Analysis of HEp-2 Images using MD-LBP and MAD-Bagging

Gerald Schaefer[1], Niraj P. Doshi[1], Shao Ying Zhu[2] and Qinghua Hu[3]

Abstract— **Indirect immunofluorescence imaging is employed to identify antinuclear antibodies in HEp-2 cells which founds the basis for diagnosing autoimmune diseases and other important pathological conditions involving the immune system. Six categories of HEp-2 cells are generally considered, namely homogeneous, fine speckled, coarse speckled, nucleolar, cytoplasmic, and centromere cells. Typically, this categorisation is performed manually by an expert and is hence both time consuming and subjective.**

In this paper, we present a method for automatically classifiying HEp-2 cells using texture information in conjunction with a suitable classification system. In particular, we extract multi-dimensional local binary pattern (MD-LBP) texture features to characterise the cell area. These then form the input for a classification stage, for which we employ a margin distribution based bagging pruning (MAD-Bagging) classifier ensemble. We evaluate our algorithm on the ICPR 2012 HEp-2 contest benchmark dataset, and demonstrate it to give excellent performance, superior to all algorithms that were entered in the competition.

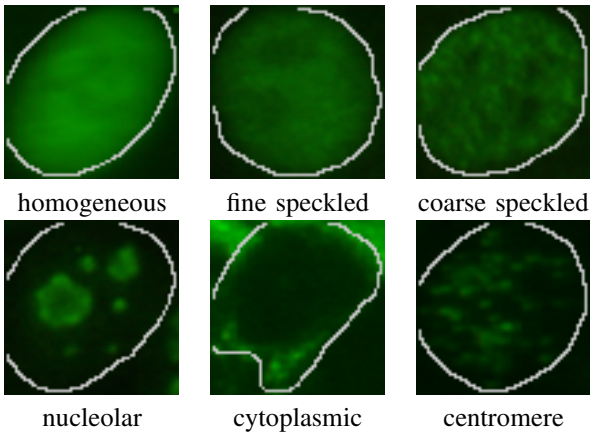

Fig. 1. Sample HEp-2 cell images (with manually defined borders in white) from the ICPR 2012 contest dataset.

I. Introduction

Indirect immunofluorescence (IIF) is commonly employed for screening of antinuclear antibodies (ANAs) based on HEp-2 cells. ANA tests allows identification of diseases such as systemic rheumatic disease, systemic sclerosis and mellitus (type-I) diabetes [1]. In IIF, cultured HEp-2 cells are observed under a fluorescence microscope and then categorised based on fluorescence intensity and on the type of staining patterns.

This classification of HEp-2 cells is crucial for diagnosis, since different patterns give indications for different autoimmune diseases. At the same time, since performed manually by an expert, it is a laborous and time consuming task. A computer-aided diagnosis (CAD) approach would hence not only speed up the task but also lead to objective, reproducible results. HEp-2 cells are typically categorised into six groups: homogeneous, fine speckled, coarse speckled, nucleolar, cytoplasmic, and centromere cells, which are also the classes we consider in this papers. Example images of each class are shown in Fig. 1.

In this paper, we utilise multi-resolution texture information for categorising HEp-2 cell images. In particular, we employ multi-dimensional local binary pattern (MD-LBP) features to characterise the cell area. LBP descriptors [2] yield relatively simple yet powerful texture features describing the relationships of pixels to their (local) neighbourhoods. MD-LBP is a multi-scale extension of LBP that also preserves the relationships between the scales in form of a multi-dimensional histogram [3]. MD-LBP features form the input for a classification stage for which we utilise a margin distribution based bagging pruning (MAD-Bagging) [4] classifier ensemble. We evaluate our algorithm on the ICPR 2012 contest dataset [5], and show that it provides very good performance, superior to all algorithms entered in the competition.

II. Related work

Automated classification of HEp-2 cell images has recently received increased attention, in particular with the running of a competition at ICPR 2012 [5]. A number of approaches were presented at the contest, of which we summarise a select few in the following.

In [6], images are contrast normalised and statistical texture features based on the grey level co-occurrence matrix (GLCM) [7] as well as frequency domain texture features based on the discrete cosine transform (DCT) [8] are extracted. To improve classification performance, a two-step feature selection method is employed where the first step is based on a minimum redundancy maximum relevance algorithm to select a candidate feature set, while a final feature set is obtained using a sequential forward selection method. A support vector machine (SVM) [9] is used for classification.

