HEp-2 Cell Classification with heterogeneous classes-processes based on K-Nearest Neighbours

Cascio Donato[1], Cipolla Marco[1], Fauci Francesco[1], and Raso Giuseppe[1]

1 Dipartimento di Fisica e Chimica, Università Degli Studi di Palermo, Palermo, Italy
Palermo, Italy
e-mail: donato.cascio@unipa.it.com

Taormina Vincenzo[2], Vasile Simone Maria[2]

CyclopusCAD srl, Palermo, Italy
Palermo, Italy

e-mail: taormina.maltese@gmail.com

Abstract— We present a scheme for the feature extraction and classification of the fluorescence staining patterns of HEp-2 cells in IIF images. We propose a set of complementary processes specific to each class of patterns to search. Our set of processes consists of preprocessing, features extraction and classification. The choice of methods, features and parameters was performed automatically, using the Mean Class Accuracy (MCA) as a figure of merit. We extract a large number (108) of features able to fully characterize the staining pattern of HEp-2 cells. We propose a classification approach based on two steps: the first step follows the one-against-all (OAA) scheme, while the second step follows the one-against-one (OAO) scheme. To do this, we needed to implement 21 KNN classifiers: 6 OAA and 15 OAO. Leave-one-out image cross validation method was used for the evaluation of the results.

Keywords—IIF images, K–Nearest-Neighbors (K-NN), multi-class, classification, one-against-all classification, leave-one out cross validation.

I. INTRODUCTION

Indirect ImmunoFluorescence (IIF) slides are considereda powerful technique for medical diagnosis.Such slides are typically examined by pathologists,however, due to the difficulty of the task, a ComputerAided Design system is desirable [1,2]. Cad systems may use several automatic classification techniques to recognize the type of IIF patterns contained into an image. Classification techniques require the location of the image regions which are considered interesting, such region are typically denoted as ROIs. Each ROI is characterized by several features based on the texture of the ROIs and on their morphological properties. A study about the features space is made and classifiers are tested to distinguish the patterns contained in a set of IIF images. In this paper we report the results obtained with a structured classification process of classification based on K-Nearest Neighbors classifier (KNN). The basic intuition behind our approach is that, instead of using a single process to discriminate each class from all other classes, it is better to combine a set of different and complementary processes. The processes were based on the KNN classifier. The main advantages of such classifier are: fast training and low sensibility to the choice of

the training parameters. We discuss details of our approach in the following sections.

II. DATASET

The development of a CAD system is intimately linked to collection of a dataset of selected images [3]. For this study we used the dataset provided for the participation to the "*Contest on Performance Evaluation on Indirect Immunofluorescence Image Analysis Systems*", hosted by the 22th International Conference on Pattern Recognition (ICPR 2014); the dataset can be downloaded after proper registration, on site http://i3a2014.unisa.it/. The participants are required to face exactly the same task of the ICIP 2013 Contest by using the same training and testing dataset. The goal is the design and implementation of a IIF pattern recognition system able to classify the cells belonging to HEp-2 images in one of the following pattern classes: homogeneous, speckled, nucleolar, centromere, golgi and nuclear membrane.The dataset has been collected between 2011 and 2013 at Sullivan Nicolaides Pathology laboratory, Australia. It utilizes 419 patient positive sera, which were prepared on the 18-well slide of HEP-2000 IIF assay from Immuno Concepts N.A. Ltd. with screening dilution 1:80. The specimens were then automatically photographed using a monochrome high dynamic range cooled microscopy camera which was fitted on a microscope with a plan-Apochromat 20x/0.8 objective lens and an LED illumination source. Approximately 100-200 cell images were extracted from each patient serum. In total there were 13,596 cell images used for training. The labeling process involved at least two scientists who read each patient specimen under a microscope. A third expert's opinion was sought to adjudicate any discrepancy between the two opinions. We used each specimen label for the ground-truth of cells extracted from it. Furthermore, all the labels were validated by using secondary tests such as ENA and anti-ds-DNA in order to confirm the presence and/absence of specific patterns. Each cell image contained in the database is annotated with the following information: cell pattern (one of the patterns defined above), cell intensity, cell mask, ID of the image which the cell belongs to. A characterization of the dataset is reported in Figure 1.
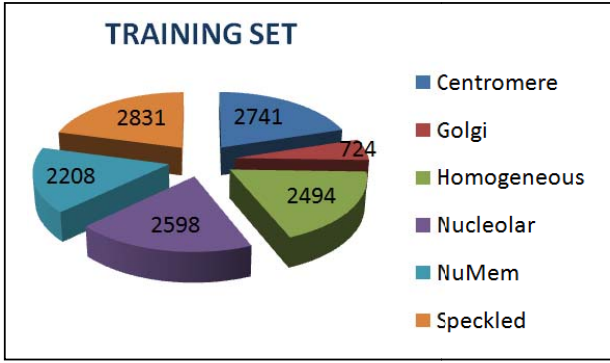
Fig. 1. Distribution of training set patterns.

