A multi-process system for HEp-2 cells classification based on SVM ☆

Donato Cascio [a,*], Vincenzo Taormina [a], Marco Cipolla [a], Salvatore Bruno [a], Francesco Fauci [b], Giuseppe Raso [a]

[a] Dipartimento di Fisica e Chimica dell, Università degli Studi di Palermo, viale delle Scienze Ed. 18, Palermo 90128, Italy
[b] CyclopusCAD ltd, Palermo, Italy

ARTICLE INFO

ABSTRACT

This study addresses the classification problem of the HEp-2 cells using indirect immunofluorescence (IIF) image analysis, which can indicate the presence of autoimmune diseases by finding antibodies in the patient serum. Recently, studies have shown that it is possible to identify the cell patterns using IIF image analysis and machine learning techniques. In this paper we describe a system able to classify pre-segmented immunofluorescence images of HEp-2 cells into six classes. 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 ICPR 2014. This system is based on multiple types of class-process and uses a two-level pyramid to retain some spatial information. We extract a large number (216) of features able to fully characterize the staining pattern of HEp-2 cells. We propose a classification approach based on the one-against-one (OAO) scheme. To do this, an ensemble of 15 support vector machines is used to classify each cell image. Leave-one-specimen-out cross validation method was used for the system optimization. The developed system was evaluated on a blind Hep-2 cells dataset performing a mean class accuracy (MCA) equal to 80.12%.

1. Introduction

Autoimmune diseases (AID) are a collection of many complex disorders of unknown aetiology resulting in immune responses to self-antigens and are thought to result from interactions between genetic and environmental factors. Antinuclear antibodies (ANAs) are significant biomarkers in the diagnosis of autoimmune diseases in humans which is done by means of indirect immunofluorescence (IIF) method, and performed by analysing patterns and fluorescence intensity. There are over 80 different AID, and overall they are among the most prevalent diseases in the US, affecting at least 7% of the population. Since most AID are chronic and incurable, from a public health perspective they constitute a major health problem that, besides causing individual suffering, has high societal costs [1]. These diseases can affect people of all ages and both sexes, with a higher frequency in women of child-bearing age. The autoimmune diseases are multi-factorial, and their risk factors are genetic and environmental. The binding of auto-antibodies on HEp-2 cells is revealed by fluorescent antibodies to human immunoglobulin. The fluorescence pattern observed with the mi-croscope (example: homogeneous, speckled, nucleolar, nentromere, etc.) is specific according to the nature of the self antigen and of its location in the cell. Such slides are typically examined by patholo-gists, however, due to the difficulty of the task, a Computer Aided Design system is desirable [2–4]. However, their pattern identifica-tion is often subjective and low standardized. Hence, it is beneficial to develop the automatic stain pattern identification algorithms which can serve as a computer-aided diagnosis (CAD) system. Due to its potential applications, the computer vision based stain patterns identification has attracted much attention [5].

Recently Elgaaied Benammar et al. [6] have optimized and tested a CAD system on HEp-2 images, and preliminary results showed that the CAD, used as second reader, resulted to better per-formance than Junior immunologists and hence may significantly improve their efficacy; compared with two Junior Immunologists, the CAD system showed higher Intensity accuracy (85.5% versus 66.0% and 66.0%), higher patterns accuracy (79.3% versus 48.0% and 66.2%) and higher mean class accuracy (79.4% versus 56.7% and 64.2%).

In this work we focus on building an automated pattern recognition system for the classification of IIF HEp-2 cell images into a set of predefined classes. The system described in this paper participated to the contest related to the special issue of pattern recognition letters titled "pattern recognition techniques for indirect immunofluorescence images analysis".

1.1. Related work

This section briefly reviews previous work related to HEp-2 cell image classification in the context of ANA testing. The problem of HEp-2 cell classification attracted major attention among researchers with the benchmarking contests [7,8]. The I3A contest (performance evaluation of indirect immunofluorescence image analysis systems) held in conjunction with ICPR 2014 was the most recent in this series of contests [9]. It received 11 submissions to Task 1 (cell classification) and 7 submissions to Task 2 (specimen classification).

In this contest Ensafi et al. [10] proposed a classification method where the SIFT and SURF features are extracted as the input features to learn a dictionary followed by spatial pyramid matching (SPM) to provide the sparse representation of the input cell images. Then a support vector machine (SVM) has been trained to classify the test images.

Manivann et al. [11] presented a system to recognize such patterns, at cellular and specimen levels, in images of HEp-2 cells. Ensembles of SVMs were trained to classify cells into six classes based on sparse encoding of texture features with cell pyramids, capturing spatial, multi-scale structure. Mean class accuracies for cell classification obtained on used test data sets were 87.1%. These were the highest achieved in the competition hosted by ICPR2014.