In [10], DCT coefficient features, local binary pattern (LBP) [2] and Gabor texture descriptors [11] as well as various global appearance statistical features (area, perimeter, average intensity and standard deviation of the cell region as well as the ratio of cell and background) are utilised. A multi-

[1]G. Schaefer and N.P. Doshi are with the Department of Computer Science, Loughborough University, U.K.

[2]S.Y. Zhu is with the School of Computing and Mathematics, University of Derby, U.K

[3]Q. Hu is with the School of Computer Science and Technology, Tianjin University, China

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

class boosting SVM [12] is employed for classification with different SVMs merged into a classifier and boosted using a modified AdaBoost.M1 algorithm [13].

Shape and texture features are combined in [14]. The cell images are thresholded at different intensity levels and shape based descriptors (perimeter, eccentricity, etc.) and intensity based features (average intensity, standard deviation, etc.) are extracted at each level. Following this, gradient magnitude features are calculated after smoothing the image using Gaussian kernels with different parameters. Finally, GLCM texture features are also calculated. The obtained features are then fed to a Random Forest classifier [15].

In [16], shape features based on the Hessian matrix are employed, where the eigenvalues of the Hessian and the related eigenvector orientations are used for shape characterisation. Edge features are extracted using an adaptive robust structure tensor and histogram of oriented gradients (ARST-HOG) [17] approach. Finally, texture information, based on LBP, is also utilised. Classification is performed using regression trees as base classifiers and a ShareBoost algorithm [18] for classifier fusion.

In [19], along with GLCM and HOG [20] features, region-of-interest (ROI)-based descriptors are used which include shape features (eccentricity, perimeter, etc.) and intensity based features (derived from intensity percentiles). For classification, an SVM is chosen as the best performing algorithm.

III. HEp-2 CELL TEXTURE ANALYSIS USING MD-LBP

Previous approaches to automatically classifying HEp-2 cells are typically based on several features, while texture features are employed in the majority of algorithms. In this paper, we utilise a single, relatively simple type of texture feature based on local binary patterns (LBP) that we show to yield very good HEp-2 cell recognition.

LBP [2] describes the local neighbourhood of a pixel by thresholding neighbouring pixels g_p with the centre pixel value g_c. The resulting sequence of 0s and 1s is then known as the local binary pattern, formally expressed as

$$\text{LBP} = \sum_{p=1}^{8} s(g_p - g_c) 2^{p-1}, \tag{1}$$

where

$$s(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}, \tag{2}$$

and a histogram of these patterns is generated to summarise texture information of an image or a region of interest. LBP is inherently invariant to monotonic intensity transformations and hence more robust than other techniques.

LBP patterns are typically obtained from a circular neighbourhood where locations in the neighbourhood that do not fall exactly at the centre of a pixel are obtained through interpolation. If a texture is rotated, essentially the patterns (that is, the 0s and 1s around the centre pixel) rotate with respect to the centre. Rotation invariance can hence be obtained by mapping all possible rotated patterns to the same

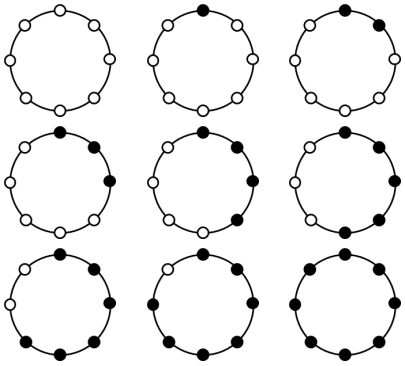

Fig. 2. Uniform LBP patterns.

descriptor. Furthermore, certain patterns are fundamental properties of texture and may account for the majority of LBP patterns. To address this, only uniform patterns can be utilised where a uniformity measure is defined by the number of transitions from 0 to 1 or vice versa in the LBP code.

Based on 8 neighbouring pixels, 9 different rotation invariant uniform patterns (with maximal two transitions) can be defined as shown in Fig. 2), while the remaining patterns are accumulated in a single bin, thus giving a histogram of 10 bins. This yields a powerful texture descriptor that was shown to work well for texture classification, especially when obtained at multiple scales [2], [21].