III. METHOD

In several multiclass classification problems, it is preferable to use a number of classifier equal to the number of classes and each classifier is trained to discriminate a class from all the others (binary approach) [4]. In this work, in addition to differentiate the classification stage by implementing six classifiers for six classes of staining patterns, the preprocessing and feature extraction steps are differentiated, too. We adopt, then, a non-standard pipeline for supervised image classification. Figure 2 shows the flow of operations adopted in this work. The generic image is simultaneously processed by six processes obtaining six separate outputs that represent how the cell resembles each one of the 6 classes analyzed in this work. The choice of methods, features and parameters was performed automatically, using the Mean Class Accuracy (MCA) as a figure of merit. The parameters involved are tuned using a cross validation scheme. The main benefit of this pipeline is that it has a good explanatory power, while its principle is simple to explain.

Fig. 2. Pipeline of developed method.
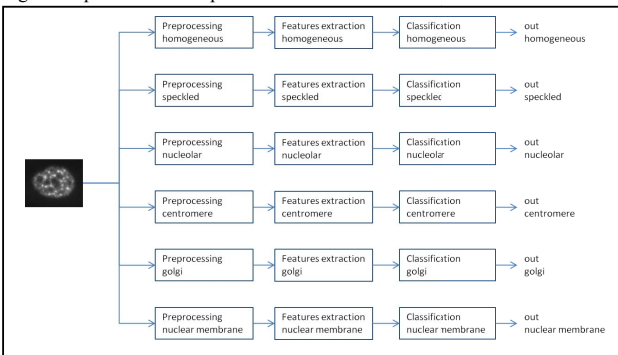
IV. PREPROCESSING

It is important that the features information is independent from variations of staining intensity, for this reason the contrast of all the images in normalized. Contrast normalization was obtained by linearly remapping the intensity values so that 1% of data is saturated at low and high intensities. For this purpose, 14 different types of preprocessing have been analyzed and for each class it has been identified the preprocessing giving the best performance in terms of final class accuracy. Table 1 shows the details of combinations of preprocessing analyzed and their possible use. The choice of the best preprocessing for a given class of staining patterns has been carried out to maximize the MCA.

TABLE I. LIST OF PREPROCCESING ANALYZED AND THEIR USE

	First function Filtering	Second Function Contrast normalization	Used for pattern:
1	None	None	---
2	None	Contrast normalization	homogeneos
3	Dilation 5x5	None	---
4	Dilation 5x5	Contrast normalization	centromere
5	Dilation 9x9	Contrast normalization	nucleolar
6	Erosion 5x5	None	---
7	Erosion 5x5	Contrast normalization	---
8	Erosion 9x9	Contrast normalization	---
9	Median 5x5	None	---
10	Median 5x5	Contrast normalization	---
11	Median 9x9	Contrast normalization	---
12	Gaussian 5x5	None	---
13	Gaussian 5x5	Contrast normalization	golgi, nuclear menbrane
14	Gaussian 9x9	Contrast normalization	speckled

V. FEATURES EXTRACTION

During the analysis of cells patterns, it appears clearly that differences between classes are mainly based based mainly on the presence and distribution of bright/dark structures: numbers, sizes (areas), intensity and colors. It seems natural to use features that provide an analysis of texture and more specifically features able to deal with bright/dark speckle-like structure description.

Different staining patterns can be characterized by a limited set of attributes describing the spatial relationships between pixels values and the main image variations occurring in each cell type; this information is generally obtained by means of textural analysis techniques. These techniques can be grouped into two major categories: (i) statistical methods describing the distribution of grey-levels in the image; and (ii) frequency domain measurements of image variations [5,6]. To achieve

the objective of robust classification, we combine several discriminative visual features known to be effective for cell classifications with a robust and scalable multi-class boosting. In our work we extract a large number (108) of features able to fully characterize the staining pattern of HEp-2 cells.

