Open Access

Allmin Pradhap Singh Susaiyah*, Suhail Parvaze Pathan and Ramakrishnan Swaminathan

Classification of indirect immunofluorescence images using thresholded local binary count features

DOI 10.1515/cdbme-2016-0106

Abstract: Computer aided classification of HEp-2 cell based indirect immunofluorescence (IIF) images is a recommended procedure for standardising autoimmune disease diagnostics. In this work a novel feature, the thresholded local binary count (TLBC) has been proposed to classify IIF images into one among six classes. The TLBC is rotational invariant and is insensitive to pixel quantization noise. It characterizes the local binary gray scale pixel information in an image. The proposed feature along with global features such as area, entropy, illumination level and mean intensity, when classified using a support vector machine gave an accuracy of 86%. This feature could help in improving the diagnostics of autoimmune diseases which is highly clinically significant.

Keywords: autoimmune diseases; HEp-2 cells; indirect immunofluorescence; thresholded local binary count feature.

1 Introduction

Autoimmune diseases such as systemic lupus erythematosus, sjogren's syndrome, multiple sclerosis and rheumatoid arthritis are caused due to abnormal functioning of the body's immune system. The antibodies of a normal person's immune system attack foreign bodies (antigens). However, the antibodies of a person with autoimmune disease attack the body's own healthy tissues and proteins (auto-antigens). Autoimmune diseases have a prevalence rate of $12.5 \pm 7.9\%$ and are one among the major concerns in healthcare. Early diagnosis is a key to successful treatment of autoimmune diseases. The traditional approach of indirect immunofluorescence (IIF) involves identification of specific patterns in microscopy slides. It is a manual process and is highly reliable. However, this process is time consuming and prone to significant inter and intra-operator variances. Recent works in literature have recommended computer aided classification of IIF images for standardising autoimmune disease diagnostics. This can be performed by digitally acquiring images of IIF slides and classifying them with the help of machine learning tools. Human epithelial type-2 (HEp-2) cells have been a standard substrate in IIF due to their large size and high reproducibility. When subjected to IIF process, the HEp2 cells exhibit staining patterns, namely, centromere, coarse-speckled, fine-speckled, homogenous, and nucleolar. The main objective of IIF is to classify the images into one among these staining patterns.

Recently, many features have been proposed for the classification of HEp-2 based IIF images. Different types of image descriptors such as global, local, region based, texture and microscopic features have been used. Wiliem et al. [1], proposed a bag of 256 visual words. IIF images were represented as a histogram of occurrences of these visual words in the image. Faraki et al. [2] extended the bag of features from Euclidean to non-Euclidean Riemannian manifolds and used it to classify IIF images. Yang et al. [3] used a filter bank trained from unlabelled IIF images using independent component analysis for classification. Ghosh and Chaudary used a concatenation of histogram of oriented gradient and region of interest based features [4]. Texture features such as local binary pattern [5] and fractal texture features [6] have also proven to perform well in discriminating staining patterns. In this work, a novel texture descriptor, the thresholded local binary count (TLBC) feature has been proposed to perform classification of HEp-2 cell images.

2 Methodology

In this work, images from a publicly available database namely the SNPHEp2 database are used. It contains 1884

***Corresponding author: Allmin Pradhap Singh Susaiyah,** Indian Institute of Technology, Madras, Chennai, Tamil Nadu - 600036, India, E-mail: allmin123@gmail.com
Suhail Parvaze Pathan and Ramakrishnan Swaminathan: Indian Institute of Technology, Madras, Chennai, Tamil Nadu 600036, India, E-mail: suhailsp@gmail.com (S. Parvaze); sramki@iitm.ac.in (R. Swaminathan)

images of all five staining pattern classes. The images also belong to positive and intermediate categories based on illumination intensities. The positive images are brighter compared to the intermediate images. The images on observation are found to be of low contrast and require pre-processing to improve it and reduce the difference in intensities between positive and intermediate images.

