Pattern Recognition 47 (2014) 2389–2399

Contents lists available at ScienceDirect

Pattern Recognition

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

Subclass Discriminant Analysis of morphological and textural features for HEp-2 staining pattern classification

Santa Di Cataldo [a,*], Andrea Bottino [a], Ihtesham Ul Islam [a], Tiago Figueiredo Vieira [a,b], Elisa Ficarra [a]

[a] Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[b] Department of Electronics and Systems, Universidade Federal de Pernambuco, Recife, Brazil

ARTICLE INFO

Available online 1 October 2013

Keywords:
Indirect Immunofluorescence
Fluorescence pattern classification
Subclass Discriminant Analysis
Morphological analysis
Textural analysis

ABSTRACT

Classifying HEp-2 fluorescence patterns in Indirect Immunofluorescence (IIF) HEp-2 cell imaging is important for the differential diagnosis of autoimmune diseases. The current technique, based on human visual inspection, is time-consuming, subjective and dependent on the operator's experience. Automating this process may be a solution to these limitations, making IIF faster and more reliable. This work proposes a classification approach based on Subclass Discriminant Analysis (SDA), a dimensionality reduction technique that provides an effective representation of the cells in the feature space, suitably coping with the high within-class variance typical of HEp-2 cell patterns. In order to generate an adequate characterization of the fluorescence patterns, we investigate the individual and combined contributions of several image attributes, showing that the integration of morphological, global and local textural features is the most suited for this purpose. The proposed approach provides an accuracy of the staining pattern classification of about 90%.

1. Introduction

Indirect Immunofluorescence (IIF) is a widespread microscopy imaging technique for the detection of antinuclear auto-antibodies (ANA), which can reveal the presence of important autoimmune pathologies such as systemic rheumatic diseases, multiple sclerosis and diabetes [1]. The ANA-screening is typically performed by visually inspecting cultured cells with a fluorescence microscope. The patient's serum is first dispensed on a HEp-2 cell substrate and then diluted and incubated. The auto-antibodies of the serum selectively bond with specific antigens on the substrate and are finally revealed by a fluorescence tag.

The specialists usually perform a three-step analysis. First, the slide is validated by checking the presence of at least one fluorescence mitotic cell. Second, the intensity of fluorescence signal is evaluated according to three levels: negative (*i.e.* absence of fluorescence), intermediate or positive. Finally, the cells of the intermediate and positive slides (Fig. 1) are classified on the basis of the pattern of the fluorescence signal, which in turn reveals the auto-antibody type. According to the literature, the different staining patterns can be classified into six main groups, namely *centromere, homogeneous, nucleolar, coarse speckled, fine speckled* and *cytoplasmatic* (Fig. 2).

The accurate classification of the staining patterns is very important for differential diagnosis, since different patterns are associated with different types of autoimmune diseases. Nevertheless, in the standard practice this process suffers from intrinsic limitations related to the visual evaluation performed by human subjects. The analysis of large volume of images is a tedious and time-consuming task and requires highly trained and specialized personnel. Moreover, human evaluation for this type of screening is affected by very high inter laboratory variability (up to 24%, as reported in [2,3]). This has tremendous impact on the reliability and reproducibility of the obtained results.

Computer-Aided Diagnosis (CAD) systems may overcome these limitations and effectively support the decision of the specialists. In particular, the automation of the staining pattern classification process may considerably reduce the time and effort required by the analysis and, at the same time, improve its repeatability. This would make IIF analysis easier, faster and more reliable.

Triggered by the growing demand for reliable CAD systems, recent research has proposed solutions for all the major steps of IIF analysis, including methods for (i) the automated segmentation of the HEp-2 cells [4–6], (ii) the recognition of the mitotic cells in the slides [7,8], (iii) the quantification of fluorescence intensity [9] and (iv) the automated classification of the staining patterns [10–16].

* Corresponding author. Tel.: +39 011 564 7020; fax: +39 011 564 7099.
E-mail addresses: santa.dicataldo@gmail.com, santa.dicataldo@polito.it (S. Di Cataldo), andrea.bottino@polito.it (A. Bottino), ihtesham.ulislam@polito.it (I. Ul Islam), tiago.figueiredo@polito.it (T. Figueiredo Vieira), elisa.ficarra@polito.it (E. Ficarra).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.09.024

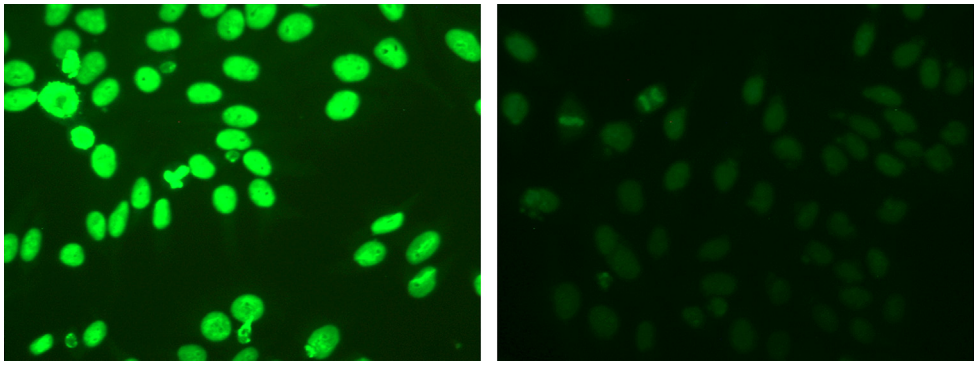

Fig. 1. HEp-2 IIF images with positive (left) and intermediate (right) fluorescence intensity.

