Pattern Recognition Letters
journal homepage: www.elsevier.com

HEp-2 staining pattern recognition at cell and specimen levels: datasets, algorithms and results

Peter Hobson[a], Brian C. Lovell[b], Gennaro Percannella[c], Alessia Saggese[c], Mario Vento[c,**], Arnold Wiliem[b]

[a]*Sullivan Nicolaides Pathology, Australia*
[b]*University of Queensland, Australia*
[c]*University of Salerno, Italy*

ABSTRACT

The Indirect Immunofluorescence (IIF) protocol applied on Human Epithelial type 2 (HEp-2) cells is the current gold standard for the Antinuclear Antibody (ANA) test. The formulation of the diagnosis requires the visual analysis of a patient's specimen under a fluorescence microscope in order to recognize the cells' staining pattern which could be related to a connective tissue disease. This analysis is time consuming and error prone, thus in the recent past we have witnessed a growing interest in the pattern recognition scientific community directed at the development of methods for supporting this complex task. The main driver of the interest towards this problem is represented by the set of international benchmarking initiatives organized in the last four years that allowed dozens of research groups to propose innovative methodologies for HEp-2 cells' staining pattern recognition. In this paper we update the state of the art on HEp-2 cells and specimens classification, by analyzing the performance achieved by the methods participating the contest on Performance Evaluation of IIF Image Analysis Systems, hosted by the 22nd edition of the International Conference on Pattern Recognition ICPR 2014, and to the Executable Thematic Special Issue of Pattern Recognition Letters on Pattern Recognition Techniques for IIF Images Analysis, and by highlighting the trends in the design of the best performing methods.

1. Introduction

The Anti-Nuclear Antibody (ANA) test is commonly used for screening connective tissue diseases, ranging from Systemic Lupus Erythematosus to Sjorgrens syndrome and Rheumatoid Arthritis (Meroni and Schur (2010); Wiik et al.). Among the methods available to detect ANA, the Indirect Immunofluorescence (IIF) protocol using human epithelial type 2 (HEp-2) cells is nowadays considered the gold standard (Wiik et al.).

In this framework, the activity typically performed by scientists consists of the following steps: each slide is examined under a fluorescent microscope so as to determine the specimen positivity (among positive, intermediate, negative). Then, intermediate and positive specimens are further analyzed and mitotic cells are recognized. If the number of mitotic cells is sufficiently high (typically at least two per field), the scientist has to determine the staining pattern to which the cells belong to. Indeed, the presence of a positive pattern may be associated with a particular autoimmune disease. Unfortunately, despite the efficacy of the IIF protocol, the process is still labour intensive, time consuming and expensive. Henceforth, there is a strong need for the definition and adoption of computer aided diagnosis (CAD) solutions.

A critical analysis of the state of the art of all the steps required for this kind of CAD system goes beyond the scope of this paper; the reader can refer to Hobson et al. (2016) for a detailed analysis of the literature and of the open challenges in this research area. In this paper we focus on the staining recognition step: indeed, this task has attracted an increasing number of scientists working in the Pattern Recognition field, mainly due to the competitions organized in the last four years

**Corresponding author
e-mail: peter_hobson@snp.com.au (Peter Hobson), lovell@itee.uq.edu.au (Brian C. Lovell), pergen@unisa.it (Gennaro Percannella), asaggese@unisa.it (Alessia Saggese), mvento@unisa.it (Mario Vento), a.wiliem@uq.edu.au (Arnold Wiliem)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

(Foggia et al. (2013, 2014); Hobson et al. (2015)), where the participants were required to design and implement a pattern recognition system able to classify a single cell belonging to HEp-2 images.

The aim of this paper is to update the state of the art in staining pattern recognition, by focusing on both the methods participating to the last *Performance Evaluation of Indirect Immunofluorescence Image Analysis Systems* competition, hosted by the *International Conference on Pattern Recognition* 2014, and to the *Executable Thematic Special Issue of Pattern Recognition Letters on Pattern Recognition Techniques for Indirect Immunofluorescence Images Analysis*. The main difference with respect to the previous initiatives lies in the problem that researchers had to deal with: in addition to the traditional cell-level classification task, a new and more challenging task was proposed to the participants, namely the classification of the whole specimen image (and not only of the single cell image). Overall, fifteen research groups joined the above two initiatives.

The paper is organized as follows: a short summary of each method is provided in Section 2; a critical analysis of the obtained results is reported in Section 3; conclusions and future directions are delineated in Section 4.

2. The methods

In this Section we provide a short summary of each of the methods considered for the benchmarking described in the experimental section. For the sake of conciseness, for each method we delineate the overall architecture with few details on the pre-processing, the adopted description and the classification paradigm, while we refer to the associated literature for more details. Hereinafter, we will refer to each method by using the surname of the first author of the associated paper.

Codrescu - The method by Codrescu (2014) is based on quadratic recurrent finite impulse response multilayer perceptron (QR-FIRMLP), a class of temporal processing neural networks where the static weights are replaced by finite impulse response filters, characterized by a quadratic neurons transfer function and a recurrent connection. Cell images are rescaled according to the following two strategies: in the first one each image is scaled from the original size to 32×32 pixels without preserving aspect ratio; in the second approach, the largest square centered into the image is rescaled to 32×32 pixels. The train set is also augmented by applying mirror and rotate operations.

Ensafi - The method by Ensafi et al. (2016) operates according to four stages: cell extraction, super pixel extraction, dictionary learning and cell classification. The first stage is aimed at providing an initial segmentation of the cells using a priori information on their average size and shape. Then in the second stage the method extracts the superpixels based on the Simple Linear Iterative Clustering (SLIC), modified so that besides the color and spatial proximities, it is also considered the gradient information for modeling texture. From each superpixel overlapping the segmentation mask, SIFT and SURF features are extracted. Then, the sparse coding with the BoW model is applied to learn the optimal dictionary and codes. The sparse code of the image represents the feature vector that is given for final cell level classification as input to a multi-class linear SVM based with one-versus-all learning strategy. Specimen level classification is obtained through majority voting on the cell labels.

Gao - Gao et al. (2014) propose to face HEp-2 cells classification adopting a deep learning approach based on convolutional neural networks (CNNs), in order to extract features directly from the pixels values of the cell image in a hierarchical way without the need of resorting to hand-crafted features. Cell images are first normalized to enhance the contrast and then resized to a fixed size of 78×78 pixels. Rotation invariance is achieved by rotating the original image in 20 angles equally spaced with step 18.

Gragnaniello - The method by Gragnaniello et al. (2016) uses the Scale-Invariant Descriptor (SID) defined in Kokkinos et al. (2012) which is able to guarantee scale and rotation invariance through the following steps: log-polar transformation, multi scale smoothing, computation of directional derivatives, Fourier transform along the angular direction. The features are calculated only on the pixels from the area of the foreground using the segmentation masks provided with the sample images. Features are encoded using the Bag of Words model with soft assignment through Gaussian weights. Then final classification of the cells is carried out by a linear SVM classifier. The authors also propose a method for specimen level classification; the latter is performed by dividing the specimen in five sub-images and applying the method described before on each sub-image. Final decision is taken by majority voting.