More precisely, were analyzed four quantization levels (256, 128, 64 and 32) [7], and for each one of them the following 27 features have been extracted [8-10]:

features Intensity based (9): mean, standard deviation , ratio of the standard deviation to the mean, entropy, moment of inertia, skewness, kurtosis, entropy of the contours gradient;

- **Geometry-based features (8):** mean radius, standard deviation of radius, maximum radius, ratio of the standard deviation to the mean, circularity, anisotropy, fractal index, eccentricity;

- **Shape-Morphological-based features (8)**: area, perimeter, convex area, convex deficiency, solidity, compactness, roundness, Euler's number;

- **Descriptors-based features (2)**: entropy of HOG (Histogram of Oriented Gradients), entropy of HAG (Histogram of Amplitude Gradients).

VI. FEATURES REDUCTION

Reducing the dimensionality of the data by selecting a subset of the original variables may be advantageous for simplifying the classification problem from different points of view. As the dimensionality of the data increases, many types of data analysis and classification problems become significantly harder. Sometimes the data also becomes increasingly sparse in the feature space. This can lead to big problems for both supervised and unsupervised learning. The main idea of feature subset selection is to remove redundant or irrelevant features from the data set as they can negatively influence the classification accuracy, the clustering quality and lead to an unnecessary increase of computational cost. The advantage of identifying a subset of features is that no information about the importance of single features is lost. Filter methods perform feature selection in two steps. In the first step, the filter method assesses each feature individually for its potential in discriminating among classes in the data. In the second step, features falling beyond some thresholding criterion are eliminated, and the smaller set of remaining features is used. This score-and-filter approach has been used in many recent publications, due to its relative simplicity. Scoring methods generally focuson measuring the differences between distributions of features. The resulting score is intended to reject the quality of each feature in terms of its discriminative power. Many scoring criteria exist, in this work we have adopted the criterion of statistical selection of Fischer (equation 1).

$$V(i) = \frac{\left(\mu_1(i) - \mu_2(i)\right)^2}{\sigma_1^2 + \sigma_2^2} \tag{1}$$

By this criterion, the quality of each i-th feature is expressed in terms of the difference among theempirical means of two distributions $\left(\mu_1(i)\ and\ \mu_2(i)\right)$, normalized by the sum of their variances ($\sigma_1\ and\ \sigma_2$).

Our choice was to sort the extracted features by using the Fisher values and select only the 15 highest values for each pattern class; each pattern class will be described by a set of 15 features (out of 108) where such set may be different among the classes.

VII. CLASSIFICATION

This approach is computationally expensive and it requires several processing steps,, thus a classifier with a simple training process is needed. Therefore in this work we decide to use the K–Nearest-Neighbors (KNN). This classifier allows the multi-class classification, however, in order to simplify the global problem and then improve the results, it was decided to use binary KNN (one per class); the conventional way to deal with a multi-class problem is to decompose an M-class problem into a series of two-class problems. The two approaches commonly used are the One-Against-One (OAO) and One-Against-All (OAA) techniques. The OAA approach represents the earliest and most commonly used multiclass approach and involves the division of an M class dataset into M two-class cases; M classifiers with the i-th one separating class i from all the remaining classes. In order to assign a class to a new sample, it is necessary to evaluate the output for all binary classifiers and, usually, choosing the pattern relative to the classifier that returned the highest classification value. The problem with the participating to the decision is assumed to be equally reliable, which is rarely the case.

The OAO approach on the other hand involves constructing a classifier for each pair of classes resulting in $M(M-1)/2$ classifier; the OAO approach is more computationally intensive. When applied to a test point, each classification gives one vote to the winning class and the point is labeled with the class having most votes. One drawback of one-against-one method, however, arises is that when the results from the multiple classifiers are counted for the final decision without considering the competence of the classifiers. To overcome this limitation, we propose in this paper an approach based on two steps of classification: the first step employs the OAA approach, while the second (and final) step uses the OAO approach. The objective of the second step is to correct any errors that occurred in the first step. In the first step, the system receives an image of the i-th roi test which it is processed and classified by six OAA classifiers (Figure 1). The two classifiers OAA which gave the first two highest values in the output identify the two classes of patterns that will be compared using a classifier OAO (one of 15 implemented), which returns the final result of the classification. To do this, we needed to implement 21 KNN classifiers: 6 one-against-all and 15 one-against-one. Table 2 shows the details of combinations of preprocessing used for the OAO classifiers. As can easily be seen from Table 2, the preprocessing obtained with 5x5 Gaussian filter and contrast normalization has been associated with most types of classifiers OAO.

