (2002) 161–173.
[13] U. Sack, S. Knoechner, H. Warschkau, U. Pigla, F. Emmrich, M. Kamprad, Computer-assisted classification of HEp-2 immunofluorescence patterns in autoimmune diagnostics, Autoimmunity Reviews 2 (5) (2003) 298–304.
[14] A. Rigon, F. Buzzulini, P. Soda, L. Onofri, L. Arcarese, G. Iannello, A. Afeltra, Novel opportunities in automated classification of antinuclear antibodies on HEp-2 cells, Autoimmunity Reviews 10 (10, SI) (2011) 647–652.
[15] P. Soda, G. Iannello, Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis, IEEE Transactions on Information Technology in Biomedicine 13 (3) (2009) 322–329.
[16] Y.-L. Huang, C.-W. Chung, T.-Y. Hsieh, Y.-L. Jao, Outline detection for the hep-2 cell in indirect immunofluorescence images using watershed segmentation, in: IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing, SUTC '08. 2008, 2008, pp. 423–427.
[17] P. Foggia, G. Percannella, P. Soda, M. Vento, Early experiences in mitotic cells recognition on hep-2 slides, in: IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS), 2010, pp. 38 –43. http://dx.doi.org/10.1109/CBMS.2010.6042611.
[18] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (1) (1979) 62–66.
[19] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297.
[20] M. Anastassova-Kristeva, The nucleolar cycle in man, Journal of Cell Science 25 (1977) 103–110.

Gennady V. Ponomarev received B.Sc. degree in Electrical Engineering from Gubkin Russian State University and M.Sc. degree in Applied Mathematics from M.V. Lomonosov Moscow State University, Russia. His current research interests include medical imaging, pattern recognition and computer vision.

Vladimir L. Arlazarov, Ph.D (tech.), D.Sc. (math.), member of European and Russian Academies of Science. Head of the Artificial Intelligence Laboratory at the Institute for System Analysis, RAS, Moscow. His research interests include chess game programming, character and speech recognition.

Mikhail S. Gelfand, Ph.D. (math.), D.Sc. (biol.), member of Academia Europaea. Specialist in bioinformatics, comparative genomics, and systems biology. Deputy director for science of the A.A. Kharkevich Institute for Information Transmission Problems, RAS, and professor of bioinformatics of the M.V. Lomonosov Moscow State University.

Marat D. Kazanov received M.Sc. degree in Applied Mathematics from Moscow Institute of Physics and Technology State University and Ph.D. degree from Institute for System Analysis, RAS, Moscow. From 2008 to 2010, he did his postdoctoral research at Sanford-Burnham Medical Research Institute, La Jolla, USA. His research interests include bioinformatics, structural biology and medical imaging.

Please cite this article as: G.V. Ponomarev, et al., ANA HEp-2 cells image classification using number, size, shape and localization of targeted cell regions, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.09.027