Han - The authors propose to recognize the HEp-2 cell patterns using texture information. To this aim they originally propose and then use a modification of the local ternary pattern, by introducing the Rotation Invariant Co-occurrence Weber-based Local Ternary Pattern (RICWLTP). In particular, they extend Local Ternary Patterns (LTPs) in the Weber-based LTPs that use adaptive thresholds inspiring to the Webers law, a principle according to which the human perception of a distinguished pattern depends not only on the absolute intensity of the stimulus but also on the relative variance of the stimulus. The final RICWLTP descriptor incorporates the contexts of spatial and orientation co-occurrences among adjacent Weber-based local ternary patterns (WLTPs) for texture representation. Classification is carried out using a linear SVM fed with histograms values of the selected features.

Liu - The authors focused only on image classification at specimen level. The proposed method densely extracts 200×200 pixel regions from each specimen image. Features are calculated over each regions and are sent to a linear SVM for region classification. The image category is estimated via the majority voting on the decision of each region classifier. The region level representation is obtained by densely extracting the SURF descriptor over the region and then using the VLAD encoding defined in Jegou et al. (2012).

Manivannan - The method by Manivannan et al. (2014a) realizes a first pixel intensity normalization, then four types of local features are extracted: multi-resolution local pattern that is a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

multi-resolution non-binarized version of the local binary pattern; root-SIFT, a variant of the well known SIFT descriptor; random projections, which are projections of patch intensity vectors from the original patch-vector space to a compressed space using randomly chosen projection vectors; intensity histograms. All features are densely extracted from the images and a dictionary was learned for each feature type. Feature are encoded using local linear coding, an efficient variant of sparse coding and max-pooling is used to aggregate the local linear codes; moreover, for each feature type, a 2-level cell pyramid is used to capture spatial structure, as in Wiliem et al. (2013). Classification is done through an ensemble of multi-class linear SVM using the one-versus-all approach. The ensemble contains four SVMs trained respectively on 0°, 90°, 180°, 270° rotated versions of the original image. At test time each image is rotated as before and provided as input to the ensemble producing 16 classification scores for each class. Scores are treated as probabilities using Platt rescaling. The final classification decision is made by averaging these rescaled scores. Manivannan et al. (2014b) address the specimen classification problem by adopting a similar approach already used for cell classification, but relying only on multi-resolution local patterns and root-SIFT. These features are extracted from patches densely extracted from images; the size of such patches varies from 12x12 to 64x64, capturing both local properties of the cells and information concerning groups of cells.

Paci - Nanni et al. (2016) propose a method for image classification at cell level. The method employs a preprocessing stage with two filters: the contraharmonic mean filter (Banerjee and Maji (2013)), for the removal of impulse noise, and the contrast limited adaptive histogram equalization (Zuiderveld (1994)) to improve the local contrast. The method uses a combination of both global features extracted from the whole image and local SIFT features densely extracted from the image and then combined using a bag of feature approach. As global descriptor the authors adopts both textural features (the Local Configuration Pattern (LCP) by Yimo Guo and Pietikinen (2011), the rotation invariant co-occurrence among adjacent LBP (RIC-LBP) by Nosaka and Fukui (2014), the extended LBP (ELBP) by Liu et al. (2012), the multiscale Pyramid LBP (PLBP) by Qian et al. (2011)) and the Strandmark morphological features (STR) Strandmark et al. (2012). For each image a set of eight images is obtained by resizing and rotating the original one. Then each image is described using the above set of features and then classified using SVM; a radial basis function kernel is used for the global features, a histogram kernel for the local features. SVM multi classification is done through the one-versus-all approach. The scores obtained using the local features are combined with those obtained using the global features over each of the eight images by the sum rule in one global score per class. The class achieving the highest global score is assigned to the sample under evaluation.

Paisitkriangkrai - The image is first preprocessed by equalizing the histogram. The description of each image is obtained by using several sets of features and specifically: the covariance of the intensity values, of the first and of the second order derivative in the vertical and horizontal directions, of the magnitude of the gradients; the Co-Occurrence of Adjacent Local Binary Patterns (CoALBPs) Nosaka et al. (2012); the local projection coefficients from raw pixel values from which it is learnt a multi-codebook based visual set of features; the morphological features (STR) defined by Strandmark et al. (2012). The dataset is augmented by rotating images at every $\pi/4$ in order to increase robustness of the classifier. Then, for cell classification it is adopted the multi-class boosting algorithm in Paisitkriangkrai et al. (2014), while for specimen level classification it is used a sliding window for selecting patches of the image to be classified and majority voting is applied over them.

Ponomarev - The method by Ponomarev and Kazanov (2016) relies on two groups of features: the first one comprises features extracted from the binary image obtained using Otsu thresholding (as number, average, variance and maximum area of the connected components and of the holes, etc), while the second group contains features that account for the local intensity peaks and valleys obtained through region growing. Classification is done through SVM with the linear kernel. Two different SVM models are trained for handling intermediate and positive fluorescence intensity images. The authors also propose to perform specimen level classification by combining the outputs obtained at the cell levels using weighted voting, where different weights are used according to the fact that the considered cell is isolated or not into the segmentation mask.

Qi - The method by Qi et al. (2016) is based on the Local Orientation Adaptive Descriptor (LOAD) by Qi et al. (2015), which is suited to describe the regional texture in an image. LOAD is obtained by aggregating on circular patches a binary sequence descriptor able to describe the relationships between selected points in the patch and their neighbors. The descriptor is based on an adaptive coordinate system that guarantees rotation invariance. In order to capture multi-resolution information the original image is Gaussian convoluted at seven scales. Then LOAD feature is calculated on round patches densely extracted from the eigth images (the original plus the seven in different scales). The Improved Fisher Vector (IFV) by Perronnin et al. (2010) is used to encode the features to preserve its discriminative power. Then, classification is performed through a linear SVM.

Roberts - In the preprocessing phase the pixels of the background are set to 0 using the provided cell mask, then the image is resized to 128×128 pixels. The feature vector is composed by aggregation of the modulus of the coefficients of the dual-tree complex wavelet transform. The classification is carried out by a multi-class SVM using the one versus all scheme.

Sarrafzadeh - The method proposed by Sarrafzadeh et al. (2016) proceeds to an initial image normalization, after which different groups of features are calculated and specifically: binary, intensity, statistical, spectral and texture features. In the last group they are considered the Haralick features, the Completed Local Binary Pattern (CLBP), wavelet based features and Gabor features. The most representative features are then selected using step-wise linear discriminant analysis, while classification is done through the Gaussian mixture model.