In conventional multi-scale LBP, the histograms for each scale are simply concatenated to form a one-dimensional feature vector. This leads to a loss of information regarding the relationships between patterns across different scales and additional ambiguity. Multi-dimensional LBP (MD-LBP) [3] addresses this by preserving the joint distribution of LBP codes at different scales in form of a multi-dimensional histogram of LBP values. To do so, for each pixel, LBP codes at different scales are obtained, while the combination of these codes identifies the histogram bin that is incremented.

Our approach to derive MD-LBP texture features from a HEp-2 cell image is illustrated in Fig. 3. We use the green channel of IIF images resized to 64×64 pixels from which we gather rotation invariant uniform LBP texture information. The MD-LBP histogram, extracted from the area of the cell and based on three scales with radii $\{1, 3, 5\}$,

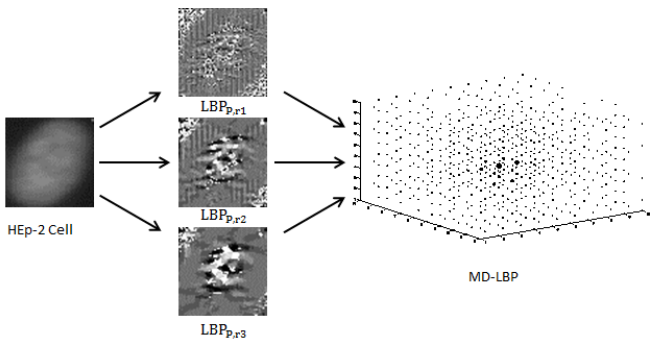

Fig. 3. MD-LBP histogram generation for HEp-2 cell image.

is then used in the subsequent classification stage.

IV. HEp-2 cell classification using MAD-Bagging

Based on the MD-LBP features derived above, we then perform classification for which we employ an ensemble learning approach. The idea of ensemble classifiers [22] is to exploit the strengths and local competencies of a pool of classifiers, while at the same time reducing their individual weaknesses. Consequently, an appropriately constructed combination of several predictors can lead to better and more robust classification performance compared to any single classifier.

In particular, we utilise a margin distribution based bagging pruning (MAD-Bagging) [4] classifier ensemble. Bagging is a well known, simple yet effective technique for generating an ensemble of classifiers [23]. A collection of base classifiers is trained on bootstrap replicates of the training set and the outputs of all trained base classifiers are combined using simple voting. In general, the error of bagging decreases as base classifiers aggregated in the ensemble increase. Eventually, the error asymptotically approaches a constant level when the size of the ensemble becomes very large.

It is well accepted that generalisation performance of base classifiers and the diversity among base classifiers greatly influence the performance of the ensemble. Bagging uses bootstrapping to generate diverse training sets which then leads to diverse classifiers. On the other hand, selecting only a subset of candidate base classifiers may lead to a significant improvement of the final ensemble. Algorithms can be developed to select diverse and accurate base classifiers in order to yield compact and powerful sub-ensembles [24].

MAD-Bagging utilises the margin of the ensemble as an optimisation objective and derives an L_1 regularised squared loss function. By solving the resulting optimisation problem, a sparse weight vector for the candidate base classifiers can be obtained. Then, only base classifiers with non-zero weights are included in the final ensemble, while the others are discarded.

Assuming a set of samples $X = \{(x_i, y_i)\}_{i=1}^{N}$ with $y_i \in \{-1, +1\}$, a base classifier h_j performs a mapping from X to $\{-1, +1\}$. The voted ensemble $f(x)$ is of the form

$$f(x) = \sum_{j=1}^{T} w_j h_j(x), \tag{3}$$

where w_j is the (non-negative) weight assigned to base classifier h_j, $\sum_{j=1}^{T} w_j = 1$, and T is the number of candidate base classifiers available. An error occurs for x_i if and only if the output of the voting classifier and the label y_i do not have the same sign. Hence, $y_i h_i(x)$ is the difference between the weights assigned to the correct label and the weights assigned to the incorrect label.