2.1 Pre-processing

Pre-processing is performed using pixel normalisation (PN) technique (see eq 1). Where, I is the raw image, I_{PN} is the pixel normalised image, $min(I)$ is the smallest pixel value in the image and $max(I)$ is the largest pixel value in the image.

$$I_{PN} = \frac{I - min(I)}{max(I) - min(I)} \times 255 \tag{1}$$

2.2 Segmentation

The improvement in contrast due to pre-processing is helpful in segmentation. Multilevel Otsu thresholding [7] is used to generate four levels of thresholds namely T1, T2, T3 and T4. Segmentation masks are generated using these thresholds on the pre-processed images. Validation is performed using Dice's coefficient, Jaccard index and accuracy on the generated masks against the ground truth mask provided in the dataset. The threshold with highest validation score is used to segment the images.

2.3 Feature extraction

A concatenation of global features, a proposed local feature and a tuple depicting intensity category (positive = 1, intermediate = 0) is extracted and used for classification.

2.3.1 Global features

Image descriptors such as area, average intensity and entropy have been extracted to describe the global characteristics of the image [6].

2.3.2 Proposed local features

The thresholded local binary count (TLBC) feature is a variant of the local binary count (LBC) feature [8]. For a 3×3 window in an image, the corresponding TLBC is the number of surrounding pixels whose values greater than the center pixel by more than the noise threshold constant τ (see eq 2 and 3). The extraction of TLBC for $\tau = 2$ is demonstrated (see Figure 1).

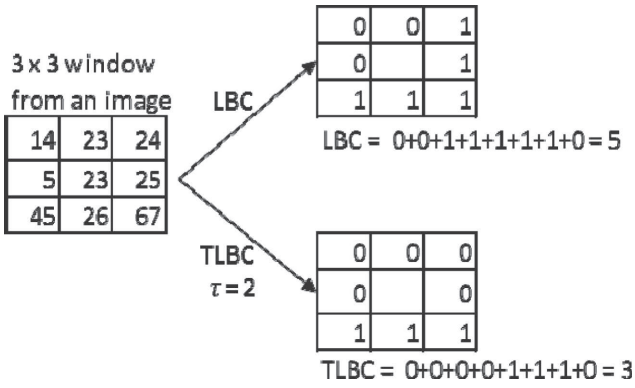

Figure 1: Demonstration of TLBC extraction against LBC extraction.

The TLBC characterizes the local binary gray scale pixel information in an image and is rotational invariant. However, unlike LBC, the TLBC is insensitive to pixel quantization noise. The noise threshold constant τ is used to specify the level of pixel variations to eliminate as noise. when τ becomes 0, the TLBC becomes equivalent to LBC [8].

$$\text{TLBC} = \sum_{i=1}^{8} f(P_i - P_c) \tag{2}$$

$$f(x) = \begin{cases} 1, & x > \tau \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Where, P_i denotes the ith pixel surrounding the center pixel P_c in a 3×3 window. The TLBC is calculated for every possible 3×3 windows in the image. The probability distribution of all TLBC's of an image is called TLBC feature. This feature contains nine elements each corresponding to the occurrence of TLBC values of 0 to 8.

2.4 Classification

In-order to optimize the value of τ, separate sets of features were extracted by varying its values from 0 to 10. The dimension of extracted features is 13 and is further reduced to six using principle component analysis. Subsequently, the features are classified using support vector machine with a radial basis kernel. Ten-fold cross-validation is performed and average classification accuracies are obtained.

3 Results

Representative images from the SNPHEp2 dataset [9] are presented (see Figures 2 and 3). Through visual inspection,

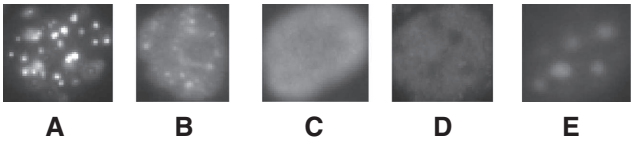

Figure 2: Representative positive-illumination IIF images used in this study. (A) centromere, (B) coarse-speckled, (C) fine-speckled, (D) homogenous and (E) nucleolar [9].

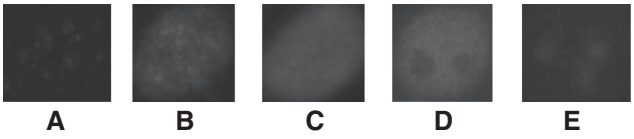

Figure 3: Representative intermediate-illumination IIF images used in this study. (A) centromere, (B) coarse-speckled, (C) fine-speckled, (D) homogenous and (E) nucleolar [9].

it can be inferred that, except the centromere, coarse-speckled and fine-speckled images of positive illumination, images are of low contrast and require pre-processing. It can also be inferred that the images are of different shape and size. It can be observed that centromere images have bright spots over a light background. The spots are of different size and numbers for different images. The coarse-speckled images have bright spots on a dense background. The fine-speckled class of images have very fine spots on dense background. The homogenous images have medium dense nucleoplasm with large dark spots. The nucleolar images have large bright spots on a dark background.