Taormina - The method proposed by Cascio et al. (2016) is based on the adoption of 15 different pattern recognition sys-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

tems where each system is designed and optimized ad hoc for the discrimination between two specific HEp-2 cells staining patterns (i.e. homogeneous versus speckled, homogeneous versus nucleolar, etc), according to the one-versus-one approach. Each pattern recognition system is organized in three stages: preprocessing, feature extraction, classification. Several combinations of different types of preprocessing are analyzed (as contrast normalization, equalization, median and Gaussian filtering, morphological operators, etc) and for each class it is identified the preprocessing giving the best performance in terms of final class accuracy. For image description four classes of features are considered, respectively, based on intensity, geometry, morphology and entropy descriptors (specifically entropy of the histogram of oriented gradients and of amplitude gradients). From each cell image, all the features were densely extracted from image patches and from the circular crown obtained by mask dilation, using for both images four quantization levels. An SVM with radial basis function kernel is used for each of the 15 binary classifiers; final class label is attributed by majority voting.

Theodorakopoulos - In Theodorakopoulos et al. (2014) median filtering is first applied to eliminate isolated intensity values, then followed by pixel values normalization. Successively, a further noise reduction step is performed: the image is divided in patches that are encoded through sparse representation using a dictionary learned on low-noise images. Then the images are reconstructed and normalized again. Images are classified using a multi class SVM with linear kernel fed with a vector including both morphological and textural features as in Theodorakopoulos et al. (2012). In order to extract the former, the image is thresholded using 14 equally spaced values. From the connected components extracted from each of the so obtained binary images the following set of features is calculated: number of detected objects, density in binary image and mean objects solidity, complexity of cells contour. Texture is represented through both the distribution of the SIFT descriptor, calculated on a dense grid over the cell image and encoded according to the VLAD scheme as in Kastaniotis et al. (2013), and the histograms of three variants of local binary patterns, i.e. the uniform, the four patches and the center-symmetric ones.

3. Perfomance evaluation

In this section we overview the results achieved by the different methods participating to the Performance Evaluation of Indirect Immunofluorescence Image Analysis Systems competition held at ICPR2014 and to the Executable Thematic Special Issue of Pattern Recognition Letters on Pattern Recognition Techniques for Indirect Immunofluorescence Images Analysis. In particular, in Subsection 3.1 we detail the two pattern recognition tasks the participants were given; Subsection 3.2 describes the datasets used for the experimentations and finally in Subsections 3.3 and 3.4 we report and analyze the results over both tasks.

3.1. The classification tasks

Two different image classification tasks have been proposed to the participants. Both tasks refer to the automatic recognition of the staining pattern of the HEp-2 cells, although in the first case the samples to be classified are the single cells, while in the second case the classifiers are requested to operate on the whole specimen which typically contains dozens to hundreds HEp-2 cells. In the following we will refer to the first problem as *Task 1* or *cell level classification* and to the second problem as *Task 2* or *specimen level classification*.

Pattern	Abbr.	Task 1		Task 2	
		Train	Test	Train	Test
Homogeneous	H	2.494	10.611	53	157
Speckled	S	2.831	10.667	52	158
Nucleolar	N	2.598	10.187	50	150
Centromere	C	2.741	10.392	51	152
Golgi	G	724	3.041	10	26
Nuclear Membrane	M	2.208	9.935	21	62
Mitotic Spindle	MS	-	-	15	44
TOTAL		13.596	54.833	252	749

Table 1: Composition of the datasets used for the benchmarking expressed in terms of the number of cell images (for Task 1) and specimen images (for Task 2) for each class for the train and test sets.

Formally speaking, for both tasks the participants were required to develop a classifier C able to recognize the class of a given image Y, belonging to the true class y. Thus, given a set S of images, each one provided with the original image I, the corresponding binary segmentation mask image M and the fluorescence intensity level $\delta \in \{intermediate, positive\}$

$$S = \{(I, M, \delta)_1, ..., (I, M, \delta)_{|S|}\}, \quad (1)$$

the aim of the classifier is to predict the class $\hat{y}$ to which Y belongs:

$$C : Y \times G \mapsto \hat{y}, \quad (2)$$

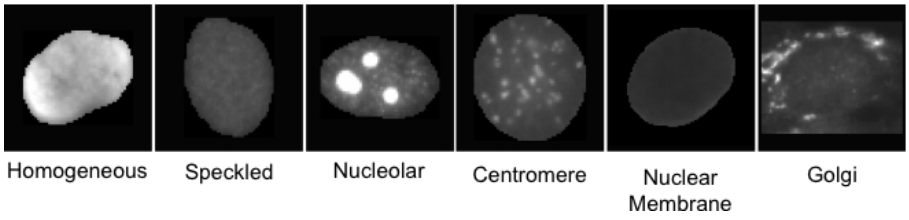
where ideally $\hat{y} = y$, with the difference that in the Task 1 the generic input image I is a single cell, while for the Task 2 it is the whole specimen image (see Figure 1 for some example input images of the two tasks).

3.2. The datasets

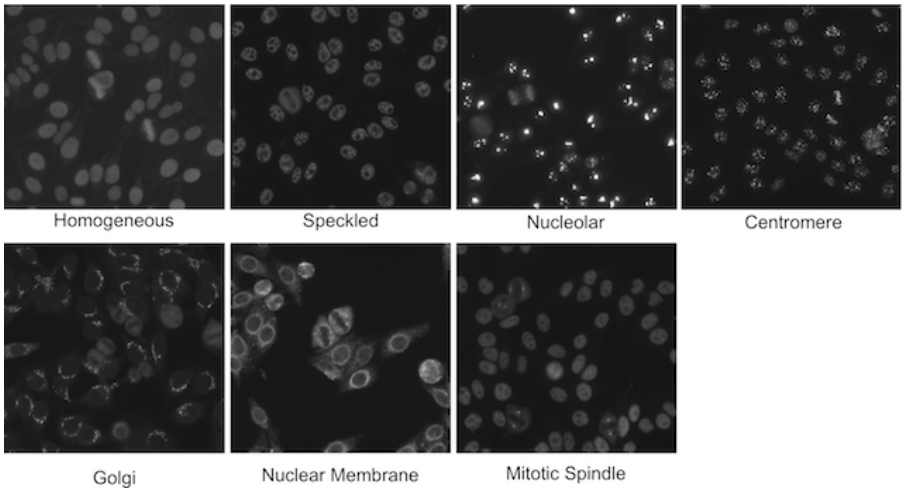
Two datasets have been made available to the participants, one for each task[1]. The dataset for Task 1 has been collected at Sullivan Nicolaides Pathology (SNP) laboratory from 2011 to 2013 by evaluating 419 patient positive sera. Each serum was prepared on the 18-well slide of HEp-2 IIF assay from Immuno Concepts N.A. Ltd. using a screening dilution of 1:80 and the related images were automatically acquired by a monochrome high dynamic range cooled microscopy camera, fitted on a microscope with a plan-Apochromat 20x/0.8 objective lens and an LED illumination source. For each patient serum, a number of cell images ranging from 100 to 200 was extracted, so resulting

[1]The train sets of both datasets are publicly available and can be downloaded by registering at the following URL: http://i3a2014.unisa.it/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters



(a) Task 1



(b) Task 2

Fig. 1: (a) Example cell images for the six staining patterns from the dataset of the Task 1. (b) Example specimen images for the seven patterns from the dataset of the Task 2.

in 68.429 images composing this dataset: approximately 20% of the dataset was used for the train set (in particular, 13.596 images) and the remaining 80% for the test set (54.833 images). The test set has not been released by the organizers during the benchmarking, so as to guarantee a fair comparison among the participants in the competition.