TABLE II.   LIST OF PREPROCCESING ANALYZED AND THEIR USE IN OAO CLASSIFIERS

	First function Filtering	Second Function Contrast normalization	Used for OAO:
1	None	None	---
2	None	Contrast normalization	(Homog-Centr) (Speac - Centr)
3	Dilation 5x5	None	---
4	Dilation 5x5	Contrast normalization	---
5	Dilation 9x9	Contrast normalization	(Homog -Nucle) (Homog - Golgi) (Homog -Membr)
6	Erosion 5x5	None	---
7	Erosion 5x5	Contrast normalization	---
8	Erosion 9x9	Contrast normalization	---
9	Median 5x5	None	---
10	Median 5x5	Contrast normalization	---
11	Median 9x9	Contrast normalization	---
12	Gaussian 5x5	None	(Speac – Nucle) (Speac - Golgi) (Speac - Membr) (Nucle - Centr) (Nucle - Golgi) (Nucle - Membr) (Centr - Membr) (Golgi - Membr)
13	Gaussian 5x5	Contrast normalization	---
14	Gaussian 9x9	Contrast normalization	(Homog - Speac) (Centr - Golgi)

TABLE III.   K-PARAMETERS FOR CLASSIFIERS OAA

Class	K
Centromere	9
Golgi	23
Homogeneous	19
Nucleolar	15
Nuclear menbrane	23
Speckled	25

TABLE IV.   K-PARAMETERSFOR FOR CLASSIFIERS OAO

Class	K
Homogeneous vs Speckled	13
Homogeneous vs Nucleolar	9
Homogeneous vs Centromere	5
Homogeneous vs Golgi	5
Homogeneous vs Nuclear membrane	11
Speckled vs Nucleolar	13
Speckled vs Centromere	25
Speckled vs Golgi	17
Speckled vs Nuclear menbrane	15
Nucleolar vs Centromere	9
Nucleolar vs Golgi	5
Nucleolar vs Nuclear menbrane	9
Centromere vs Golgi	25
Centromere vs Nuclear menbrane	11
Golgi vs Nuclear menbrane	21

A. K-parameter

In K-NN classification, an object is classified by a majority vote of its neighbors, with the object being assigned to the most voted class among its K nearest neighbors (K is a positive integer). Obviously the K parameter can not be greater than the number of examples N of a class. The best choice of k depends upon the data; generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. Usually the order of magnitude to be assigned to K is $N^{1/2}$ [11]. In this work an intensive study for the determination of optimal parameters K was carried out. In Table 3 are reported the values obtained for the parameter K regard to six classifiers OAA implemented. While, in Table 4 are reported the values obtained for the parameter K regard to fifteen classifiers OAO implemented.

VIII. RESULTS

For each classifier the training and validation sets are used for the optimization and cross validation phases. The leave-one-out cell cross validation [12] technique is used to exploit the highest possible number of patterns during the training phase, without invalidate the results.

In order to minimize the effects of bias of the method, it was decided to leave out all ROIs belonging to the same image, rather than leave out a single ROI for the construction of the training set; ROI of the same image are similar in terms of the average intensity and contrast and introduce bias.

The classification results obtained in our experiments have been summarized in Table 5, and organized by staining pattern class; for each of them, we show the accuracy obtained.

The estimated mean class accuracy for the proposed method is 84.58%; the class accuracy (not mediated) of the method is 86.37%, with a maximum per-class accuracy of 89.09% for homogeneous patterns. From the table we can notice that the poorest classification accuracy is achieved by Golgi pattern class.

TABLE V. PER-CLASS ACCURACY

Class	Accuracy (OAA+OAO)
Centromere	88.84%
Golgi	71.27%
Homogeneous	89.09%
Nucleolar	86.76%
Nuclear menbrane	86.64%
Speckled	84.88%
MAC	**84.58%**

In the Table 6 we report the results organized by staining pattern class by using classifiers based only on OAA approach and OAO approach.

TABLE VI. PER-CLASS ACCURACY: OAA AND OAO CLASSIFIERS

Class	Accuracy OAA	Accuracy OAO
Centromere	86.54%	90.11%
Golgi	66.99%	62.71%
Homogeneous	85.99%	79.23%
Nucleolar	81.14%	75.98%
Nuclear menbrane	85.19%	81.57%
Speckled	85.09%	83.36%
MAC	**81.82%**	**78.83%**

The average running time for analyzing a cell image and its mask is 1.58 s, about 4 min per IIF image (using a 3.4 GHz Intel i7 CPU). The Training time, using fisher analisys, take about 4 hours. Note that is possible reduce the total running time by parallelizing each of the six indipendent processes.