Pre-processing using PN has been performed (see Figures 4 and 5). It is observed that PN increases image contrast and reduces the difference in contrast between positive and intermediate images. The images

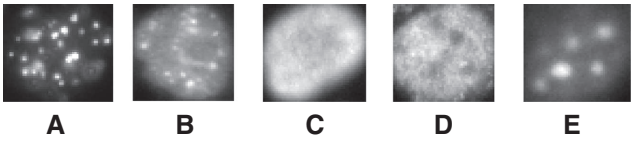

Figure 4: Representative PN enhanced positive-illumination IIF images used in this study. (A) centromere, (B) coarse-speckled, (C) fine-speckled, (D) homogenous and (E) nucleolar.

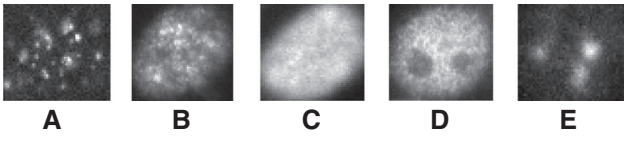

Figure 5: Representative PN enhanced intermediate-illumination IIF images used in this study. (A) centromere, (B) coarse-speckled, (C) fine-speckled, (D) homogenous and (E) nucleolar.

pre-processed using above method are segmented using four multilevel Otsu thresholds. Comparison between the thresholds (T1, T2, T3 and T4) is performed using Dice's coefficient, Jaccard Index and accuracy metrics (see Figure 6A and B). It is observed that threshold T1 performs well in segmenting the regions of interest than other thresholds. TLBC features are extracted from T1 segmented images and classified using SVM classifier by varying τ (see Figure 7). It is observed that when $\tau = 0$, the TLBC becomes LBC and the classification accuracy is 84.5%. The classification accuracy reaches a maximum of 86% at a τ value of 2. It is observed that the classification accuracy decreases with a further increase in τ. The accuracy reaches a constant value of 84.7 at a τ range of 4–6. It is seen that a further increase in τ causes a decrease in accuracy. The confusion matrix for the highest classification accuracy

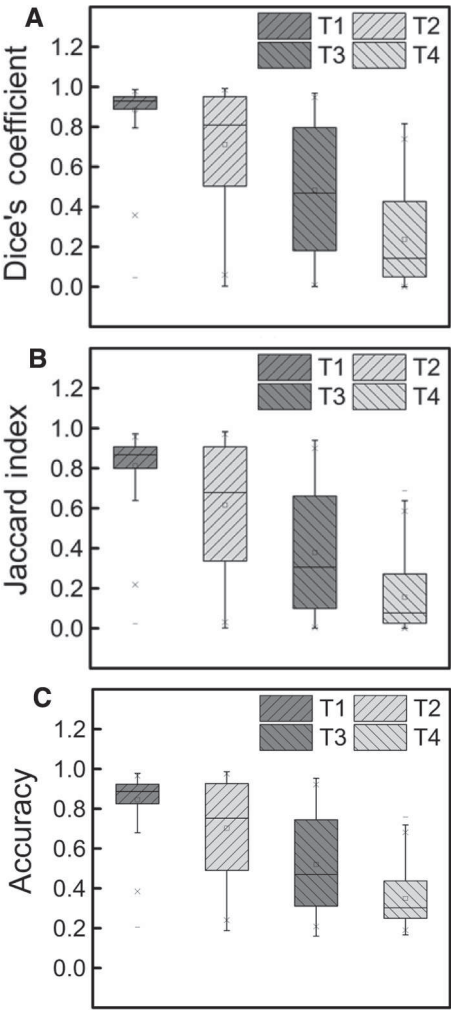

Figure 6: Comparison of segmentation performance among different multi Otsu threshold levels T1-T2 using (A) dice's coefficient, (B) jaccard index (C) accuracy metrics.