For each image contained in the dataset, the following information has been provided:

- cell staining pattern, among the six included in the dataset (homogeneous, speckled, nucleolar, centromere, nuclear membrane and mitotic spindle);
- cell mask, automatically obtained by segmenting the specimen using the DAPI channel;
- cell fluorescence intensity level, namely positive or intermediate;
- identifier of the specimen image to which the cell belongs to.

In order to attribute the staining pattern and the intensity of the cells used as the ground truth, each specimen was analyzed under a microscope by at least two scientists. In case of a disagreement between the two opinions, a third expert was involved in the specimen labeling procedure. Then, all the cells were labeled according to the class of the specimen they belong to. Furthermore, secondary tests, such as ENA or anti-ds-DNA have been employed so as to confirm the presence of the pattern identified by the scientists.

The dataset for Task 2 comprised by 1.001 images acquired in 2013 at SNP, using the same tools already used for the acquisition of the images of Task 1: approximately 25% are used for training (in particular, 252 images) while the remaining 75% for testing (749 images). Each specimen was photographed in four different locations rendering, so resulting in four different images.

For each specimen image, the following information has been provided:

- specimen pattern, among the seven included in the dataset (homogeneous, speckled, nucleolar, centromere, golgi, nuclear membrane and mitotic spindle);
- cell mask, automatically obtained by segmenting the specimen using the DAPI channel; it is important to highlight that the mask does not contain only the cells in interphase, but also those cells in mitosis as well as artifact potentially present in the specimen (such as air bubbles);
- cell intensity level, namely positive or intermediate.

More detailed information about the dataset for Task 2 can be found in Hobson et al. (2014).

The composition of the datasets is detailed in Table 1, and some example images are shown in Figure 1. The number of cells for each class is variable, ranging from 3.765 cells exhibiting the Golgi pattern to 13.498 images with speckled staining

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

Method	Reference	Task1		Task2	
		Train set	Test set	Train set	Test set
Codrescu	Codrescu (2014)	98.94	74.60	-	-
Gao	Gao et al. (2014)	99.95	83.23	-	-
Liu	Lovell et al. (2014)	-	-	97.14	86.09
Manivannan	Manivannan et al. (2014a) Manivannan et al. (2014b)	98.04 (98.93)	84.24 (**87.10**)	100.00	**88.48**
Paisitkriangkrai	Lovell et al. (2014)	100.00	81.55	99.05	79.26
Roberts	Lovell et al. (2014)	72.99	67.00	-	-
Theodorakopoulos	Theodorakopoulos et al. (2014)	93.82	83.33	-	-
Ensafi	Ensafi et al. (2016)	98.07	80.82	100.00	80.77
Gragnaniello	Gragnaniello et al. (2016)	95.47	83.64	100.00	87.07
Han	Han et al. (2016)	94.68	64.30	-	-
Paci	Nanni et al. (2016)	100.00	79.85	-	-
Ponomarev	Ponomarev and Kazanov (2016)	95.85	75.46	75.14	74.29
Qi	Qi et al. (2016)	99.91	**84.63**	-	-
Sarrafzadeh	Sarrafzadeh et al. (2016)	88.59	73.33	-	-
Taormina	Cascio et al. (2016)	90.25	80.12	82.11	59.78

Table 2: Results expressed in terms of MCA achieved by the participants to both the ICPR2014 contest (first seven rows of the table) and this journal special issue (remaining eight rows of the table) over the Tasks 1 and 2 .

witihn the dataset of Task 1 and from 36 specimen images belonging to Golgi class to the 210 images of homogeneous and speckled patterns. The skewness of the class distributions of the two datasets resembles the real clinical evidence with some very common patterns (such as homogenous, speckled, centromere, nucleolar) and rare patterns (such as Golgi and mitotic spindle). The train and the test sets for the two tasks preserve the overall class distribution of the respective original datasets.

3.3. Experimental results

The performance index adopted for scoring and comparing the methods is the Mean Class Cccuracy (MCA). For each class k, we compute the Ccorrect Classification Rate (CCR_k), as:

$$CCR_k = \frac{1}{N_k}(TP_k + TN_k) \tag{3}$$

being TP_k and TN_k the number of true positive and true negatives for the class k, respectively, and N_k the number of images belonging to the kth class. Thus, the MCA is computed as the average value of CCR_k:

$$MCA = \frac{1}{K}\sum_{k=1}^{K} CCR_k. \tag{4}$$

The results achieved by the participants are shown in Table 2 for both Tasks 1 and 2. All the results reported in the table are obtained by the participating research teams using the same experimental protocol, i.e. by training their pattern recognition system on the public train set of the respective task, and then providing their executable to the organizers of the two benchmarking initiatives, who independently tested such systems on the test sets. It is worth pointing out that there was an exception, represented by the contribution from Mannivannan; in fact for the Task 1, they provided two systems: the former trained using only images from the respective train set, while for the second submission they augmented the train set with images extracted from the Task 2 dataset. We decide to report both results in Table 2, namely $MCA = 84.24\%$ using the basic train set and $MCA = 87.10\%$ using the augmented set; in fact, the first result is useful to make a fair comparison of all the methods allowing us to make considerations over the proposed approaches; the second one draws attention over the role of the train set size for the specific classification problem.

Furthermore, in order to assess the statistical significance of the achieved performance, we applied the Student's t-tests over all the couples of methods. This analysis has shown that each improvement is significant with a probability of 100%.

Finally, we report the confusion matrix of each method, in Tables 3-15 and 17-23 for Task 1 and Task 2, respectively. For each confusion matrix, the generic element $c(i, j)$ indicates the number of samples belonging to the class i and classified as belonging to the class j.

3.4. Discussion

We note that all methods achieve highest accuracy over the centromere, nucleolar and nuclear membrane classes, and in particular the centromere class results to be the most accurate one in Task 1 for all the methods. Furthermore, we can note that the reduced number of training samples of the Golgi class has a very negative impact over the accuracy of some methods, namely Paci (Table 9), Sarrafzadeh (Table 14), Han (Table 7) and Roberts (Table 13), where the accuracy for this pattern is lower than 50%. Conversely, the remaining methods are able to cope with the limited number of samples achieving a good generalization that guarantees an accuracy over such class comparable to that obtained on more common classes. It is also interesting to note that, independently of the features used and classifier, the cells belonging to the class Golgi are typically

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

confused as belonging to the nuclear membrane and nucleolar classes, while the errors over the speckled cells are typically due to confusion with the centromere and homogeneous classes. We also note that homogeneous samples are often misclassified as nuclear membrane or speckled. This common behavior is mainly due to the high similarity, both in terms of shape and texture, among such classes.