$y_i h_i(x)$ is considered the sample margin r_i with respect to the voting classifier f. In order to obtain a good margin for each sample, a loss function $\sum_i C(y_i h_i(x))$ is designed

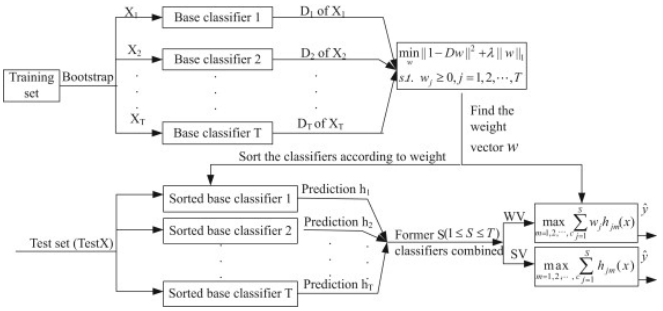

Fig. 4. MAD-Bagging framework.

in MAD-Bagging, leading to solving for

$$\min \sum_i C(y_i h_i(x)) + \lambda ||w||_1 \tag{4}$$

subject to $W_j \geq 0$. This objective function leads to an optimal margin distribution over the training samples.

The overall procedure of MAD-Bagging is illustrated in Fig. 4.

V. Experimental results

For evaluation, we use the ICPR 2012 HEp-2 classification contest dataset [5] which is based on 28 HEp-2 images acquired by means of a fluorescence microscope under 40-fold magnification coupled with a 50W mercury vapour lamp. Images were taken with a SLIM system digital camera, and stored in 24-bit true-colour format with a resolution of 1388×1038 pixels. Cells were manually segmented and annotated by a specialist to obtain a ground truth for the competition.

The training dataset provided to contestants comprises 721 samples of individual cells, extracted from part of the captured images. There are 150 homogeneous, 94 fine speckled, 109 coarse speckled, 102 nucleolar, 58 cytoplasmic, and 208 centromere cells (an example of each class is given in Fig. 1). The testing dataset, extracted from different images, contains 734 cells in total of which 172 are homogeneous, 114 fine speckled, 101 coarse speckled, 139 nucleolar, 51 cytoplasmic, and 149 centromere cells.

For MAD-Bagging, 100 bags were used, and support vector machines, in particular one-against-one multi-class SVMs [25] with linear kernels and kernel parameters optimised as in [26], served as base classifiers.

We first evaluate the performance on the training dataset, by performing 10-fold cross validation (10CV), where the dataset is split into 10 partitions and training is performed

TABLE I
HEp-2 cell classification results on ICPR12 training data.

algorithm (evaluation)	accuracy
Cataldo *et al.* [6] (10CV)	86.96
Li *et al.* [10] (5CV)	98.34
Strandmark *et al.* [14] (LOOCV)	97.40
Ersoy *et al.* [16] (5CV)	92.80
Ghosh and Chaudhary [19] (10CV)	91.13
Proposed (10CV)	97.22

TABLE II

CLASSIFICATION RESULTS ON ICPR12 TEST DATA.

method	accuracy [%]
Cataldo *et al.* [6]	48.50
Li *et al.* [10]	64.17
Strandmark *et al.* [14]	47.82
Ersoy *et al.* [16]	49.18
Ghosh and Chaudhary [19]	59.81
top ICPR contest entry (Nosaka and Fukui, unpublished)	68.66
Proposed	71.39

on all but one partition while testing is conducted on the remaining one, reporting the average classification accuracy over all 10 folds. The results are reported in Table I, which also lists the classification accuracies on the same dataset of the methods discussed in Section II[1].

From Table I, we can see that our MD-LBP based HEp-2 cell classification approach affords good performance, giving a classification accuracy exceeding 97% and thus outperforming three of the other five algorithms.

As the ICPR contest revealed [5], while several of the 28 submitted entries obtained high classification performance (95+%) on the training dataset, accuracy on the test dataset was significantly lower, suggesting that the test data is much more challenging. The best approach, by Nosaka and Fukui, was reported to give a classification accuracy of 68.66%, while about half of the submitted approaches reached less than 50% [5] including some of those discussed in Section II (most submitted approaches were not published). Interestingly, also a medical doctor, a specialist with 12 years experience in immunology, did not fare much better with a correct recognition rate of 73.30%.

In Table II, we report the results obtained on the test dataset. As we can see from there, impressively our method outperforms all competition entries on the test dataset with a classification accuracy of 71.39%.