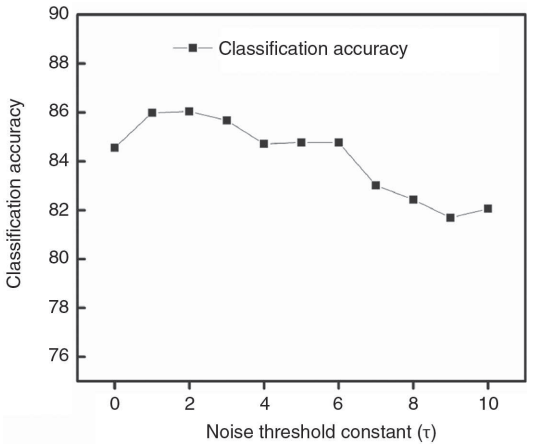

Figure 7: Comparison of classification accuracy using different values of noise threshold constant τ.

	Ce	CS	FS	H	N
Ce	80.40	6.34	0.58	3.75	8.93
CS	5.49	87.09	0.27	2.75	4.40
FS	0.75	0.25	90.50	6.25	2.25
H	2.41	3.22	6.97	85.79	1.61
N	5.00	5.75	1.25	2.25	85.75

Figure 8: Average confusion matrix of ten-fold cross validation using $\tau = 2$.

obtained at $\tau = 2$ is shown (see Figure 8). It can be observed that fine-speckled and coarse-speckled images are segmented with maximum accuracies of 90.5% and 87.0%, respectively.

4 Conclusion

IIF images were pre-processed, segmented and classified using the proposed framework. It appears that pixel normalisation performs well as a pre-processing method. Apart from improving image contrast, it also decreases difference in contrast levels between positive and intermediate images. Segmentation using multi-level Otsu thresholding is performed and validation using Dice's coefficient, Jaccard index and accuracy suggest that first level among four levels of Otsu performed well in segmenting the nucleus. A concatenation of thresholded local binary count features (TLBC), area, entropy, mean intensity and illumination level were extracted and a support vector machine was used for classification. For varying values of the noise thresholding constant of TLBC, the highest accuracy was observed for a noise threshold constant of 2. Among various staining patterns, fine-speckled and coarse-speckled classes have been classified with highest accuracies of 90.5% and 87%, respectively. Since computer aided techniques are recommended to standardise the diagnosis of autoimmune disease, the proposed feature seems to be clinically significant.

Author's Statement

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Material and Methods: Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The conducted research is not related to either human or animal use.

References

[1] Wiliem A, Sanderson C, Wong Y, Hobson P, Minchin RF, Lovell BC. Automatic classification of human epithelial type 2 cell indirect immunofluorescence images using cell pyramid matching. Pattern Recognit. 2014;47:2315–24.

[2] Faraki M, Harandi MT, Wiliem A, Lovell BC. Fisher tensors for classifying human epithelial cells. Pattern Recognit. 2014;47:2348–59.

[3] Yang Y, Wiliem A, Alavi A, Lovell BC, Hobson P. Visual learning and classification of human epithelial type 2 cell images through spontaneous activity patterns. Pattern Recognit. 2014;47:2325–37.

[4] Ghosh S, Chaudhary V. Feature analysis for automatic classification of HEp-2 florescence patterns?: computer-aided diagnosis of auto-immune diseases. In: 21st International Conference on Pattern Recognition; 2012. p. 174–7.

[5] Schaefer G, Doshi NP, Krawczyk B. HEp-2 cell classification using multi-dimensional local binary patterns and ensemble classification. In: 2nd IAPR Asian Conference on Pattern Recognition; 2013. p. 951–5.

[6] Suganthi SS, Pradhap SSA, Ramakrishnan S. An approach to diagnosis of auto-immune diseases using HEP-2 staining pattern and fractal texture features. Biomed Sci Instrum. 2015;51:349–54.

[7] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern. 1979;9:62–6.

[8] Zhao Y, Huang DS, Jia W. Completed local binary count for rotation invariant texture classification. IEEE Trans Image Process. 2012;21:4492–7.

[9] Wiliem A, Wong Y, Sanderson C, Hobson P, Chen S, Lovell BC. Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors. In: IEEE Workshop on Applications of Computer Vision (WACV); 2013. p. 95–102.