In the following we analyze the behavior described above in light of the design choices made by the participants for the different stages of their method, namely the low level representation, the high level representation and the classification methods.

Low level representation. The authors explored a very wide range of different low level representations, both based on local and global descriptors. Although it is not possible to find out in absolute terms the best descriptor for HEp-2 cell classification, some general findings can be delineated. The discriminant between the methods achieving very promising performance and the others lies in the adoption of descriptors which are both invariant to scale and rotation: to this concern we can refer to the SIFT (or SURF)-based approach by Ensafi and Manivannan as well as to the SID descriptors proposed by Gragnaniello. We also note that when these properties are not intrinsically guaranteed by the descriptor, a common approach is to extract the descriptor not only on the original image, but also over a set of derived images (rotated and/or scaled), as in Gao, Paci and Qi. It is worth pointing out that often even those methods based on scale and rotation invariant descriptors prefer to evaluate the cells not only on the original image, but on a set of derived images as well, as in Manivannan.

The importance of accounting for the scale and the rotation invariance is also confirmed by the analysis of Qi et al. (2016), where the authors study the impact on the accuracy achieved by their method based on the LOAD descriptor while varying parameters related to scale and rotation showing that the *MCA* degrades up to 7% if these factors are not taken into account. Similarly, Gao et al. (2014) evaluated the performance of their method in two different situations, namely by adding rotated images to the train set and with no data augmentation. Interestingly, also in this case an improvement of 7% was observed, that further confirms the above observation.

High level representation. Most methods for HEp-2 cell classification typically do not perform classification using directly the extracted features; in fact, they often resort to different strategies for feature encoding or aggregation in order to derive a more compact and representative high level representation. As confirmed by Hobson et al. (2015), a typical approach is based on the bag of words paradigm. We also note that the introduction of techniques exploiting enhanced versions of bag of words, such as those based on Fisher vector (Qi), on the the soft assignment (Gragnaniello), as well as on the VLAD (Theodorakopoulos): these approaches appear very promising, scoring among the best four, due to their ability to enrich information conveyed in the high level representation with respect to traditional bag of words, thus allowing a better discrimination among the different classes.

Classification. The SVM is the most widely used classifier for HEp-2 image classification, as well as the most promising one; indeed, it is used by ten methods over fourteen analyzed in this review, including most of the methods with an accuracy higher than 80%.

Specimen level classification: from Task 1 to Task 2. Two main strategies can be identified for the classification of specimen images. The former is based on cell level classification: cells are first detected in the image and then classified independently from the others. Then, a majority voting strategy is applied, as in Ensafi, Ponomarev and Taormina. The second group of methods (composed by Liu, Manivannan, Paisitkriangkrai and Gragnaniello) partitions the image into a set of regions and extracts the features directly from each region, without previously detecting the single cells). The regions are typically obtained by using sliding patches densely extracted from the image, with the only exception of Gragnaniello, which partitions the image into five non overlapped regions.

It is interesting to note that in general the methods belonging to the second class outperform those belonging to the first class. This suggests that analyzing the image in the whole, without extracting the single cells, makes the system more robust with respect to the analysis of multiple cells, which may produce inconsistent results if analyzed with traditional cell recognition systems.

Where are we now? The experimental protocol adopted by the various HEp-2 image classification competitions with the test set privately held by the organizers had the merit of guaranteeing a true advancement of the state of the art and a fair comparison of the methods proposed so far. From Table 2, we can note that 57% of the methods reported in this review (8 over 14) achieve a mean class accuracy higher than 80% on the Task 1. This result is very encouraging, especially when it is compared with performance achieved by the methods participating in the previous HEp-2 Cells Classification contest organized in 2013 (Hobson et al. (2015)). Indeed, in that competition only 2 methods over 14 submissions achieved an *MCA* above 80% on the same dataset. The advancement in the state of the art in the last two years can be also noted by analyzing the accuracy of the top performing methods: 81.22% in 2013 vs 84.63% in 2015.

Nevertheless, it can be expected that in the next years it will become more and more difficult to significantly increase the accuracy with the pattern recognition systems adopted so far, by just using different texture and/or shape descriptors calculated either at a local or at a global level. We deem that the future adoption of deep architectures might be a viable way for doing a step forward, due to their ability to learn the discriminant properties of the cells without requiring any features engineering. Although deep learning methods have not been widely used for this task, it is worth mentioning the methods by the team of Malon et al. as reported in Foggia et al. (2013) and by Gao et al. (2014), both based on convolutional neural networks. On the one hand, the method by Malon reached an accuracy of 60% over a small dataset, the MIVIA HEp-2 Images Dataset (Foggia et al. (2013)), comprised approximatively 720 cells images for the training and 730 for the test set. Over the same dataset, the highest score was achieved by Nosaka (69%), while the scien-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

tist achieved an accuracy of approximatively 73%. On the other hand, the method by Gao, over a significantly larger dataset reported a promising accuracy (83%), which allows the method to score within the first five positions, only 1.5% less than the best performing method. The results achieved by Gao highlight the potential of deep architectures and suggests further investigation in this direction.

However, it is well known that the accuracy of this kind of method is strongly dependent on the amount of data available for training: the larger the size of the training set, the higher the possibility to better represent the data and thus to achieve good results. Of course, this is true even with traditional pattern recognition systems, as confirmed by the results achieved by Manivannan, with an improvement of 3% when the train dataset was augmented with data from an external dataset. The exploitation of new classification architecture, as those based on deep learning, and more importantly the availability of larger datasets might pave the way for significant advances in this area and the possibility of adopting such systems in real computer aided diagnosis.

4. Conclusions

The international benchmarking initiatives organized in the last years for the recognition of the staining pattern of the HEp-2 cells received significant attention from the scientific community with several dozens of research groups that proposed original methods. In this paper we have reported and analyzed the results achieved by 15 methods submitted to the last two initiatives of this series, namely the *Performance Evaluation of Indirect Immunofluorescence Image Analysis Systems* competition, hosted by the *International Conference on Pattern Recognition* 2014, and to the *Executable Thematic Special Issue of Pattern Recognition Letters on Pattern Recognition Techniques for Indirect Immunofluorescence Images Analysis*.

We had confirmation of results and also some new findings which are both summarized in the following in the form of the *key ingredients* that are shared by the best methods:

- modeling texture: best performing methods typically model features using various derivations of the basic LBP descriptor;
- rotation invariance: taking into account in the description stage the rotation invariance explicitily, by augmenting the train set with rotated versions of the patterns, or implicitily, by adopting a rotation invariant descriptor, may guarantee significant improvements with respect to the case when rotation invariance is not accounted for;
- feature encoding: the bag of words approach with its enhanced versions seems to guarantee a better discrimination among the different classes;
- SVM for classification: the support vector machine is the most commonly used and probably most effective tool for final classification;
- specimen level classification: methods that partition the whole specimen image in sub-regions and extract features on them typically outperforms approaches that simply aggregate the classification outputs obtained on the single cells.