At present, there are many mature image recognition and classification algorithms, and bag of features (BOF) algorithm is a popular one among them; BOF algorithm came from the bag of words (BOW) and was originally used for documents matching. In recent years, BOF algorithm is also been applied in the field of HEp-2 cell classification and has achieved results [12]. However, in the BOF algorithm there exists some problem in the application of image recognition and classification such as run is not fast enough and the classification accuracy is not high enough. So it needs to be optimized.

Recent research has focused on extracting various morphological features and texture features such as local binary patterns (LBP) have been widely applied for classification of segmented images of HEp-2 cells [14]. Due to poor quality images, utilizing low-level features directly may be inefficient.

Nosaka et al. [13] propose co-occurrence of adjacent local binary patterns (CoALBP) to extract textural features. Using linear support vector machine (SVM), their method won the first prize in the contest ICPR2012.

Recently [15] a specimen-level image classification system was proposed, in which a specimen-level image descriptor is learned, based on cell-level attributes.

In another system, proposed by Soda et al. [16], each specimen image is classified based on the strength of the recognition of the cells in that specimen image.

Han et al. [17] have employed a parametric probability process to model local image patches (textons: microstructures in the cell image) and extract the higher-order statistics of the model parameters for the image description. They have used a simple linear support vector machine for cell pattern identification. Cascio et al. [18] proposed 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, they have implemented 21 KNN classifiers: 6 OAA and 15 OAO.

Perner et al. [19] presented an early attempt on developing an automated HEp-2 cell classification system. Cell regions were represented by a set of basic features extracted from binary images obtained at multiple grey level thresholds. Those features were then classified into six categories by a decision tree algorithm.

Theodorakopoulos et al. [20] combined a set of local features, including LBP and rotation-invariant SIFT, with vectors of locally aggregated descriptors (VLAD).
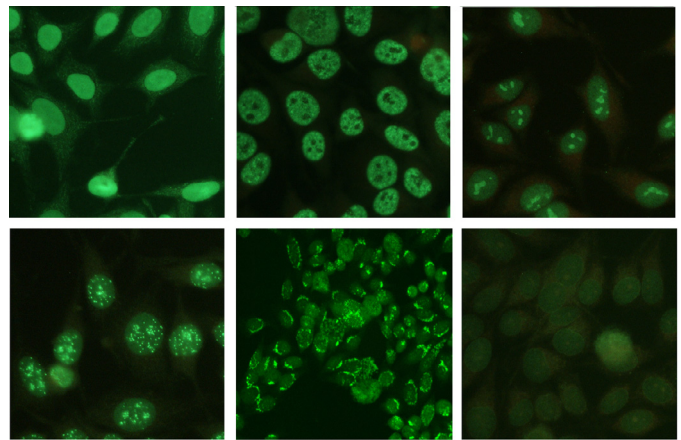

Fig. 1. IIF images with different staining patterns (from left to right and from up to down: homogeneous, speckled, nucleolar, centromere, Golgi and nuclear membrane).

1.2. Our contributions

In this paper we describe a system for the classification of pre-segmented immunofluorescence images of HEp-2 cells into six classes: homogeneous, speckled, nucleolar, centromere, Golgi, and nuclear membrane. Fig. 1 shows examples of each class. This system is based on multiple types of class-process which uses a two-level pyramid to retain some spatial information. 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. For this purpose, an intensive analysis of preprocessing types has been conducted and for each process it has been identified the preprocessing method giving the best performance in terms of final class accuracy. That analysis was carried out by distinguishing between the mask cells and the adjacent circular crown, and 30 different best preprocessing were obtained.

We extracted a large number of features (216), and for each classification process a phase of features reduction, based on Fisher criterion, was performed in order to select the best features, able to fully characterize the staining pattern of HEp-2 cells.

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

As final consideration, we want to highlight that the method here presented has the advantage of allowing an easy extension to the problem of aided diagnosis, and hence to the development of a CAD for the support to diagnosis, which is the ultimate goal of this research [6], this is possible by including an automatic segmentation phase in the system, also differentiated (given the enormous variety of ways in which the pattern may occur, this is probably the only way to achieve an efficient system).

2. Materials and methods

In several multiclass classification problems, it is preferable to use a number of classifiers equal to the number of classes and each classifier is trained in order to discriminate a class from all the others (binary approach) [21]. In this work, the classification stage is differentiated by using 15 one-against-one classifiers for six classes of staining patterns. In addition, the preprocessing and feature extraction steps are differentiated too. Therefore, we adopt a non-standard pipeline for supervised image classification. The generic image is simultaneously processed by 15 processes obtaining 15 separate outputs that represent how the cell resembles each one of the 6 classes analysed in this work. This system