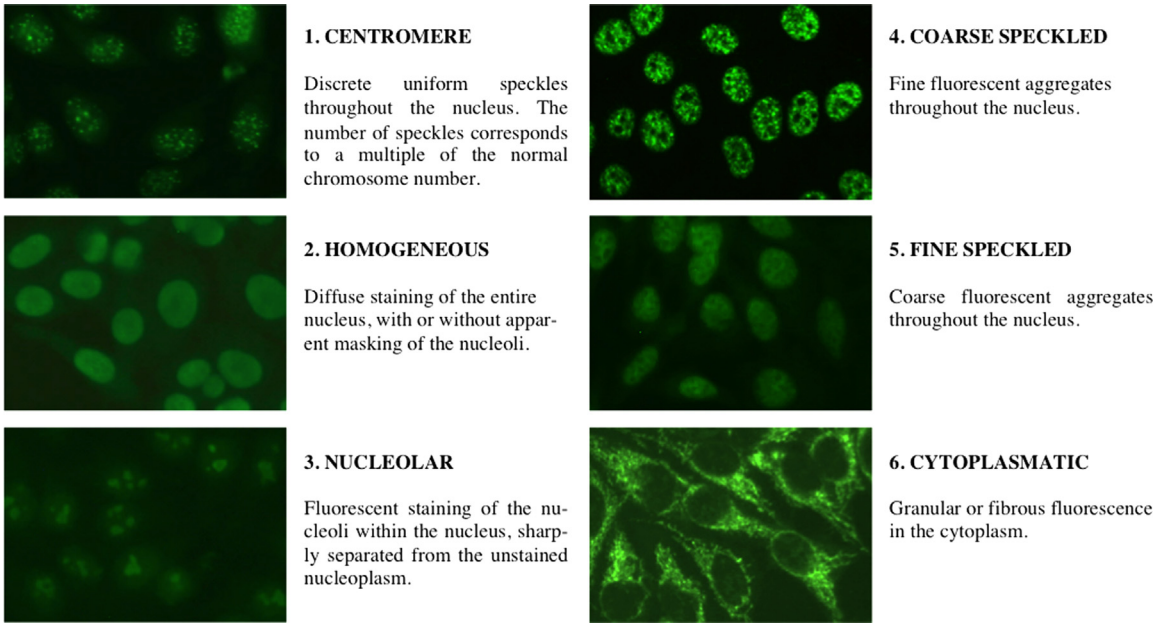

Fig. 2. HEp-2 staining patterns that are considered relevant to diagnostic purposes.

In particular, most of the recent efforts are focused on the latter task and the proposed classification schemes span the entire spectrum of machine learning (*e.g.*, learning vector quantization [10], decision tree induction algorithms [11,12], support vector machines [15], random forests [16], self-organizing maps [14] and multi-expert systems [13]). The image attributes used to characterize the fluorescent patterns include several types of descriptors, with special focus on morphological and textural features [17,18].

In spite of the recent extensive research, the accurate classification of the staining patterns still remains a challenge. Moreover, the comparison of the solutions presented in the literature is extremely difficult because they are based on different datasets and different experimental protocols. Conversely, the recent availability of a public dataset of HEp-2 cell images that bounds the researchers to a standardized experimental procedure enables the direct comparability of different approaches [19].

In this paper we describe a new technique for the automated characterization of the staining patterns in HEp-2 IIF images. Our contribution is twofold:

- we propose a set of features to characterize cell images that are highly discriminative with respect to their fluorescence patterns. Such a feature set is obtained through a detailed analysis of single image attributes, as well as of their integration, and of the selection of the most relevant feature variables. In particular, we investigated image attributes derived from morphological, global and local texture analyses. For each of these categories, we first performed experiments with different formulations of features to evaluate their individual accuracy. Then, we combined attributes of different categories showing that the aggregation of descriptors of different nature provides a more comprehensive solution to characterize the HEp-2 fluorescence patterns;
- we propose a Subclass Discriminant Analysis (SDA, [20]) based strategy to remap the cell representations into a novel feature space that provides a better separation of the classes aimed at simultaneously improving the intra-class similarities and inter-class dissimilarities and, thus, at making the classification task easier and more accurate.

A preliminary version of this technique was presented in [21], where we first analyzed the use of SDA to approach the staining pattern classification problem, using textural descriptors based on Gray-Level Co-occurrence Matrices (GLCM) and Discrete Cosine Transform (DCT). Here we present a deeper insight into the problem aimed at obtaining a better characterization of the staining patterns and a larger set of experiments supporting our findings.

The rest of the paper is organized as follows. After providing a full characterization of the HEp-2 image dataset in Section 2, in