We note a general performance improvement with respect to methods from previous initiatives confirming the fact that there was, and probably still is, room for improvements in accuracy. On one side we deem that this might be achieved by exploiting a larger dataset, as confirmed by some experiments done by a contest participant giving insights on this point. On the other side we notice that the deep learning strategies have not yet been sufficiently explored in this applicative area, thus we might expect new relevant results using such approaches. In our opinion the latter two points will guide next research efforts.

Acknowledgements

This work has been partly funded by Sullivan Nicolaides Pathology, Australia and the Australian Research Council (ARC) Linkage Projects Grant LP130100230.

References

Banerjee, A., Maji, P., 2013. Computer Analysis of Images and Patterns: 15th International Conference, CAIP 2013, York, UK, August 27-29, 2013, Proceedings, Part I. Springer Berlin Heidelberg, Berlin, Heidelberg. chapter Contraharmonic Mean Based Bias Field Correction in MR Images. pp. 523–530. doi:10.1007/978-3-642-40261-6_63.

Cascio, D., Taormina, V., Cipolla, M., Bruno, S., Fauci, F., Raso, G., 2016. A multi-process systems for HEp-2 cells classification based on svm. Pattern Recognition Letters this issue.

Codrescu, C., 2014. Quadratic recurrent finite impulse response mlp for indirect immunofluorescence image recognition, in: Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on, pp. 49–52. doi:10.1109/I3A.2014.14.

Ensafi, S., Lu, S., Kassim, A.A., Tan, C.L., 2016. Accurate HEp-2 cell classification based on sparse coding of superpixels. Pattern Recognition Letters this issue.

Foggia, P., Percannella, G., Saggese, A., Vento, M., 2014. Pattern recognition in stained HEp-2 cells: Where are we now? Pattern Recognition 47, 2305 – 2314. doi:http://dx.doi.org/10.1016/j.patcog.2014.01.010.

Foggia, P., Percannella, G., Soda, P., Vento, M., 2013. Benchmarking HEp-2 cells classification methods. Medical Imaging, IEEE Transactions on 32, 1878–1889. doi:10.1109/TMI.2013.2268163.

Gao, Z., Zhang, J., Zhou, L., Wang, L., 2014. HEp-2 cell image classification with convolutional neural networks, in: Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on, pp. 24–28. doi:10.1109/I3A.2014.15.

Gragnaniello, D., Sansone, C., Verdoliva, L., 2016. Cell image classification by a scale and rotation invariant dense local descriptor. Pattern Recognition Letters this issue.

Han, X.H., Chen, Y.W., Xu, G., 2016. Integration of spatial and orientation contexts in local ternary patterns for HEp-2 cell classification. Pattern Recognition Letters this issue.

Hobson, P., Lovell, B., Percannella, G., Vento, M., Wiliem, A., 2014. Classifying anti-nuclear antibodies HEp-2 images: A benchmarking platform, in: IEEE ICPR 2014, pp. 3233–3238. doi:10.1109/ICPR.2014.557.

Hobson, P., Lovell, B.C., Percannella, G., Saggese, A., Vento, M., Wiliem, A., 2016. Computer aided diagnosis for anti-nuclear antibodies HEp-2 images: Progress and challenges. Pattern Recognition Letters this issue.

Hobson, P., Lovell, B.C., Percannella, G., Vento, M., Wiliem, A., 2015. Benchmarking human epithelial type 2 interphase cells classification methods on a very large dataset. Artificial Intelligence in Medicine 65, 239 – 250. doi:http://dx.doi.org/10.1016/j.artmed.2015.08.001.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

	C	G	H	N	M	S
C	91.81	0.51	0.06	3.59	0.60	3.44
G	0.95	55.21	10.65	17.00	13.45	2.73
H	1.00	0.79	66.53	4.18	7.58	19.91
N	4.27	1.17	1.26	86.12	1.17	6.02
M	0.22	1.60	5.99	3.25	82.80	6.14
S	14.07	0.46	12.94	4.88	2.54	65.11

Table 3: Codrescu Task1: Confusion Matrix, MCA = 74.60

	C	G	H	N	M	S
C	94.44	0.69	0.90	1.91	0.64	1.40
G	0.36	70.80	5.13	9.04	12.76	1.91
H	0.38	0.72	74.04	5.00	7.84	12.03
N	2.51	1.85	1.57	90.18	2.01	1.87
M	0.14	0.60	7.25	0.61	89.41	1.98
S	16.54	0.95	11.33	2.84	2.30	66.04

Table 4: Ensafi Task1: Confusion Matrix, MCA = 80.82

	C	G	H	N	M	S
C	96.03	0.18	0.05	1.50	0.48	1.76
G	0.03	73.20	5.75	10.42	9.14	1.45
H	0.19	0.84	78.29	5.97	7.52	7.20
N	0.72	1.33	1.86	93.72	1.17	1.22
M	0.08	0.83	4.22	0.73	91.27	2.87
S	11.31	0.59	14.61	4.80	1.85	66.85

Table 5: Gao Task1: Confusion Matrix, MCA = 83.23

	C	G	H	N	M	S
C	95.52	0.42	0.21	1.15	0.05	2.66
G	0.03	71.82	4.74	7.27	14.60	1.55
H	0.05	0.80	78.57	4.94	8.07	7.58
N	0.75	1.58	1.96	92.55	1.70	1.46
M	0.05	0.76	3.14	0.85	93.39	1.81
S	13.35	0.71	11.11	2.65	2.17	70.01

Table 6: Gragnaniello Task 1: Confusion Matrix, MCA = 83.64

	C	G	H	N	M	S
C	92.73	1.95	0.17	1.05	0.14	3.95
G	0.16	31.57	12.89	26.60	25.91	2.86
H	0.92	1.38	65.71	3.09	20.45	8.44
N	1.17	10.17	3.84	70.92	7.09	6.81
M	0.67	4.05	14.02	5.60	69.55	6.11
S	16.39	0.97	18.84	2.61	5.89	55.30

Table 7: Han Task1: Confusion Matrix, MCA = 64.30

	C	G	H	N	M	S
C	96.75	0.09	0.51	1.15	0.19	1.31
G	0.20	72.94	3.68	6.12	16.18	0.89
H	0.17	0.60	77.33	7.40	6.40	8.10
N	0.75	0.35	1.51	94.25	1.79	1.35
M	0.06	0.52	6.13	0.92	91.07	1.30
S	11.09	0.49	10.75	2.57	2.02	73.09

Table 8: Manivannan Task 1: Confusion Matrix, MCA = 84.24

	C	G	H	N	M	S
C	95.17	0.02	0.90	2.16	0.09	1.66
G	0.62	46.20	10.65	14.17	27.95	0.39
H	0.07	0.05	75.89	4.21	11.53	8.26
N	0.76	0.40	1.96	94.25	2.09	0.54
M	0.26	0.17	3.30	0.60	95.03	0.63
S	11.88	0.09	11.31	2.04	2.14	72.54