VI. CONCLUSIONS

In this paper, we have presented an effective approach to the automated classification of HEp-2 cell images obtained through indirect immunofluorescence imaging. Our method is based on extracting useful texture information from the cell area. In particular, multi-dimensional local binary pattern (MD-LBP) descriptors are employed and then classified using a margin distribution based bagging pruning (MAD-Bagging) classifier ensemble. Based on the ICPR 2012 competition dataset, our approach is shown to deliver very good classification performance and to outperform all entries submitted to the contest.

REFERENCES

[1] K. Egerer, D. Roggenbuck, R. Hiemann, M. G. Weyer, T. Büttner, B. Radau, R. Krause, B. Lehmann, E. Feist, and G. R. Burmester, "Automated evaluation of autoantibodies on human epithelial-2 cells as an approach to standardize cell-based immunofluorescence tests," *Arthritis Research and Therapy*, vol. 12, no. 2, p. R40, 2010.

[1]Note, that not all papers use 10CV as evaluation method, and that hence the cited numbers are only partially comparable.

[2] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.

[3] G. Schaefer and N. P. Doshi, "Multi-dimensional local binary pattern descriptors for improved texture analysis," in *21st Int. Conference on Pattern Recognition*, 2012, pp. 2500–2503.

[4] Z. Xie, Y. Xua, Q. Hu, and P. Zhu, "Margin distribution based bagging pruning," *Neurocomputing*, vol. 85, pp. 11–19, 2012.

[5] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking HEp-2 cells classification methods," *IEEE Trans. Medical Imaging*, vol. 32, no. 10, pp. 1878–1889, 2013.

[6] S. Di Cataldo, A. Bottino, E. Ficarra, and E. Macii, "Applying textural features to the classification of HEp-2 cell patterns in IIF images," in *21st Int. Conference on Pattern Recognition*, 2012, pp. 3349–3352.

[7] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[8] G. Sorwar, A. Abraham, and L. S. Dooley, "Texture classification based on DCT and soft computing," in *10th IEEE International Conference on Fuzzy Systems*, 2001, pp. 545–548.

[9] V. N. Vapnik, *Statistical Learning Theory*. John Wiley & Sons, 1998.

[10] K. Li, J. Yin, Z. Lu, X. Kong, R. Zhang, and W. Liu, "Multiclass boosting SVM using different texture features in HEp-2 cell staining pattern classification," in *21st Int. Conference on Pattern Recognition*, 2012, pp. 170–173.

[11] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, August 1996.

[12] M. Gonen, A. Tanugur, and E. Alpaydm, "Multiclass posterior probability support vector machines," *IEEE Trans. Neural Networks*, vol. 19, no. 1, pp. 130–139, 2008.

[13] X. Li, L. Wang, and E. Sung, "Adaboost with SVM-based component classifiers," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 5, pp. 785–795, 2008.

[14] P. Strandmark, J. Ulen, and F. Kahl, "HEp-2 staining pattern classification," in *21st Int. Conference on Pattern Recognition*, 2012, pp. 33–36.

[15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[16] I. Ersoy, F. Bunyak, J. Peng, and K. Palaniappan, "HEp-2 cell classification in IIF images using ShareBoost," in *21st Int. Conference on Pattern Recognition*, 2012, pp. 3362–3365.

[17] K. Palaniappan, F. Bunyak, P. Kumar, I. Ersoy, S. Jaeger, K. Ganguli, A. Haridas, J. Fraser, R. Rao, and G. Seetharaman, "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video," in *13th Int. Conference on Information Fusion*, 2010.

[18] J. Peng, C. Barbu, G. Seetharaman, W. Fan, X. Wu, and K. Palaniappan, "ShareBoost: boosting for multi-view learning with performance guarantees," in *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2011, pp. 597–612.

[19] S. Ghosh and V. Chaudhary, "Feature analysis for automatic classification of HEp-2 florescence patterns : Computer-aided diagnosis of autoimmune diseases," in *21st Int. Conference on Pattern Recognition*, 2012, pp. 174–177.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Int. Conference Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.

[21] N. P. Doshi and G. Schaefer, "A comparative analysis of local binary pattern texture classification," in *Visual Communications and Image Processing*, 2012.

[22] L. I. Kuncheva, *Combining pattern classifiers: Methods and algorithms*. Wiley-Interscience, New Jersey, 2004.

[23] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[24] Z. H. Zhou, J. X. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," *Artificial Intelligence*, vol. 137, no. 1–2, pp. 239–263, 2002.

[25] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[26] C.-C. Chang and C.-J. Lin, "libSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.