IX. CONCLUSIONS

In this paper we proposed an approach for the automatic classification of staining patterns in HEp-2 cell IIF images, which is critical for the diagnosis of immune diseases. We adopted a non-standard pipeline for supervised image classification that provides a diversified process for all types of class-patterns to classify. For this purpose, 14 different types of preprocessing have been analyzed and for each class it has been identified the preprocessing giving the best performance in terms of final class accuracy. We extract a large number of features (108), able to fully characterize the staining pattern of HEp-2 cells. Then, for each class, a sub-set of 15 features has been selected.

Also, we proposed an classification-approach based on two steps: the first step adopts the OAA classification rule, while the second step adopts the OAO rule. Our classification system is thus composed by 21 KNN classifiers: 6 one-against-all and 15 one-against-one.

Leave-one-out image cross validation method was used for the evaluation of the results.

We evaluated the method performances on the HEp-2 Cells dataset (manually segmented and annotated). Although being the early results of our methodology for such a challenging application, performances are really satisfactory (mean class accuracy of 84.58%). The choice to pair the two approaches (OAA + OAO), therefore, produces an improvement, even if not exciting, of the mean class accuracy.

The performance of the proposed method can be easily improved. In fact, the Fisher criteria, used for the selection, is simplistic as it does not take into account the possible correlation between different features. Further analysis is being performed for evaluating the robustness and reliability of other features selection methods. This includes the evaluation of exhaustive or almost-exhaustive selection approaches, by focusing on the feature space proposed in this work.

REFERENCES

[1] S. Ciatto, D. Cascio, F. Fauci, R. Magro, G. Raso, R. Ienzi, F. Martinelli, M. Vasile, "Computer assisted diagnosis (CAD) in mammography. Comparison of diagnostic accuracy of a new algorithm (Cyclopus®, Medicad) with two commercial systems", Radiol Med, 114, pp 626-635, 2009

[2] D. Cascio, F.Fauci, M. Iacomi, G. Raso, R. Magro, D. Castrogiovanni, G. Filosto, R. Ienzi, M. Vasile, "Computer-aided diagnosis in digital mammography: comparison of two commercial systems", Imaging in Medicine, vol 6 issue 1, pp 13-30, 2014

[3] S. Tangaro, et al, "MAGIC-5: an Italian mammographic database of digitized images for research". Radiol Med 2008, 113:477–485.

[4] Furnkranz J, "Round Robin Classification", Journal of Machine Learning Research 2, 721-747, 2002.

[5] G. L. Masala , et al, "Comparative study of feature classification methods for mass lesion recognition in digitized mammograms", Nuovo Cimento C, 30, 305-316, 2007

[6] F. Fauci F, D. Cascio, A. La Manna, R. Magro, G. Raso, M. Vasile, M. Iacomi, "A Fourier Based Algorithm for Microcalcifications Enhancement in Mammographic Images", IEEE Nuclear Science Symposium and Medical Imaging Conference; 2012:4388–4391. art. No. 4774254.

[7] P. Perner, H. Perner, B. Muller, "Mining knowledge for Hep-2 cell image classification" , Artificial Intelligence in Medicine, 26, 161-173, 2002.

[8] F. Fauci, et al, "Mammogram segmentation by contour searching and massive lesion classification with neural network". IEEE Nuclear Science Symposium Conference Record 2004, 5:2695–2699.

[9] L. Vivona, D. Cascio, F. Fauci, G. Raso, "Fuzzy technique for microcalcifications clustering in digital mammograms", BMC Medical Imaging, 14:23, pp 1-14, 2014.

[10] M. Iacomi, D. Cascio, F. Fauci, G. Raso, "Mammographic images segmentation based on chaotic map clustering algorithm", BMC Medical Imaging, 14:12, pp 1-11, 2014

[11] G. L. Masala, B. Golosio, P. Oliva, D. Cascio, F. Fauci, S. Tangaro, M. Quarta, S. C. Cheran, E. Lopez Torres, "Classifiers trained on dissimilarity representation of medical pattern: A comparative study", Nuovo Cimento C, 28, 905-912, 2005.

[12] D.Cascio, R. Magro, F. Fauci, M. Iacomi, G. Raso. "Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models. Computers in Biology and Medicine", 42 (11) , pp. 1098-1109, 2012.