Table 9: Paci Task1: Confusion Matrix, MCA = 79.85

	C	G	H	N	M	S
C	94.96	0.12	0.18	1.63	0.61	2.51
G	0.92	65.11	4.21	15.13	11.97	2.66
H	0.13	0.21	76.83	6.61	7.56	8.67
N	1.02	0.51	1.17	92.92	2.56	1.82
M	0.15	0.65	4.89	1.42	91.08	1.80
S	13.59	0.23	11.91	3.25	2.60	68.42

Table 10: Paisitkriangkrai Task1: Confusion Matrix, MCA = 81.55

	C	G	H	N	M	S
C	93.29	0.54	0.17	2.18	0.62	3.19
G	0.95	51.56	6.18	9.93	28.05	3.32
H	0.64	2.33	63.34	4.24	17.03	12.42
N	2.30	2.70	1.57	86.09	4.97	2.38
M	0.40	2.84	5.41	2.32	85.59	3.45
S	15.03	2.01	12.52	3.49	5.64	61.31

Table 11: Ponomarev Task 1: Confusion Matrix, MCA = 75.46

	C	G	H	N	M	S
C	96.95	0.14	0.16	1.13	0.17	1.44
G	0.20	70.73	9.08	9.08	9.83	1.09
H	0.23	0.46	78.94	4.67	7.18	8.52
N	0.87	0.47	1.55	94.05	1.74	1.32
M	0.09	0.30	5.28	0.69	91.49	2.13
S	7.74	0.44	12.70	1.94	1.57	75.60

Table 12: Qi Task 1: Confusion Matrix, MCA = 84.63

	C	G	H	N	M	S
C	90.70	1.41	0.53	4.21	0.39	2.75
G	0.85	49.49	8.25	11.48	25.49	4.44
H	0.99	3.10	58.53	6.16	15.48	15.73
N	2.84	7.21	2.15	72.89	7.63	7.29
M	1.17	10.47	7.03	2.86	76.50	1.98
S	11.30	2.71	15.56	11.74	4.83	53.87

Table 13: Roberts Task 1: Confusion Matrix, MCA = 67.00

	C	G	H	N	M	S
C	95.14	0.72	0.47	1.94	0.20	1.52
G	0.10	48.11	5.82	3.52	41.73	0.72
H	0.06	2.61	73.75	1.21	15.36	7.01
N	4.40	6.66	1.50	78.95	5.43	3.06
M	0.05	6.93	4.95	1.16	86.02	0.90
S	16.72	0.64	17.09	2.14	5.41	58.00

Table 14: Sarrafzadeh Task 1: Confusion Matrix, MCA = 73.33

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI:https://doi.org/10.1016/j.patrec.2016.07.013, Pattern Recognition Letters

	C	G	H	N	M	S
C	93.65	0.16	0.70	2.29	0.27	2.93
G	0.03	76.39	1.97	8.06	11.31	2.24
H	0.02	0.72	73.81	4.43	13.05	7.97
N	1.58	0.73	2.20	85.05	2.62	7.82
M	0.14	2.25	5.20	4.26	86.65	1.49
S	14.99	0.47	13.30	2.17	3.90	65.17

Table 15: Taormina Task 1: Confusion Matrix, MCA = 80.12

	C	G	H	N	M	S
C	94.74	0.25	1.31	1.68	0.15	1.87
G	0.30	71.03	5.03	7.53	15.65	0.46
H	0.00	0.98	74.31	3.36	13.21	8.14
N	0.84	0.92	1.60	92.85	2.24	1.54
M	0.17	1.46	3.64	1.11	91.99	1.63
S	8.18	0.59	12.16	1.69	2.30	75.08

Table 16: Theodorakopoulos Task 1: Confusion Matrix, MCA = 83.33

	C	G	H	MS	N	M	S
C	94.74	0.00	0.66	0.00	0.00	0.00	4.61
G	0.00	80.77	7.69	3.85	0.00	0.00	7.69
H	0.00	0.00	82.17	2.55	0.00	1.91	13.38
MS	0.00	4.55	36.36	43.18	4.55	6.82	4.55
N	0.00	0.00	0.67	0.00	98.67	0.67	0.00
M	0.00	1.61	11.29	0.00	0.00	85.48	1.61
S	5.06	0.00	14.56	0.00	0.00	0.00	80.38

Table 17: Ensafi Task 2: Confusion Matrix, MCA = 80.77

	C	G	H	MS	N	M	S
C	97.37	0.00	0.66	0.00	0.00	0.00	1.97
G	0.00	73.08	0.00	23.08	0.00	0.00	3.85
H	0.00	0.00	86.62	7.01	0.00	1.27	5.10
MS	0.00	0.00	20.45	75.00	0.00	2.27	2.27
N	0.00	0.00	0.00	1.33	98.00	0.00	0.67
M	0.00	0.00	1.61	0.00	0.00	98.39	0.00
S	0.63	0.00	16.46	1.90	0.00	0.00	81.01

Table 18: Gragnaniello Task 2: Confusion Matrix, MCA = 87.07

	C	G	H	MS	N	M	S
C	98.68	0.66	0.00	0.00	0.00	0.00	0.66
G	0.00	80.77	7.69	3.85	3.85	3.85	0.00
H	0.00	0.00	92.99	0.00	0.00	1.27	5.73
MS	0.00	2.27	34.09	52.27	2.27	6.82	2.27
N	0.00	0.00	0.67	0.67	98.00	0.67	0.00
M	0.00	0.00	6.45	0.00	0.00	91.94	1.61
S	0.00	0.00	11.39	0.63	0.00	0.00	87.97

Table 19: Liu Task 2: Confusion Matrix, MCA = 86.09

	C	G	H	MS	N	M	S
C	98.68	0.00	0.00	0.66	0.00	0.00	0.66
G	0.00	80.77	0.00	3.85	3.85	11.54	0.00
H	0.00	0.00	92.99	0.00	0.00	1.91	5.10
MS	0.00	0.00	18.18	61.36	2.27	11.36	6.82
N	0.00	1.33	0.67	0.67	96.00	0.67	0.67
M	0.00	0.00	1.61	0.00	0.00	98.39	0.00
S	0.00	0.00	6.96	0.63	0.00	1.27	91.14

Table 20: Manivannan Task 2: Confusion Matrix, MCA = 88.48

	C	G	H	MS	N	M	S
C	98.68	0.00	0.00	0.66	0.00	0.00	0.66
G	0.00	80.77	0.00	3.85	3.85	11.54	0.00
H	0.00	0.00	92.99	0.00	0.00	1.91	5.10
MS	0.00	0.00	18.18	61.36	2.27	11.36	6.82
N	0.00	1.33	0.67	0.67	96.00	0.67	0.67
M	0.00	0.00	1.61	0.00	0.00	98.39	0.00
S	0.00	0.00	6.96	0.63	0.00	1.27	91.14

Table 21: Paisitkriangkrai Task 2: Confusion Matrix, MCA = 87.07

	C	G	H	MS	N	M	S
C	96.05	0.00	0.00	0.00	0.00	0.00	3.95
G	0.00	80.77	0.00	0.00	0.00	19.23	0.00
H	0.00	0.00	74.52	0.00	0.00	5.73	19.75
MS	0.00	0.00	34.09	0.00	4.55	59.09	2.27
N	0.00	1.33	0.00	0.00	94.67	2.00	2.00
M	0.00	0.00	3.23	0.00	0.00	96.77	0.00
S	8.86	0.00	12.03	0.00	0.00	1.90	77.22

Table 22: Ponomarev Task 2: Confusion Matrix, MCA = 74.29

	C	G	H	MS	N	M	S
C	94.08	0.00	0.00	0.00	1.97	0.00	3.95
G	0.00	38.46	15.38	19.23	11.54	11.54	3.85
H	0.00	0.00	87.26	0.64	0.00	0.64	11.46
MS	0.00	4.55	38.64	20.45	4.55	25.00	6.82
N	4.00	1.33	1.33	0.67	80.00	2.67	10.00
M	0.00	0.00	41.94	4.84	12.90	38.71	1.61
S	9.49	0.00	26.58	0.00	4.43	0.00	59.49

Table 23: Taormina Task 2: Confusion Matrix, MCA = 59.78

Jegou, H., Perronnin, F., Douze, M., Sanchez, J., Perez, P., Schmid, C., 2012. Aggregating local image descriptors into compact codes. IEEE Trans. Pattern Anal. Mach. Intell. 34, 1704–1716. doi:10.1109/TPAMI.2011.235.

Kastaniotis, D., Theodorakopoulos, I., Economou, G., Fotopoulos, S., 2013. HEp-2 cells classification using locally aggregated features mapped in the dissimilarity space, in: Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on, pp. 1–4. doi:10.1109/BIBE.2013.6701591.

Kokkinos, I., Bronstein, M., Yuille, A., 2012. Dense Scale Invariant Descriptors for Images and Surfaces. Research Report RR-7914. INRIA. URL: https://hal.inria.fr/hal-00682775.

Liu, L., Zhao, L., Long, Y., Kuang, G., Fieguth, P., 2012. Extended local binary patterns for texture classification. Image and Vision Computing 30, 86 – 99. doi:http://dx.doi.org/10.1016/j.imavis.2012.01.001.

Lovell, B.C., Percannella, G., Vento, M., Wiliem, A., 2014. Performance evaluation of indirect immunofluorescence image analysis systems, in: Report on the contest. URL: http://i3a2014.unisa.it/.

Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., McKenna, S., 2014a. HEp-2 cell classification using multi-resolution local patterns and ensemble svms, in: Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on, pp. 37–40. doi:10.1109/I3A.2014.18.

Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., McKenna, S., 2014b. HEp-2 specimen classification using multi-resolution local patterns and svm, in: Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on, pp. 41–44. doi:10.1109/I3A.2014.20.

Meroni, P.L., Schur, P.H., 2010. ANA screening: an old test with new recommendations. Annals of the Rheumatic Diseases 69, 1420 –1422.

Nanni, L., Lumini, A., dos Santos, F.L.C., Paci, M., Hyttinen, J., 2016. Ensembles of dense and dense sampling descriptors for the HEp-2 cells classification problem. Pattern Recognition Letters this issue.

Nosaka, R., Fukui, K., 2014. HEp-2 cell classification using rotation invariant co-occurrence among local binary patterns. Pattern Recognition 47, 2428–2436. doi:10.1016/j.patcog.2013.09.018.

Nosaka, R., Ohkawa, Y., Fukui, K., 2012. Feature extraction based on co-occurrence of adjacent local binary patterns, in: Ho, Y.S. (Ed.), Advances in Image and Video Technology. Springer Berlin Heidelberg. volume 7088 of *Lecture Notes in Computer Science*, pp. 82–91. doi:10.1007/978-3-642-25346-1_8.

Paisitkriangkrai, S., Shen, C., van den Hengel, A., 2014. A scalable stagewise approach to large-margin multiclass loss-based boosting. Neural Networks and Learning Systems, IEEE Transactions on 25, 1002–1013. doi:10.1109/TNNLS.2013.2282369.

Perronnin, F., Sánchez, J., Mensink, T., 2010. Improving the fisher kernel for large-scale image classification, in: Proceedings of the 11th European Conference on Computer Vision: Part IV, Springer-Verlag, Berlin, Heidelberg. pp. 143–156.

Ponomarev, G.V., Kazanov, M.D., 2016. Classification of ana HEp-2 slide images using morphological features of stained patterns. Pattern Recognition Letters this issue.

Qi, X., Zhao, G., Chen, J., Pietikainen, M., 2016. HEp-2 cell classification: The role of gaussian scale space theory as a pre-processing approach. Pattern Recognition Letters this issue.

Qi, X., Zhao, G., Shen, L., Li, Q., Pietikäinen, M., 2015. LOAD: local orientation adaptive descriptor for texture and material classification. CoRR abs/1504.05809. URL: http://arxiv.org/abs/1504.05809.

Qian, X., Hua, X.S., Chen, P., Ke, L., 2011. Plbp: An effective local binary patterns texture descriptor with pyramid representation. Pattern Recognition 44, 2502 – 2515. doi:http://dx.doi.org/10.1016/j.patcog.2011.03.029.

Sarrafzadeh, O., Rabbani, H., Dehnavi, A.M., 2016. Analyzing features by swlda for the classification of HEp-2 cell images using gmm. Pattern Recognition Letters this issue.

Strandmark, P., Ulen, J., Kahl, F., 2012. HEp-2 staining pattern classification, in: Pattern Recognition (ICPR), 2012 21st International Conference on, pp. 33–36.

Theodorakopoulos, I., Kastaniotis, D., Economou, G., Fotopoulos, S., 2012. HEp-2 cells classification via fusion of morphological and textural features, in: Bioinformatics Bioengineering (BIBE), 2012 IEEE 12th International Conference on, pp. 689–694. doi:10.1109/BIBE.2012.6399750.

Theodorakopoulos, I., Kastaniotis, D., Economou, G., Fotopoulos, S., 2014. HEp-2 cells classification using morphological features and a bundle of local gradient descriptors, in: Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on, pp. 33–36. doi:10.1109/I3A.2014.16.

Wiik, A.S., Hier-Madsen, M., Forslid, J., Charles, P., Meyrowitsch, J., . Antinuclear antibodies: A contemporary nomenclature using HEp-2 cells. Journal of Autoimmunity In Press, Corrected Proof.

Wiliem, A., Wong, Y., Sanderson, C., Hobson, P., Chen, S., Lovell, B., 2013. Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors, in: Applications of Computer Vision (WACV), 2013 IEEE Workshop on, pp. 95–102. doi:10.1109/WACV.2013.6475005.

Yimo Guo, G.Z., Pietikinen, M., 2011. Texture classification using a linear configuration model based descriptor, in: Proceedings of the British Machine Vision Conference, BMVA Press. pp. 119.1–119.10. Http://dx.doi.org/10.5244/C.25.119.

Zuiderveld, K., 1994. Graphics gems iv, Academic Press Professional, Inc., San Diego, CA, USA. chapter Contrast Limited Adaptive Histogram Equalization, pp. 474–485.